A Model for Studying Inventory Flexibility Advantages of a Direct Market

Kyle Cattani
The Kenan-Flagler Business School
The University of North Carolina at Chapel Hill

Gilvan C. Souza
The Robert H. Smith School of Business
University of Maryland

Abstract

We develop a modeling framework to study inventory flexibility advantages possible in a direct market channel. In a lost-sales environment we explore the possibility of blocking shipments to less-profitable customers before inventory levels drop to zero. We analyze the effect of introducing a backlogging policy. We explore the use of multiple modes of transportation to meet customer requirement dates. Modeling stochastic demands and replenishment times, we identify scenarios in which it is beneficial to use more expensive transportation than the customer specified in order to postpone the actual shipment.

Introduction

E-commerce firms have loudly proclaimed the cost advantage of the direct channel that operates without the cost of traditional "bricks and mortar" retail outlets. Less visible have been the inventory management benefits that are possible from the structure of the channel. A direct channel allows for statistical economies of scale through the pooling of demand over many geographical markets. In addition, subtle manufacturing and shipping flexibility is possible with customers who are willing to wait hours or days for receipt of their product. This paper develops a model to explore some of the inventory benefits that are possible with a distribution channel found in companies like Amazon.com or the numerous other firms that have experienced phenomenal growth in the recent past by using a direct channel. We propose a relatively simple framework from queuing theory for analysis of some of the tradeoffs that arise in e-business supply chains.

The basic model in this paper considers a direct channel supply chain with two players. In particular, it models a direct market firm with in-house production to finished goods inventory (FGI). Alternatively, the model structure also arises when production for the (e-commerce) firm is outsourced but the supplier is located close to the e-commerce firm's fulfillment warehouse such that the shipment leadtime from the supplier to the warehouse is negligible. This situation might arise, for example, if the supplier is co-located on the warehouse site. The model also applies directly to catalog sellers or any firm that ships its product directly to the customer. We assume there is stochastic demand from the customer and stochastic production time at the supplier.

In this study we develop a framework to explore answers to the following questions:

1. Can expected profits increase by refusing some orders even if there is available inventory?
2. When might backlogging (if there is no available inventory) increase profits? Decrease profits?
3. Can expected profits increase by backlogging some orders even before inventory is depleted?
4. When might it be appropriate to ship by air (at higher cost) an order for ground shipment in order to gain flexibility through postponement?

We first model as a basis for comparison a default strategy in which orders are shipped as they arrive unless there is zero inventory, in which case orders are lost. We call this policy 1. Then, we consider policy 2, a potential improvement to the default that holds in reserve some units for the more profitable orders. The third policy we consider backlogs orders if there is no available inventory. An unintended result of such a backlogging policy is that it may lead to reduced profits if the customers who are willing to wait have lower-margin orders. We next consider a policy that can increase profits if one class of customers is more likely to be willing to have their order backlogged. In this policy, which we call policy 4, some orders are backlogged even before inventory levels reach zero. Finally, we analyze a fifth policy that uses a more expensive transportation means in order to postpone the actual shipment.

Related Research

Our research is founded on the use of queuing models in an inventory problem. Queuing models have been employed in the study of inventory systems for many years. Early applications are found in Morse (1958) and Gross and Harris (1971). Yet, as Buzacott and Shanthikumar (1993) stated, "…despite the enormous literature on inventory models, it is surprising how rarely queuing models are used." The assumption of Poisson demands, typical of retailer environments, and recent advances in information systems that allow continuous-time review to be efficient, naturally make queuing models suitable for use in analysis. In this paper, we use queuing models to study inventory systems of e-commerce firms. Since information and inventory systems are the core elements of their operations, the e-commerce firms that most successfully capitalize on supply chain strategies uniquely available to direct market firms may sustain a competitive advantage. Our models are motivated by Cachon (1999), who used a similar framework of queuing models in an inventory system to analyze a very different problem.

