Inventory Deployment and Market Area Segmentation in a Two-Echelon Distribution Network Design

by

Nebibe Varol

B.Sc. Industrial Engineering (2002)
Boğaziçi University, Istanbul, Turkey

Submitted to the Sloan School of Management in Partial Fulfillment of the Requirements for the Degree of Master of Science in Operations Research at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY June 2004

© Massachusetts Institute of Technology 2004. All rights reserved.

Signature of Author.......................................................... Sloan School of Management
May 7, 2004

Certified by........................................................................
Donald B. Rosenfield
Senior Lecturer, Sloan School of Management
Director, Leaders for Manufacturing Fellows Program
Thesis Supervisor

Accepted by................................................................. Jim Orlin
Co-director, Operations Research Center
Inventory Deployment and Market Area Segmentation in a Two-Echelon Distribution Network Design

by

Nebibe Varol

Submitted to the Sloan School of Management on May 7, 2004
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Operations Research

Abstract

Most of the logistics systems involve a multi-level distribution system structure due to value added by a multi-level configuration. Interactions of these levels, i.e. echelons, should be considered while making strategic decisions regarding the choice of the size, number and location of stocking sites as well as the tactical decision regarding the choice of inventory policy to be used. We analyze a two-echelon distribution network to characterize the market segmentation of each echelon and inventory deployment between the two-levels. Allocation of stock under a stochastic demand structure is considered simultaneously with warehousing and transportation decisions, which is an extension of the General Optimal Market Area (GOMA) Model developed by Erlenkotter. The distribution of inventory is investigated under different stock policies and the sensitivity of this distribution to various system parameters is analyzed.

Thesis Supervisor: Donald B. Rosenfield
Title: Senior Lecturer, Sloan School of Management
Director, Leaders for Manufacturing Fellows Program
Acknowledgements

I would like to thank my advisor, Donald Rosenfield, for his valuable guidance, advice and support throughout my experience at MIT. I feel that having the opportunity to work with him was a great honor for me. Working with him was a really rewarding experience through which I gained a lot.

Many thanks to the ORC staff, students and faculty members for making the center and the program a welcoming, motivating, interesting and friendly place to live, and for providing me the opportunity to study in this program. My sincere thanks to Prof. Richard Larson and Prof. Orlin for their encouragement and support throughout my studies at the Operations Research Center.

I wish to thank my family and all of my friends for helping, motivating and teaching me along the way. Their love, time, and support have made these two years an unforgettable learning experience in my life. I will always be grateful for what they have brought into my life in such a short time.
Contents

Chapter 1 .................................................................................................................. 11
Introduction .......................................................................................................... 11
  1.1 Major Decisions in Logistics Network Design .............................................. 12
  1.1.1 Facility Location and Allocation ............................................................ 13
  1.1.2 Inventory Policy ..................................................................................... 13
  1.1.3 Transportation and Delivery Means ....................................................... 14
  1.1.4 Customer Service Standards .................................................................. 14
  1.2 Customer Service ....................................................................................... 16
  1.3 The Total Cost Concept ............................................................................ 17
  1.4 Thesis Outline and Contribution ................................................................ 18

Chapter 2 .............................................................................................................. 21
General Optimal Market Area (GOMA) Model .................................................. 21
  2.1 Two-Echelon GOMA Model Incorporating Inventory Costs ..................... 24
  2.2 Literature Review ....................................................................................... 27
  2.2.1 Facility Location .................................................................................... 27
  2.2.2 Multi-Stage Systems – Lot Sizing and Inventory Policies ....................... 28
  2.2.3 Two Echelon Designs ............................................................................ 32
  2.2.4 Distribution System Design .................................................................. 40

Chapter 3 .............................................................................................................. 43
Logistics Cost Components ............................................................................... 43
  3.1 Transportation Costs ............................................................................... 43
  3.2 Distribution Warehousing and Warehousing Costs .................................... 46
  3.3 Inventory Costs and Replenishment Systems ............................................ 52
  3.3.1 Inventory Policies ............................................................................... 55
  3.3.2 Continuous versus Periodic Review Policies under Probabilistic Demand .. 56
  3.3.2.1 Continuous Review ........................................................................ 59
  3.3.2.2 Periodic Review .............................................................................. 64
  Order-Point, Order-Quantity (s, Q) System .................................................. 59
  Order-Point, Order-Up-to-Level (s, S) System ............................................. 63
  Periodic-Review, Order-up-to-Level (R, S) System ..................................... 64
  (R, s, S) System ......................................................................................... 68

Chapter 4 .............................................................................................................. 71
Strategic Decisions in a Two-Echelon Network ............................................... 71
  4.1 Inventory in a Two-Echelon Distribution System ....................................... 71
  4.1.1 Continuous Review Approach .............................................................. 75
  4.1.2 Periodic Review Approach .................................................................. 77
  4.1.2.1 Echelon Stock Approach ................................................................. 79
  4.1.2.2 Installation Stock Approach ............................................................. 88
  4.2 Total Logistics Cost .................................................................................. 91
  4.2.1 Single-Echelon Distribution System vs. Two-Echelon Distribution System 92
  4.2.2 Total Logistics Cost in the Two-Echelon Network .................................. 98
  Logistics Costs with Respect to k in a Two-echelon Network with Given Market Areas .... 99
  Logistics Costs in a Two-echelon Network with Given Market Areas for the DC ... 103
  Logistics Costs in a Two-echelon Network with Given Market Areas for the WH ... 109
Chapter 5 .......................................................... 117
Conclusions and Further Research Opportunities .............. 117
Appendix ...................................................................... 119
   I. The Economic Order Quantity (EOQ) Model .............. 119
   II. Expected Shortages per Cycle (Continuous Review) ... 120
   III. Expected Distribution Center to Warehouse Distance ... 121
Bibliography ............................................................... 125
List of Figures

Figure 1-1: A Multi-Stage Manufacturing and Distribution System ................................................. 11
Figure 1-2: Major Decisions in Logistics Strategic Planning .......................................................... 15
Figure 1-3: A Simplified Illustration of the Total Cost Concept ..................................................... 17
Figure 2-1: Flow of goods in the Two-Echelon System ................................................................. 25
Figure 2-2: Market Areas in the Two-Echelon System ................................................................. 26
Figure 3-1: Relationship between Number of Facilities and Inventory Cost .................................. 47
Figure 3-2: Relationship between Number of Facilities and Transportation Cost ......................... 48
Figure 3-3: Relationship between Number of Facilities and Facility Costs ................................... 48
Figure 3-4: Response Time and Total Logistics Cost vs. Number of Facilities ......................... 48
Figure 3-5: (s, Q) Continuous Review Policy ............................................................................... 60
Figure 3-6: Average Behavior of On-hand Stock in an (s, Q) System ......................................... 61
Figure 3-7: The (R, S) Periodic Review System ........................................................................... 65
Figure 3-8: Average Behavior of On-hand Stock in (R, S) System ............................................. 66
Figure 4-1: Two-level Arborescent System ................................................................................. 71
Figure 4-2: Two-Echelon System and the Lead Times within the System ................................. 73
Figure 4-3: Variance of Lead Time for the Warehouses ............................................................... 80
Figure 4-4: The Behavior of Echelon-Stock at the Warehouses ................................................... 82
Figure 4-5: Expected Behavior of Echelon 2 Aggregate Stock Level and Echelon 1 Stock Level (k=2) .................................................................................................................. 85
Figure 4-6: Stock-out Probabilities for k = 7 and k = 15 ............................................................... 87
Figure 4-7: The Expected behavior of Installation Stock at the DCs ........................................... 90
Figure 4-8: Total Logistics Cost vs. Market Area in a Single-Echelon System ......................... 96
Figure 4-9: Logistics Cost Components vs. Number of Facilities in a Single-Echelon System .... 96
Figure 4-10: Logistics Cost Components vs. Number of Facilities in a Single-Echelon System ... 97
Figure 4-11: Logistics Cost Components with Respect to k .......................................................... 100
Figure 4-12: Inventory and Total Logistics Cost per unit when K_{DC} is increased ................. 101
Figure 4-13: Inventory and Total Logistics Cost per unit when F_{DC} is increased .................. 103
Figure 4-14: Logistics Cost Components When Echelon 1 Areas are fixed .............................. 104
Figure 4-15: Logistics Cost Components When Echelon 1 Areas are fixed and the uncertainty in demand, i.e. σ, is increased ................................................................. 106
Figure 4-16: Logistics Cost Components When Echelon 1 Areas are fixed and the Review Period R is increased .............................................................................................. 107
Figure 4-17: Logistics Cost Components When Echelon 1 Areas are fixed and the order cost for the warehouses is increased to 100........................................................................ 109
Figure 4-18: Logistics Cost Components vs. N (number of warehouses) when Echelon 2 Areas A_{w1} are fixed ............................................................................................... 111
Figure 4-19: Logistics Cost Components vs. N (when A_{w1} is fixed and R is increased) ......... 112
Table 4-1: Sensitivity Analysis with Respect to Parameters in a Two-Echelon Network.......... 115
Figure A-1: Inventory versus time in the EOQ Model .................................................................... 119
Figure A-2: Two-echelon Configuration ...................................................................................... 121
List of Tables

Table 4-1: Sensitivity Analysis with Respect to Parameters in a Two-Echelon Network 115
Chapter 1

Introduction

Logistics strategic planning of a multi-stage manufacturing and distribution system, whose general structure is shown in Figure 1-1 involves the determination of the number and size of the facilities in the system, the spatial location of the plants and warehouses within the market area and the allocation of inventory among the stages so as to achieve the minimum cost configuration of the overall system.

![Diagram of a Multi-Stage Manufacturing and Distribution System](image)

Figure 1-1: A Multi-Stage Manufacturing and Distribution System

The importance of logistics planning of a production and distribution system as depicted above lies in the fact that an efficient design can serve as a weapon to gain and maintain a sustainable "competitive advantage", and thus, has a critical strategic role within a company. Competitive advantage is gained through differentiation of a firm from its competitors. The differentiation can take several forms including new product innovation, technological leadership, superior quality, differentiation along customer service lines, marketing superiority (e.g., brand identification), and overall cost leadership.

The elements of a logistics system must be mutually consistent and in accord with the company's strategy. There are different criteria that can be applied to match a firm's logistics system to its strategy. The most dominant three of these criteria can be summarized as follows:
1. **Efficiency**: The criterion of traditional cost minimization is efficiency, i.e. achieving a desired level of performance at the lowest total cost.

2. **Flexibility**: Flexibility is an important criterion for the innovators and requires the ability to adjust quickly to changes in volume, markets and products. The factors that make flexibility a major concern are: rapid growth or changing geographic patterns of demand, changing product specifications, and rapid pace of new product introduction.

3. **Service**: Service includes fast response, short order lead times, dependable delivery premises, broad line availability, and accurate order tracking. The most important strategic choices that should be made related to service are: the aspects of service on which the firm should concentrate and the level of services that the firm and customers consider adequate in light of the firm’s corporate and competitive goals [13].

“The total cost concept” states that designing a minimum cost configuration requires balancing the relevant costs within the system. Another important issue is the sensitivity of the optimal configuration to parameters in the system and the identification of the factors which drive the need for opening facilities at a particular stage and their effect on the system wide cost. The key costs that impact the number and location of the facilities in different stages are:

- Inventory holding costs,
- Fixed and variable facility costs,
- Facility inbound and outbound transportation costs.
- Warehouse storage and handling costs,
- Production/Purchase costs,
- Order processing costs [5].

### 1.1 Major Decisions in Logistics Network Design

The cost factors presented above show that logistics strategic planning is a multi-faceted decision problem. The major decision areas and their role within the logistics network and in the competitive strategy of the firm are as follows:
1.1.1 Facility Location and Allocation

Facilities within a production-distribution network are the places where the product is stored or assembled. The two major facilities in such a system are production and storage sites. Decisions regarding the geographic placement, capacity and flexibility of facilities create an outline for the logistics strategy. Fixing the number, size and location of the facilities and allocating the market demand to them determine the paths by which products are directed to the market.

The facility location problem should consider all product movements and related costs of movement from plant or vendors through stocking points to customer locations. Customer demand can be served directly from plants or vendors or through stocking points such as warehouses and distribution centers. The assignment of customer demand to the elements of the logistics system affects the total distribution cost. The essence of facility location planning is to find the least cost assignment [6].

Facilities and their capacities are a key driver of the system performance in terms of responsiveness and efficiency. For instance, economies of scale can be gained by producing or storing only at one location; in other words, centralization increases efficiency by reducing the cost. However, this detracts from the responsiveness of the company as customers may be located at different locations far from each other.

1.1.2 Inventory Policy

Inventory within a logistics system represents all raw materials, work in process and finished goods. Inventories are carried as a buffer between transportation, manufacturing and processing operations for an effective and economical system operation. Products can be stored at the source or at stocking points closer to the customer. Inventories permit the system to handle the unexpected variations in demand or output at any stage and allow the operation of manufacturing and transportation activities on a time cycle or with quantities adapted to their characteristics [37]. Among other drivers for holding inventory are uncertainty in the quality and quantity of the supply and delivery times, economies of scale offered by transportation and production.
Inventory policy determines how inventory levels within a logistics system are managed. Two such strategies are pushing inventories into the warehousing system and pulling them in by monitoring warehouse inventory levels. The choice of the inventory policy has an important role in the logistics strategy as it affects the facility location decision, and the overall efficiency and responsiveness of the logistics system.

Inventory has an important role in supporting a firm’s competitive strategy. For instance, if the firm’s competitive strategy requires high level of responsiveness, this can be achieved by locating large amounts of inventory close to the customer. On the other hand, a low-cost competitive strategy may increase its efficiency through centralized stocking.

1.1.3 Transportation and Delivery Means

Transportation includes the flow of raw materials from suppliers to the plants and the movement of products between different manufacturing stages, from plants to warehouses and from warehouses to customers. When the facility location decision is made, customer demand is allocated to stocking points in the system, which affects the utilization of the transportation equipment and the effectiveness of the distribution system. Transportation is also used as a driver to make the logistics network more efficient or responsive depending on the competitive strategy of the firm. Faster transportation increases the responsiveness; however, it also reduces the cost efficiency of the network [13].

1.1.4 Customer Service Standards

The customer service standards have a major effect on the design of the logistics system. Low-level service allows for a less costly design through centralized inventories at few locations and the use of less expensive transportation means. Customer service involves inventory availability, speed of delivery and the speed and accuracy in order-filling. As the service levels are increased, costs associated with these factors increase and the logistics costs rise disproportionately. For this reason, the determination of a proper customer service level based on customer needs and the competition within the market must be the starting point in the strategic planning of a logistics system.
Logistics System Design

Inventory Policy:
- What turnover ratio to maintain?
- Which products should be maintained at which stocking points?
- What is the best method of inventory control?
- What should be the product availability in inventory?
- Should push or pull based strategies be used?

Facility Location:
- What is the best number, location and size of plants and stocking points?
- Which plants should serve which stocking points?
- Which products should be shipped directly from plants and which through the warehousing system?

