# Optimizing Inventory Replenishment of Retail Fashion Products

Marshall Fisher • Kumar Rajaram • Ananth Raman

*The Wharton School, University of Pennsylvania, 3620 Locust Walk, 3207 SH-DH, Philadelphia, Pennsylvania 19104-6366*

*University of California, Los Angeles, Anderson School of Management, Los Angeles, California 90095-1481*

*Harvard Business School, Soldiers Field, Boston, Massachusetts 02163*

We consider the problem of determining (for a short lifecycle) retail product initial and replenishment order quantities that minimize the cost of lost sales, back orders, and obsolete inventory. We model this problem as a two-stage stochastic dynamic program, propose a heuristic, establish conditions under which the heuristic finds an optimal solution, and report results of the application of our procedure at a catalog retailer. Our procedure improves on the existing method by enough to double profits. In addition, our method can be used to choose the optimal reorder time, to quantify the benefit of leadtime reduction, and to choose the best replenishment contract.

(*Retailing; Inventory Replenishment; Stochastic Dynamic Programming; Heuristics*)

## 1. Introduction

Retail inventory management is concerned with determining the amount and timing of receipts to inventory of a particular product at a retail location. Retail inventory-management problems can be usefully segmented based on the ratio of the product's lifecycle $T$ to the replenishment leadtime $L$. If $T/L < 1$, then only a single receipt to inventory is possible at the start of the sales season. This is the case considered in the well-known newsvendor problem. At the other extreme, if $T/L \gg 1$, then it's possible to assemble sufficient demand history to estimate the probability density function of demand and to apply one of several well-known approaches such as the $Q$, $R$ model.

The middle case, where $T/L > 1$ but is sufficiently small to allow only a single replenishment or a small number of replenishments, has received much less attention both in the research literature and in retail practice. As we describe in § 2, there is a small but growing literature on limited-replenishment inventory problems. Perhaps because of the lack of published analysis tools, we have found that retailers often ignore the opportunity to replenish when $T/L$ is close to one and treat this case as though it were a newsvendor problem. This is unfortunate, because, as we show with the numeric computations in this paper, planning for even a single replenishment, can, in this case, dramatically increase profitability.

In this paper, we consider limited lifecycle retail products in which only a single replenishment is possible. We model the problem of determining the initial and replenishment order quantities (to minimize the cost of lost sales, backorders, and obsolete inventory at the end of the product's life) as a two-stage stochastic dynamic program. We show that the second-stage cost function of this program may not be convex or concave in the inventory position after the reorder is placed, which means that simulation-based optimization techniques (Ermoliv and Wetts 1998) typically used to solve problems of this type are not guaranteed to find an optimal solution. For this reason, and also for computational efficiency, we formulate a heuristic for this problem. We show that this heuristic finds an optimal solution if demand subsequent to the time a reorder is placed is perfectly correlated with demand prior to this time. While perfect



1523-4614/01/0303/0230$05.00
1526-5498 electronic ISSN

correlation between early and late demand is unlikely, we believe this result indicates that our heuristic will work well if this correlation is high. Thus, in practice, it seems reasonable to expect good performance from this heuristic because the logical basis of implementing replenishment based on early sales is that demand during the later season is highly correlated with early demand. In our application, the correlation between early and late demand was 0.95. We also apply simulation-based optimization techniques (Ermoliv and Wetts 1998) and find that our heuristic is much faster and finds solutions within 1% of the optimization procedure if the correlation between early and late demand is at least 60%. For lower correlations, the solutions are within 1% to 5%.

We have applied this process at a catalog retailer and find that it improved over their current process for determining initial and replenishment quantities by enough to essentially double profits. Remarkably, compared to no replenishment, a single-optimized replenishment improves profit by a factor of five. A key challenge in implementing short lifecycle replenishment is estimating a probability density function for demand with no demand history. To circumvent this problem in our application, we applied the committee-forecast process in Fisher and Raman (1996) and found that it worked well.

The most important difference between catalog and traditional retail management is that a catalog customer will generally accept a backorder if an item is stocked out. Because our application was at a catalog retailer, our model and heuristic are given for this version of the problem, but it is straightforward to modify the model, heuristic, and proof of optimality for a case where backorders are not allowed.

In § 2 of this paper, we review the literature on short lifecycle inventory replenishment. In § 3, we formulate the problem; in § 4, we state our heuristic and establish optimality conditions; in § 5, we show how to modify the process when customers may return merchandise, and in § 6, report results of our application.

## 2. Literature Review

Analytical models for managing inventory for short lifecycle products share many common features. First, all are stochastic models, because they consider demand uncertainty explicitly. Second, they consider a finite selling period at the end of which unsold inventory is marked down in price and sold at a loss. In this sense, these models are similar to the classic newsvendor model. Third, they model multiple production commitments such that sales information is obtained and used to update demand forecasts between planning periods. The last two characteristics, finite-selling periods and multiple production commitments, differentiate style goods inventory models from other stochastic inventory models. Examples of papers that consider style goods inventory problems include Murray and Silver (1966), Hausman and Peterson (1972), Bitran et al. (1986), Matsuo (1990), and Fisher and Raman (1996). A detailed review of these papers can be found in Raman (1999).

Recent work that deals specifically with the retailer's inventory-management problem for short lifecycle products includes Bradford and Sugrue (1990), Eppen and Iyer (1997a), and Eppen and Iyer (1997b). Bradford and Sugrue model a decision that is similar to the one we study, but they do not consider the impact of replenishment leadtimes. In addition, their solution procedure consists of complete enumeration, which works efficiently for smaller problems but could be difficult to implement in larger, practical-sized problems. Eppen and Iyer (1997a) consider a problem that is substantially different from ours. Even though their model allows the retailer to ''buy'' and ''dump'' at the beginning of each period, the solution method they propose applies only when no ''buy'' decisions are permitted after the first period.

