Inventory Decision in Hybrid Distribution Channels Involving Partial Postponement Strategy

Ying Lu and Junping Xie
School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang, China

Abstract: Manufacturers today are increasingly adopting hybrid distribution channels to sell their different types of products. This study focuses on the inventory decision problem in hybrid distribution channels involving partial postponement strategy. We develop a decision model for finding the optimal inventory levels of different types of finished products and generic products. Then, a procedure for solving this model is proposed. By using a numerical example we discuss the impact of demand characteristics on the performance of partial postponement. The results show that partial postponement can reduce the decline of the total expected profit in hybrid distribution channels resulting from demand uncertainty and effectively cut down the total inventory cost. Finally, we set future research directions on partial postponement in the context of hybrid distribution channels suggested by our results.

Key words: Hybrid distribution, postponement strategy, inventory decision, partial postponement

INTRODUCTION

Our research is motivated by the practices of some Chinese furnish manufacturers. Nowadays, besides offering more product options, many Chinese furnish manufacturers distribute their products via multiple channels including store-based channels such as department stores, discount stores, outlet stores, specialty retailers, superstores; and direct channels such as interactive home shopping, tele-marketing and the Internet. However, due to the fact that the prices in direct channels are commonly lower than those in store-based channels, more and more customers would like to buy the products in direct channels, especially in the Internet, rather than in store-based channels which makes the stores become so-called experience stores where the customers see the products but buy nothing. As a result, some of the stores have to quit the sale channels. To guarantee the revenue of the retailers and stores, some Chinese furnish manufacturers decide to differentiate the products sold in different channels. A decision these manufacturers need to make is how many products should be manufactured and which inventory policy for the semi-finished products that are used to produce finished products needs to be chosen. This can be regarded as an inventory decision problem involving partial postponement.

With the development of the online and store-based channels, hybrid distribution system which means that the products would be sold in more than one distribution channel becomes an increasingly attractive research subject. Xu and Liu (2013) investigated how price and delivery lead time decisions affect channel configuration strategy. Hua et al. (2010) examined the optimal decisions of delivery lead time and prices in a centralized and a decentralized dual-channel supply chain using Stackelberg game. Dan et al. (2012) examined the optimal decisions on retail services and prices in a centralized and a decentralized dual-channel supply chain using the two-stage optimization technique. Yan and Pei (2009) analyzed the strategic role played by the retail services in a dual-channel competitive market. Chen et al. (2013) developed a pricing decision framework in a dual-channel supply chain and examined the optimal prices under centralized and decentralized decision models. Zhou et al. (2011) proposed three different information-sharing models in the dual-channel supply chain based on whether the two retailers take part in the information sharing or not and analyzed the stable Game Equilibrium. From the literature mentioned above, we can see that much research on hybrid channels focuses on pricing strategies and the information sharing in the channels and there is little research that focuses on the inventory decision in the hybrid channels.

Partial postponement means to stock products both in generic and customized forms so that customer demands can be satisfied either from customized-goods inventory or generic-goods inventory. If the generic goods are used to fill the customers’ orders, they need to be customized before being sent to the customers. A

Corresponding Author: Ying Lu, School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang, China
number of studies in the literature deal with the problem of finding the optimal inventory policy for generic and customized goods. Swaminathan and Tayur (1998) built a mixed integer programming model for analyzing the final assembly process with limited production capacity where the inventory is stored in the intermediate form (i.e., vanilla box). Graman (2010) developed a single-period, two-product, order-up-to-cost model to aid in setting the inventory levels of finished-goods and postponement capacity. Silver and Minter (2005) considered a problem where the satisfaction of a limited portion of the demand for finished products can be postponed by holding stocks of semi-finished items and showed how to choose the appropriate order quantities of finished products and the finishing capacity. Fu et al. (2012) proposed a closed-form formula to obtain the optimal procurement quantities for customized products and common components. In the context of hybrid channels involving partial postponement, the key question is how to determine the inventory levels of finished products and semi-finished products.

The rest of this study is organized as follows. In Section 2 we describe the operation of partial postponement in hybrid channels and build a mathematical model for finding the optimal inventory levels of finished and generic products to maximize the total expected profit. Section 3 proposes a procedure for finding the optimal solution to the model. Section 4 discusses the impact of demand characteristics on the performance of partial postponement. Section 5 presents summary comments and discusses the areas for future research.