Transportation:
- Which customers should be served out of which stocking points?
- Which vehicle types should be assigned to which customers?
- Which modes of transportation should be used?

Figure 1-2: Major Decisions in Logistics Strategic Planning

The major decision factors in the design of the production and distribution network have conflicting cost patterns. Transportation versus inventory, production versus distribution, and customer service versus all logistics costs are in trade-off or in conflict with each other. For example, reducing the number of stocking points throughout the logistics system introduces a disproportionate reduction in the amount of inventory carried in the system, which is known as the "inventory consolidation effect". In general, the cost reduction associated with inventory consolidation is in trade-off with higher transportation costs. Analogous to the inventory consolidation, another significant economic force in logistics is "shipment consolidation", i.e. creating large shipments from potentially small ones. Therefore, the facility location-allocation decisions, the deployment of inventory in the system and the transportation means used are closely interrelated.

The main goal of an effective logistics design is to find a balance between these conflicting factors in a cost-effective manner. For this reason, the formulation of a mathematical model to represent this system should capture the interdependency of the major decisions and the related costs of facility location, inventory policy and transportation. In addition, the model should be able to incorporate the interdependency of different stages or echelons within the supply chain.

A classic approach that tries to capture the interdependency of the cost factors is the "Total Cost Concept", which we will briefly introduce. First, let’s review the “Customer Service” concept, which is the main driver of the competitive strategy, and thus, system cost.
1.2 Customer Service

Customer service is a primary objective of a logistics system, as well as an overall goal for a company. Customer service is a multi-faceted concept, embracing a series of activities to achieve product flows from manufacturers to ultimate users. Viewed as a marketing and corporate concept, customer service is seen as an “organizational goal” that contributes to “profit realization”.

“Product availability” is considered the primary factor of customer service. Getting the right products to the right place at the right time in a consistent, accurate and damage-free manner is the heart of customer service. The order cycle process, service flexibility, and information services represent other major customer service elements. Some of the important measures of product availability are as follows:

1. **Product fill rate** is the fraction of product demand that is satisfied from product in inventory. In other words, it is the probability of being able to supply product demand from available inventory.

2. **Order fill rate** is the fraction of orders that are filled from available inventory. This fill rate is important in a multi-product situation since an order can be filled only if all products can be supplied from the available inventory. Order fill rates tend to be lower than product fill rates because all products must be in stock for an order to be filled. However, the distinction between product and order fill rates in a single-product environment is not significant.

3. **Cycle service level** is the fraction of replenishment cycles that end with all the customer demand being met. A replenishment cycle is the interval between two successive replenishment deliveries [13].

A major value of customer service, as mentioned before, is its “profit impact”. By minimizing lost sales through “product availability”, customer service can contribute to increased profits. A reduction in order cycle time, order accuracy improvements, together with inventory adjustments and premium transportation are factors that contribute to profit gains. For this reason, the service level cost, i.e. the cost of not being able to meet commitments or customer demand, should be considered in strategic planning and operational decisions.
1.3 The Total Cost Concept

The total cost concept recognizes that the cost of providing "time and place utility" through the logistics system is minimized by trading off the various cost components of that system, and that concentrating on minimizing the cost of a single component may well lead to a higher system cost than if all the inter-connected elements of the system are considered simultaneously.

![Graph showing the relationship between logistics costs and total costs](image)

**Figure 1-3: A Simplified Illustration of the Total Cost Concept**
(Hypothetical Trade-offs among Transportation, Inventory and Order-Processing Costs)

The figure above gives a simple illustration of the total cost concept. Assume that the firm’s goal is to achieve a particular level of service at minimum cost (customer service can relate to the percentage of on-time deliveries, the number of complete orders shipped, or number of backorders compared to total orders). At one extreme of the available logistics policies is a configuration based on centralized storage—few warehouses located close to the point of manufacture, with adequate delivery speed provided by premium transport and high-speed, expensive order-processing capability. At the other extreme,
customer service is assured by a multitude of storage locations located close to the
customer, with low-cost transport.

If the firm concentrates only on the cost of transportation, the second extreme is
chosen as the cost-minimizing approach. However, this incurs high inventory costs. On the
other hand, if the focus is solely on inventory, the first extreme will be the cost-minimizing
approach resulting in large transportation costs. The set of policies with the least system
cost is characterized by a position in the middle of the continuum. This implies that the
system minimum can be achieved only by trading off the various costs of transport, storage
and order-processing.

To summarize, the total cost concept states that all relevant costs—transportation,
inventory and facility costs—should be considered in reaching a decision that will most
effectively utilize logistics and firm resources in order to meet customer needs most
economically, consistent with the firm’s objectives and competitive strategy. Total system
cost combines all of the individual factors into a single cost curve for the system [53].

1.4 Thesis Outline and Contribution

This thesis analyzes the allocation of inventory between the levels of a two-echelon multi-
stage distribution network. In the analysis that we make we use the “Total Cost Concept”
to come up with a cost efficient characterization that captures different trade-offs within
the logistics system. The system analyzed is an extension of the single-echelon GOMA
model developed by Erlenkotter [24], where we consider a two-echelon distribution
network design and incorporate inventory costs as well. Before defining the cost
components of the system that we analyze, we will present some background on the
literature of inventory lot sizing in multi-echelon systems, distribution network planning
and two-echelon models developed so far where a central facility serves $N$ local storage
facilities. Next, we will make an overview of the GOMA model and define the system that
we consider as an extension of this model in Chapter 2.

Chapter 3 introduces different cost components involved in the logistics cost
function. Warehousing, transportation and inventory costs are developed for the two-
echelon setting and a basic review of inventory heuristics used in practice is provided.
Chapter 4 presents the total logistics cost for a single-echelon and two-echelon distribution network. We develop some insights about the sensitivity of the system configuration and stock allocation scheme to various system parameter values for single and two-echelon systems.

Chapter 5 summarizes the conclusions and gives some suggestions for further possible research opportunities in this field.
Chapter 2

General Optimal Market Area (GOMA) Model

According to Geoffrion [28], the true purpose of mathematical programming in strategic applications is to develop insights into system behavior which can be used to guide the development of effective plans and decisions, rather than just figuring out what the optimal solution is. He advocates reducing the level of detail and complexity of the full mathematical model until it can be solved in closed form or by simple arithmetic. The use of such simplified analytic models and diagnostic tools help explain the “whys” behind the solutions of mathematical programming.

Diagnostic tools that promote understanding of numerical results are especially important from managerial point of view as managers are unlikely to accept analysis they do not satisfactorily understand. The General Optimal Market Area Model (GOMA) is a diagnostic model in operations research literature that has potential range of applications in distribution system design. Market area models are used to analyze the relationship between market size and performance measures and to determine the optimal market size of a facility [20].

Erlenkotter [24] reviews market area models literature and formulates the General Optimal Market Area Model, which models a distribution system with uniform and unchanging demand. The GOMA model has a number of desired properties that make it important in the operations research literature. To begin with, the operating cost of the system can be expressed as a continuous function that can further be analyzed. Another important fact is that for a wide range of facility operation cost structures and product transportation cost structures, this model provides a closed form expression for the optimal market area per facility [28]. Finally, although demand in practice is often lumpy, research
has found out that the basic properties of the GOMA extend to cases where demand spaces are much more complex [61]. The assumptions of the GOMA model are as follows:

1. Demand is distributed uniformly over an infinite plain, with density \( \rho \) per unit area.

2. The cost for a facility producing the amount \( w \) is \( k \cdot w^\alpha + c \cdot w \), where \( k > 0 \) represents fixed cost, \( 0 \leq \alpha \leq 1 \) is the economies-of-scale parameter for facilities, and \( c \geq 0 \) is the variable facility operating cost.

3. Unit transport costs are related to the distance traveled, \( \delta \), by the expression \( t \cdot \delta^\beta \), where \( t > 0 \) is the per unit transportation cost, and \( 0 < \beta \). For the case where \( \beta < 1 \), economies of scale in distance exist.

4. Various regular two-dimensional market shapes may be specified: circle, regular hexagon, square, diamond. For each of these shapes the facility is located at the center of the market.

5. Various distance measures may be designated, which include the “Manhattan metric” (or rectilinear norm) and the “Euclidean distance”. Regular hexagons have been shown to be the most efficient market shapes for the Euclidean distance norm and diamond-shaped market areas are the most efficient in the case of rectilinear norm.

6. The objective is to minimize cost.

Under these assumptions the average unit cost for any GOMA model with a regular two-dimensional market shape has the form:

\[
C(\bar{w}) = k \cdot \bar{w}^{\alpha - 1} + \sigma(\beta, s, d) \cdot t \cdot (\bar{w} / \rho)^{\beta / 2} + c,
\]

where \( \sigma(\beta, s, d) \) is a configuration factor reflecting the market shape \( s \), the distance metric \( d \), and the transport cost exponent \( \beta \). For many combinations of \( \beta \), market shape and distance norm, transportation cost is proportional to the market area raised to the \( \beta / 2 \)th power, i.e. average transportation cost per unit is \( t \cdot \sigma \cdot \bar{A}^{\beta / 2} \). For \( \beta = 1 \), \( \sigma \) is the average distance from the facility to a random customer for a market area of unit area. For diamond-shaped market area and the rectilinear distance norm \( \sigma = \left(2^{1 - \beta / 2}\right) / (2 + \beta) \). The optimal facility size is obtained as

\[
\bar{w}^* = \left[ \frac{2 \cdot (1 - \alpha) \cdot k \cdot \rho^{\beta / 2}}{\beta \cdot t \cdot \sigma(\beta, s, d)} \right]^{\frac{2}{1 + \beta}}
\]

and the optimal market area is given by

\[
\bar{A}^* = \frac{\bar{w}^*}{\rho} = \left[ \frac{2 \cdot (1 - \alpha) \cdot k \cdot \rho^{\alpha(1 - \alpha)}}{\beta \cdot t \cdot \sigma(\beta, s, d)} \right]^{\frac{2}{1 + \beta}}.
\]
Erlenkotter shows that the relative impact of a non-optimal choice of \( w \) on average unit cost, in other words the ratio \( C(w)/C(w^*) \), depends only on the cost functions' exponents \( \alpha \) and \( \beta \), and is independent of the other cost coefficients, the level of demand, the market shape and the distance norm.

Although the assumptions of the GOMA model are restrictive, the value of the model lies in the fact that it yields insights that extend to complex and realistic problems as in Erlenkotter [24] and Geoffrion [27]. The GOMA model can be extended by incorporating other features into its basic such as inbound transportation costs [27], demand that is dependent upon distance from the facility [24], a network comprising primary and secondary facilities [62], and demand that changes over time [61].

Webster and Robinson [62] investigate a two-echelon distribution system design, where the model is an extension of the single-echelon analytical GOMA model described above with uniform demand. The two-echelon GOMA model is viewed using echelon distance and cost, analogous to the echelon inventory and holding cost introduced by Clark and Scarf [49] for the analysis of multi-echelon inventory systems. Echelon distance is defined as the delivery distance from an echelon to the final customer. Analytical results pertaining to the question of when two-echelon system is beneficial are developed.

Webster and Gupta [61] provide another extension of the GOMA model by introducing dynamic demand variation into the problem of designing a distribution system. They show that under general conditions for the distribution of demand over time, the dynamic demand problem is equivalent to minimizing the expected distribution cost rate at some point during the planning horizon. They derive a number of results for the relationship between the dynamic demand problem and its equivalent stationary demand problem for general and specific demand functions.

One feature of the GOMA model and other diagnostic models that make simplifying assumptions to reduce complexity is that they deal with the determination of the optimal number of facilities and their optimal locations simultaneously, whereas a major portion of facility location literature deals with the identification of the locations of the facilities given the number of facilities to be located.

In the single-echelon GOMA model defined by Erlenkotter [24], all facilities are assumed to have identical operating characteristics and distribution service capabilities. In this study we will analyze a two-echelon GOMA model that incorporates the inventory
costs into the system cost in addition to facility and transportation costs as suggested by the total cost concept.

2.1 Two-Echelon GOMA Model Incorporating Inventory Costs

A multi-echelon distribution system has several advantages over a single echelon distribution system such as the potential to provide better customer service from regional locations. In addition, a multi-echelon system offers transportation economies due to consolidated shipments, provides risk pooling over the manufacturing or procurement lead time, and reduces the amount of inventory required for a specific customer service level. A basic question in a multi-echelon environment relates to the operating policy, i.e. whether it should be decentralized or centralized. In a “decentralized” system, each stocking location operates with its own inventory policy. On the other hand, a “centralized” control system has visibility of all stocks and all demands, and a central agent plans inventory movements, replenishments, allocations. In practice, real systems are somewhere in between these two extremes.

In general, a distribution system design must specify the number of echelons in the distribution system, the number and location of facilities in each echelon of the system, the flow of product between facilities at different echelon levels, and the assignment of customers (retailers or consumers) to supplying facilities. Another design and planning issue is the choice of replenishment policy, which items to stock at what location and the type of information system to be used. The objective is to determine the least total cost system such that a particular customer service level is maintained. This requires an analysis of the trade-off among the component costs of the system which include: fixed and variable costs for operating the facilities at each echelon level in the system, total facility throughput and inventory holding costs, and total inbound and outbound transportation costs for moving products through the distribution system. In the two-echelon model we develop, we will make use of the GOMA model to characterize the spatial distribution of demand and come up with the transportation and facility operating costs.

In a typical logistics network, products commonly pass through multiple echelons of distribution facilities as they flow from production plants to final consumers. From distribution point of view, a typical facility arrangement is to consolidate products at regional distribution centers (DCs). The DCs normally perform break bulk and product
mixing operations, and supply district, market-oriented warehouses (WH). The warehouses in turn serve retailers or final consumer demands.

The major benefit of utilizing such a multi-echelon distribution system is that it promotes the use of efficient, large volume transportation movements throughout the physical distribution channel while minimizing inventory investment [44]. In addition, the DC-level inventory inexpensively provides increased range and depth of stock over that carried by the warehouse. The stock at the warehouse has the role of filling the pipeline and supporting the warehouse. Muckstadt and Thomas [40] show that if a time-weighted objective is used, a multi-echelon model has an incentive to increase the range of stock so as to provide quick replenishment following a stock-out, and they conclude that it is efficient to centralize safety stock for relatively high cost, low demand items.

In the two-echelon model we consider there are two potential levels, i.e. echelons, of facility types. Facilities at the first echelon are called “distribution centers” (DC) and facilities at the second echelon are called “warehouses” (WH). Distribution centers receive product from suppliers and supply warehouses. Each distribution center supplies a certain set of warehouses that do not receive items from any other distribution center.

![Figure 2-1: Flow of goods in the Two-Echelon System](image)

Demand in this system is spatially uniformly distributed. We will assume that demand over an area $A$ is normally distributed with a mean of $\mu_d = \rho \cdot A$ and a variance of $\sigma_d^2 = \sigma^2 \cdot A$, where $\rho$ is the demand intensity factor and $\sigma^2$ is the variance per unit area. Demand over two non-intersecting regions with identical areas is independent and identically distributed. In addition, demands within non-overlapping time intervals are also assumed to be independent.