Eppen and Iyer (1997b) model a backup agreement in place at a catalog retailer. A backup agreement is one of the mechanisms by which a retailer achieves replenishment of branded merchandise supplied by a manufacturer to several retailers. In a backup agreement, a retailer places an initial order before the start of the sales season and commits to reorder a certain quantity during the season. After assessing sales during the early part of the season, if the retailer chooses to reorder less than this commitment, there is a penalty cost assessed for each unit not ordered. In this model, replenishment leadtimes are assumed to be zero, which is reasonable because the manufacturer would typically have produced the product and held it in inventory for this and other retailers. Because the replenishment leadtime is zero, it is not necessary for the retailer to accept backorders from consumers.

In this paper, we consider the case where replenishment is achieved without backup agreements. After receiving an updated order, the manufacturer produces and delivers products to the retailer after a significant leadtime. The retailer is not required to commit to any of the reorders. To compensate for the long leadtime, consumer backorders are accepted by the retailer. This case occurs when the manufacturers are either captive suppliers or wholly owned by the retailer and the retailer sources from several such manufacturers. Thus, in this case, it is crucial to model the impact of leadtimes and backorders, although this significantly complicates the analysis leading to a nonconvex optimization model. In addition to incorporating replenishment leadtimes and backorders, our work differs from all these papers in the process that we use to estimate demand densities and to compare our method to actual practice.

## 3. Model

We model the supply decisions faced by a catalog retailer for a product with random demand over a sales season of fixed length. The retailer must determine an initial order $Q_1$ available at the start of the sales season. At a fixed time $t$ during the season, the retailer updates the demand forecast, based on observed sales, and places a reorder quantity $Q_2$ that arrives after a fixed leadtime $L$ at time $t + L$.

Price is fixed throughout the season. Inventory left over at the end of the season is sold at a salvage price below cost. Customers who encounter a stockout will backorder if there will be sufficient supply at some point in the future to satisfy the backorder. Specifically, the opportunity to backorder is not offered to a customer once the total supply quantity $(Q_1 + Q_2)$ has been committed through sales or prior backorders. A lost sale is incurred when an item requested by a customer is not in stock or not backordered.

We first model this problem and formulate a solution heuristic assuming that the reorder time $t$ is fixed. Then, we determine an optimal reorder time $t$ empirically for a given data set by parametrically solving this problem with varying $t$. We are given:

$C_u$ = Cost per unit of lost sale. This is set to the difference between the per-unit sales price and cost of the product.
$C_o$ = Cost per unit of leftover inventory. This is set to the difference between the per unit cost and the salvage price of the product.
$C_b$ = Cost per unit of backorders. This is set to the additional costs incurred in procurement and distribution when an order is backlogged, plus an estimate of the cost of customer ill will.
$L$ = Length of replenishment leadtime.

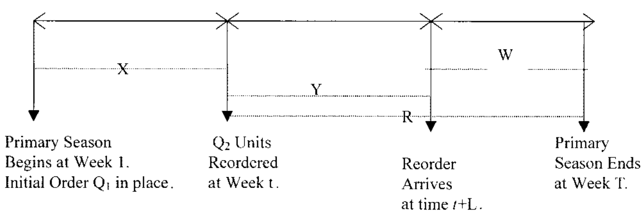
Define the following variables:

$X$ = Random variable representing total demand until the reorder is placed.
$Y$ = Random variable representing total demand during the replenishment leadtime.
$W$ = Random variable representing total demand *after* the reorder arrives until the end of the season.
$R$ = Random variable representing total demand after the reorder *is placed* until the end of the season, where $R = Y + W$.
$Q_1$ = Initial-order quantity.
$Q_2$ = Reorder quantity.
$I$ = Inventory position after reorder is placed. $I = Q_1 + Q_2 - x$.

For the given reorder time $t$, the decision process involves choosing $Q_1$, observing $x$, and then determining the inventory position $I$ for the remainder of the season to minimize total backorder, understock, and overstock costs. This sequence of decisions is shown in Figure 1.

We consider random variables $\Theta$ and $\Psi$ with joint density function $f(\theta, \psi)$. Let $g(\theta)$ be the marginal density on $\Theta$ defined by $f(\theta, \psi)$ and $h(\psi|\theta)$ be the conditional density on $\psi$ given $\theta$ defined by $f(\theta, \psi)$. We define $E_{\Psi/\Theta}(\Omega(\psi)) = \int_0^\infty \Omega(\psi)h(\psi|\theta)\,\partial\psi$ and $E_\Theta(\Omega(\theta)) = \int_0^\infty \Omega(\theta)g(\theta)\,\partial\theta$, where $\Omega(\,)$ is any real-valued scalar

**Figure 1 The Replenishment Planning Process**




function. Let $(a)^+ = \max(a, 0)$. We model this problem by the two-stage stochastic dynamic program (P1).

$$Z(t) = \min_{Q_1 \geq 0} C(Q_1) = E_x[C_1(Q_1, x) + C_2(Q_1, x)] \quad \text{(P1)}$$

where

$$C_1(Q_1, x) = C_b(x - Q_1)^+$$

$$\begin{aligned} C_2(Q_1, x) &= \min_I C_2(I, x) \\ &= E_{Y/X}\{C_b \min((y - (Q_1 - x)^+)^+, \\ &\qquad I - (Q_1 - x)^+)\} \\ &\quad + E_{Y/X}E_{W/X}\{C_u(y + w - I)^+ \\ &\qquad + C_o(I - y - w)^+\} \end{aligned}$$

$$I \geq Q_1 - x, \qquad I \geq 0 \qquad \text{(P2)}$$

$C_1(Q_1, x)$ represents the backorder costs $C_b(x - Q_1)^+$ during the period before the reorder is placed. $C_2(I, x)$ represents expected cost as a function of the inventory position $I$ after the reorder is placed and consists of two terms. The first term, $E_{Y/X}\{C_b \min((y - (Q_1 - x)^+)^+, I - (Q_1 - x)^+)\}$, represents the expected costs of backorder during the replenishment leadtime. Because backorders are accepted only if they can be filled from replenishment, it is important to recognize that backorders during the replenishment leadtime can never exceed the effective inventory position after the first period backlog is cleared (i.e., $I - (Q_1 - x)^+$). This condition is enforced by the operator $\min((y - (Q_1 - x)^+)^+, I - (Q_1 - x)^+)$. The second term of $C_2(I, x)$ is $E_{Y/X}E_{W/X}\{C_u(y + w - I)^+ + C_o(I - y - w)^+\}$, represents the expected overstock and understock costs in the periods after the reorder is placed until the end of the season.