The Model

Consider a direct channel firm that produces, stocks, and sells products such as books or compact discs (CDs). We examine the inventory management of a single sku of such a product in a two-player supply chain. The supplier produces the product to FGI at a finite production rate μ with exponentially distributed build times and a fixed per-unit production cost. The product is sold and shipped from FGI directly to the customer at a fixed retail price. The direct channel firm offers the customer a choice of shipment options at different prices that also may result in different margins. For example, overnight air might be offered for a price of $12 while ground shipment might be priced at $4. We consider both lost-sales and backorder cases; in either, a penalty may be incurred if inventory is not available. Inventory is held only in FGI and incurs a fixed holding cost per unit. In this section we determine the profit rate that results from various base stock policies; inventory levels in FGI determine whether or not the supplier is producing and potentially whether or not to ship or backlog specific orders as described by the various policies below. In the following section we analyze the respective performance of the various policies.

A base stock policy is optimal for this supply chain (i.e., it maximizes supply chain profit per unit time), as is shown in the literature (see Beckmann, 1961; Hadley & Whitin, 1963). This is seen intuitively by noting that if production is initiated when there are x units in stock, then the expected profit from that unit in production is decreasing in x since the marginal benefits arise from the probability that an additional unit in stock will increase sales. Marginal costs arise from the probability that an additional unit will incur holding costs and are increasing in x. Thus, production should cease for all x greater than or equal to some base stock level S.

Various Policies

We explore different settings and policies to gain insights to the questions posed above. Superscripts denote policy number. There are two customer types: those desiring air shipment (subscript a) and those desiring ground shipment (subscript g). Orders from each customer type arrive according to a Poisson process with rate λ_k, where $k \in \{a, g\}$. The profit rate is calculated as a function of prices and costs and is determined by modeling inventory as a continuous time Markov chain, similarly to Cachon (1999).

We use the following notation for all policies:

r_k = price (revenue), including shipping, charged to a customer of type $k \in \{a, g\}$

c = production cost per unit

c_l = incremental shipping cost for transportation type $l \in \{a, g\}$

$m_{k,l}$ = margin due to a unit sale to customer type k shipped via transportation type l
= $r_k - c - c_l$

h = unit holding cost

b = unit lost-sale cost (goodwill)

b_g = unit backlogging cost (goodwill)

λ_k = mean demand rate for customer type $k \in \{a, g\}$

λ = total mean demand rate = $\lambda_a + \lambda_g$

μ = mean production (supply) rate

ρ = utilization of system = λ / μ

ρ_k = utilization of system regarding only type *k* customers = λ_k / μ

$p_i^{(j)}$ = stationary probability of state *i* for Markov chain under policy *j*

$I^{(j)}(\square)$ = average inventory under policy *j*

$D_{k,l}^{(j)}(\square)$ = expected sales rate for customers of type *k* shipped via transportation type *l* under policy *j*

$L^{(j)}(\square)$ = expected lost sales rate under policy *j* = $\lambda - \sum_{k\in\{a,g\}} \sum_{l\in\{a,g\}} D_{k,l}^{(j)}(\square)$

$B^{(j)}(\square)$ = expected backorder rate under policy *j*

Profit can be determined as a rate whose function for policy *j* is

$$\begin{aligned}\prod^{(j)}(\square) &= \sum_{k\in\{a,g\}} \sum_{l\in\{a,g\}} m_{k,l} D_{k,l}^{(j)}(\square) - hI^{(j)}(\square) - bL^{(j)}(\square) - b_g B^{(j)}(\square) = \\ &= \sum_{k\in\{a,g\}} \sum_{l\in\{a,g\}} \left(m_{k,l} + b\right) D_{k,l}^{(j)}(\square) - hI^{(j)}(\square) - b_g B^{(j)}(\square) - \lambda b.\end{aligned} \tag{1}$$

Lost Sales Policies

Policy 1 represents the default strategy of shipping orders exactly as requested as long as inventory exists.

Policy 1: Ship orders as they arrive, unless there is zero inventory (lost sales). Shipments are made only by the mode requested, i.e., air or ground.