Each warehouse has a pre-specified market area within which it provides distribution services and maintains inventory. The market areas of different warehouses do not overlap, and facilities are centered within their corresponding market areas. The following figure represents the market areas of both echelons.

25
The market areas are assumed to be a perfect, four-sided diamond (i.e. a square rotated by 45° with respect to the travel directions in the case of right-angle travel), and a rectilinear distance norm will be adopted in calculating the transportation costs. Larson and Odoni [35] show that for the case where demand is uniformly distributed throughout the plane and travel distances are right-angled, the facilities should be positioned in a regular lattice to achieve minimal mean travel distance (sections 3.7, 3.8 and 6.5). In addition, they prove that a diamond shaped market area (i.e., a square rotated by 45°) results in minimum mean travel distance, given that a facility's district must contain an area $A$ [35]. Note that the right-angle distance between the two points on the plane $(x_1, y_1)$ and $(x_2, y_2)$ is defined to be $|x_1 - x_2| + |y_1 - y_2|$.

Every time the warehouse requests an order from the corresponding distribution center, a fixed cost $K_{wh} = K_2$ and a variable ordering cost of $c_{wh} = c_2$ proportional to the amount ordered is incurred. Analogously, each time a distribution center places an order to the supplier a fixed cost of $K_{dc} = K_1$ and a variable ordering cost of $c_{dc} = c_1$ proportional to the amount ordered is incurred.
Inventory holding cost is charged per item per unit time. Assume that the unit inventory holding cost per unit is $h_{WH} = h_2$ at the warehouses and $h_{DC} = h_1$ at the distribution centers. Then, the corresponding echelon inventory holding costs for the DCs and the WHs are $h'_{DC} = h_{DC} = h_1$ and $h'_{WH} = h_1 - h_2$ assuming that the inventory holding cost for the warehouses is higher than the inventory holding cost at the distribution centers.

The lead time from the supplier to the DC is $L_{DC} = L_1$ and the lead time from the DC to the warehouses is $L_{WH} = L_2$. In addition, each echelon in the system has a prespecified customer service level. We will investigate the logistics cost components of such a two-echelon system in Chapter 3.

2.2 Literature Review

2.2.1 Facility Location

Cavalier and Sherali [8] consider location-allocation problems in which the region to be served is a convex polygon having a uniform demand distribution, and which may also contain discrete demand points, whereas the GOMA (General Optimal Market Area) model assumes that the market area is a regular two-dimensional shape. A given number of facilities have to be located, and the sub-region that each facility serves has to be determined, so as to minimize total transportation costs using a Euclidean distance measure. They develop a method to optimally locate a single facility in a convex polygon with a uniform demand distribution and with discrete demand points by using an efficient implementation of the Weiszfeld technique [63].

Drezner and Weslowsky [21] consider the Weber single-facility location where the demands are not only points but may be areas as well. In the ordinary Weber problem, a facility is to be located so that the sum of weighted distances from the facility to $n$ demand points is minimized. Drezner and Weslowsky provide an iterative procedure for solving the problem with area demands and with $lp$ distances when $p > 1$. Their approach is also based on the Weiszfeld procedure as in Cavalier and Sherali [8].
Chen [12] considers the location-allocation problem where \( m \) service facilities have to be located with respect to \( n \) demand points, which involves the minimization of a non-linear unconstrained function. Chen’s work introduces a differentiable approximation to the location-allocation objective function, which can be solved efficiently using a quasi-Newton method to solve the non-linear problem. However, the main difficulty in solving the location-allocation problem is that the problem is neither convex nor concave. Therefore, any solution found by using any nonlinear programming method is usually only a local minimum. After approximating the objective function with a differentiable version, the problem is still neither convex nor concave; hence, the solutions obtained by Chen’s method are also not guaranteed to be global optima.

The analogous problem of locating \( P \) service points on a network with \( n \) vertices is known as the “\( P \)-median problem”. Hakimi has shown that there exists a \( p \)-median whose points are all vertices. Hakimi and Kariv [34] show that the problem of finding a \( p \)-median of a network is an \( NP \)-hard problem even when the network has a simple structure. However, they present results leading to efficient algorithms when the network is a tree.

2.2.2 Multi-Stage Systems – Lot Sizing and Inventory Policies

Crowston, Wagner and Williams consider the optimal lot size problem for multi-stage assembly systems where each facility has only a single successor but may have many predecessors. Assumptions they make include constant continuous final product demand, instantaneous production, and an infinite planning production. Costs at each facility consist of a fixed charge per lot and a linear holding cost. Under the constraint that the lot sizes remain time invariant, they prove that the optimal lot size at each facility is an integer multiple of the lot size at the successor facility. Then, using this fact they construct a dynamic programming algorithm which uses echelon stock for the computation of optimal lot sizes.

Roundy (1993) considers a multi-stage, multi-product production/inventory system in discrete time, where it is assumed that the external demand for a component is non-constant, deterministic and must be met without backlogging. Roundy proposes two cluster-based heuristic procedures, and shows that the heuristics are within 0.7% and 1.3% of optimal, and that their performance is highly insensitive to the size of the system and to other input parameters [47].
Taha and Skeith (1970) develop a model for a single-product, multi-stage production/inventory system with static deterministic demand, where the product moves between the stages in a serial fashion. They determine the optimal production scheduling of the product within the multi-stage production system where backlogging is allowed. The batch size and the shortage quantity of the finished product are determined by minimizing the total cost per unit time which includes the inventory holding cost, the shortage cost and the setup cost. They also consider the case with storage constraints at different stages [60].

Schwarz and Schrage (1975) propose optimal and near-optimal policies for multi-echelon production/inventory assembly systems under continuous review with constant demand over an infinite planning horizon. Costs at each stage consist of a fixed charge per order or production setup plus a linear holding cost on echelon inventory. The goal is to minimize the average cost per unit time. They provide a branch-and-bound algorithm that usually finds the optimal solution quickly. Schwarz and Schrage come up with a “system myopic solution”, where a given objective function is optimized with respect to any two stages and multi-stage interaction effects are ignored. Their major conclusion regarding the system myopic policies is that these policies are easy to understand, require less information, and are fast and easy to compute [51].

Love (1972) considers an \(n\)-period single-product inventory model with known requirements and concave production and storage costs. The model is a multi-echelon structure in which \(N\) facilities are arranged in series. Love shows that if the storage and production costs are non-decreasing in order of facilities and non-increasing in time, then an optimal schedule has the property that if in a given period, facility at a certain stage \(j\) produces, then facility at stage \(j+1\) which uses the output of stage \(j\) as an input also produces. This nested structure is exploited in an algorithm for finding an optimal schedule [36].

De Bodt and Graves (1985) consider a multi-stage, serial inventory system. The demand for the end item is stochastic and stationary [17]. The relevant costs include a fixed ordering cost and an echelon inventory holding cost for each stage, and a backordering cost for the end item. The objective is to find a continuous-review inventory policy that minimizes the expected average costs. The approximation presented by De Bodt and Graves is analogous to that for the traditional single-item, continuous-review inventory model that assumes a re-order point, re-order quantity policy. The most closely related work to DeBodt and Graves (1985) is by Clark and Scarf [14], who analyze a
system under periodic review when the end item demand is stochastic. If all stages have fixed ordering costs, their solution is approximate but provides upper and lower bounds for the optimal solution. The method by Clark and Scarf computes the optimal policy for each echelon separately, with successive echelons being linked by a penalty function that represents the cost of not being able to fill a replenishment request. Federgruen and Zipkin (1984) extend the Clark and Scarf approach to the infinite horizon case and give a new computational approach [26].

A key aspect to the Clark-Scarf (1960) approach is the use of echelon inventory in the inventory control policies for the multi-echelon system [14]. They define the “echelon-stock” of a component to be the inventory of the component plus the entire inventory of downstream items that use or require the component such as sub-assemblies or end-items. This is different from the “installation stock”, which is the inventory of the component neglecting all downstream inventories. Federgruen and Zipkin [26] extend the results of Clark and Scarf (1960) for a two-echelon, two-location inventory model to the infinite-horizon case (for both discounted and average costs). The computations required turn out to be far easier than for the finite horizon case. They provide further simplification for normal demands. They also consider the case of multiple locations at the lower echelon, and they show that under certain conditions this problem can be closely approximated by a model with one such location. These conditions include assuming demand at each outlet is normal, with identical coefficients of variation, costs and lead times.

Crowston et al. (1973) and Schwarz and Schrage (1975) have studied lot-sizing in assembly systems with a constant, known demand rate. Both papers assume nested policies such that whenever one stage orders, all downstream stages also order [16].

F. Chen and Y.-S. Zheng (1994) establish lower bounds on the minimum costs of managing certain production-distribution networks with setup costs at all stages and stochastic demands. These networks include serial, assembly, and one-warehouse multi-retailer systems. The bounds are obtained through cost allocation schemes. They evaluate the bounds’ performance for one-warehouse multi-retailer systems by comparing them with simple, heuristic policies. F. Chen and Y.-S. Zheng show that the bounds are quite tight for systems with a small number of retailers and present simplified proofs of known optimality results for serial and assembly systems [11].

S. Axsater and K. Rosling (1993) compare installation and echelon stock policies for multi-level inventory control. Their major results are for serial and assembly systems.
They conclude that echelon stock policies are in general superior to installation stock policies for \((Q, r)\) policies, where \(r\) is the reorder point and \(Q\) is the lot size. They identify a Kanban policy as a restricted type of installation stock \((Q, r)\) policy [3].

F. Chen and Y.-S. Zheng (1998) study echelon-stock \((R, nQ)\) policies in a multi-stage, serial inventory system with compound Poisson demand. They provide a simple method for determining near-optimal control parameters. First, they establish lower and upper bounds on the cost function by over-charging and under-charging a penalty cost to each upstream stage for holding inadequate stock. Second, they minimize the bounds, which are separable functions of the control parameters to obtain heuristic solutions. They also provide an algorithm that guarantees an optimal solution at the expense of additional computational effort. The numerical study they make suggests that the heuristic solutions are easy to compute and are close to the optimal [9].

Multi-echelon inventory systems are often managed using adaptations of single-location models. However, J. A. Muckstadt and L. J. Thomas [40] show that these models can perform poorly in comparison with methods designed to take advantage of the system structure. They point out that taking advantage of system structure is important when there are many items with low demand and high relative cost, since these items may be best managed by having stock only at a central supply point. The system they study is a two-echelon inventory system with no lateral supply. A distribution center is at the upper level or echelon and customer-service warehouses are on the lower level. They use \((S - 1, S)\) continuous review policy at each location, which is appropriate for low demand items at the warehouse level and for most items at the distribution center (DC). The demand process for each item at the warehouse (WH) level is assumed to be a Poisson process. Since each demand is immediately forwarded to the DC, the DC also faces a Poisson demand process. Their objective is to investigate the potential worth of applying multi-echelon inventory models in certain practical situations.

One of the fundamental results in stochastic inventory theory is the optimality of an \((s, S)\) policy in several standard infinite-horizon single-item models. Under a stationary \((s, S)\) policy, an order is placed to raise the item’s inventory position to the level \(S\) as soon as this inventory position drops to or below the level \(s\). Scarf [49] shows that the optimal ordering policy for an \(n\)-period dynamic inventory problem in which the ordering cost is linear plus a fixed reorder cost and the other one-periodic costs are convex is characterized by a pair of critical numbers \((s_n, S_n)\). Iglehart (1963) gives bounds for the sequences \(s_n\) and
and discusses their limiting behavior. Iglehart shows that the limiting (s, S) policy characterizes the optimal ordering policy for the infinite horizon problem. Similar results are obtained if there is a time-lag in delivery [32].

2.2.3 Two Echelon Designs

Roundy (1985) considers the case where a warehouse supplies N retailers, where external demand occurs at the retailers and shortages are not allowed. At each retailer and at the warehouse there are linear holding costs and fixed ordering costs. The objective is to minimize the long-run average cost over an infinite time horizon. Roundy defines a simple-structured policy, namely power-of-two policy, which facilitates both computation and implementation in practice, and shows that the optimal power-of-two policy is within 2% of the cost of an optimal policy for the original problem [48].

L. B. Schwarz (1973) examines a one-warehouse N-retailer deterministic inventory system [50]. The objective is the minimization of average system cost per unit time over an infinite time horizon. L. B. Schwarz derives necessary properties of an optimal policy, and gives optimal solutions for the one-retailer and N identical retailer case. In addition, heuristic solutions for the general problem are suggested, tested against analytical lower bounds and, on the basis of these tests, found to be near optimal. L. B. Schwarz considers two classes of basic policies for the multi-retailer system case: separate retailing policies and single cycling policies. In a “separate retailing” policy the N-retailer problem is simply divided into N one-retailer problems. The average cost of this policy is the sum of the average costs of the N one-retailer problems. Schwarz defines a “single cycle” for the deterministic case to be a basic policy where all retailers have zero inventory every time the warehouse receives a delivery, and a “basic policy” to be any feasible policy with the following assumptions:

- Deliveries are made to the warehouse only when the warehouse has zero inventory and at least one retailer has zero inventory.
- Deliveries are made to any given retailer only when that retailer has zero inventory
- All deliveries made to any given retailer between successive deliveries to the warehouse are of equal size.

L. B. Schwarz shows that for the N-identical retailer problem single-cycling is optimal. We extend this case for the one DC N-identical warehouse system case under stochastic
demand, and characterize single-cycle type of policies such that the DCs receive the stock they order at the same time they face replenishment order from the lower echelon, which corresponds to the time where both echelons have their lowest inventory levels. Note that in a setting with stochastic demand and non-zero lead times allowing the inventory to drop to zero before placing the order creates stock-out situations during the lead time; therefore, the order is placed before the inventory position drops to zero.

S. C. Graves and L. B. Schwarz (1977) examine optimal and near-optimal continuous review policies for a deterministic arborescent inventory system where known and constant outside demand must be met without backlogging or loosing sales at minimum average system cost per unit time. Each stage has a fixed order cost and a proportional holding cost on each stage’s echelon inventory. They describe the characteristics of optimal policies and prove that if the overall optimal solution is a stationary cycling policy then it is a single cycle policy and that given zero initial inventory, if there exists an optimal policy which is stationary that policy is a single cycle policy. They present an efficient branch-and-bound algorithm for determining optimal single-cycle policies for arborescent systems and examine the near-optimality under system myopic single-cycle policies. A stationary policy is one where each stage orders the same lot size each time it orders [30].

Muckstadt and Thomas (1980) point out that multi-echelon, multi-item inventory systems based on adaptations of single-location models can perform poorly in comparison to methods that take advantage of the system structure [40]. Taking advantage of the system structure is important when there are many items with low demand and high relative costs, since such items can best be managed by having stock only at central supply point. The multi-echelon method Muckstadt and Thomas use involves an \( (S - 1, S) \) ordering policy and Poisson demand, both of which are appropriate for low-demand items. Their method requires much lower total investment than the single-location method requires achieving the same average level of performance. Thus, they conclude that multi-echelon methods are worthwhile in many situations. The system they study is a two-echelon inventory system, where a distribution center is at the upper level and customer-service warehouses are at the lower level. Muckstadt and Thomas conclude that centralizing safety stock is efficient for relatively high-cost, low demand items. The larger the number of low-demand items, the more important a multi-echelon approach becomes. In addition, the tighter the inventory budget, the more important it becomes to use multi-
echelon methods to avoid duplication of significant amounts of safety stocks at both the distribution centers and the warehouses.