It is important to recognize that $C_2(I, x)$ is neither convex nor concave in $I$. We illustrate this property using the following example.

Example 1. Let $Q_1 = 50$ and $x = 10$. Let the conditional probability distribution for $Y/(X = 10)$ be $P(Y/(X = 10) = 100) = 0.5$ and $P(Y/(X = 10) = 200) = 0.5$, while the conditional probability distribution for $W/(X = 10)$ is $P(W/(X = 10) = 100) = 0.5$ and $P(W/(X = 10) = 200) = 0.5$.

Let $C_b = 15$, $C_o = 20$, $C_u = 40$, $\lambda = 0.9$, $I^1 = 80$, and $I^2 = 110$, $I^\lambda = \lambda I^1 + (1 - \lambda)I^2 = 83$. By substituting these values, using the values of $Q_1$ and $x$, and distributions $Y/X$ and $W/X$ to calculate expectations, it is easy to verify that:

$$\begin{aligned} C_2(I^\lambda, x) &= E_{Y/X}[C_b \min((y - (Q_1 - x)^+)^+, \\ &\qquad I^\lambda - (Q_1 - x)^+)] \\ &\quad + E_{Y/X}E_{W/X}[C_u(y + w - I^\lambda)^+ \\ &\qquad + C_o(I^\lambda - y - w)^+] \\ &= 9325 > \lambda C_2(I^1, x) + (1 - \lambda)C_2(I^2, x) \\ &= 9317.5. \end{aligned}$$

This shows that $C_2(I, x)$ is not convex in $I$.

Next, let $I^1 = 80$ and $I^2 = 300$, so that $I^\lambda = \lambda I^1 + (1 - \lambda)I^2 = 107$, while all the other values remain unchanged. Now,

$$\begin{aligned} C_2(I^\lambda, x) &= E_{Y/X}[C_b \min((y - (Q_1 - x)^+)^+, \\ &\qquad I^\lambda - (Q_1 - x)^+)] \\ &\quad + E_{Y/X}E_{W/X}[C_u(y + w - I^\lambda)^+ \\ &\qquad + C_o(I^\lambda - y - w)^+] \\ &= 8672.5 < \lambda C_2(I^1, x) + (1 - \lambda)C_2(I^2, x) \\ &= 8775. \end{aligned}$$

This shows that $C_2(I, x)$ is not concave in $I$.

In view of this characteristic of $C_2(I, x)$, simulation-based optimization techniques (Ermoliv and Wetts 1998), typically used to compute the solution to this class of problems, are not guaranteed to solve Problem P2 and subsequently Problem P1 to optimality. Consequently, for this reason and run-time considerations, we elected to develop a heuristic. This heuristic is described in the next section.

Once we have developed a scheme to solve this problem, to find the optimal reorder time, we would perform a line search on $t$ to solve (P):

$$Z^* = \min_{0 \le t \le T-L} Z(t) \qquad \text{(P)}$$

## 4. The Two-Period Newsvendor Heuristic

The purpose of this heuristic is to set $Q_1$. In this regard, it is useful to understand the costs affected by the choice of $Q_1$. Firstly, a portion of $Q_1$ may remain unsold at the end of the season, generating an overstock. Secondly, during the interval 0 to $t + L$, if $Q_1$ is too small, one may incur backorder costs. During the interval $t$ to $t + L$, one may also incur stockouts if satisfied and backordered demand exceeds $Q_1 + Q_2$, but it seems more natural to think of this cost as resulting from the choice of $Q_2$, not the choice of $Q_1$. Given this, we let $S = X + Y$, $U = X + Y + W$, and choose $Q_1$ to solve:

$$Z_h(t) = \min_{Q_1 \ge 0} \bar{C}(Q_1)$$
$$= E_S C_b(s - Q_1)^+ + E_U C_o(Q_1 - u)^+ \qquad \text{(PH)}$$

To solve this problem, let $F_1(s)$ and $F_2(u)$ be the distribution functions of random variables $S$ and $U$, respectively. The first-order condition for problem (PH) is:

$$\frac{\delta Z_h(t)}{\delta Q_1} = -C_b(1 - F_1(Q_1)) + C_o F_2(Q_1) = 0$$

We set the heuristic order quantity $Q_1^h$ to the value of $Q_1$ that satisfies this condition. Rearranging terms, this is calculated as the solution to the following equation:

$$F_1(Q_1) + \frac{C_o}{C_b} F_2(Q_1) = 1$$

Let $f_1(s)$ and $f_2(u)$ be the density functions of random variables $S$ and $U$, respectively. Because

$$\frac{\delta^2 Z_h(t)}{\delta Q_1^2} = +C_b f_1(Q_1) + C_o f_2(Q_1) \ge 0,$$

the first-order conditions are sufficient to establish the optimality of $Z_h(t)$ at $Q_1^h$. Note that our choice of $Q_1^h$ minimizes expected backordering costs during the period before replenishment and minimizes expected overstock cost at the end of the season because of $Q_1^h$. The following result establishes conditions under which this heuristic finds an optimal solution.

