Routing and Inventory Allocation Policies in Multi-Echelon Distribution Systems

Sangwook Park
College of Business Administration
Seoul National University

1. Introduction

This paper reviews the previous research in the area of logistics systems, especially in the area of the multi-echelon distribution system with stochastic demand. Many researches in this area emphasize the value of real-time information (which is more readily available nowadays by use of EDS or satellite systems) in distribution-related decisions. Also a great deal of efforts has been taken to study the risk-pooling effect of the various distribution policies in the multi-echelon systems. The objectives of this review are two-folded; (1) to help readers to understand the research paradigm in the multi-echelon area, and (2) to help them to find future research topics not yet explored.

Logistics is a very important component of the economy and includes a wide variety of managerial activities. There has been growing interest in logistics systems since World War II, when large
quantities of men and materials needed to be moved across large distances in a relatively short time. We can attribute this growing interest to various reasons. First, logistics costs (both at the company level and at the national level) are huge. At the level of individual firms, the distribution costs represent 10 to 30 percent of the total costs of goods sold (Robeson and Copacino (1994)). Nationally, logistics costs have been estimated at about 21 percent of the gross national product (Ballou (1987)). Second, the logistical considerations are crucial in determining a firm's strategic priority; that is, distribution policies of a firm determine its response time to changing market conditions. Lastly, the latest developments in economy require different logistics systems. Examples of such developments include (i) increased transportation costs as a result of rising fuel and labor costs, (ii) escalation in the inventory-holding costs, and (iii) the emergence of computer-integrated manufacturing systems (CIM) and Just-in-Time production system (JIT). These reasons have accelerated research effort in logistics.

In response to these changes, a vast body of research has appeared in the area of logistics. However, most of these works have focused on optimizing the individual functions of the logistics system such as transportation, inventory allocation, location, etc., which could result in a sizable degree of suboptimality in the operational policies.
Therefore, there is a great need for efforts to integrate three of the logistical functions: system replenishment, delivery routing, and inventory allocation. The systematic review of the previous works on the multi-echelon distribution system will give readers opportunities to understand the major results of the research efforts in this area and help them to initiate their own works. Since World War II, there has been a large body of research in logistics-related activities. We selectively review that work that relates to distribution systems or inventory-routing logistics systems operating in a periodic-review environment.

2. Single-Level, Periodic-Review Inventory Systems

The work on single-level inventory systems appeared in early 1950's and an excellent review is available in Aggarwal (1974). Arrow et al. (1951) model a classical single-period problem which maximizes single-period expected profit. Under specific conditions (basically the convexity of total expected purchasing, holding, and backorder costs), Arrow et al. show that the optimal replenishment policy is a base-stock policy.

Arrow, Karlin, and Scarf (1958) extend the single-period model in the following ways: (i) they consider a finite horizon of periods and (ii) they allow for a fixed delivery leadtime between the
order placement and arrival. The model is formulated as a stochastic dynamic program. The key assumption is complete backordering of the unfilled demand, which yields optimality of base-stock policies in each period. Veinott (1965) showed much more: He made an assumption on the end-of-problem net inventory: If there are leftovers at the end of the last period, then it has value of the original purchasing cost per unit. If there are backorders outstanding, then they are met by purchasing additional units at the same purchasing cost per unit. Under this assumption, the optimal policy is a stationary myopic base-stock policy. Furthermore, the optimal base-stock level can be found as the solution to a single-period "newsboy" problem. Scarf (1960) analyzed the case where the purchasing cost function is of the following type:

\[
c(z) = \begin{cases} 
0 & \text{if } z = 0 \\
K + cz & \text{if } z > 0 
\end{cases}
\]

He showed the optimality of \((S,s)\) policy in this case: At the beginning of each period, stock-level is checked. If the level is above \(s\), no new order is placed. If, however, the level is below \(s\), then order up to \(S\) is placed.

3. Allocation Assumption and Risk-Pooling

Two of the most important concepts in multi-echelon literature
Routing and Inventory Allocation Policies in Multi-Echelon Distribution Systems

are closely related to our research: the Allocation Assumption and Risk-Pooling. Each work addressed below uses different terminology for the upper level and lower levels in the system. To be consistent with our terminology in this dissertation, we will use the term 'warehouse' for the upper level and 'retailers' for the lower levels regardless of the original terminology.