Graves (1985) addresses the problem of how to determine the inventory levels in a multi-echelon inventory system for a repairable item. The multi-echelon system in its simplest form consists of a set of operating sites supported by a centrally located repair depot. Each operating site requires a set of working items and maintains an inventory of spare items. The repair depot also holds an inventory of spare items. Item failures are infrequent and are replaced on a one-for-one basis. Graves presents an exact model for finding the steady state distribution of the net inventory level at each site. This model assumes that the failures are generated by a compound Poisson process and that the shipment time from the depot to each site is deterministic. Based on the exact model, an approximation for the steady-state distribution for the case with ample servers at the repair depot is presented [31].

B. L. Deuermeyer and L. B. Schwarz [19] present an analytical model for estimating the expected service level, e.g. fill-rates and back-orders, of a one-warehouse N retailer distribution system as a function of that system's parameters, such as warehouse and retailer lot sizes, nominal fixed lead times and known stochastic demand parameters. The system they examine is one involving N identical retailers facing stationary Poisson demand and operating with stationary (Q, r) continuous replenishment policies. The model they develop is a generalization of the exact single-facility (Q, r) model developed by Hadley and Whitin. They test the model via computer simulation and conclude that the model's estimates closely approximate the empirically observed service levels of the simulated systems.

Svoronos and Zipkin (1988) [59] propose several refinements of the technique used by B. Deuermeyer and L. B. Schwarz who provide a simple approximation of a complex multi-echelon system. Svoronos and Zipkin come up with models that are nearly as simple and considerably more robust. Their analysis suggests guidelines for the design of large-scale logistics systems. Svoronos and Zipkin (1988) consider a two-echelon inventory system where several lower-echelon facilities (retailers) experience direct customer demand. The retailers order batches from a single higher-echelon facility (the warehouse) to replenish their stocks. All stock-outs in both echelons are backordered. They assume that the inventory policies of all facilities are of the continuous review (R, Q) type, i.e., when the inventory position (inventory level-stock on hand minus backorder-plus stock on
order) declines to the reorder point $R$, a batch of size $Q$ is ordered. They estimate the most important measures of performance; namely, the average backorders at the retailers and the average inventory at each location. These estimates are used within a cost-minimization framework to choose among different policies. Their approach, which they define as a "decomposition technique" based on the METRIC technique by Sherbrooke (1968), approximates each facility as a single-location inventory system. The innovation in their model is the use of second-moment information in the approximations. To approximate the distribution of the warehouse lead-time demand, they use a mixture of translated Poisson variables. Then, they develop formulas to express the average on-hand inventory and average backorder level at the warehouse. By using the average backorder level, they approximate the average retailer retard through the use of Little's Formula. This, in turn, is used to find the expectation and variance of the retailer lead time. Finally, they estimate the mean and variance of the retailer lead time demand. Fitting a negative binomial distribution to this mean and variance, they develop expressions to approximate the average on-hand inventory and backorders at a retailer. Their cost function accounts for the holding and order cost both at the retailers and the warehouse, and the backorder cost at the retailers.

Deuermeyer and Schwarz [19] consider the same problem as Svoronos and Zipkin (1988). However, they approximate the warehouse lead-time-demand distribution by a normal distribution by using the asymptotic values of the real mean and variance, whereas Svoronos and Zipkin use the mixed translated Poisson distribution. In addition, Deuermeyer and Schwarz treat the retailer lead time as deterministic. In Svoronos and Zipkin, the variance of the retailer lead time is approximated using the moments of the backorder distribution.

Axsater (1990) considers an inventory system with one warehouse and $N$ retailers [1]. Lead times are constant and the retailers face independent Poisson demand. Replenishments are one-for-one, which implies that ordering costs are low and can be disregarded. Axsater provides simple recursive procedures for determining the holding and shortage costs of different control policies. Axsater’s model and minor variations of it have been previously analyzed. Sherbrooke’s METRIC model (1968) approximates outstanding orders at the retailers by Poisson random variables [54]. Graves (1985) determines exactly both the average and the variance of outstanding orders at a retailer. Then a negative binomial distribution is used to approximate the distribution of outstanding orders. The
two-parameter approximation of Graves is in general more accurate than the METRIC approximation. Simon (1971) derives the steady-state distribution for the inventory levels at each site [58]. Evaluation methods for more general batch ordering policies are more complex because the demand process seen by the warehouse is then a superposition of Erlang renewal processes, whose characterization is difficult. Different approximate methods for this more general problem have been suggested by Deuermeyer and Schwarz (1981), Svoronos and Zipkin (1988) and Moinzadeh and Lee (1986).

Axsater (1993) again considers a two-level inventory system with one warehouse and \( N \) identical retailers. Lead times are again constant and the retailers face independent Poisson demand [3]. In his previous work Axsater (1990) derives a recursive procedure for determining the policy costs for an average item in case of one-for-one replenishment policies. As an extension of this work, Axsater (1993) shows how these results can be used for the exact or approximate evaluation of more general policies where both the retailers and the warehouse order in batches, i.e., all facilities use continuous inventory policies (when the inventory position drops to the reorder level, a batch of size \( Q \) is ordered). Axsater provides a comparison of the results with the method suggested by Svoronos and Zipkin (1988).

S. Axsater [2] examines a two-level system (a central warehouse and multiple retailers) under continuous review policies, where the inventory levels are followed continuously and replenishments may be ordered at any time. Axsater states that in case of low-demand it is suitable to apply continuous review policies. On the other hand, in the case of very high demand, continuous review can lead to high information costs and it may be more practical to review the inventory system periodically. According to Axsater the reason for assuming a Poisson demand process when dealing with a relatively low stochastic demand is the fact that in many situations it is a very realistic assumption and the fact that it simplifies the analysis considerably. In case of low demand, high holding costs and moderate ordering costs it is often optimal to let the deliveries to be one-for-one. An important and simplifying consequence of such policies is that Poisson demand at the retailers will give a demand process at the WH that is also Poisson. In many practical situations, however, ordering costs are sufficiently high to motivate deliveries in batches.

Simon (1971) considers a two-echelon inventory model for consumable or repairable parts in which the first-echelon facilities use one-for-one replenishment policies and the second-echelon facility uses a continuous review \((s, S)\) policy. Repair and
transportation times are assumed to be deterministic, and the failure processes that generate the demand are assumed to be Poisson. Simon derives exact expressions for the stationary distributions of stock on hand, stock in repair, and backlogged demand at each facility, and indicates how these expressions may be utilized for optimization purposes.

F. Chen and Y.-S. Zheng [9] consider distribution systems with one warehouse and \( N \) retailers. The warehouse (stage 0) receives stock from an outside supplier with infinite stock and replenishes the retailers (stages 1, 2 \( \ldots \) \( N \)). Customer demands occur only at the retailers. They denote this system by Distribution \( (N) \). If only the warehouse has a setup cost \( K \), they denote the system by Distribution \( (N, K) \). F. Chen and Y.-S. Zheng (Jan. 1984) state that optimal policies for Distribution \( (N, K) \) are unknown. The difficulty is due to stock imbalances among different retailers. Federgreen and Zipkin provide a lower bound on the minimum cost of the system by allowing a “free” inventory (position) rebalance among the retailers. Under such a relaxation, the original system reduces to a “single-location system” whose minimum cost can be easily computed. This minimum cost is a lower bound on the minimum cost of the original system. Chen and Zheng establish lower bounds on the minimum cost of Distribution \( (N) \) through cost allocation and physical decomposition. They imagine that the product at a retailer consists of two components: a common component and a retailer-specific component.

Federgreen and Zipkin [25] consider a central depot which supplies several locations experiencing random demands. Orders are placed periodically by the depot. The order arrives after a fixed lead time, and is then allocated among the several locations. The depot itself does not hold inventory. The allocations are finally received at the demand points after another lag. Unfilled demand at each location is backordered. Linear costs are incurred at each location for holding inventory and for backorders. The object is to minimize the expected total cost of the system over a finite number of periods. This system gives rise to a dynamic program with a state space of very large dimension. Federgreen and Zipkin show that this model can be approximated by a single-location inventory problem. Their model and the results obtained extend those considered by Eppen and Schrage (1981).

G. Eppen and L. Schrage (1981) model a depot-warehouse system with independent normally distributed stationary demand (known parameters) at the warehouses, identical proportional costs of holding and backordering at the warehouses, and no transshipment [23]. The operating policy is periodic review and the depot does not
hold inventory. Under an assumption that the incoming order is large enough so that an equal probability of stock-out can be achieved at each warehouse, G. Eppen and L. Schrage derive approximately optimal policies and costs of a base stock policy at the depot, assuming no fixed order costs, and an \((m, y)\) policy at the depot, assuming fixed order costs at the depot. An \((m, y)\) policy is one in which every \(m\) periods the system inventory position is raised to a base stock of \(y\). The difference in these two operating policies is motivated by the absence or presence of fixed ordering cost at the depot. In both models, the depot uses a base stock policy. In the latter situation, every \(m\) periods the depot orders enough items to bring the inventory position up to a pre-determined level. G. Eppen and L. Schrage state that when a setup cost is incurred at the depot for placing an order it no longer need be optimal to order every period. Thus, they assume that the depot places an order every \(m\) periods so as to bring the system stock up to a number \(y\). Upon receipt of goods, the depot distributes the goods to the warehouses. They note that the \((m, y)\) policy may be non-optimal, and that the optimal policy may be a function of the inventories on-hand and on-order for all warehouses. The form of the optimal policy is not known. The \((m, y)\) policy is easy to implement and the derivation of this policy provides insight into reorder point, reorder quantity policies.

F. Chen and Y.-S. Zheng (1997) consider a distribution system with a central warehouse and multiple retailers [11]. The warehouse orders from an outside supplier and replenishes the retailers which in turn satisfy customer demand. The retailers are non-identical, and their demand processes are independent compound Poisson. There are economies-of-scale in inventory replenishment, which is controlled by an echelon-stock, batch transfer policy. For the special case with simple Poisson demand, they develop an exact method for computing the long-run average holding and backorder costs of the system. Based on this exact method, they provide approximations for compound Poisson demand. They also present a numerical comparison between the average costs of a heuristic, echelon-stock policy and an existing lower bound on the average costs of all feasible policies.

Schwarz, Deuermeyer and Badinelli (1985) examine the system fill-rate of a one-warehouse \(N\)-identical retailer distribution system as a function of warehouse and retailer safety stock [52]. Using the approximation of Deuermeyer and Schwarz (1981), they examine the problem of maximizing system fill-rate subject to a constraint on system safety stock. Optimal safety stock policy is characterized to be the intersection of a fill-rate
policy line and the safety stock budget line. They state that the properties of the fill-rate policy line may be used to provide managerial insight into system optimization and as the basis for heuristics. A heuristic based on the corresponding deterministic version of the problem-called the "vertical heuristic"-provides simple and near-optimal policies. N-identical retailers each face independent, stationary Poisson demand, and follow a continuous review stock replenishment policy. Retailer orders are filled by a single warehouse following a similar policy, where the order quantity of the warehouse is an integer multiple of the order quantity of the retailers. They define a "policy line" to be the locus of points [warehouse safety stock vs. total retailer safety stock], each one of which is a solution to the fill-rate maximization problem for some fixed budget. The safety stock budget line is the locus of points [warehouse safety stock vs. total retailer safety stock], such that all safety stocks add up to a fixed budget, i.e. a straight line with slope of -1.

P. L. Jackson (1988) considers a single order cycle of a warehouse, serving N retailers where the only shipments allowed during the cycle are from the warehouse to the retailers [33]. For a simple ship-up-to S allocation policy, Jackson develops both the exact cost model and a computationally tractable approximate cost model for the case of identical retailers, and demonstrates empirically the benefits of centralizing at least a portion of the total system stock. Jackson addresses the problem of stock allocation in a two-echelon distribution system for a consumable item. Stock is acquired by the warehouse and allocated to several local retailers. The information system is assumed to allow periodic review of all stock levels. Control of the system is centralized. The manager of the system must decide the timing and quantities of orders to place with the outside supplier and for shipment to the warehouse (the order policy) and the timing and quantities of shipments from the warehouse to each of the retailers (the allocation policy). Jackson allows the warehouse to hold stock in contrast to other papers such as Eppen and Schrage (1981) and Federgruen and Zipkin (1984) that restrict the role of the warehouse to a central ordering policy and transshipment point. The management problem is one of minimizing the expected total holding and shortage costs over a single warehouse order cycle.

Clark and Scarf (1960) identify the complexity of managing such pure distribution systems as arising out of the possibility that the distribution of stock in the system can become unbalanced. In other words, if one computes the ideal stock level for each location based on knowing only the total amount of stock in the system, then it could be the case that some retailers have stock in excess of the ideal levels for those locations. Whenever
the stocks at the retailers are out of balance, the optimization problem requires knowledge of more than just the total amount of stock in the system. Hence, the computational burden involved in finding the optimal solution becomes impractically large. Several papers investigate extending the Eppen and Schrage model to allow the warehouse to hold stock during the cycle. One purpose of holding stock is to permit the warehouse to make dynamic allocation to the retailers rather than one single allocation at the beginning of the cycle.

Most of the inventory models for multi-echelon inventory systems assume a single-cycle policy. K. S. Park and D. H. Kim state that this type of policy is not completely satisfactory over a wide range of problems. They propose the "congruential policy" which is a set of policies that includes the single-cycle policy. They found out that the congruential policy is very close to the overall optimal policy, relatively fast and easy to compute and easy to implement (periodic ordering). A general belief is that under the optimal policies in a multi-echelon distribution system, a facility must order more frequently than its predecessors, as lots are distributed from the initial stage to the final stage. Park and Kim state that it is an unnecessary restriction to consider only single-cycle (nested) policies for a multi-echelon distribution system, where each time any stage orders all of its successor stages also order. Park and Kim develop an inventory model for a two-echelon inventory system under the assumption that the central warehouse and retailers order periodically. Based on the characteristics of the optimal policy, an iterative solution procedure is presented to find the optimal or near-optimal operating policy variables minimizing the average annual variable cost. Solutions of the model to a large number of test examples show that the model outperforms other existing models in the literature without sacrificing the computation time. They assume that deliveries are instantaneous, initial inventory is zero, no shortages are allowed, demand over time is known and constant and the time horizon is infinite.

2.2.4 Distribution System Design

D. B. Rosenfield and M. E. Pendrock focus on the problem of investigating aggregate inventory levels for multi-echelon distribution strategies, where they consider two types of inventory control systems: independent systems and coupled systems. Factors that favor one system over the other are discussed and a case study involving the evaluation of
inventory levels under each system is presented. In an “independent system”, the remote warehouses are controlled independently of the procurement process. In the “coupled system” the remote warehouses are coupled to the procurement process. Some of the factors they consider that affect the choice of control strategy are: the relative value added between the central stocking point and remote locations, the number of warehouses, the proportion of market served by the central warehouse, the transit time from the central warehouse to remote warehouses relative to the procurement lead time, the lumpiness of demand and sizes of order lots from customers, the correlation of regional demands, transshipment cost, the absolute sales level, and the management and control factors.