PROPOSITION 1. *Suppose $Z(t)$ is the optimal solution to Problem (P1) when random variables $X$, $Y$, and $W$ are perfectly correlated. Then, $Z_h(t) = Z(t)$.*

PROOF. If random variables $X$, $Y$, and $W$ are perfectly correlated, then $Y = \alpha X$ and $W = \beta X$, where $\alpha$, $\beta$ are positive constants. Thus, $E(Y/X = x) = \alpha x$, $V(Y/X = x) = 0$, $E(W/X = x) = \beta x$, and $V(W/X = x) = 0$. When all customers backorder, the optimal reorder quantity is $Q_2^* = [x(1 + \alpha + \beta) - Q_1]^+$. If $Q_2^* > 0$, then one incurs no overstock and understock costs in the third period after the reorder arrives. The only costs incurred will be possible backorder costs during the first two periods represented by $C_b[x(1 + \alpha) - Q_1]^+$. If $Q_2^* = 0$, in addition to the backorder costs in the first two periods, one could incur an overstock of $[Q_1 - x(1 + \alpha + \beta)]^+$ because of the initial order with associated costs $C_o[Q_1 - x(1 + \alpha + \beta)]^+$. Consequently, total expected costs in the season when one has perfectly correlated demand can be expressed as

$$C(Q_1) = E_X\{C_b[x(1 + \alpha) - Q_1]^+ + C_o[Q_1 - x(1 + \alpha + \beta)]^+\}.$$

Because by definition, $S = X + Y = X(1 + \alpha)$ and $U = X + Y + W = X(1 + \alpha + \beta)$, $C(Q_1) = E_S\{C_b(s - Q_1)^+\} + E_U\{C_o(Q_1 - u)^+\} = \bar{C}(Q_1)$. Thus, $Z(t) = \min_{Q_1 \ge Q} C(Q_1) = \bar{C}(Q_1) = Z_h(t)$. Q.E.D.

It can be shown that Proposition 1 also holds under the assumption of no customer backorders. In light of this proposition, it is reasonable to expect good performance from this heuristic because the logical basis of implementing replenishment based on early sales is that demand during the later season is highly correlated with early demand. In our application, we found across all the products the correlation between $X$ and $Y$ to be around 0.96 and between $X$ and $W$ to be about 0.95. This suggests that this heuristic could provide a simple and efficient basis to model the required decisions in this application.

Once one uses the two-period newsvendor heuristic to determine $Q_1$ and observes demand $x$ during the first period, the optimal solution to the minimization problem P2 can be approximated by setting $I = \max(I^*, Q_1 - x)$, where $I^*$ is the newsvendor quantity defined on $H_{R/X}$, the cumulative distribution of $R$ updated by $X = x$ (i.e., $I^* = H^{-1}_{R/X}[(C_u - C_b)/(C_u - C_b + C_o)]$. The quality of this approximation is also assessed in the application while evaluating the performance of the heuristic.

## 5. Modifications to Account for Returns of Merchandise

In catalog retailing, because customers place orders based on photographs displayed in catalogs, purchased merchandise is often returned if the actual product differs from what the customer expected from the catalog. In this section, we describe how to extend our model to include merchandise returns.

Returned items can be resold if they are received before the season ends. This means that backorders in the interval $(0, t + L)$ and stockouts during the interval $(t + L, T)$ may be reduced by the availability of returns, but returns that are received too late to be resold can contribute to overstock. Based on the practice followed by the catalog retailer described in the application, we assume that a known fraction $\omega$ of customers return products, where $0 \le \omega \le 1$. These returns are immediately reusable if necessary to satisfy either a backorder or demand. We also assume that recycled returns (i.e., returns on returns and so on) are not reusable during the sales season. These assumptions ensure that we make an unbiased comparison with existing practice at this retailer.

Consequently, to adapt the heuristic to include returns of merchandise, we first consider the period until the reorder arrives. If $Q_1$ is the initial-order quantity and $s$ is the demand during this period, the total number of reusable returns is $\omega \min(s, Q_1)$. If $s > Q_1$, the total backorders that occur during this period are $[s - (Q_1 + \omega \min(s, Q_1))]^+ = [s - Q_1(1 + \omega)]^+$. Similarly, if $u$ is the demand during the entire season, total reusable returns because of the initial-order quantity are $\omega \min(u, Q_1)$. If $Q_1 > u$, the total overstock that occurs during the entire season is

$$[Q_1 - (u - \omega \min(u, Q_1))]^+ = [Q_1 - u(1 - \omega)]^+.$$

Using these results, we redefine (PH) to:

$$\begin{aligned} Z_r(t) &= \min_{Q_1 \ge 0} C_r(Q_1) \\ &= E_S C_b(s - Q_1(1 + \omega))^+ \\ &\quad + E_U C_o(Q_1 - u(1 - \omega))^+. \qquad (\text{PH}_r) \end{aligned}$$

We use the procedure outlined in the previous section to determine $Q_1^h$ as the solution to the following equation.

$$F_1[Q_1(1 + \omega)] + \frac{C_o}{(1 + \omega)C_b} F_2\left(\frac{Q_1}{(1 - \omega)}\right) = 1$$

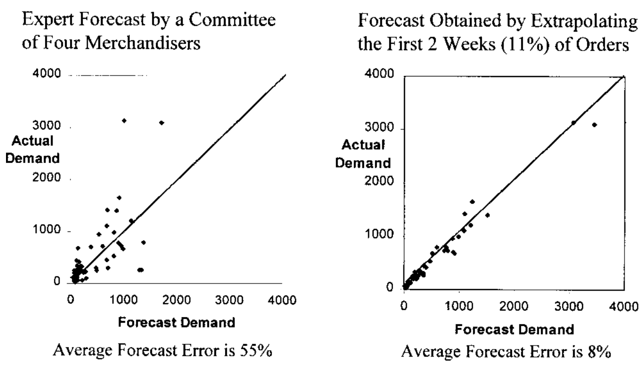
At the end of the first period, we observe realized demand $x$ and use it to set the reorder quantity $Q_2 = I - (Q_1^h - x)^+ - \omega \min(Q_1^h, x)$, where $I = \max(I^*, (Q_1^h - x)^+ + \omega \min(Q_1^h, x))$, and $I^*$ is the newsvendor quantity defined on $H_{R'/X}$, the cumulative distribution of $R' = R(1 - \omega)$ updated by $X = x$ (i.e., $I^* = H^{-1}_{R'/X}[(C_u - C_b)/(C_u - C_b + C_o)]$).