**MODEL FORMULATION**

In this model, we consider two types of products supplied by a manufacturer and sold in a direct channel and a retail channel, respectively. At the beginning of a period, the finished products are stored in both channels, while the inventory of generic products that can be manufactured into different types of finished products is only held by the manufacturer. Once customers’ orders are received in the direct channel, they would first be filled by the finished products stored by the manufacturer. Then, the orders left would be fulfilled by generic items after being converted into finished products sold in the direct channel. Similarly, when the inventory quantity in the retail channel exceeds the number of finished products available in that channel, the retailers would ask the manufacturer to produce finished products for them which come from the generic items (Fig. 1).

![Fig. 1: Hybrid distribution channels involving partial postponement](image)

If in both channels exist the customers’ orders needing to be satisfied by the generic items, a first-in-first-out issuing policy is used since none of the channels has the priority over the other. We focus on modeling the benefits and costs of partial postponement strategy adopted in hybrid distribution channels. The objective is to determine the inventory level of each type of finished products in each channel and the inventory quantity of generic items held by the manufacturer to maximize the expected profit.

**Notations:**

- \( X_{dm}, x_i \): demand for finished products in the direct channel and the retail channel, respectively.
- \( p_m, p_i \): Unit selling price of a finished product in the direct channel and the retail channel, respectively.
- \( g_m, g_i \): Unit shortage cost of a finished product in the direct channel and the retail channel, respectively.
- \( s_m, s_i \): Unit salvage value of a finished product in the direct channel and the retail channel, respectively.
- \( \pi_m, \pi_i \): Total profit from the sales of finished products in the direct channel and the retail channel, respectively.
- \( a_m, a_i \): Fixed cost for holding a unit of a finished product in the direct channel and the retail channel, respectively.
- \( b_m, b_i \): Variable cost for holding a unit of finished product in the direct channel and the retail channel, respectively.
- \( c \): Unit cost of a finished product.
- \( w \): Unit wholesale price of a finished product.
- \( t_m \): Unit distribution cost in the direct channel.
- \( c_e \): Extra unit cost for converting a unit of generic product to a unit of finished product, including additional manufacturing cost and distribution cost, in the direct channel and the retail channel, respectively.
- \( Q_m, Q_i \): Inventory level of finished products in direct channel and retailer channel, respectively.
- \( W \): Inventory quantity of generic products.
- \( i \): index for channels, \( i = m, r \).
Assumptions: The assumptions of our model are as follows:

- The customization process converting generic products to finished products takes little time so that the waiting time caused by this process is not taken into account in our model.
- The probability distribution of demand for finished products in the direct channel is a normal distribution, with mean $\mu_m$ and standard deviation $\sigma_n$. The negative tail is typically negligible. Similarly, the probability distribution of demand for finished products in the retailer channel is a normal distribution, with mean $\mu_r$ and standard deviation $\sigma_r$.
- If there are more than one customer’s order in both channels needs to be satisfied by the generic goods, a first-in-first-out issuing policy is used.
- The extra unit costs for converting generic products to different types of finished products in the channels are the same.

Mathematical model: In our model, the decision involves choosing $Q_m$, $Q$, and $W$ to maximize the total expected profit given by:

$$
\text{ETP} = \pi_m(Q_m, Q) + \pi_r(Q) + \text{EP}(W)
$$

where, $\pi_m(Q_m, Q)$ is the expected profit from the sales of $Q_m$ and $Q$ units of the finished products in the direct channel, $\pi_r(Q)$ is the expected profit from the sales of $Q$ units of the finished products in the retail channel, EP(W) is the expected profit from sales of W units of generic items.