Ballou presents a large-scale computer model (DISPLAN) which is used for the strategic planning of supply and distribution networks such as the determination of the number, size and location of facilities. This model handles non-linear inventory costs. They state that a proper analysis of network design to achieve the minimum cost configuration of facilities should balance transportation cost, warehouse storage and handling costs, order processing costs, facility fixed costs, and inventory carrying costs. An important observation Ballou makes is that inventory carrying costs are usually treated as a linear function of the number of facilities in the network. However, if the inventory policy is based on the concept of economic order quantity, then the relationship between total inventory in the system and the number of facilities turns out to be non-linear, which is referred to as the “consolidation effect”. DISPLAN uses a heuristic iterative procedure to converge on the minimum cost network configuration subject to facility capacity and customer service constraints.

Graves and Geoffrion (1974) consider a common problem in distribution system design, i.e. the optimal location of intermediate distribution facilities between plants and customers. They formulate a multi-commodity capacitated single-period version of this problem as a mixed-integer linear program, and use a solution technique called Benders’ Decomposition, which turns out to be very effective as a computational strategy for static multi-commodity intermediate location problems.
Chapter 3

Logistics Cost Components

3.1 Transportation Costs

Transportation, which is a significant component of logistics costs, refers to the movement of a product from one location to another within the supply chain. Major modes of transportation are air, package carriers, truck, rail, water, pipeline and inter-modal. All transportation decisions in a logistics network must be made taking into account their impact on inventory costs, facility and processing costs, and the level of responsiveness provided to customers. Major trade-offs that need to be considered when making transportation decisions are:

- Transportation and inventory cost trade-off, and
- Transportation cost and customer responsiveness trade-off.

_Transportation and inventory cost trade-off:_ The trade-off between transportation and inventory costs is important at the design stage of the logistics/supply chain network and involves decisions regarding the choice of transportation mode and inventory aggregation. The mode of transportation that results in the lowest transportation cost does not necessarily lower total costs for a supply chain. Cheaper modes typically have longer lead times and larger minimum shipment quantities both of which result in higher levels of inventories. On the other hand, modes that allow for shipping in small quantities lower inventory levels but tend to be more expensive. For products with a high value-to-weight ratio where reducing inventories is important, faster modes of transportation are preferred. Slower modes are more suitable for products with a small value-to-weight ratio where reducing transportation costs is important.
Firms can significantly reduce the safety stock they require by physically aggregating inventory in one location. This has been used by most e-businesses, such as Amazon.com, to gain over their competitors with facilities in many locations. However, when inventory is aggregated transportation cost increases. For this reason, inventory should be aggregated if inventory and facility costs account for the majority of the logistics cost. Inventory aggregation is especially useful for products with a large value-to-weight ratio and for products with high-demand uncertainty as in the PC industry. However, if the product has a low value-to-weight ratio and customer orders are relatively small, inventory aggregation may deteriorate the performance of the supply chain by increasing the transportation costs.

*Transportation cost and customer responsiveness trade-off:* If a firm aims for high levels of responsiveness and ships all orders quickly, the outbound shipments will be small and will result in high transportation cost. If responsiveness is decreased by aggregating the orders over a longer time period before shipping, larger shipments will enable the firm to exploit economies of scale and lower transportation costs. The process of combining orders across time is “temporal aggregation”. Temporal aggregation decreases responsiveness due to shipment delays but decreases the transportation cost as well which results from economies in scale in larger shipment quantities.

In short, the transportation function should be evaluated based on a combination of transportation cost, other costs such as inventory affected by transportation decisions, and the degree of responsiveness offered to customers. For the two-echelon distribution network that we described before, we will make use of the spatial uniformity of demand in the GOMA model that allows a simple formulation of the transportation cost per unit.

Under this setting we will assume as in the GOMA model that unit transportation costs are related to the distance traveled, $\delta$, by the expression $\Gamma \cdot \delta^\beta$, where $\Gamma > 0$ is the transportation cost, and $0 < \beta$. If $\beta < 1$, then economies-of-distance in transportation exist. As mentioned in the description of the GOMA model, for many combinations of $\beta$, market shape and distance norm, average transportation cost per unit is $\Gamma \cdot \kappa \cdot A^{\beta/2}$, where $\kappa$ is the configuration factor and $A$ is the market area. The value of $\kappa$ depends on $\beta$, the market shape and the distance norm. For $\beta = 1$, $\kappa$ is the average distance from the facility to a random customer for a market area of unit area. For the two-echelon distribution
network that we consider, $\kappa = \frac{2^{1-\beta/2}}{2 + \beta}$ since the market areas are diamond shaped and the distance norm is rectilinear.

From the GOMA model we know that for a spatially uniformly distributed demand transportation cost per unit is given by $\Gamma \cdot \kappa \cdot A^{\beta/2}$. For $\beta = 1$ and a diamond shaped market area $A$ with rectilinear distance norm $\kappa = \sqrt{2}/3$, and transportation cost per unit is $\Gamma \cdot \sqrt{2}/3 \cdot A^{1/2}$. This implies that the transportation cost per unit at a warehouse that serves a diamond shaped market area with rectilinear norm and with uniformly distributed demand over the area $A_{WH}$ is:

$$\Gamma_{WH} \cdot \frac{\sqrt{2}}{3} \cdot A_{WH}^{1/2}.$$

To find the average transportation cost per unit at the distribution center we need to identify the average distance between a distribution center and the warehouses within its market area. This is because the distribution centers serve the warehouses and thus the demand that they face does not have spatially uniform characteristics. Let $\delta$ be the average distance from a DC to warehouses within the market area of the DC. For the setting where market areas are diamond shaped the average DC-WH distance is given by:

$$\delta = \sqrt[3]{\frac{A_{WH}}{A_{DC}}} \left(\frac{A_{DC}}{A_{WH}} - 1\right) \cdot \frac{\sqrt{2}}{3} \cdot A_{WH} (\text{see the Appendix}).$$

Thus, transportation cost per unit for $\beta = 1$ at a DC is:

$$\Gamma_{DC} \cdot \delta^\beta = \Gamma_{DC} \cdot \delta = \Gamma_{DC} \cdot \sqrt[3]{\frac{A_{WH}}{A_{DC}}} \left(\frac{A_{DC}}{A_{WH}} - 1\right) \cdot \frac{\sqrt{2}}{3} \cdot A_{WH}.$$

Note that alternatively we can use the corresponding echelon distance and cost figures for the above terms as in Webster and Robinson [62], where they define the echelon distance to be the delivery distance from an echelon to the final customer. Per unit transportation cost rates are defined to be $\Gamma_1 \cdot \delta^\beta$ for DC to warehouse shipments and $\Gamma_2 \cdot \delta^\beta$ for WH to customer shipments. The corresponding echelon costs are
\[ \gamma_{DC} = \gamma_1 = \Gamma_1 \] for Echelon 1 and \[ \gamma_{WH} = \gamma_2 = \Gamma_2 - \Gamma_1 \] for Echelon 2. In this case, average Echelon 2 cost per unit is \[ \gamma_{WH} \cdot \frac{\sqrt{3}}{3} \cdot A_{WH}^{1/2} \]. The average Echelon 1 distance, however, is equal to the sum of the average Echelon 2 distance plus the average DC-warehouse distance, i.e. average Echelon 1 distance is given by:

\[ \left[ \frac{A_{WH}}{A_{DC}} \left( \frac{A_{DC}}{A_{WH}} \right) + 1 \right] \cdot \frac{\sqrt{2} \cdot A_{WH}}{3} \],

and the transportation cost per unit for Echelon 1 is

\[ \gamma_{DC} \cdot \left[ \frac{A_{WH}}{A_{DC}} \left( \frac{A_{DC}}{A_{WH}} \right) + 1 \right] \cdot \frac{\sqrt{2} \cdot A_{WH}}{3} \].

3.2 Distribution Warehousing and Warehousing Costs

"Distribution" refers to the steps in a logistics network that involve the movement and storage of products from the supplier level to the customer level. "Distribution warehousing", thus, is the activity related to product storage and product flows through and from the firm. Distribution warehousing activities fulfill time and place utilities. Proper goods storage means that customer needs are satisfied promptly and lost sales are kept to a minimum. Strategically placed distribution warehouses can minimize total order filling time for customers. Thus, one of the major roles of distribution warehousing is to be responsive to customer service needs.

Seasonality and uncertainty of demand, combined with production run economies contribute to the need for storage. Storage facilities guard against the reduction in sales by making goods available for sale. The use of the storage function permits the decoupling of production and marketing to increase flexibility, thus enabling the firm to overcome demand fluctuations and uncertainties in the production process. Combating future uncertainties is another need which is filled by storage.

Inventory reporting and recording represent additional services of warehouses. Warehouses implement inventory control policies by providing information about stock receipts, inventory levels and customer feedback. Inventory replenishment points are noted, along with associated inventory costs.

Warehousing is essential in meeting the firm’s present and future production and marketing needs, and also serves the needs of customers and suppliers. Improvements in
customer service performance, lower production costs, smoother production scheduling, protection against uncertainty and increased revenues are possible. The role of distribution warehousing within the logistics network includes facilitating movement flows and coordinating demand and supply functions. A more efficient distribution center results when increased customer service is accompanied by decreased costs.

The performance of a distribution network is usually evaluated based on two dimensions: customer needs and the cost of meeting these customer needs. As mentioned before the way a distribution network is designed has impacts on the following logistics system cost components: inventories, transportation, facilities and handling.

In general, for a single-echelon system as the number of facilities in a distribution network is increased, the resulting inventory costs also increase as shown in the figure below. Limiting the number of facilities and consolidating inventory helps to keep the inventory cost lower.

![Figure 3-1: Relationship between Number of Facilities and Inventory Cost](image)

On the other hand, increasing the number of warehouse or stocking locations decreases the average outbound distance and makes outbound transportation distance a smaller portion of the total distance traveled by the product (outbound transportation is related to sending the product out of a facility, whereas inbound transportation refers to bringing a product into a facility). Therefore, as long as the increase in the number of facilities still maintains the economies of scale in inbound transportation, increasing the number of facilities decreases the total transportation cost (Inbound lot sizes tend to be larger than outbound lot sizes, which implies that outbound transportation cost per unit tends to be higher than inbound transportation cost per unit). However, increasing the number of facilities results in a cost increase at the point where the increase results in small inbound lot sizes and economies of scale is lost for inbound transportation (see Figure 3-2).
As the number of facilities is decreased, the facility operating and handling costs are reduced due to the fact that economies of scale can be exploited due to the consolidation effect.

The behavior of the total logistics cost, which is the sum of the inventory, transportation and facility cost, generally shows a convex pattern with respect to the number of facilities as shown in the figure below.
The above figure suggests that facilities should be added beyond the cost minimizing point only if the increase in revenues due to better response times outweighs the increase in costs because of additional facilities [13].

Warehousing Decisions

The major warehousing decisions are related to the choice of location, size and number of the warehouses in the distribution system.

1. Location: Three basic alternatives exist for warehouse locations: production-oriented, market oriented, and intermediate oriented. Combinations are also possible. Location near production facilities is likely to occur when additional material processing is required because transport costs are lower for the movement of unassembled or unprocessed goods. Considerable differences between production output and immediate demand requirements also would tend to favor a warehouse location near production plants in order to absorb these differences quickly and effectively.

Market-oriented warehouses stem from high customer service requirements. Customers desire to carry minimum inventory but they demand quick replenishment. To meet these needs as well as to minimize local delivery costs, firms will locate distribution warehouses close to their major market areas. Food products, consumer shopping items, building supplies, and drug items are among the products that should be market-positioned in terms of storage. Locating the warehouses near the market is also advantageous from the cost point of view. Movement costs would be reduced by consolidating shipments from the plant to the warehouse and making small deliveries short distances to the retailers. Similarly, different products from geographically-dispersed plants can be dispatched to a warehouse located in a sales concentration area, with subsequent break bulk operations for final delivery.

Increasing revenues and decreasing costs are the basic objectives of establishing market oriented warehouses. Additional sales can be generated when the firm can provide timely delivery and quick inventory replenishment. Increased market penetration is also possible when superior customer service standards can be maintained. Cost economies result from consolidated shipments and short local delivery movements. Distribution costs
can be kept at a minimum when movement economies result from consolidated shipments and short delivery movements.

When neither production nor market considerations are compelling intermediate-oriented warehouses are selected. Geographical dispersion of production plants, coupled with a broad diffusion of markets also favors intermediate location. By locating at an intermediate point firms can minimize their total distribution costs and still provide increased customer service. In summary, warehouse location is a function of the production and marketing characteristics as well as the logistics system of a firm. The greatest determinant, however, is the profit impact, i.e. the combined results of sales generation and cost control.

2. Size and Number: There exists a direct relationship between size and number. Increasing the number of warehouses will reduce the space needs for an individual warehouse, and vice versa. Total size is affected primarily by present and projected volume throughput. Expected sales volumes for inbound and outbound shipments dictate warehouse capacities.

Size requirements can be identified in several ways. Size can refer to cubic measurements (length, width and height), or size can be measured in terms of square feet requirements. It can also be determined on the basis of product throughput, measured in terms of hundredweight, packages, barrels etc.

The number of warehouses required is determined through cost comparisons, although space utilization and customer service requirements also play critical roles. As the number of warehouses increases fixed costs rise because of duplication, and size economies are unattainable. Another assumption is that the degree of utilization affects unit costs because warehouse costs are both fixed and variable. A single warehouse should be able to achieve a higher utilization ratio than several smaller warehouses, even though the total square footage is the same. The greater storage volume in a single warehouse allows for increased storage flexibility. Utilization is defined as the percentage of available space which is occupied on an average basis over a period of time. The higher the utilization ratio, the lower the unit costs.

From a sales and marketing standpoint, the number of warehouses is primarily influenced by customer service requirements. The delivery cycle time is the criterion. The length of time from order placement to merchandise receipt by customers is the performance measure. Marketing considerations would favor many warehouses, widely dispersed, in order to maintain high levels of customer satisfaction. The sales/marketing
department’s desire for high customer service levels produces a classical trade-off situation; that is, the determination of the optimum number of warehouses to maintain in order to provide the most efficient customer service at the least possible cost.

Warehousing costs are infrastructure and operations costs associated with the warehousing function other than the inventory costs. Location, capacity, technologies and operation costs are the main cost drivers of a warehouse or distribution center. Location affects the real estate costs, labor costs and transportation requirements. A great percentage of the cost, however, is due to infrastructure investments such as technologies, cross-docking capabilities, systems and equipment for sorting and handling items. Operation costs define the costs of loading, unloading, sorting and retrieving. These costs usually have economies of scale.