## 6. Application

We have tested the ideas presented in this paper at a large catalog retailer. We applied the model and the two-period newsvendor heuristic to make purchase decisions for 120 styles/colors from the women's dress department appearing in a particular catalog. We chose this division because it represented a significant portion of the business. The sales season for these products is $T = 22$ weeks, and the replenishment leadtime is $L = 12$ weeks. Because these products are sold through mailorder catalogs, the price during the season is fixed. Around 35% of sales are returned, i.e., $\omega = 0.35$.

In the process currently in place at this retailer, initial-order quantities are set to forecast demand for the 22-week season adjusted for anticipated merchandise returns. Forecasts for each style/color are updated after two weeks by dividing observed sales by the historical fraction of total-season sales for the department, which have been observed in the past to

**Figure 2 Comparison of Early and Updated Forecasts**




normally occur in the first two weeks. Reorders are placed to make up the difference from an updated forecast adjusted for returns. Specifically for a given style/color, if $f$ is the total forecast sales and $\omega$ is the anticipated fraction of returns, then $Q_1 = (1 - \omega)f$. Letting $x_2$ be the actual sales observed at the end of two weeks and $k_2$ be the fraction of total demand historically observed at this point for a group of similar products, then we set $Q_2 = ((1 - \omega)x_2/k_2 - Q_1)^+$. Note that this procedure sets $Q_1$ to the forecast sales net of anticipated returns during the entire season and, hence, reorders are used as a reaction to larger-than-anticipated sales rather than something that is planned for in advance.

Figure 2 shows the improvement in forecast accuracy because of updating at the retailer. Each point shows forecast and actual demand for a particular style/color combination. The left graph compares demand forecasts with actual demand for the average of forecasts made by four expert buyers prior to the beginning of the season. In the right graph, the forecasts equal actual sales after two weeks into the season divided by a factor representing the fraction of total sales historically observed after two weeks.

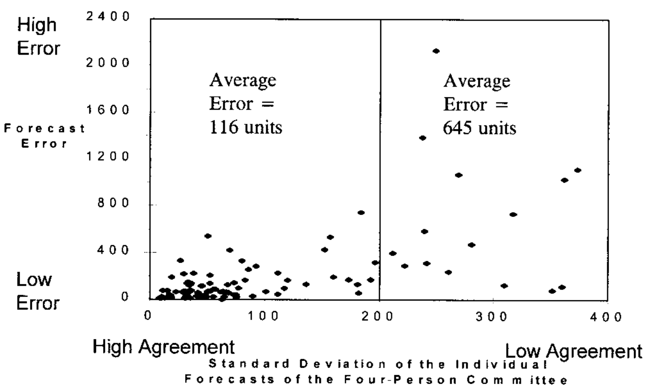
Application of our model requires a method to estimate demand-probability distributions. This is particularly challenging because there was no sales history for any of the new dresses. However, we were able to calculate forecast errors, defined as the difference between buyer forecast and actual sales for similar products appearing in the same catalog from the past two years. We used this information to conclude that the distribution of forecast errors was normally distributed with a large degree of confidence ($\chi^2$ test holds at $\alpha = 0.01$ level). We assumed that forecast errors would follow a similar distribution in past and future seasons. This seemed reasonable because the same individuals who forecasted product demand in the past were also forecasting current season demand.

Because the demand for any given product is equal to its forecast plus the associated forecast error, this implies that the demand distribution for $U$ for a given product during the entire season is normally distributed. While probability distributions for retail products seem to have long tails, these result from plotting *actual* demand for products that seem indistinguishable (or at least similar) ex ante. However, in contrast, $U$ represents the demand distribution for a given product.

To estimate normal parameters $\mu$ and $\sigma$ of this distribution, we implemented the procedure developed by Fisher and Raman (1996). In this method, the members of a committee (comprised of four buyers in our case) independently provide a forecast of sales for each product. The mean $\mu$ is set to the average of these forecasts. The standard deviation of demand $\sigma$ is set to $\theta\sigma_c$, where $\sigma_c$ is the standard deviation of the individual committee member's forecasts and the factor $\theta$ is chosen so that the average standard deviation of historical forecast errors equals the average standard deviation assigned to new products. In our application, we found $\theta$ to be 1.4.

To estimate the parameters of distribution $X$ and $R$, where $U = X + R$, we assume that $(X, R)$ follows a nondegenerate bivariate normal distribution. For this distribution, it is well known (Bickel and Docksum 1977) that the marginal distribution of $X$ is an univariate normal distribution with mean $\mu_x$ and standard deviation $\sigma_x$, while the marginal distribution of $R$ is also normally distributed with mean $\mu_R$ and standard deviation $\sigma_R$. Let $k_t$ represent the proportion of total sales until reorder point $t$, $\delta_t$ the correlation between $X$ and $R$, and $\rho_t$ the correlation between $X$ and $U$. We estimate $k_t$, $\delta_t$, and $\rho_t$ from historical data and use the formulas developed in Fisher and Raman (1996) to calculate

**Figure 3 Committee Standard Deviation Versus Forecast Error**




$$\mu_x = k_t\mu, \quad \sigma_X = \sigma\left[\rho_t - \delta_t\sqrt{\frac{(1-\rho_t^2)}{(1-\delta_t^2)}}\right],$$

$$\mu_R = k_t(1-\mu), \quad \text{and} \quad \sigma_R = \sigma\sqrt{\frac{(1-\rho_t^2)}{(1-\delta_t^2)}}.$$

For the bivariate normal distribution $(X, R)$, note that the updated distribution $R/X = x$ is also normally distributed with mean $\mu_{R/X} = \mu_R + \delta_t(x - \mu_X)\sigma_R/\sigma_X$ and standard deviation $\sigma_{R/X} = \sigma_R\sqrt{1-\delta_t^2}$. Because $0 \le \delta_t < 1$, this implies that $\sigma_{R/X} \le \sigma_R$. Thus forecast updating based on actual sales $x$ reduces variance in the distribution of demand during the remaining season and permits a more accurate forecast. By using replenishment, the retailer can take advantage of this improved forecast by placing a more precise reorder that directly contributes to higher expected profits during the remaining season.