Allocation Assumption

The allocation assumption is frequently used in multi-echelon optimization models to obtain some analytical tractability of problem (see Eppen and Schrage (Schwarz, 1981)). The allocation assumption relaxes the non-negativity constraints on allocations; that is, it permits negative allocations to any given retailer provided that the sum of the allocations to all the retailers is equal to a system-replenishment quantity. Eppen and Schrage define the allocation assumption as follows:

"In each allocation period $t$, the warehouse receives sufficient goods from the supplier so that each retailer can be allocated goods in sufficient quantity to ensure that the probability of stockout in period $t + \lambda + m - 1$ is the same at all retailers."

Here is $\lambda$ the delivery leadtime from the warehouse to each retailer and $m$ is the interval between successive allocations. Under the
allocation assumption, if all retailers are identical in unit backorder cost, unit holding cost, and delivery leadtime, then, the optimal allocation brings each retailer to the same fractile of the $\lambda + m$ period demand distribution. In particular, when the demand at retailer $i$ is normally-distributed with mean $\mu_i$ and standard deviation $\sigma_i$, the optimal allocation equalizes $\frac{Y_i - \mu_i(m + \lambda)}{\sigma_i \sqrt{m + \lambda}}$ for each retailer in the system, where $Y_i$ is the net inventory of retailer $i$ at the time of the allocation decision plus the amount allocated to that retailer. Eppen and Schrage show that the probability of allocation assumption holding true given that it held in the previous period decreases progressively as the coefficient of variation of the demands at retailers increases.

The Risk-Pooling Phenomenon

Primarily, there are two kinds of risk-pooling phenomenon that occur in the context of the distribution system: (i) risk-pooling through the centralization of demand, and (ii) risk-pooling over the outside supplier's leadtime.

Risk-Pooling Through Centralization of Demand: This kind of risk-pooling occurs because the random demands in any given period
at different locations are perceived by the system as a single demand equal to the sum of these random variables.

Eppen (1979) quantifies the cost implications of this kind of risk-pooling. He compared two systems: (i) a completely decentralized system that maintains a separate inventory to meet the demand from each source and (ii) a completely centralized system that meets all demands from one central warehouse. The analysis assumes identical, normally-distributed demands, identical shortage and holding costs per unit across the retailers, and a periodic-review system. It is shown that the expected cost incurred by the centralized and the non-centralized systems are equal when the demands are perfectly correlated, but the expected cost for the centralized system decreases as this correlation decreases. In particular, when the demands at each retailer are totally uncorrelated (i.e., totally independent), the expected cost of the centralized system is reduced by a factor of $\sqrt{N}$ as compared to the expected cost of the decentralized system. Subsequently, Schwarz (1981) investigates a system of identical retailers in a continuous-review, centralized distribution system and shows the validity of the $\sqrt{N}$ effect for such a system when the demands are independent across retailers.

Risk-Pooling Over the Outside Supplier's Leadtime: This type of
risk-pooling occurs due to the random demands convoluting during the supplier's leadtime.

This phenomenon was first noted by Simpson (1959), and later by Schwarz (1989). Schwarz constructs two systems: (i) in the decentralized system (System 1), retailers place an order directly to the outside supplier with no opportunity for risk-pooling. The leadtime for the order arrival at the retailer is $LS + Ltr$ where $LS$ is the supplier's processing time and $Ltr$ is the delivery leadtime from supplier to retailer. (ii) in the centralized system (System 2), the system order is placed and allocated through a central warehouse. The leadtime for the order to arrive at the retailer is $LS + Ltw + Lpw + Ltr$ where $Lpw$ is the processing time at the warehouse and $Ltw$ is the leadtime needed for routing the order through the warehouse.

The overall reduction in variance of the net-inventory process of System 2 compared with that of System 1 is denoted 'Risk-Pooling Incentive' or RPI. Furthermore, System 2 will incur higher holding cost compared to System 1 due to the extra internal leadtime $Lpw + Ltw$ which is denoted as the "Price of Risk-Pooling". Each of these measures of risk-pooling can be evaluated in terms of the extra leadtime that System 2 needs to have to break-even with System 1 for the same specified service level and the same safety-stock level (the safety-stock break-even leadtime) or for the same specified
service level and the same safety-stock holding costs including extra pipeline cost in case of System 2 (the inventory-cost break-even leadtime). The break-even leadtimes provide a measure of the value of risk-pooling. The significant findings are as follows:

(i) Pipeline inventory-holding cost has significant impact on the value of risk-pooling: when the inventory costs in the extra pipeline can be ignored, the extra leadtime that would break-even with the performance of System 1 is quite large. However, when the inventory costs in the extra pipeline can not be ignored, the break-even leadtimes are small. Equivalently, the value of risk-pooling is small.