Choice of capacity requires making a trade-off between flexibility and costs. Excess capacity allows the facility to be very flexible and to respond to swings in the demand placed on the facility. However, excess capacity costs money, and therefore, can decrease efficiency.

Let $V$ be the variable facility operating cost per unit flowing through the warehouse. Let the annual fixed facility cost be $F \cdot \xi^\theta$, where $F$ is the fixed cost for operating the facility, $\xi$ is the maximum capacity required by the facility. In practice, the choice of $\xi$ for a warehouse might be based on both the annual volume and the expected maximum stock kept during the year. However, for the purposes of simplicity we will take $\xi$ to be the expected maximum stock in the warehouse. $0 \leq \theta \leq 1$ is the economies-of-scale parameter for the warehouse or the storage facility [20]. Smaller values of $\theta$ imply greater relative economies of scale. Thus, for an expected annual volume of $\mu_D$ and an expected maximum capacity requirement of $\xi$, the warehousing costs for warehouses and distribution centers are given by the relation:

$$\text{Warehousing Cost} = F \cdot \xi^\theta + V \cdot \mu_D, \text{ and}$$

$$\text{Warehousing Cost per unit} = \frac{F \cdot \xi^\theta}{\mu_D} + V.$$

If as in the GOMA model demand has an intensity of $\rho$ per unit area, then the warehousing cost per unit for a facility serving a market area of $A$ is $\frac{F \cdot \xi^\theta}{\rho \cdot A} + V$. If the
GOMA model is defined over the parameter set such that per unit transportation cost proportional to $A^{\beta/2}$ is used, the average transportation and facility operation cost per unit is $f(D) = F \cdot \xi^0 / D + V + \Gamma \cdot \kappa \cdot (D / \rho)^{\beta/2}$ or $f(A) = F \cdot \xi^0 / (\rho \cdot A) + V + \Gamma \cdot \kappa \cdot A^{\beta/2}$.

3.3 Inventory Costs and Replenishment Systems

A logistics system integrates and co-ordinates the elements within the supply chain to ensure the best possible flow of materials and information in order to meet customer requirements in the most efficient manner and at the lowest possible cost. One of the primary material flows is the movement of physical products between stages in the supply chain.

Inventory decisions—one of the most challenging decisions in a logistics system—involve the site where inventory is located, the level of inventory to be carried, reorder policies in terms of quantity and frequency, and the choice of which products to store at each facility. Managing inventory in supply chains is a process that has significant impact on the customer service level and the overall system cost.

Inventory in a supply chain may appear in different forms and locations. “Raw material inventory” that flows from suppliers to manufacturers are the resources required in the production or processing activity of the firm. “Components” or “subassemblies” correspond to items that have not yet reached completion in the production process. “Work-in-process (WIP) inventory” located at the manufacturers is inventory either waiting in the system for processing or being processed. WIP inventories include component inventories and some raw materials inventory. The level of WIP is often used as a measure of the efficiency of a production scheduling system. “Finished product inventory”, also known as “end items”, represents the final products of the production process and is distributed from distribution centers or warehouses to customers. Each of these forms of inventory in the supply chain needs its own inventory control mechanism. However, interactions of the different layers in a supply chain make the determination of these mechanisms a complex process. For this reason, in practice some heuristics have been developed to deal with these complexities.
Inventory is important because it requires a large financial resource commitment. Controlling and reducing inventory costs play an important role in increasing system wide savings and profits. On the other hand, inventory has system-wide benefits as well. Adequate inventory—the ability to give the customers what they want—means increased sales. In addition, inventory facilitates the production, marketing and logistics functions and acts as a safety valve for fluctuations in demand and supply. The basic reasons for holding inventory can be summarized in the following way:

- **Safety valve against uncertainties**: Inventory protects the system against the inherent uncertainties in the supply chain. To begin with, unexpected changes might occur in customer demand. Especially in the last years, the short life cycle of increasing number of products and the increase of the competing products in the market place have increased the uncertainty of customer demand. Significant uncertainty may also reside on the supply side in terms of the quantity and quality of the supply, supplier costs and supply lead times.

- **Economies of scale**: Inventory permits economies of scale from longer production runs and decreased setup costs per unit. In addition to the production function, supply and transportation functions also favor inventory by providing volume discounts that encourage large-size shipments and purchases.

- **Flexibility**: Inventory provides flexibility to marketing by having merchandise for sale without being totally dependent upon production schedules and output. Also, it acts as a decoupling function between production and sales so that the two functions do not act as a deterrent to each other.

- **Speculation and Smoothing/Stabling production**: Inventory allows for stable production throughout the year, especially in the case where products have high seasonality. Building inventories in low-demand periods helps to alleviate the disruptions caused by changing production rates and work force levels. Anticipated price variations also favor building inventory to gain profit. For instance, if the value of a resource is expected to increase, it may be more economical to store items for future use rather than purchasing later at a much higher price.

- **Work-in-Process**: Work-in-process inventory is required in a system to maintain continuity in the manufacturing process. Also, the fact that transportation requires time generates a need for *in-transit or pipeline* inventories. In the case of an overseas production, for instance, high transportation times increase pipeline inventories.
Minimum order-quantity type of constraints related to purchase, production or
distribution of items may also force the system to maintain inventory.

To summarize, inventory control is one of the key elements in logistics
management. Inventory significantly affects investment, sales, operations, costs/profits and
affects other logistics activities in achieving system and corporate objectives. A basic
objective of inventory control is to meet a desired level of customer service consistent with
the associated inventory costs in order to maximize profits. Balancing the trade-offs
between the value created by holding inventory and the costs incurred is a key issue in a
successful inventory policy.

Incorporating inventory costs in the strategic planning of the logistics system
requires the choice of a specific replenishment policy. In most cases, the determination of
an optimal policy is very hard, or even if an optimal policy is found, its complexity may
preclude implementing this policy in practice. In this respect, we will investigate the
effects of adopting replenishment strategies that are commonly used in practice.

Before developing the models, it is helpful to give an overview of some important
terminology relating to the categorization of inventory:

- **On-hand Stock (OHS):** This is the stock which physically exists on the shelf, and thus,
cannot be negative. It determines whether a customer demand is satisfied directly or
not.

- **Net stock (NS):** Net stock is equal to on-hand stock minus the backorders. In the case of
backorders, it can become negative.

- **Inventory position (IP):** The inventory position is defined as net stock plus on-order
stock. The on-order stock represents the stock which has been ordered but not yet
received. The inventory position is a key determinant of when to replenish the
inventory.

- **Safety Stock (SS):** By definition, safety or buffer stock is the average level of net stock
just before a replenishment order is received. A safety stock above zero acts like a
buffer against cases where demand may turn out to be above average during the
effective replenishment lead time. The amount of safety stock to be kept in a system
depends on what happens to demand in case of a stock-out.
In case of a stock-out situation, there are two extreme cases that can occur: complete backordering or complete lost sales. In complete backordering, the demand is backordered and can be filled when enough stock accumulates. Under complete backordering, if demand occurs during the stock-out, the net stock just before the next replenishment arrives is negative. On the other hand, in complete lost sales the customer demand is lost and cannot be met when enough stock has arrived and the net stock will remain at zero level throughout the stock-out period. Thus, the value of safety stock, the average net stock just before a replenishment order is received, is influenced by whether backordering can occur or not [56] (see also E. Silver, D. Pyke and R. Peterson [57]).

3.3.1 Inventory Policies

The purpose of a replenishment control system is to provide answers to the following policy questions:

- How often to determine the inventory status?
- When to place a replenishment order?
- How much to replenish?

It takes resources to determine the inventory status. On the other hand, determining the status less frequently implies that the period over which the system must protect against unforeseen demand is longer. When to place an order involves trade-offs between the costs of ordering early (thus carrying extra stocks) and the costs of inadequate customer service.

In practice, though, demand is highly unlikely to be deterministic. However, the deterministic demand models (see the Appendix for EOQ model) are important because they offer a simple tool for understanding the basic inherent trade-offs, and they usually form the basis for the analysis of the stochastic demand case. In the remainder of this work, we will mainly work with continuous and periodic review policies highly used in practice. The complexity involved in computing the parameters in these policies is not extremely demanding. In addition, their ease of implementation is what makes them so popular in practice.
3.3.2 Continuous versus Periodic Review Policies under Probabilistic Demand

Under probabilistic demand, the determination of the distribution of the lead-time demand is important in finding the replenishment levels. In general, the distribution is first characterized by computing the first and second moments of lead time demand, and then fitting a suitable distribution to these parameters. In this respect, the nature of the lead time—whether it is constant or stochastic—also affects this characterization.

To characterize the demand, we need to make some assumptions related to its behavior. For both continuous and periodic review policies, we assume that demands in non-overlapping intervals are independent and identically normally distributed, which is valid in most cases. Let $\mu_d$ and $\sigma_d^2$ be the mean and variance of demand per unit time (e.g., unit time can be taken as a day). Define the customer service level to be the fraction of replenishment cycles in which the total customer demand is met from stock; stated alternatively, let customer service be the probability of being able to meet all customer demand from stock within a cycle. Assume that the customer service level is determined to be $\alpha$; that is, the probability of stocking out is $1 - \alpha$. To characterize the form of the safety stock to meet this customer service level, it will be assumed that the demand per unit time is normally distributed with mean $\mu_d$ and variance $\sigma_d^2$ as stated before. The form of the safety stock is determined by the behavior of the demand during the lead time. The first and second moments of demand that determine the lead-time-demand distribution depend on whether the lead time is constant or stochastic.

**Case 1: Constant Lead Time**

Let the lead time be $L$ time units. If $D_i$ is the demand that occurs during time unit $i$ and LTD is the lead time demand, then:

$$LTD = \sum_{i=1}^{L} D_i = D_1 + D_2 + \ldots + D_L$$

$$E(LTD) = E(\sum_{i=1}^{L} D_i) = \sum_{i=1}^{L} E(D_i) = L \cdot E(D_1) = L \cdot \mu_d$$

$$Var(LTD) = Var(\sum_{i=1}^{L} D_i) = \sum_{i=1}^{L} Var(D_i) = L \cdot Var(D_1) = L \cdot \sigma_d^2$$

Since the pre-specified customer service level is $\alpha$, we require that:

56
Prob \( \{LT D \geq \text{Safety Stock}\} = 1 - \alpha \)

Assuming that the daily demands are IID (independent and identically distributed) and normally distributed with mean \( \mu_D \) and variance \( \sigma_D^2 \), lead time demand is also normally distributed with mean \( L \cdot \mu_D \) and variance \( L \cdot \sigma_D^2 \). Then, the safety stock (SS) which is kept to protect the system against deviations from the average demand \( L \cdot \mu_D \) is characterized as:

\[
SS = z \cdot \sigma_{LTD} = z \cdot \sigma_D \cdot \sqrt{L},
\]

where \( z = \Phi^{-1}(1 - \alpha) \) and \( \Phi^{-1}(\cdot) \) is the Normal (Gaussian) inverse cumulative distribution function.

Assume the yearly demand is uniformly distributed over an area of \( A \), as in the GOMA model. Let \( \sigma_D^2 = \sigma^2 \cdot A \) and \( \mu_D = \rho \cdot A \), where \( \sigma^2 \) is the variance and \( \rho \) is the demand intensity factor per unit area. This implies that the variance and mean of daily demand are linearly dependent on the market area. In this case:

\[
SS = z \cdot \sigma_{LTD} = z \cdot \sigma \cdot \sqrt{L \cdot A}
\]

**Case 2: Random Lead Time**

The previous case assumed that uncertainty is present only in the lead-time demand. If we introduce an additional uncertainty in the lead-time, it is evident that more safety stock is required to protect against this extra uncertainty. In cases where the pattern of variability is known as in seasonally varying lead times, the safety stock can be accordingly adjusted since the lead time at any given calendar time is known.

Assume that the lead time and demand rate are independent random variables. This is a reasonable approximation to reality except for the cases where high demand is likely to be associated with long lead times because of the heavy workload placed on the supplier, or low demand can be associated with long lead times because the supplier has to wait longer to accumulate sufficient orders for the desired run size. Assuming lead time and demand rate are independent:

\[
LT D = \sum_{i=1}^{\ell} D_i = D_1 + D_2 + ... + D_{\ell}
\]
However, $L$ in this case is a random variable. By using the Law of Iterated Expectations, i.e. $E(X) = E[E(X|Y)]$, the mean of the lead time demand is determined as follows:

$$E(LTD) = E[E(LTD|L)] = E[L\cdot \mu_D] = \mu_D \cdot E(L) = \mu_D \cdot \mu_L$$

In a similar fashion, the variance of the lead time demand can be determined by using the Law of Total Variance, i.e. $Var(X) = E[Var(X|Y)] + Var(E(X|Y))$.

$$Var(LTD) = E[Var(LTD|L)] + Var[E(LTD|L)] = E[L\cdot \sigma_D^2] + Var[L \cdot \mu_D]$$

$$Var(LTD) = \sigma_D^2 \cdot E(L) + \mu_D^2 \cdot Var(L)$$

$$Var(LTD) = \sigma_D^2 \cdot \mu_L + \mu_D^2 \cdot \sigma_L^2$$

In this case, the distribution of lead time demand is not necessarily normal since $L$ is random. However, if a normal distribution is fit to the lead time demand distribution by using the first and second moments given by $E(LTD)$ and $Var(LTD)$, then the safety stock to guard the system against unexpected deviations from the average value of $\mu_L \cdot \mu_D$ can be computed as:

$$SS = z \cdot \sigma_{LTD} = z \cdot \sqrt{\frac{\sigma_D^2 \cdot \mu_L + \mu_D^2 \cdot \sigma_L^2}{}}$$

where $z = \Phi^{-1}(1 - \alpha)$ and $\Phi^{-1}(\cdot)$ is the Normal (Gaussian) inverse cumulative distribution function as before. As in the previous case, if the demand is uniformly distributed over an area of $A$, as in the GOMA model with $\sigma_D^2 = \sigma^2 \cdot A$ and $\mu_D = \rho \cdot A$, safety stock in the case of stochastic lead time turns out to be:

$$SS = z \cdot \sigma_{LTD} = z \cdot \sqrt{\sigma_D^2 \cdot \mu_L + \mu_D^2 \cdot \sigma_L^2} = z \cdot \sqrt{\sigma_D^2 \cdot A \cdot \mu_L + \rho^2 \cdot A^2 \cdot \sigma_L^2}$$

In the following sections, the most commonly used continuous and periodic control systems will be introduced. For continuous review major review policies are:

- Order-Point, Order-Quantity ($s$, $Q$) System, and
- Order-Point, Order-Up-to-Level ($s$, $S$) System.

On the other hand, for periodic review major review policies are:

- Periodic-Review, Order-up-to-Level ($R$, $S$) System, and
- ($R$, $s$, $S$) System.
3.3.2.1 Continuous Review

Under continuous review the inventory position is continuously monitored, and a replenishment order is placed when the inventory position falls below a particular level (the reorder point). The major advantage of continuous review over periodic review is that in general it provides the same level of customer service requiring less safety stock and hence lowers carrying costs. This is due to the fact that the period over which safety protection is required is longer under periodic review since the stock level has the opportunity to drop appreciably between successive review-instants and no ordering action is possible.