To better understand the nature of forecast errors, we compared the standard deviation of the committee forecast for individual products at the beginning of the season (i.e., $\sigma_c$) with its corresponding forecast error. These results, shown in Figure 3, suggest that when the committee agrees, they tend to be accurate, and that the committee process is a useful way to determine what you can and cannot predict.

With the exception of the backorder penalty $C_b$, all the cost parameters required for our analysis were readily available. Estimating the backorder penalty is challenging in practice because, in addition to the $1 per-unit extra-transaction cost for procurement and distribution associated with a backorder, there is an intangible cost because of customer ill will. The company was uncertain as to the exact value of the ill-will cost, but felt a value of $C_b$ in the range $5 to $15 was reasonable. We applied our analysis to three cases using $5, $10, and $15 per unit as values of $C_b$. We also analyzed the case $C_b = 1$ to insure that our heuristic did not outperform the current rules because we assessed an ill-will cost that was not used in the current rule.

Note that although ill-will costs can also be added to $C_u$, we did not add them because an ill-will cost in this application was charged only because management was mainly interested in insuring that flexibility to backorder was not abused. Given the values of $C_u$ and $C_o$, if $C_b = 0$, then it is optimal to set $Q_1 = 0$ and backorder all first-period demand. But, these excessive backorders would likely reduce market share in the long term. The omission of ill-will costs in $C_u$ does not affect the analysis because, depending on the product, $C_u$ was two to four times greater than backorder costs, and consequently, it would never be optimal to not satisfy demand to avoid a backorder.

As a practical matter, we found that historically around 5% of customers chose not to accept the offer to backorder at this retailer. Consequently, we adjusted the backorder cost to account for this fraction of lost sales by defining an effective backorder cost, $C_b' = 0.95 \cdot C_b + 0.05 \cdot C_u$, representing the costs of a backorder and stockout weighted by the expected fraction of customers who would choose either option. We replaced $C_b$ with $C_b'$ in the definition of problem ($PH_r$).

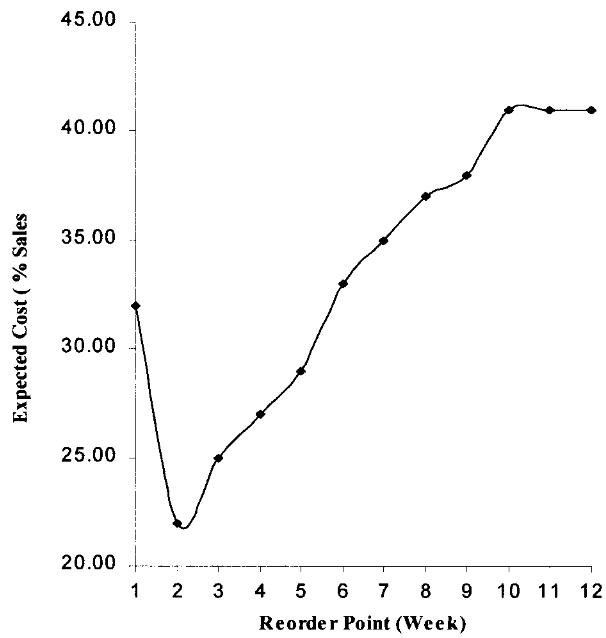
The first step in our methodology is to determine the reorder time. It is important to accurately choose this time because, if chosen too early in the season, actual sales will not be sufficient to provide an accurate revision of the second-period demand forecast. On the other hand, if the reorder time is too far into the season, the benefit of replenishment is diminished because it is then used to service only a small portion of the season. The specific choice of reorder point depends on the proportion of total sales observed during the initial weeks and the length of the replenishment leadtime. For instance, if this proportion is high, there is a long leadtime or both, one would choose the reorder point early in the season to ensure that a reasonable proportion of total sales is serviced by the reorder.

To determine the best reorder time, we used Monte Carlo simulation with the estimated distributions of demand to calculate $Z(t)$ for Week $t$. In this procedure, for a given reorder time, we estimate the initial-order quantity using our heuristic. We simulate $x$ (as a realization of $X$), the distribution of total demand until the reorder is placed. We use $x$ to calculate the backorder costs before the reorder is placed, update $R/X = x$, and calculate the expected costs during the remaining season. We repeat this procedure for several simulated realizations of $X$ and calculate the expected costs during the entire season associated with a reorder time by averaging the costs associated with each realization. As discussed previously, using the bivariate normal distribution to model demand ($X$, $R$) ensures that both $X$ and $R$ are univariate normal distributions and the variance of updated second-period demand $R/X = x$ is also univariate normal whose variance is now reduced from $\sigma_R$ to $\sigma_{R/X} = \sigma_R\sqrt{1 - \delta_t^2}$).

We repeat the simulation for several choices of reorder time. The results of this simulation are summarized in Figure 4. Because $Z(t)$ attains its minimum at $t = t^* = 2$, the reorder time is chosen to be at the end of Week 2. Note that the length of the replenishment leadtime assumed in this analysis is 12 weeks. As the season lasts only 22 weeks, we cannot reorder after Week 10. Consequently, the value of $Z(t)$ for $t \geq 10$ is set equal to the expected costs incurred for a single period buy if we set $Q_1$ to $Q_1^s$, the newsvendor quantity defined on the total distribution of demand (i.e., $Q_1^s = F_2^{-1}[C_u/(C_u + C_o)]$.

A key factor that influences the level of profits gained by replenishment is the proportion of total demand over time observed during the early part of the season and the time until the reorder arrives. Clearly, if this proportion is very high, then the benefits of replenishment are limited, as the reorder would only serve a small proportion of total-season demand. In our application, we found that historically, for similar product lines, 10% of total demand is observed when the reorder is placed after two weeks, and 50% of

**Figure 4 Reorder Time and Expected Costs**




demand is observed at Week 10 when the reorder arrives. These values confirmed that replenishment based on actual sales was a viable strategy for the chosen product line and motivated us to apply our method to determine initial- and replenishment-order quantities.