(ii) Holding-cost break-even leadtimes decreases; as $N$, the number of retailers, decreases; as $L_{tr}$ increases; as $\frac{\sigma}{\mu}$ decrease; as $H$, the number of time periods per cycle, increases.

(iii) For System 2 to outperform System 1, $L_{pw}$ must be quite small compared to $L_s$, and $L_{tw}$ may be considerably larger than $L_s$.

Schwarz and Weng (1990) further analyze the risk-pooling value of System 2. In this work, the basic configurations of System 1 and 2 are retained but leadtimes are modeled as Poisson-distributed.
The main findings of this study are:

(i) Value of risk-pooling, as measured by the safety-stock break-even leadtime, remains unchanged when the leadtimes are Poisson-distributed.

(ii) Value of risk-pooling as measured by the holding-cost break-even leadtimes, is considerably larger in cases of Poisson-distributed leadtimes.

(iii) For both the deterministic as well as the Poisson-distributed leadtimes, the holding-cost break-even leadtimes are insensitive to supplier-to-warehouse leadtime but sensitive to warehouse-to-retailer leadtime.

(iv) Holding-cost break-even leadtimes are insensitive to the retailer demand uncertainties.

4. Static Allocation Policies

Many articles deal with the issue of system replenishment and inventory allocation for centralized distribution systems following static allocation policies. Simpson (1959) deals with the issue of static allocation of a given quantity amongst several retailers for two distinct scenarios: the emergency replenishment case (an emergency replenishment is ordered every time the inventory level at a retailer hits a predetermined emergency trigger level) and the emergency
non-replenishment case. He shows that in both cases, some appropriate function of the system parameters is equalized across the retailers. The author does not consider the possibility that for a given Q, the proposed equalization may not be feasible.

Clark and Scarf (1960) develop optimal replenishment policies for each stage of a serial system. The assumptions of this work are as follows:

(i) Demand occurs at the lowest echelon.
(ii) Purchasing cost and transportation cost between the stages are linear.
(iii) Holding and shortage costs are convex on echelon inventory.
(iv) Excess demand is completely backordered.
(v) Delivery to any stage is instantaneous, but amount shipped can not exceed on-hand inventory.

Under these assumptions, they proved that the optimal policy for the highest stage is a base-stock policy. The result can not be extended for multiple successors because of the possibility of "out of balance" situations in retailers' inventories.

Eppen and Schrage (Schwarz, 1981) model a centralized distribution system consisting of an outside supplier, a warehouse, and several retailers (respectively called supplier, depot, and the
warehouses in their model). Three different modes of operation are considered: (i) the centralized system: replenishment and allocation functions are performed at the supplier's site in a centralized manner, (ii) the depot system: allocation and replenishment is done centrally at the depot located between the supplier and the retailers. (iii) the decentralized system: each retailer directly and independently places orders to the outside supplier. The depot model allows flexibility in the replenishment policy since the orders can be placed every period or every \( m \) periods. The following features characterize their model:

(i) Proportional holding and shortage costs which are also identical across retailers.

(ii) Stochastic, normally-distributed, independent demands at retailers. The demands distributions are not necessarily identical across the retailers.

(iii) Stationary demands and costs.

(iv) Identical delivery leadtimes between the supplier and each retailer.

(v) System orders up to a base stock at the beginning of each period/cycle.

(vi) The warehouse holds no inventory.

The following are the key assumptions of the model:
Demand is backordered if not met in any given period.

Service level is sufficiently high to limit backorders only in the last period of each cycle.

Myopic allocation - minimize the expected cost of each $m$-period cycle.

Allocation assumption - eliminates "out of balance" situations described by Clark and Scarf.

The following are the significant results/findings of the analysis:

(i) A computationally simple method for determining allocations and replenishments.

(ii) The total inventory on-hand plus on-order is greater for the decentralized system than for the depot system for the same total leadtime between the supplier and any retailer. In turn, the total inventory on-hand plus on-order is greater for the depot system than for the centralized system.

(iii) The expected inventory cost for the decentralized system is greater than for the depot system. In turn, the expected inventory cost for the depot system is greater than for the centralized system.
Federgruen and Zipkin (1984a) relax several assumptions of the Eppen and Schrage model and construct a more general model as follows:

(i) Marginal holding and backorder costs are not necessarily identical across the retailers.

(ii) Stochastic demands, while normally-distributed, are not assumed to be stationary across periods. Further, the analysis allows for some other distributions of demands such as Gamma or Weibull.