Major Continuous Review Policies

Order-Point, Order-Quantity \((s, Q)\) System

In this system a fixed quantity \(Q\) is ordered whenever the inventory position drops below the reorder point \(s\). To trigger an order the inventory position is used since it already includes the on-order stock and takes account of the material already requested but not yet received. Ordering with respect to net stock level might result in unnecessarily large orders when shipments are already due in the immediate future.

The \((s, Q)\) system is simple and has advantages in decreasing the likelihood of error and on part of the supplier it creates predictability for the production requirements. A disadvantage of the \((s, Q)\) system occurs in cases where a replenishment size of order \(Q\) is not sufficient to raise the inventory to the desired level when the transaction that triggers the replenishment is very large.

In this system, the time between orders varies but the amount ordered is fixed. The reorder point needs to be chosen in such a way that sufficient inventory is available to cover the demand over the replenishment lead time \(L\). This model is especially appropriate for high value items and fixed order sizes dictated by the supplier or manufacturing process.

Now, let's characterize the behavior of cost in this system. We assume that replenishment order of size \(Q\) is placed when inventory position is exactly equal to \(s\).
This implies that demand transactions are of unit-size or undershoots below the order-point are of negligible magnitude compared to the total lead time demand.

Assume that if two or more orders are outstanding simultaneously for the same item, they must be received in the same order they were placed. In other words, crossing of orders is not allowed. This assumption is satisfied if the replenishment lead time is constant. If an order is placed at time \( t \) when the inventory position at level \( s \), then all previous orders outstanding at that time must have arrived prior to \( t + L \) (point of time at which the current order arrives) because of the assumption of no crossing of orders. Furthermore, any orders placed after time \( t \) cannot arrive before \( t + L \). This implies no other stock is received by time \( t + L \). Therefore, what determines the service impact of placing the current order when the inventory position at the level \( s \) is whether the lead-time demand exceeds \( s \) or not. Note that if crossing of orders occurs, then the characterization of the service impact of placing the replenishment order when the inventory position is at the level \( s \) becomes more complex.

![Diagram](image)

---

**Figure 3-5: \((s, Q)\) Continuous Review Policy**

From the figure and argument given above, it can be seen that the reorder level \( s \) consists of two components: the expected lead-time demand and the safety stock.

Reorder level \( s \) = Expected Demand during Lead Time \( L \) + Safety Stock (SS)
The first component makes sure that the system is able to meet the expected demand during the lead-time and the second component, i.e. safety stock, protects the system against deviations from the expected demand during lead time. From the analysis where lead time was assumed to be constant:

\[ s = L \cdot \mu_d + z \cdot \sigma_d \cdot \sqrt{L} \]

\[ SS = E(OHS^-) = z \cdot \sigma_{LTD} = z \cdot \sigma_d \cdot \sqrt{L} \]

\[ E(OHS^+) = E(OHS^-) + Q = z \cdot \sigma_d \cdot \sqrt{L} + Q \]

\[ E(OHS) = \frac{E(OHS^-) + E(OHS^+)}{2} = \frac{Q}{2} + z \cdot \sigma_d \cdot \sqrt{L}, \]

where \( E(OHS) \) is the expected on-hand stock, \( E(OHS^-) \) is the expected on-hand stock just before a replenishment order is received and \( E(OHS^+) \) is the expected on-hand stock just after a replenishment order is received. On the other hand, if lead time is stochastic:

Safety Stock = \( SS = E(OHS^-) = z \cdot \sigma_{LTD} = z \cdot \sqrt{\sigma_d^2 \cdot \mu_d + \mu_d^2 \cdot \sigma_L^2} \)

\[ E(OHS^+) = E(OHS^-) + Q = z \cdot \sqrt{\sigma_d^2 \cdot \mu_d + \mu_d^2 \cdot \sigma_L^2} + Q \]

\[ E(OHS) = \frac{E(OHS^-) + E(OHS^+)}{2} = \frac{Q}{2} + z \cdot \sqrt{\sigma_d^2 \cdot \mu_d + \mu_d^2 \cdot \sigma_L^2} \]

Reorder Level = \( s = \mu_d \cdot \mu_d + z \cdot \sqrt{\sigma_d^2 \cdot \mu_d + \mu_d^2 \cdot \sigma_L^2} \)

This analysis assumes that the level of \( Q \) is known; for example, it can be taken to be the economic order quantity \( Q^* = \sqrt{\frac{2 \cdot K \cdot \mu_d}{h}} \).

Figure 3-6: Average Behavior of On-hand Stock in an \((s, Q)\) System
Assuming that the mean rate of demand is constant in time (although demand itself is stochastic), the average inventory cost in an \((s, Q)\) system can be computed by using Figure 3-6 as follows:

\[
E(OHS) = \frac{Q \cdot T}{2} + SS \cdot \frac{T}{2} = \frac{Q}{2} + SS = \frac{Q}{2} + z \cdot \sigma_D \cdot \sqrt{L}
\]

Expected Cost per Replenishment Order = \(K + h \cdot \left(\frac{Q}{2} + SS\right) \cdot T + c \cdot Q\)

Expected Cost per Unit Time = \(\frac{K}{T} + h \cdot \left(\frac{Q}{2} + SS\right) + \frac{c \cdot Q}{T} = \frac{K \cdot \mu_D}{Q} + h \left(\frac{Q}{2} + SS\right) + c \cdot \mu_D\).

\(T\) is the cycle time and is given by \(T = \frac{Q}{\mu_D}\). Plugging in \(Q^* = \sqrt{\frac{2 \cdot K \cdot \mu_D}{h}}\), we get:

Expected Cost per Unit Time = \(\sqrt{2 \cdot K \cdot \mu_D \cdot h + h \cdot SS + c \cdot \mu_D}\),

where \(SS = z \cdot \sigma_D \cdot \sqrt{L}\) if the lead time is constant and \(SS = z \cdot \sqrt{\sigma_D^2 \cdot \mu_L + \mu_D^2 \cdot \sigma_L^2}\) if the lead time is stochastic. \(K\) is the cost per order and \(h\) is the cost of holding one unit of item for one day. This analysis does not include backordering costs in the objective function. However, expected backorders per order in an \((s, Q)\) system can be computed as follows:

Let \(Y\) be the demand during lead time of duration \(L\).

\(\mu_Y = \mu_D \cdot L\) and \(\sigma_Y = \sqrt{\sigma_D^2 \cdot L} = \sigma_D \cdot \sqrt{L}\), if lead time is constant.

\(\mu_Y = \mu_D \cdot \mu_L\) and \(\sigma_Y = \sqrt{\sigma_D^2 \cdot \mu_L + \mu_D^2 \cdot \sigma_L^2}\), if lead time is stochastic.

\[
E[\text{Backorders per replenishment}] = \int_{y=s}^{y=\infty} (y-s) \cdot f_Y(y) \cdot dy.
\]

This integral is known as the “partial loss function”. Assume \(Y\) has a normal distribution with mean \(\mu_Y\) and standard deviation \(\sigma_Y\), and \(s = \mu_Y + z \cdot \sigma_Y\), where \(z\) is the safety factor. The partial loss function can equivalently be expressed as (see the Appendix):
\[ \int_{y=s}^{\infty} (y-s) \cdot f_Y(y) \cdot dy = \sigma_Y \int_{u=z}^{\infty} (u-z) \cdot \phi(u) \cdot du, \]  
where \( \phi(u) \) is the probability density function for the standard normal random variable \( u \). The integral \( \int_{u=z}^{\infty} (u-z) \cdot \phi(u) \cdot du \) is called "the partial expectation function", denoted by \( L(z) \). Thus:

\[ F [\text{Backorders per replenishment}] = \sigma_Y \cdot L(z) \]  
and the fraction of demand not met from stock is \( \frac{\sigma_Y \cdot L(z)}{Q} \). The fill rate is thus \( 1 - \frac{\sigma_Y \cdot L(z)}{Q} \). (Tables for \( L(z) \) can be found in Nahmias [42]. For example for the safety factor \( z = 1.64, L(z) = 0.0211 \).)

**Order-Point, Order-Up-to-Level (s, S) System**

This system also involves continuous review and replenishment is made whenever the inventory position drops below the order point \( s \). However, in contrast to the (s, Q) system the replenishment quantity is variable. The amount of the replenishment order is such that it increases the inventory position to the order-up-to-level \( S \). (s, S) system is identical to (s, Q) system if all demand transactions are unit sized since in this case the replenishment requisition will be made when the inventory level is at \( s \) and a constant amount \( Q = S - s \) will be ordered. The replenishment quantity in an (s, S) system, however, is variable when the transactions can be larger than unit size. The (s, S) system is also referred to as a min-max system because the inventory position is almost between a minimum value of \( s \) and a maximum value of \( S \), except for possible momentary drops below the always reorder point.

The best (s, S) system is no worse than the best (s, Q) system in terms of the total costs of replenishment, storage and shortage. However, the computational effort is prohibitive except for items where potential savings in costs are appreciable. A possible disadvantage of this system is the danger of errors in placing replenishment orders caused by the variable order quantity.

In this system if all transactions are of unit size then every order is of size \( S - s \), and is placed exactly when the inventory position is \( s \). In this case, \( Q = S - s \) and the assumptions of an (s, Q) system hold.
The analysis is complicated in the case of non-unit sized transactions. In addition, order cycles are of random length. If the undershoots (how far below \( s \) the inventory position is located when an order is placed) are neglected, \( s \) and \( S \) can be determined in a simple sequential manner. \( Q = S - s \) is set equal to the economic order quantity. Then, given the value of \( Q \), \( s \) is found by the procedure described in \((s, Q)\) systems. Given \( Q \) and \( s \), \( S \) is set to \( s + Q \).

### 3.3.2.2 Periodic Review

Periodic review is appealing in cases when coordination of replenishments may be attractive for instance when the items are purchased from the same supplier, shipped in the same transportation mode or produced on the same piece of equipment. In these cases, items in a coordinated group can be given the same review period. Periodic review also allows the prediction of the workload in issuing replenishment orders. In addition, it is generally cheaper compared to continuous review policies in terms of reviewing costs and reviewing errors.

### Major Periodic Review Policies

**Periodic-Review, Order-up-to-Level \((R, S)\) System**

In this system, which is also known as a replenishment cycle system, the control policy is to order every \( R \) units of time to raise the inventory position to the level \( S \). The amount ordered may vary among different order cycles whereas the time between successive replenishment requisitions is fixed.

This system is often preferred to order-point systems because of the simplicity and regularity of the periodic review property. In addition, it offers an opportunity every \( R \) units of time to adjust order-up-to-level \( S \), which is desirable if the demand pattern is changing with time. The main disadvantage of this system is that the carrying costs are generally higher than in continuous review systems.
This system is also known as base-stock policy. In base stock policy, at each review instant (every $R$ units of time) the inventory position is reviewed and enough is ordered to raise the inventory position to the target inventory level $S$, i.e. the base-stock level.

![Diagram](Image)

Net stock or both the inventory position and the net stock when they are identical

Inventory position

Figure 3-7: The $(R, S)$ Periodic Review System

Note that the replenishment order placed is equal to the demand between two consecutive reorder points in time. For example, assume that a replenishment order is placed at time $t$. At the next reorder point $t + R$, the inventory position has dropped from the reorder level $S$ by an amount equal to the demand during the period $(t, t + R)$.

Let’s assume as in the continuous review case that if two or more replenishment orders for the same item are simultaneously outstanding, then they are received in the same order in which they were placed; i.e. crossing of orders is not permitted.

Consider a review instant $t$ at which an order of $Y$ is placed to raise the inventory position to the level $S$. The next order $Z$ is not placed till time $t + R$. Suppose order $Z$ arrives at $t + R + L$. Because of the assumption of no crossing of orders, all previous orders including $Y$ have arrived. In other words, all $S$ of the inventory position at time $t$, and no other stock, must have reached the stocking point by a time just before $t + R + L$. Therefore, the service impact of using an order-up-to level of $S$ in placing the order $Y$ is
determined by whether or not the total demand in a review interval plus a replenishment lead time, \( R + L \), exceeds \( S \).

In other words, the level of the inventory position when an order is placed should be enough to cover the demand during a period of \( R + L \) days until the next order arrives. Therefore, the base-stock level \( S \) has two components: the average demand during an interval of \( R + L \) units of time and the safety stock to protect the system against deviations from average demand during a period of length \( R + L \).

Base Stock Level \( S \) = Average Demand during \( (t, t + R + L) \) + Safety Stock (SS)

If lead time is constant:

\[
S = (R + L) \cdot \mu_D + z \cdot \sigma_D \cdot \sqrt{R + L}
\]

\[
SS = E(OHS^-) = z \cdot \sigma_D \cdot \sqrt{R + L}
\]

\[
E(OHS^+ \prime) = E(OHS^-) + E[D(t, t + R + L)] = SS + R \cdot \mu_D
\]

![Figure 3-8: Average Behavior of On-hand Stock in \((R, S)\) System](image)

Assuming that the average demand rate is more or less constant in time (although demand is probabilistic and deviates around this mean), the expected on-hand stock can be calculated as:

\[
E(OHS) = \frac{E(OHS^-) \cdot T + [E(OHS^+) - E(OHS^-)] \cdot T/2}{T} = \frac{E(OHS^-) + E(OHS^+)}{2}
\]

\[
E(OHS) = \frac{E(OHS^-) + E(OHS^+)}{2} = \frac{R \cdot \mu_D}{2} + SS = \frac{R \cdot \mu_D}{2} + z \cdot \sigma_D \cdot \sqrt{R + L}
\]
In the case of a stochastic lead time, let $Y$ be the demand during an interval of length $R + L$. Thus, we have:

$$E(Y) = E[E(Y | L)] = E[\mu_D \cdot (R + L)] = \mu_D \cdot (R + \mu_L)$$

$$Var(Y) = Var[E(Y | L)] + E[Var(Y | L)] = E[\sigma_D^2 \cdot (R + L)] + Var[\mu_D \cdot (R + L)]$$

$$S = \mu_D \cdot (R + \mu_L) + z \cdot \sqrt{\sigma_D^2 \cdot (R + \mu_L) + \mu_D^2 \cdot \sigma_L^2}$$

$$SS = E(OHS^-) = z \cdot \sigma_Y = z \cdot \sqrt{\sigma_D^2 \cdot (R + \mu_L) + \mu_D^2 \cdot \sigma_L^2}$$

$$E(OHS^*) = E(OHS^-) + E[D(t, t + R + L)] = SS + R \cdot \mu_D$$

Using the assumption of constant rate of demand, the expected on-hand stock is as before:

$$E(OHS) = \frac{E(OHS^-) + E(OHS^*)}{2} = \frac{R \cdot \mu_D}{2} + z \cdot \sqrt{\sigma_D^2 \cdot (R + \mu_L) + \mu_D^2 \cdot \sigma_L^2}$$

Note that unlike in the $(s, Q)$ case, the assumption of unit-sized transactions is not needed since in the $(R, Q)$ model we are not concerned with possible undershoots below the reorder point. Let’s choose $R$ such that the order quantity is approximately equal to the economic order quantity. To characterize the average cost of the $(R, Q)$ policy we need to make the additional assumption that the chance of no demand between reviews is negligible; consequently, replenishment is placed at every review. This implies that number of reviews per year and the number of replenishment orders per year are both $1/R$.