For the 120 styles/colors in this department, we determined the initial-order quantity by solving problem ($PH_r$) using the heuristic modified to include returns. We then observed $x$, the sales until the second week, and set the reorder quantity $Q_2$ using the procedure developed in § 5. Because at the end of the season we knew total sales and actual sales per week for each dress, we were able to calculate the stockouts, overstock, backorders, dollar sales, and profits that would have resulted from our ordering policy. To compare our method with the current ordering rules, we also calculated these values for the current policy. The results consolidated across all the 120 dresses are tabulated in Table 1.

Observe that although the total orders placed by our method and existing practice are similar, the composition of these orders across the two periods is different. Our heuristic reduces overstocks, stockouts,

**Table 1 Comparison of the Two-Period Newsvendor Heuristic with Current Practice**

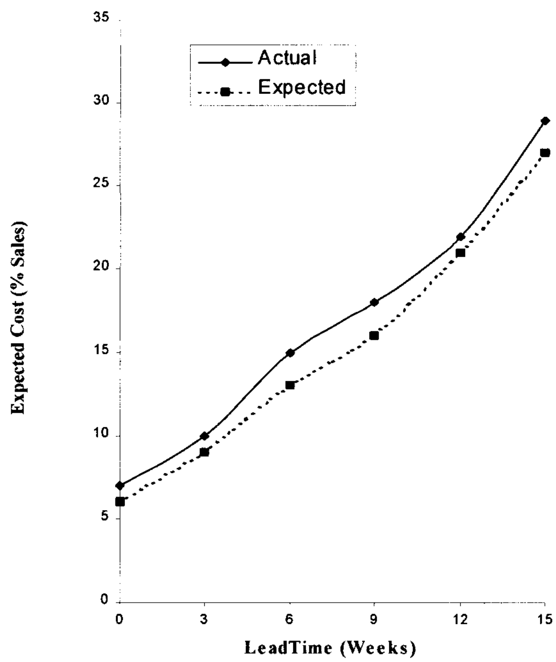
| | Current Rule | Model $C_{iB}$ = \$5 | Model $C_{iB}$ = \$10 | Model $C_{iB}$ = \$15 |
|---|---|---|---|---|
| Initial Order | 19050 | 14479 | 18015 | 20680 |
| Reorder | 2816 | 5229 | 3931 | 3179 |
| Total Buy | 21866 | 19708 | 21946 | 23859 |
| Overstock | 7689 | 4924 | 7061 | 8925 |
| Omits | 3712 | 2998 | 2566 | 2423 |
| Backorders | 6534 | 8643 | 6969 | 5989 |
| Profit (\$) | 431696 | 496597 | 445384 | 395782 |
| Sales (\$) | 1317889 | 1301579 | 1372926 | 1422108 |
| Profit Increase By Model (As % Current Sales) | | 4.92 | 3.52 | 2.23 |

and backorders enough to increase profits compared with the current rule from 2.23% to 4.92% of current sales, depending on the value of $C_b$. Profit before tax for this retailer is around 3% of sales. Consequently, our heuristic offers the potential to approximately double profit.

Our results also show the impact of $C_b$ on the solution. Order quantities for the current rule do not change because the current rule does not consider $C_b$ in determining order quantities. When $C_b$ = \$5, we order a smaller initial quantity because backorders are now relatively less expensive. This in turn increases backorders and stockouts but reduces overstocks, which increases profit improvement from 3.52% to 4.92%. On the other hand, when $C_b$ = \$15, we order a larger initial quantity because backorders are relatively more expensive. This reduces backorders and stockouts but increases overstocks, which reduces the profit improvement from 3.52% to 2.23%.

To insure that our heuristic did not outperform the current rules because we assessed an ill-will cost that was not used in the current rules (possibly because ill-will costs were not recorded in the books), we also considered the case when $C_b$ = 1. Here, $C_b$ only consists of the additional transaction cost per unit of procurement and distribution associated with a backorder. As expected, our heuristic ordered a substantially smaller amount initially than the cases with larger values of $C_b$. This in turn increased backorders and stockouts, but reduced overstocks. Over all, the profit improvement over the current rules increased to 5.52%.

**Figure 5 Replenishment Costs and Leadtimes**




This is consistent with the general pattern in Table 1, which shows profit improvement increasing as $C_b$ decreases.

This analysis assumes a replenishment leadtime of 12 weeks. It is easy to understand that reducing this time could potentially increase the benefits of replenishment, because a greater portion of the season can be serviced from the more accurate reorder. However, it is important to precisely calculate this benefit to justify the costs of leadtime reduction. Our methodology provides a framework to analyze these benefits both before and after sales are realized. To perform this analysis before actual sales are realized, we use a simulation to calculate expected costs for different values of leadtimes. Using actual sales, and for the case in which $C_b$ = 10, we performed an analysis identical to the one used to obtain the results reported in Table 1, but using the new choices of the leadtime. These results are summarized in Figure 5 and indicate that the length of the replenishment leadtimes significantly influences the benefits of replenishments. This type of analysis could be used to decide between a domestic supplier with typically higher costs but shorter leadtimes and a foreign supplier with relatively low costs but long leadtimes.

To further evaluate the quality of this heuristic, we solved (P1) using a simulation-based optimization method.[1] In this technique, we use simulation to numerically compute $C(Q_1)$ for selected values of $Q_1$ in the range $[0, Q^*]$, where $Q^* = F_2^{-1}[C_u/(C_u + C_o)]$ is the newsvendor quantity[2] defined across the whole distribution of demand. Finally, we set $Z(t) = \min_{0 \le Q \le Q_1^*} C(Q_1)$. In the case where $C_b = 10$, this technique improved profit relative to the current rule by around 3% of current sales, a lower improvement than was achieved with our heuristic. In the cases where $C_b$ equaled 5 or 15, the profit improvement was also marginally less than achieved with our heuristic.