This policy can be modeled as a birth and death process with $S+1$ states, corresponding to the inventory levels in FGI (ranging from 0 to *S*). The birth rate μ corresponds to the server completion rate, and the death rate λ corresponds to the sum of air and ground order rates. (The notation may seem backwards because the state space is the inventory level, which increases with a service completion and decreases with a customer arrival. In traditional queuing models, the state space is the number of individuals in the system, which decreases with service completions and increases with customer arrivals.) The only optimization in policy 1 is the determination of the optimal base stock level *S**, above which production is ceased. Policy 1 is identical to the base case from Cachon (1999). Thus, for $S \in \{0,1,2,\ldots\}$,

$$p_i^{(1)} = \frac{\lambda^{S-i}\mu^i}{\sum_{j=0}^{S} \lambda^{S-j}\mu^j} = \frac{1-\rho}{\rho^{i-S} - \rho^{i+1}}. \tag{2}$$

For this system, $D_{k,l}^{(1)}(S) = 0$ if $k \neq l$, $D_{k,k}^{(1)}(S) = \lambda_k\left(1 - p_0^{(1)}\right)$, $B^{(1)}(S) = 0$, and

$$I^{(1)}(S) = \sum_{i=0}^{S} i p_i^{(1)} = \frac{S - \left(\frac{\rho}{1-\rho}\right)\left(1-\rho^S\right)}{1-\rho^{S+1}}. \tag{3}$$

Example 1: Consider an example where capacity is limited and there are lost sales with some additional lost goodwill. Let λ_a = 3, λ_g =7, μ = 12, r_a = \$28, r_g = \$20, *c* = \$10, c_a = \$6, c_g = \$4, $h = \$1$, and *b* = \$1. The profit margins are \$12 and \$6 for air and ground orders, respectively. Then, $p_0^{(1)} = .04$, *S**=9, $I^{(1)}(S^*) = 5.9$, $D_{a,a}^{(1)}(S^*) = 2.9$, $D_{g,g}^{(1)}(S^*) = 6.7$, and $\prod^{(1)}(S^*) = \$68.68$.

The first alternative to the default case that we explore is a policy that reserves some units for higher-margin customers.

Policy 2: If inventory drops to S_a or lower, ship only higher-margin (air shipment) orders.

This policy causes lost sales of lower-margin (ground) orders whenever inventory is S_a or lower, and lost sales of both types when inventory is zero.

For a given S and S_a, this policy constitutes a birth and death process, as seen in figure 1, whose stationary probabilities are:

$$p_0^{(2)} = \rho_a^{S_a} \left\{ \frac{1-\rho_a^{S_a+1}}{1-\rho_a} + \frac{\rho^{S_a-S}-1}{1-\rho} \right\}^{-1}, \text{ and} \tag{4}$$

$$p_i^{(2)} = \begin{cases} p_0^{(2)} \rho_a^{-i}, & i = 0,1,\ldots,S_a \\ p_0^{(2)} \left(\frac{\rho}{\rho_a}\right)^{S_a} \rho^{-i}, & i = S_a+1,\ldots,S \end{cases} \tag{5}$$

Define $P_i^{(2)} = \sum_{k=0}^{i} p_k^{(2)}$. Then,

$$P_{S_a}^{(2)} = \frac{\rho^{-S_a} - \rho_a}{1-\rho_a} p_0^{(2)}.$$

Here, $D_{k,l}^{(2)} = 0$ if $k \neq l$, $D_{g,g}^{(2)}(S,S_a) = \lambda_g \left(1 - P_{S_a}^{(2)}\right)$, $D_{a,a}^{(2)}(S,S_a) = \lambda_a \left(1 - p_0^{(2)}\right)$, $B^{(2)}(S,S_a) = 0$, and

$$I^{(2)}(S,S_a) = \sum_{i=0}^{S} i p_i^{(2)} = S - \frac{p_0^{(2)}}{\rho_a^{S_a}} \left\{ S \frac{\rho - \rho_a^{S_a+1}}{1-\rho} + \frac{\rho^{1+S_a-S} - \rho^{S_a+1}}{(1-\rho)^2} \right\}.$$

Backlogging Policies

Policies 1 and 2 assume lost sales. We now consider backlogging cases, which take advantage of the on-line customers' acceptance of non-instantaneous fulfillment. Backlogging is reasonable only when capacity utilization is less than one.