By analyzing the structure of the cost and revenue in each channel, we can obtain the expressions for $\pi_m(Q_m, Q)$ and $\pi_r(Q)$ as follows:

$$
\pi_m(Q_m, Q) = p_n \times E\{\min\{Q_m, x_m\}\} + w \times Q_r \\
+ s_n \times E\{\max\{Q_m - x_m, 0\}\} - c \times (Q_m + Q) \\
- t_m \times E\{\min\{Q_m, x_m\}\} \\
- g_m \times E\{\max\{x_m, Q_m - 0\}\} - (a_m + b_m \times Q_m)
$$

$$
\pi_r(Q) = p_r \times E\{\min\{Q_r, x_r\}\} - w \times Q_r \\
+ s_r \times E\{\max\{Q_r - x_r, 0\}\} \\
- g_r \times E\{\max\{x_r, Q_r - 0\}\} - (a_r + b_r \times Q)
$$

In Eq. 2, item 1 represents the revenue from selling finished products in the direct channel, item 2 is the revenue from selling finished products to the retailers, item 3 is the salvage value of the leftover finished products in the direct channel, item 4 is the cost for manufacturing the finished products which would be sold in both channels, item 5 is the cost for the manufacturer’s distributing the finished products to the customers, item 6 is the shortage cost and item 7 is the cost for holding the finished-products inventory in the direct channel that includes fixed costs which are independent of the inventory level and the variable costs which are affected by the inventory level. In Eq. 3, item 1 represents the revenue from selling finished product in the retail channel, item 2 is the cost for purchasing finished products from the manufacturer, item 3 is the salvage value of the leftover finished products in the retail channel, item 4 is the shortage cost and item 5 is the cost for holding the finished-products inventory in the retail channel that includes fixed costs which are independent of the inventory level and the variable costs which are affected by the inventory level.

Note that we have assumed that the demand for finished products in each channel follows the normal distribution, so we can transform (2) and (3) into the following forms:

$$
\pi_m = \mu_m (p_m - t_m - c - b_m) \\
+k_m \sigma_m (s_m - c - b_m) + (w - c)(\mu_r + k_r \sigma_r) \\
-a_m + (s_n + t_m - p_m - g_m) \sigma_m G(k_m)
$$

$$
\pi_r = \mu_r (p_r - w - b_r) + k_r \sigma_r (s_r - w - b_r) \\
-a_r + (s_r - p_r - g_r) \sigma_r G(k_r)
$$

where, $G(k) = \Phi(k) - k(1 - \Phi(k))$ for $i = m, r$, $k = (Q_m - Q)/\sigma_m, \Phi(.)$ is the unit normal probability density function, $\Phi(.)$ is the cumulative distribution function for the unit normal distribution (Axsenet, 2006).

Let $y_m = \max(x_m, Q_m, 0)$ and $y_r = \max(x_r, Q_r, 0)$ be the demands for generic products resulting from demands for finished products not being satisfied by $Q_m$ and $Q_r$ respectively. Also let $Y = y_m + y_r$ be the total requirement for generic products.

If we assume the total requirement for generic products follows a normal distribution, we can derive the mean and variance as follows:

$$
\mu_Y = E(Y) = E(y_m) + E(y_r) \\
= \mu_m G(k_m) + \sigma_m G(k_m)
$$

$$
\sigma_Y^2 = \text{Var}(y_m) + \text{Var}(y_r) \\
= \sigma^2_m \left\{ G(k_m) - [G(k_m)]^2 \right\} \\
+ \sigma^2_r \left\{ G(k_r) - [G(k_r)]^2 \right\}
$$

where, $J(k) = (1 + k^2)[1 - \Phi(k)] - k \Phi(k)$ for $i = m, r$. 

8258
According to the central limit theorem, the above assumption can hold only if there are a sufficiently large number of sales channels, whereas in our mode only two channels are taken into consideration. As a result, there is no simple representation of the probability distribution of \( Y \). Note that the total requirement for generic products has a spike at zero and a distribution of non-zero values that is normally distributed. The probability that the total requirement for generic products is zero (spike at zero) is given by:

\[
K_0 = P(Y = 0) = \Phi(k_0) \Phi(k_0) \tag{8}
\]

The means \( \mu_Y \) and the variance \( \sigma_Y^2 \) of the conditional random variable \( Y' = Y|Y > 0 \) are obtained from \( \mu_Y \) in (6) and \( \sigma_Y^2 \) in (7) as:

\[
\mu_Y = \frac{\mu_y}{1 - P(Y = 0)} = \frac{\sigma_m G(k_m) + \sigma_G(k_r)}{1 - K_0} \tag{9}
\]

\[
\sigma_Y^2 = \frac{\sigma_Y^2}{1 - P(Y = 0)} = \frac{P(Y = 0)(\mu_Y)^2}{[1 - P(Y = 0)]} = \frac{\sigma_Y^2}{1 - K_0} \frac{K_0 \{ \sigma_m G(k_m) + \sigma_G(k_r) \}^2}{(1 - K_0)} \tag{10}
\]