In addition to resulting in a smaller profit gain than the two-period newsvendor heuristic, we found that the solution time for this technique was around 40 hours on a Dell Pentium II PC, as compared with less than a minute for the two-period newsvendor heuristic. These results provide strong justification for using this heuristic in this application.

We also considered the impact on performance of correlation between early and later demand. In our application, across all products, we found the correlation between $X$ and $Y$ to be around 0.96 and between $X$ and $W$ to be about 0.95. In view of Proposition 1, such high correlation suggests that the heuristic solution is very close to the optimal solution for this problem.

We performed a computational study to evaluate the performance of the heuristic for different levels of correlation between early and later season demand. Define $\rho_1$ as the correlation between $X$ and $Y$ and $\rho_2$ as the correlation between $X$ and $W$. For simplicity, we set $\rho_1 = \rho_2 = \rho$ and vary $\rho$ from 0 to 0.99 in steps of 0.1. For a given value of $\rho$, and using the committee estimates of $\mu$ and $\sigma$ for each product, we calculated $Q_1$ using the heuristic and using the simulation-based optimization method. We then used these values of the initial-order quantities and Monte Carlo simulation with the estimated distribution of $U$ for each product to calculate expected costs for each technique.

**Table 2** **Percentage Performance Gap Between the Two-Period Newsvendor Heuristic and a Simulation-Based Optimization Method for Different Levels of Demand Correlations**

| Correlation | Performance Gap (%) | Correlation | Performance Gap (%) |
|---|---|---|---|
| 0 | 5.2 | 0 | 5.2 |
| −0.1 | 4.15 | 0.1 | 4 |
| −0.2 | 3.1 | 0.2 | 3 |
| −0.3 | 1.8 | 0.3 | 1.7 |
| −0.4 | 1.5 | 0.4 | 1.4 |
| −0.5 | 1.2 | 0.5 | 1.3 |
| −0.6 | 1.1 | 0.6 | 0.9 |
| −0.7 | 0.8 | 0.7 | 0.7 |
| −0.8 | 0.5 | 0.8 | 0.6 |
| −0.9 | 0.3 | 0.9 | 0.2 |
| −0.99 | 0.1 | 0.99 | 0.07 |

For a given value of $\rho$, let $C_h$ represent the total expected cost of the heuristic across all products, and let $C_s$ represent the corresponding total cost of the simulation-based optimization procedure. The percentage-performance gap of the heuristic is defined as $(C_h - C_s)/C_s \times 100\%$. We report the percentage-performance gap across a range of values for demand correlation in Table 2. For each level of demand correlation, the run times for the heuristic across all the products was less than a minute, while the equivalent run time for the optimization approach was around forty hours.

The results in Table 2 show that this gap varies from 0.06% to 5.2%, with the highest gaps occurring at the lowest levels of correlation. These results suggest that the heuristic provides an efficient basis to address this problem, even for cases that have modest levels of correlation (e.g., $|\rho| > 0.3$) between early and later sales. Because fashion replenishment makes most sense for products with some degree of correlation between early and later sales, this heuristic

[1]Please refer to Ermoliv and Wetts (1998), ‘‘Numerical Techniques for Stochastic Optimization,’’ Springler Verlag, New York, for a theoretical justification and a detailed description of this technique. The same simulation-based technique is used to numerically estimate $C_2(Q_1, x)$ required in the computation of $C(Q_1)$. Here we vary $I$ over the range $[0, 10Q^*]$ and set $C_2(Q_1, x) = \min_{0 \le I \le 10Q^*} C_2(I)$.

[2]We choose this value because it is highly unlikely that an initial-order quantity greater than this quantity would not be sufficient to cover sales during the periods before the replenishment arrives.

seems to be an accurate, simple, and intuitive way for a retailer to implement this strategy.

In conclusion, we believe that the method described here provides a useful framework to improve the accuracy and analyze several crucial aspects of replenishment-based planning.

## Acknowledgments
The authors would like to thank Professor Leroy B. Schwarz, a Senior Editor, and two anonymous referees for several excellent suggestions during the review process.

## References
Bickel, P., K. Doksum. 1977. *Mathematical Statistics.* Holden Day Publishers, San Francisco, CA.

Bitran, G. R., E. Haas, H. Matsuo. 1986. Production planning of style goods with high setup costs and forecast revisions. *Oper. Res.* **34**(2) 221–246.

Bradford, J. W., P. K. Sugrue. 1990. A Bayesian approach to the two-period style-goods inventory problem with single replenishment and heterogeneous poisson demands. *J. Oper. Res. Soc.* **43**(3) 211–218.

Eppen, G. D., A. V. Iyer. 1997a. Improved fashion buying using Bayesian updates. *Oper. Res.* **45**(6) 805–819.

——, ——. 1997b. Backup agreements in fashion buying—The value of upstream flexibilty. *Management Sci.* **43**(11) 1469–1484.

Ermoliv, Y., R.J.B. Wetts. 1998. *Numerical Techniques for Stochastic Optimization.* Springler Verlag, New York.

Fisher, M., A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44**(4) 87–99.

Hausman, W. H., R. Peterson. 1972. Multiproduct production scheduling for style goods with limited capacity, forecast revisions and terminal delivery. *Management Sci.* **18**(7) 370–383.

Matsuo, H. 1990. A stochastic sequencing problem for style goods with forecast revisions and hierarchical structure. *Management Sci.* **36**(3) 332–347.

Murray, G. R., Jr., E. A. Silver. 1966. A Bayesian analysis of the style goods inventory problem. *Management Sci.* (11) 785–797.

Raman, A. 1999. Managing inventory for fashion products. *Quantitative Models for Supply Chain Management.* S. Tayur, R. Ganeshan, and M. Magazine, eds. Kluwer Academic Publishers, Norwell, MA.

*The consulting Senior Editor for this manuscript was Gary Eppen. This manuscript was received on November 30, 1999, and was with the authors 488 days for 2 revisions. The average review cycle time was 45 days.*