Policy 3: Ship orders as they arrive. When inventory drops to zero, backlog customer orders and ship them as inventory arrives.

Policy 3 represents a natural policy to follow when inventory drops to zero and customers are willing to wait for delivery. This straightforward backlogging policy may be intuitively appealing but could have an unfortunate outcome in an e-commerce market. Customers with ground orders likely will be more willing to backlog than those with air orders since most people will not pay to ship by air something on backorder, given competitive suppliers that are only a mouse click away. If air orders are relatively more profitable, then policy 3 may reduce profits compared to policy 1.

For a given S, we model policy 3 as a birth and death process, with states $\{\ldots,-1,0,1,\ldots,S\}$. We assume that only ground orders backlog, and thus the death process rate is λ for states 1 through S, and λ_g for non-positive states. For a given S, the stationary probabilities are:

$$p_0^{(3)} = \left\{ \frac{1}{1-\rho_g} + \frac{\rho^{-S}-1}{1-\rho} \right\}^{-1}, \text{ and} \tag{6}$$

$$p_i^{(3)} = \begin{cases} p_0^{(3)} \rho^{-i}, & i \geq 0, \\ p_0^{(3)} \rho_g^{-i}, & i \leq 0. \end{cases} \tag{7}$$

Here $D_{k,l}^{(3)} = 0$ if $k \neq l$, $D_{a,a}^{(3)}(S) = \lambda_a p_0^{(3)} \dfrac{\rho^{-S} - 1}{1 - \rho}$, $D_{g,g} = \lambda_g$, $B^{(3)}(S) = p_0^{(3)} \dfrac{\rho_g}{(1-\rho_g)^2}$, and

$$I^{(3)}(S) = p_0^{(3)} \frac{\rho}{(1-\rho)^2} \left[1 - \rho^{-S} \left(1 + S\left(1 - \rho^{-1}\right)\right)\right].$$

The next policy reserves some units for the higher-margin customers by backlogging lower-margin orders when inventory levels are low.

Policy 4: If inventory drops to S_a or lower, ship only (higher-margin) air orders; backlog ground orders and ship them by ground after inventory returns to S_a.

For computational purposes, we limit the number of orders that can be backlogged to *K*. This limit might be interpreted as the size of the backlog beyond which long lead times cause customers to balk. In the various experiments we find the optimal value for *K*. The Markov chain associated with this policy is shown in figure 2. A state is defined as (*i*, *u*), where *i* is the inventory level and *u* is the backlogging level. The solution for the stationary probability distribution of the Markov chain in figure 2 is calculated numerically by solving a system of equations as follows. Let **Q** denote the chain's rate matrix, **p** the vector of $p_{i,u}^{(4)}$, and **0** the vector of zeros. Then, $p_{i,u}^{(4)}$ is the solution to the following system of equations:

$$\begin{aligned} &\mathbf{pQ} = \mathbf{0}, \\ &\sum_{i=0}^{S} \sum_{u=0}^{K} p_{i,u}^{(4)} = 1. \end{aligned} \tag{8}$$

Then, $D_{k,l}^{(4)} = 0$ if $k \neq l$, $D_{g,g}^{(4)}(S, S_a, K) = \lambda_g \left[\sum_{i=S_a+1}^{S} p_{i,0}^{(4)} + \sum_{i=0}^{S_a} \sum_{u=1}^{K-1} p_{i,u}^{(3)} \right]$,

$D_{a,a}^{(4)}(S, S_a, K) = \lambda_a \left(1 - \sum_{u=0}^{K} p_{0,u}^{(4)} \right)$, $I^{(4)}(S, S_a, K) = \sum_{i=0}^{S} \sum_{u=0}^{K} i p_{i,u}^{(4)}$, and $B^{(4)}(S, S_a, K) = \sum_{i=0}^{S_a} \sum_{u=1}^{K} u p_{i,u}^{(4)}$.

Backlogging ground orders and shipping them by ground may create undesirable lead times. A more service-oriented policy would be the following.

Policy 5: If inventory drops to S_a or lower, ship only air orders; backlog ground orders and ship them by air after inventory returns to S_a.