(iii) The problem horizon can be finite or infinite.

The equality of delivery leadtimes of the retailers is still the limiting feature of their model although can be relaxed. Also, the allocation assumption is assumed to hold with probability 1.0. The key results of the model are:

(i) The system can be reduced to a single-location, newsboy-type model for the purpose of computing the replenishment policy. Considering the very general nature of their parameterization, this result is particularly significant. This implies optimality of base-stock policies which is actually an assumption in Eppen and Schrage (Schwarz, 1981).

(ii) The Myopic Allocation Assumption is shown to be
non-restrictive for systems with relatively low coefficient of variations.

Jönsson and Silver (1987a) consider a centralized distribution system comprised of one warehouse and several retailers. The objective is to determine the optimal initial system stock that yields a specified service-level over a replenishment cycle, where the service level is defined as 
\[
1 - \frac{B_{H-1} + B_H}{D},
\]
\( (B_{H-1} + B_H) \) is the amount of backorders in the last two periods of the replenishment cycle, and \( D \) is the average cycle demand. Under the basic policy, the warehouse operates as follows: It orders some stock (instantly available) \( I_0 \) at the beginning of the cycle and \( (I_0 - I_c) \) is optimally distributed at the beginning of the cycle to maximize the service level. Under the allocation assumption and identical retailers, this requires equal allocation to each retailer. The remaining stock \( I_c \) is then optimally allocated at the beginning of penultimate period of the cycle (period \( H-1 \)). The performance of this basic policy is compared with two extreme cases: (i) Ship-All policy: The system distributes the entire stock available at the beginning of the cycle amongst the retailers in an optimal manner. (ii) Extreme Push Policy in which the entire stock is redistributed at the second allocation opportunity. The key result is
that the performance of the basic policy is vastly superior to the ship-all policy and not too inferior to the complete redistribution policy.

In a related work, Jönsson and Silver (1987b) investigate the effect of total redistribution of inventory among retailers one period before the end of the replenishment cycle, and compare the expected backorders of this system with that of the system without redistribution. The key assumptions involve a high service level assumption that limits backorders in the last two periods of the cycle and the allocation assumption. This redistribution is intended to achieve the benefits of warehouse risk-pooling between system replenishments. Computational tests show that the system with redistribution can provide the same service level (as the system without redistribution) with a considerably reduced inventory investment.

McGavin, Schwarz, and Ward (1993) construct a model for a system of one warehouse and $N$ identical retailers to determine warehouse inventory-allocation policies which minimize system lost sales per retailer between system replenishments. An allocation policy is specified by: (i) the number of withdrawals from warehouse stock; (ii) the intervals between successive withdrawals; (iii) the quantity of stock to be withdrawn from the warehouse in each withdrawal; and
(iv) the division of withdrawn stock among the retailers. They show that in case of two withdrawals, available stock in each interval should be allocated to "balance" retailer inventories. They also develop an infinite-retailer model and use it to determine two-interval allocation heuristics for $N$-retailer systems. Simulation tests suggest that the infinite-retailer heuristic policies are near-optimal for as few as two retailers, and that the risk-pooling benefits of allocation policies with two well-chosen intervals are comparable to those of base-stock policies with four equal intervals.

Graves (1996) introduces a new scheme for allocating stock in short supply in multi-echelon systems where each site in the system orders at preset times according to an order-up-to policy. The new allocation scheme is called the "virtual allocation" and permits significant tractability. Under the virtual allocation, whenever a unit demand occurs, each site on the supply chain commits or reserves a unit of its inventory, if available, to replenish the downstream site. He applies the model to a set of test problems for two-echelon systems and finds that both the central warehouse and the retailer sites should hold safety stock, but that most of the safety stock should be at the retailer sites. Consequently, the central warehouse will stock out with high probability. Furthermore, he shows that the virtual allocation rule is near-optimal for the set of test problems.
5. Dynamic Allocation Policies

Kumar, Schwarz, and Ward (1995) examine static and dynamic policies for replenishing and allocating inventories amongst $N$ retailers located along a fixed-delivery route. Each retailer faces independent, normally-distributed period demand and incurs a proportional inventory-holding or backorder costs on end-of-period net-inventory. A warehouse places a system-replenishment order every $m$ periods which is received after a fixed leadtime. Immediately upon receipt, a delivery vehicle leaves the warehouse with the system-replenishment quantity and travels to the retailers along a fixed route with fixed leadtimes between successive retailers. The warehouse holds no inventory. Under the static allocation policy, allocations are determined for all retailers simultaneously at the moment the delivery vehicle leaves the warehouse. Under the dynamic allocation policy, allocations are determined sequentially upon arrival of the delivery vehicle at each retailer. The objective is to minimize the sum of total expected inventory-holding and backorder costs per cycle under the two types of allocations.