Moreover, with \( Y' \) being normally distributed and with an available stock of \( W \) generic units, the expected total number of units sold from generic products is:

\[
EUS(W) = [1 - P(Y = 0)] \left[ \mu_Y - \sigma_Y G(k') \right] = (1 - K_0) \left[ \mu_Y - \sigma_Y G(k') \right] \tag{11}
\]

where, \( k' = (W - \mu_Y) / \sigma_Y \). Under the rule of the first-come-first-served rationing, each channel gets a share of \( E(Y|Y) \) for \( i = m, r \). The expected profit from sales of \( W \) generic units can be obtained from:

\[
EP(W) = \left( p_m - c - c_r \right) EUS(W) E(Y) - c_m W \tag{12}
\]

Until now, all the three items in (1) have been well defined and we can express the total expected profit as a function of the decision variables \( k_m \) and \( W \) for \( i = m, r \). Note that there is an interaction of the \( k_i \)’s through the \( \sigma_i \) term so that the effects of the \( k_i \)’s are no longer separable. Thus, an attempt to set the partial derivatives (with respect to the \( k_i \)’s) equal to zero would lead to a very complicated set of non-linear equations. Moreover, there is no guarantee that \( ETP \) is a concave function of the \( k_i \)’s and \( W \). Thus, it appears that some type of search procedure is needed to find the best (or at least a very good) solution.

**Algorithm Solution Procedure for the Model**

Unless it is proved that the objective function is indeed a concave function of decision variables \( k_m, k_r \) and \( W \), we have to find the optimal solution by enumerative search of all possible combinations for \( k_m, k_r \) and \( W \) which tends to be an exhausting work. However, instead of evaluating all combinations of \( k_m, k_r \) and \( W \) in the channels, we suggest that we first define bounded feasible regions for \( Q_m \) and \( Q_r \) by means of working out the lower and upper bounds for \( Q_m \) and \( Q_r \).

Furthermore, we can easily formulate the expected profit \( \pi_i(W) \) in the situation where all the finished products are sold in the single direct channel and \( \pi_i(W) \) in the situation where there is only the retail channel, respectively.

\[
\pi_m(k_m, W) = \mu_m (p_m - t_m - c - b_m) + k_m \sigma_m (s_m - (c + b_m)) - a_m + (s_m + t_m - p_m - g_m) \sigma_m G(k_m) + (p_m - c - c_r) G(k_m) - G(k_m + W / \sigma_m) - c_m W \tag{13}
\]

\[
\pi_r(k_r, W) = \mu_r (p_r - c - b_r) + k_r \sigma_r (s_r - c - b_r) - a_r + (s_r - p_r - g_r) \sigma_r G(k_r) + (p_r - c - c_r) G(k_r) - G(k_r + W / \sigma_r) - c_r W \tag{14}
\]

We can prove that \( \pi_m(k_m, W) \) and \( \pi_r(k_r, W) \) are concave functions of \( k_m \) and \( k_r \). If all of \( W \) is used, by setting \( d\pi_m(k_m, W)/dk_m = 0 \), we can obtain a lower bound solution of the optimal \( k_m \) denoted by \( k_m^L \) which must satisfy:

\[
\Phi(k_m^L) + \frac{p_m - c - c_r}{g_m + c + c_r - s_m - t_m} \Phi(k_m^L + W / \sigma_m) = \frac{p_m + g_m - c - b_m - t_m}{g_m + c + c_r - s_m - t_m} \tag{15}
\]

If setting \( W = 0 \), we can obtain an upper bound solution of the optimal \( k_m \) denoted by \( k_m^U \) from (15) which is given by:

\[
\Phi(k_m^U) = \frac{c + b_m + t_m - p_m - g_m}{s_m + t_m - b_m - s_m - p_m} \tag{16}
\]
Proceeding in the same way for \( \pi_r(k_r, W) \), we can obtain the lower bound and upper bound solution of the optimal \( k_m^* \) which are denoted by \( k_m^L \) and \( k_m^U \), respectively.