This policy may be attractive if the cost of air shipment is not significantly greater than ground shipment. The Markov chain associated with this policy is the same as in figure 2, and thus the stationary probabilities can be computed as in (8). The higher shipping costs are factored in through the following terms: $D_{a,a}^{(5)}(\cdot) = D_{a,a}^{(4)}(\cdot)$, $D_{g,g}^{(5)}(S, S_a, K) = \lambda_g \sum_{i=S_a+1}^{S} p_{i,0}^{(5)}$, $D_{a,g}^{(5)} = 0$,

$D_{g,a}^{(5)}(S, S_a, K) = \lambda_g \sum_{i=0}^{S_a} \sum_{u=1}^{K-1} p_{i,u}^{(5)}$, $I^{(5)}(\cdot) = I^{(4)}(\cdot)$, and $B^{(5)}(\cdot) = B^{(4)}(\cdot)$.

Results

Our model allows us to explore inventory flexibility nuances available to firms that ship products directly to customers. We analyzed the performance of the five policies using a full-factorial experimental design. We do not present the details of the experiment here, but provide a brief summary of the key results.

Given the customers' willingness to wait for delivery, it is possible to increase profitability by using policies that take advantage of differences in customer profitability (e.g., air versus ground shipment) as well as differences in the costs and speeds of different transportation. For lost-sales situations, in some cases it is optimal to cut off shipments to lower-margin customers if inventory drops below a threshold in order to reserve inventory for possible orders for higher-margin customers. Even if customers are willing to backlog their order, in some cases it may make sense not to allow them to backlog if the orders more likely to backlog are lower-margin ones. Alternatively, these lower-margin orders might be backlogged before inventory is depleted in order to reserve units for possible orders from high-margin customers.

In lost-sales or backlogging scenarios, the use of air shipments to meet demand for ground orders may increase profits by allowing the shipment to be delayed without being late. The ground customer may expect the order to arrive in a certain number of days, such as five. If the order is backlogged for four days and then shipped by air, the customer is satisfied while the company has been able to buy time to get past an especially busy period.

Conclusion

As e-commerce matures and concerns over profitability increase, these more subtle inventory flexibility opportunities may allow some e-commerce firms to operate more efficiently, and might provide a means for sustaining a cost advantage over other e-commerce competitors. A simple modeling framework from queuing theory allows analysis of some of the inventory nuances that arise in e-business supply chains.

References

Alderson, W.. "Marketing efficiency and the principle of postponement". *Cost and Profit Outlook*, Vol 3 (1950).

Beckmann, M.. "An inventory model for arbitrary interval and quantity distributions of demand". *Management Science*, Vol. 8 (1961). pp. 35-57.

Buzacott, J., and S. Shanthikumar. *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, NJ: Prentice Hall, 1993.

Cachon, G. "Competitive and cooperative inventory management in a two-echelon supply chain with lost sales". Working Paper, The Fuqua School of Business, Duke University, Durham, NC. (1999).

Gross, D. and C. Harris. "On one-for-one ordering inventory policies with state-dependent leadtimes". *Operations Research*, Vol. 19. (1971). pp. 735-760.

Hadley, G. and T. Whitin. *Analysis of Inventory Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1963.

Morse, P.. *Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply*. New York, NY: John Wiley and Sons, 1958.

Figures

Figure 1: Markov chain associated with policy 2

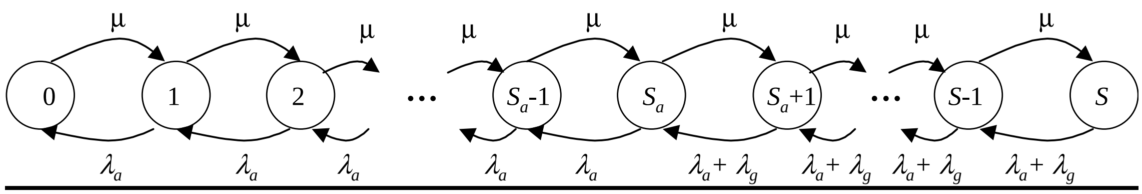

Figure 2: Markov chain associated with policies 4 and 5