Their major analytical results, under appropriate dynamic (static) allocation assumptions, are: (i) optimal allocations under each policy involve bringing each retailer's "normalized-inventory" to a
corresponding "normalized" system inventory; (ii) optimal system replenishments are base-stock policies; (iii) the minimum expected cost per cycle of dynamic (static) policy can be derived from an equivalent dynamic (static) "composite retailer". Given this, they prove that the "Risk-Pooling Incentive", a simple measure of the benefit from adopting dynamic allocation policies, is always positive. Simulation tests confirm that dynamic allocation policies yield lower costs than static policies, regardless of whether or not their respective allocation assumptions are valid. However, the magnitude of the cost savings is very sensitive to some system parameters.

Park et al. (1998) models a dynamic delivery-routing and allocation problem in a one-warehouse $N$-retailer distribution system operating in a periodic-review mode to study the cost-reduction effect of dynamic routing. With dynamic routing, the delivery vehicle travels along a route that is determined sequentially. In particular, just before the delivery vehicle leaves the warehouse or each retailer, management decides which retailer to visit next, based on the inventory status of the subsystem of retailers not yet visited. They first prove that the optimal routing policy in a one-warehouse $N$-retailer "symmetric" system is to go to the retailer with the least inventory first (LIF). They formulate the finite horizon problem as a dynamic-programming problem and show that under the "allocation
assumption", myopic allocation is optimal. The myopic allocation problem is not easy to solve even in the two-retailer case. Several important properties of the optimal myopic allocation for the two-retailer case, including the first-order optimality condition, are presented. Through a numerical study, they show that the benefit of using dynamic routing is significant in the "medium-to-large" demand variance cases. Also, some heuristics for allocation are shown to be very efficient. They also show the universality of the first-order optimality condition of the system-replenishment problem in the two-retailer case. A numerical study suggests that using the optimal system-replenishment policy for the fixed-route case is a good heuristic.

6. Combined Inventory-Routing Models

There are many research works in the area of integrated logistics system in general and in inventory and routing in particular. While the studies involving the combined modeling of inventory allocation and delivery routing have been few and far between up until 1982, there has been acceleration of interest since 1982. Readers may like to refer to Bodin et al. (1983) and Golden and Assad (1986) for a survey of inventory-routing literature pertaining to deterministic-demand systems. A common point in all these studies has been that the
interactions between two modules (allocation and routing) of the logistics system are significant enough to warrant integrated modeling.

Federgruen and Zipkin (1984b) analyze a combined vehicle-routing and inventory-allocation problem with stochastic demand. In their model, both allocation and routing are static; that is, the route for each vehicle and allocation for each location once determined are fixed. They assume (i) zero outside-supplier leadtime; (ii) instantaneous delivery to the retailers; and (iii) a one-period planning horizon. Their objective is to determine a joint route-allocation strategy that minimizes the sum of expected inventory cost and transportation cost for the entire system. The interdependence between routing and inventory allocation arises from the fact that while the optimal allocation may prescribe a positive allocation to some retailer, the cost of routing the vehicle through that retailer may exceed the savings achieved by the allocation. Another source of interdependence is the vehicle capacities. Overall savings of 5-6% is reported, accruing from the joint consideration of the inventory-allocation and routing decisions. Anily and Federgruen (1990) study the dynamic vehicle-routing and inventory problem in one-warehouse multi-retailer systems when demand is deterministic.

7. Concluding Remarks
We reviewed a wide range of works on the multi-echelon distribution system. To summarize, the early researches focus on either finding a form of optimality (either exact or approximated) for well-known problems, while the latter works explain the risk-pooling effects of the various system designs, which include using the most up-to-date information on inventory levels at various locations. We hope that this review provide readers with the big picture of the research efforts on the multi-echelon system and help them to start their own research.
REFERENCE


Eppen, G.D., "Note-Effects of Centralization on Expected Costs in a


Jackson, P., Stock Allocation in a Two-Echelon Distribution System
Routing and Inventory Allocation Policies in Multi-Echelon Distribution Systems


Park, S., Schwarz, L. B., and Ward, J. E., Dynamic Routing and