\[
\Phi(k_m^L) = \frac{p_r - c - \epsilon_r}{g_r + c + \epsilon_r - s_r} \Phi(k_m^L + W / \sigma_r) = \frac{p_r + g_r - c - b_r}{g_r + c + \epsilon_r - s_r}.
\]

\[
\Phi(k_m^U) = \frac{c + b_r - p_r - g_r}{s_r - g_r - p_r}.
\]

(17)

(18)

Hence, the lower solution for \( Q_m \), denoted by \( Q_m^L \), is \( Q_m^L = \mu + k^L \sigma \), for \( i \in m, r \). The upper solution for \( Q_m \), denoted by \( Q_m^U \), is \( Q_m^U = \mu + k^U \sigma \) for \( i \in m, r \).

Since having obtained the limited range for finding the optimal solution, we can provide an efficient algorithm based on the steepest ascent method. The idea is that for a given \( W \), we can start from a given solution where \( Q_m \) and \( Q_r \) are set at their upper limits and find the best downward adjustment of a single quantity until no further improvement is found. Record the corresponding ETP for the given \( W \), denoted by ETP. Then search for the optimal \( W \), denoted by \( W^* \), that leads to the maximum ETP. The solution procedure can be summarized as follows:

- **Step 1**: Set the initial value of \( W^m \) and ETP as zero. Let \( Q_m^0 \) be a 2-vector whose elements all are zeros.
- **Step 2**: For current given \( W \), solve optimal solution \( Q_m^*, Q_r^* \) and ETP with following procedures.
  - **Step 2.1**: Obtain initial \( Q_m = Q_m^U \) via (16), \( Q_r = Q_r^U \) via (18).
  - **Step 2.2**: Subtract 1 from \( Q_m \) while keeping \( Q_r \) and \( W \) unchanged, then calculate the corresponding value of ETP\( Q_m^*, Q_r^*, W \) via (1). Obtain ETP\( Q_m^*, Q_r^*, W \) for \( Q_r \) in the same way. Compare ETP\( Q_m^*, Q_r^*, W \) with ETP\( Q_m^*, Q_r^*, W \), if the former is larger than latter, set METP = ETP\( Q_m^*, Q_r^*, W \)and \( u = m \). Otherwise, set METP = ETP\( Q_m, Q_r, W \) and \( u = r \).
  - **Step 2.3**: If METP < ETP\( Q_m, Q_r, W \), then current \( (Q_m, Q_r) \) is the optimal set with current \( W \). We denote \( Q_m, Q_r \) by \( Q^* \) and the corresponding ETP by ETP.- If METP > ETP\( Q_m, Q_r, W \), then let \( Q_m = Q_m^L - 1 \) and return to Step 2.2.
- **Step 3**: If ETP\( W^m \), then \( Q_m^* \) and \( W^m \) are the optimal solution. Set \( W^* = W^m \), \( Q^* = Q^m \), ETP* = ETPm.

If ETP\( W^m > \)ETP*, set \( W^m = Q^m \), ETP* = ETPm and increase current \( W \) by 1. Return to Step 2.

**NUMERICAL EXAMPLES**

In this section, we consider an example with two type of products sold in a direct channel and a retail channel respectively. The probability distributions of the products in the channels are normal distributions with mean \( \mu_m = 200 \), variance \( \sigma_m^2 = 400 \) and mean \( \mu_r = 300 \), variance \( \sigma_r^2 = 400 \), respectively. The parameters we used are listed as follows:

- \( \mu_m = 10 \), \( \mu_r = 10 \), \( c = 3 \), \( w = 4 \), \( g_m = 2 \), \( g_r = 1 \), \( s_m = 2 \), \( s_r = 1 \), \( \sigma_m = 2 \), \( \sigma_r = 1 \), \( b_m = 0.5 \), \( b_r = 0.5 \), \( \epsilon_m = 0.5 \), \( \epsilon_r = 0.5 \).

We have programmed the algorithm described in Section 3 with MATLAB 7.1. Figure 2 shows the total expected profit with different values of \( W \). In this example, when \( W \) equals 85, the corresponding ETP is the global maximal value. Thus, we can get the optimal solution as \( W^* = 81 \), \( Q_m^* = 174 \), \( Q_r^* = 284 \) and ETP* = 262.

**THE IMPACT OF DEMAND CHARACTERISTICS ON THE PERFORMANCE OF PARTIAL POSTPONEMENT**

In this section we would investigate the effect of demand characteristics on the performance of partial postponement. Although the demand characteristics typically consist of mean, standard deviation and probability distribution of demand, here we only consider the standard deviation of the demand distribution. For simplicity, we assume that the standard deviations of the demand distributions in the two channels have the same value. This assumption can be reasonable since there is little difference between the two types of products in the channel so that the fluctuations of demands in channels tend to be approximately equal. Keeping the other parameters unchanged, for different values of \( \sigma_m \) and \( \sigma_r \), we calculate \( Q_m^*, Q_r^* \), and \( W^* \) as well as the corresponding ETP*. In order to evaluate the performance of partial postponement, we also consider the situation where no partial postponement is adopted in the channels. Let \( Q_m \) and \( Q_r \) denote the optimal inventory levels of finished products in direct channel and retail channel without partial postponement, respectively. For convenience, we also set \( k_m = (Q_m - \mu_m) / \sigma_m \) and \( k_r = (Q_r - \mu_r) / \sigma_r \). Note that \( k_m \) is the same as \( k_m^0 \) and \( k_r \) can be calculated by substituting \( W \) for \( c \) in (18), so we can easily obtain \( Q_m \) and \( Q_r \). By inserting \( k_m \) and \( k_r \) into (4) and (5), we can calculate the expected...
Fig. 2: Total Expected Profit

Fig. 3: Total expected profits in the hybrid channels with partial postponement and without partial postponement

Fig. 4: Optimal inventory levels of finished products in the hybrid channels with partial postponement and without partial postponement

Fig. 5: Optimal inventory levels of generic products in the hybrid channels with partial postponement

would decrease as the demand uncertainty becomes high, partial postponement can reduce the decline of the total expected profit resulting from demand uncertainties.

From Fig. 4 and 5, it can be seen that as the value of standard deviation is increased, the optimal inventory levels of finished products would increase in the hybrid distribution channels without partial postponement, while in those with partial postponement the optimal inventory levels of finished products would decrease, instead the inventory quantity of generic products would increase. Moreover, the optimal inventory levels of finished products in the hybrid distribution channels involving partial postponement are always lower than those in the hybrid distribution channels without partial postponement and the higher the demand uncertainty is,
the larger the gap between $Q_{n'}$ and $Q_n$ as well as that between $Q_{n'}$ and $Q_n$ is. This result shows that once partial postponement is introduced, it tends to increase the inventory of generic products rather than that of finished products to deal with demand uncertainty. Note that the unit holding cost of a generic product is usually lower than that of a finished product. Consequently, partial postponement can effectively cut down the total inventory cost. This is the main advantage of partial postponement applied in hybrid distribution channels.

**CONCLUSION**

In this study we analyze the inventory decision problem in hybrid distribution channels involving partial postponement strategy. We develop a mathematical model for evaluating the total expected profit in the hybrid distribution channels and the impact of demand characteristics on the performance of partial postponement. The results of a numerical example show that as the demand uncertainty increases, the total expected profit in hybrid distribution channels decreases, the optimal inventory levels of finished products decrease and the inventory quantity of generic products increases. As a result, partial postponement can reduce the decline of the expected profit resulting from demand uncertainty and effectively cut down the total inventory cost due to the fact that the unit holding cost of a generic product is usually lower than that of a finished product.

There are several possibilities for further our research on this topic area. First, one could consider the situation where the finished products are sold in more than two channels. Second, one could explore whether arbitrary continuous or discrete distributions of demand (instead of normal distributions) could be easily handled by modifying our model presented in this study. A third extension would be to include demand substitution which means a certain proportion of customers would like to buy similar products in another channel if in the initially desired channel none of finished products is available.

**ACKNOWLEDGMENTS**

This study is supported by National Natural Science Foundation of China (51208232), Science Foundation of Jiangsu University (09JDG078).

**REFERENCES**