From the above analysis we have found the expected on-hand stock to be:

$$E(OHS) = \frac{R \cdot \mu_D}{2} + z \cdot \sqrt{R + L} = \frac{R \cdot \mu_D}{2} + SS$$

**Expected Cost per Order**

$$h \cdot \left( \frac{R \cdot \mu_D}{2} + SS \right) \cdot T + c \cdot R \cdot \mu_D$$

**Expected Cost per Unit Time**

$$\frac{K}{T} + h \cdot \left( \frac{R \cdot \mu_D}{2} + SS \right) + \frac{c \cdot R \cdot \mu_D}{T}, \text{ where } T \text{ is the cycle time } R. \text{ Thus, Expected Cost per Unit Time is given by }$$

$$\frac{K}{R} + h \cdot \left( \frac{R \cdot \mu_D}{2} + SS \right) + c \cdot \mu_D.$$

Since we choose $R$ such that the order quantity is approximately equal to the economic
order quantity \( (Q^*) \), \[ R = \frac{Q^*}{\mu_d} = \sqrt{\frac{2 \cdot K}{h} \cdot \frac{1}{\mu_d}} = \sqrt{\frac{2 \cdot K}{h \cdot \mu_d}}. \] Plugging in this value of \( R \), we obtain the following expression for the cost function:

\[
\text{Expected Cost per Unit Time} = \sqrt{2 \cdot K \cdot \mu_d \cdot h} + h \cdot SS + c \cdot \mu_d, \quad \text{where}
\]

\[ SS = z \cdot \sigma_d \cdot \sqrt{R + L} \text{ if the lead time is constant, and} \]
\[ SS = z \cdot \sqrt{\sigma_d^2 \cdot (R + \mu_L) + \mu_d^2 \cdot \sigma_L^2} \text{ if the lead time is stochastic.} \]

As in the continuous review case, \( K \) is the cost per order and \( h \) is the cost of holding one unit of item for one day. Note that the backorder costs are not included in the objective function. However, expected backorders per cycle can be calculated by following the same procedure described in the continuous case. One difference, however, is that in the \((R, Q)\) model what determines whether a stock-out occurs or not is the demand during a time interval of \( R + L \) as explained before. Thus, expected backorders for the \((R, Q)\) periodic review policy are given by \( \sigma_{R+L} \cdot L(z) \), where \( L(z) = \int_{u=z}^{\infty} (u-z) \cdot \phi(u) \cdot du \) is “the partial expectation function”, and \( \phi(u) \) is the probability density function for the standard normal random variable \( u \) [42].

\((R, s, S)\) System

This system is a combination of \((s, S)\) and \((R, S)\) systems. The inventory position is checked every \( R \) units of time. If the inventory position is at or below \( s \), enough is ordered to raise it to the level of \( S \). If the position is above \( s \), nothing is done until at least the next review instant. The \((s, S)\) system is the special case where \( R = 0 \), and the \((R, S)\) system is the special case when \( s = S - 1 \). \((R, s, S)\) can be viewed as the periodic version of the \((s, S)\) system.

It has been shown that under quite general assumptions concerning the demand pattern and the costs involved, the best \((R, s, S)\) system results in the lowest total replenishment, carrying and order costs [49]. However, the computational effort to obtain the best values of the three parameters is in most cases prohibitive. The difficulty results partly due to the fact that undershoots of the reorder point \( s \) are present even if transactions
are unit-sized. This is because inventory level is observed only every $R$ units of time.

Another aspect of the difficulty is the fact that if we set $Q = S - s$, then the optimum value of $Q$ is not a uni-modal function of the demand rate. In practice, $R$ is chosen large for convenience (e.g. 1 day) even when point-of-sale equipment permits continuous review of the inventory position. Another most common procedure in practice is to set $s = S - 1$. In this case, as mentioned above the analysis reduces to the one of an $(R, S)$ system.
Chapter 4

Strategic Decisions in a Two-Echelon Network

4.1 Inventory in a Two-Echelon Distribution System

When products are distributed over a large geographical region it is often practical to have several local stocking locations that are reasonably close to the final customers. These local inventories may, in turn, receive stock from a higher level site such as a central distribution center. We will consider only a network of facilities comprised of two-levels. The system below is a two-level (or two-echelon) arborescent system, where each inventory has a single predecessor. (The concept of echelon stock is first introduced by Clark and Scarf [14]. Echelon stock of a level is defined as the inventory position of the subsystem consisting of the level itself and all its downstream levels.)

![Diagram of a two-level arborescent system]

Figure 4-1: Two-level Arborescent System

Multi-level methods require some centralized information and control to take advantage of the system structure. In a completely centralized system, all control actions should be carried out by a central decision unit and be based on costs and inventory positions at all facilities. Centralizing inventory reduces both safety stock and average inventory. This is mainly because reallocations can be made in the central stock in the cases where demand from one market area is higher than average while demand in another market area is lower than average. In addition, the higher the variance of demand the
greater is the benefit obtained from centralizing stock because the stock held usually
depends on the average of the demand and the standard deviation of the demand (safety
stock is kept to account for the deviations of demand from average). However, a
disadvantage of a centralized system is the additional costs required to move data between
levels and the complexity of the control policy. In this respect, it is generally more suitable
to limit the degree of centralization. Multi-level methods in most cases can be described as
methods of coordination of local decentralized inventory control systems [2].

Finding the best balance between central stock and local stock is the fundamental
problem in connection with the two-level inventory systems. Inventories at different levels
should support each other. For instance, a large central inventory makes sure that orders
from the lower level are not delayed and may make it possible to reduce the local safety
stocks. Local inventory is advantageous because it is closer to the customers. The central
inventory, on the other hand, can in a way provide service to customers in all locations.
The optimal solution of a multi-level inventory problem depends on several factors such as
the system structure, demand distributions, lead times, and costs.

Material flow from one level to the next one requires a lead time and incurs a set-up
cost, in addition to a variable cost proportional to the quantity. In multi-level environments
the optimal solution does not have a simple structure. Thus, even if an optimal policy
exists and is identified, it would not be easy to implement. In other words, the optimal
policy is no longer optimal or attractive when the managerial effort of implementation is
taken into account. Therefore, the use of simple, cost-effective heuristics such as \((R, S)\)
policies is favorable in practice. Consider the following two-echelon system:
The literature on multi-echelon systems with stochastic demand is extensive. Deuermeier and Shwarz (1981), De Bodt and Graves (1985), Moinzadeh and Lee (1986), Lee and Moinzadeh (1987), Svorones and Zipkin (1988), Badinelli (1992), Axsater (1993), and Chen and Zheng (1994, 1997) are some of the major papers in this field. Chen and Zheng (1994) state that in the case when only the distribution center has a set-up cost the optimal polices are unknown. For this reason, the optimal policy for the more complex system where the warehouses also have a set-up cost is unknown too.

Assume that the entire system is controlled by a single decision maker whose goal is to satisfy the customer demand and to minimize the long-run average system-wide cost. This implies that the decision maker has information on inventory levels both at the warehouses and at the distribution centers.

The total market area to be covered is $M$. In this system each distribution center serves a number of warehouses within its market area and also carries some stock to protect the system against stock-outs. Let Echelon 1 represent the distribution centers (DC) and Echelon 2 represent the warehouses (WH). In addition, let $A_{DC} = A_1$ be the area of the warehouses that one distribution center supplies, and $A_{WH} = A_2$ be the market area covered by a warehouse. So, the number of distribution centers in the system is $n_{DC} = M / A_{DC}$ and the number of warehouses per distribution center is $n_{WH} = A_{DC} / A_{WH}$, where $n_{DC}$ and $n_{WH}$ are non-negative integers. Assuming that no lateral flow of inventory is allowed between
market areas covered by different distribution centers, the system is composed of \( n_{DC} \) independent market units of size \( A_{DC} \). Therefore, the average cost in the system will be equal to the average cost of one such unit. Remember that we make the following assumptions regarding the demand and the cost structure for holding and ordering stock:

- The spatial distribution of yearly demand is uniform with a density of \( \rho \) and variance of \( \sigma^2 \) per unit area. In this case, the mean rate of demand and the uncertainty inherent in the demand process depends on the area of the market that generates the demand. Thus, the first and second moments for the demand generated by a market of area \( A \) are \( \mu_D = \rho \cdot A \) and \( \sigma_D^2 = \sigma^2 \cdot A \).

- Market areas are diamond shaped and the rectilinear distance norm is used. In addition, the facilities are centered within their corresponding market areas. Note that this assumption may not be the optimum choice due to differences in inbound and outbound transportation cost rates. However, it allows for simplicity in calculating the average distances.

- Demand per unit time follows normal distribution, and demand over two non-intersecting regions with identical areas is independent and identically distributed.

- Demand is stationary and independent over non-intersecting time periods.

- Every time the warehouse requests an order from the corresponding distribution center, a fixed cost \( K_{W_H} = K_2 \) and a variable ordering cost of \( c_{w_H} = c_2 \) proportional to the amount ordered is incurred. Analogously, each time a distribution center places an order to the supplier a fixed cost of \( K_{DC} = K_1 \) and a variable ordering cost of \( c_{DC} = c_1 \) proportional to the amount ordered is incurred.

- Inventory holding cost is charged per item per unit time. Assume that the unit inventory holding cost per unit is \( h_{w_H} = h_2 \) at the warehouses and \( h_{DC} = h_1 \) at the distribution centers. Then, the corresponding echelon inventory holding costs for the DCs and the WHs are \( h'_{DC} = h_{DC} = h_1 \) and \( h'_{w_H} = h_1 - h_2 \) assuming that the inventory holding cost for the warehouses is higher than the inventory holding cost at the DCs.

- The lead time from the supplier to the DC is \( L_{DC} = L_1 \) and the lead time from the DC to the warehouses is \( L_{w_H} = L_2 \). In addition, each echelon in the system has a pre-specified customer service level.
• Each echelon has a pre-specified customer service level to limit the occurrence of stock-outs within the system.

• The annual facility cost for a facility with an annual throughput of \(D\) is \(F \cdot \xi^\theta + V \cdot D\), where \(F > 0\) is the fixed cost for operating the facility, \(V \geq 0\) is the variable facility operating cost, \(0 \leq \theta \leq 1\) is the economies-of-scale parameter for facilities (\(\theta = 0\) implies extreme economies of scale), and \(\xi\) is the maximum capacity required by the facility, which we will take to be the maximum inventory held at a stocking facility. Expected annual volume \(D\) is equal to \(\rho \cdot A\), where \(A\) is the market area. Thus, average facility cost per unit is \(F \cdot \xi^\theta / D + V\).

• Unit transport costs are related to the distance traveled, \(d\), by the expression \(\Gamma \cdot \delta^\beta\), where \(\Gamma > 0\) is the transportation cost and \(0 < \beta < 1\) allows for economies-of-distance in transportation costs (if \(\beta = 1\), no economies of distance exist). From the GOMA model, for many combinations of \(\beta\), market shape and distance norm, average transportation cost per unit is \(\Gamma \cdot \kappa \cdot A^{\beta/2}\), where \(\kappa\) is the configuration factor and \(A\) is the market area. For \(\beta = 1\), \(\kappa\) is the average distance from the facility to a random customer for a market area of unit area. For diamond-shaped market area and the rectilinear distance norm \(\kappa = 2^{1-\beta/2} / (2 + \beta)\).

### 4.1.1 Continuous Review Approach

Clark and Scarf [14] show that for the multi-period stochastic inventory problem, \((s, S)\) policy is optimal due to the \(K\)-convexity of the cost functions. Iglehart [32] investigates the optimality of the \((s, S)\) policy in the infinite horizon dynamic inventory problem.

Assuming that the inventory decisions are made by a single decision maker with the objective of minimizing system wide cost, an effective way to manage the system by using continuous review is to use a heuristic inventory policy based on the echelon inventory. Echelon inventory at the DC is equal to the inventory at the warehouse plus the entire inventory in transit to and in stock at the warehouses. The echelon inventory position at the DC is the echelon inventory at the DC plus items ordered by the DC but that have not yet arrived minus the backorders.
An effective approach in managing a two-echelon one DC multi-warehouse system is to manage the warehouses by an \((s, S)\) approach or identically \((s, Q)\) if the transactions are unit sized and to make the distribution center decisions based on the echelon inventory position at the distribution center. Whenever the inventory position falls below the reorder point \(s_{dc}\), an order of \(Q\) units is placed to raise the inventory position up to the level \(S_{dc}\). Whenever the echelon inventory position falls below \(s_{dc}\), an order is placed to raise it to the order-up-to level \(S_{dc}\). From the analysis made for the continuous review policies:

\[
\text{Expected Cost per unit time} = \frac{K \cdot \mu_D}{Q} + h \left( \frac{Q}{2} + SS \right) + c \cdot \mu_D.
\]

Plugging in \(Q^* = \frac{2 \cdot K \cdot \mu_D}{h}\):

\[
\text{Expected Cost per unit time} = \sqrt{2 \cdot K \cdot \mu_D \cdot h + h \cdot SS + c \cdot \mu_D}
\]

\[
\text{Expected Cost per unit time} = \frac{2 \cdot K \cdot \mu_D \cdot h + h \cdot SS + c \cdot \mu_D}{\mu_D}, \text{ where}\]

\[
SS = z \cdot \sigma_d \cdot \sqrt{L} = z \cdot \sigma \cdot \sqrt{A \cdot L}
\]

if the lead time is constant and

\[
SS = z \cdot \sigma \cdot \sqrt{\mu_d^2 + \mu_d^2 \cdot \sigma_l^2}
\]

if the lead time to warehouses is stochastic.

\[
\text{Avg. Echelon 2 Cost per unit} = \eta_{w1} \left[ \frac{1}{\rho} \left( \sqrt{2 \cdot K_{w1} \cdot \mu \cdot h_{w1} + h_{w1} \cdot z \cdot \sigma \cdot \sqrt{L_{w1}}} \right) \cdot A_{w1}^{1/2} + c_{w1} \right]
\]

Once a distribution center places an order it takes \(L_{dc} + L_{w1}\) units of time for the stock to reach the warehouses. Thus, since we are considering the echelon stock, the effective echelon 1 lead time is \(L_{dc} + L_{w1}\). The reorder point for the distribution center is, therefore, composed of two parts: the average aggregate demand (i.e. average Echelon 1 demand) during the effective lead time for Echelon 1 plus the safety stock to protect the system against the deviations of the aggregate demand from the average in \(L_{dc} + L_{w1}\) time units [55]. Therefore,

\[
S_{dc} = Q_{dc} + SS_{dc} = Q_{dc} + z_{dc} \cdot \sigma \cdot \sqrt{A_{dc} \cdot (L_{dc} + L_{w1})}, \text{ where}
\]

\[
Q_{dc} = \sqrt{\frac{2 \cdot K_{dc} \cdot \rho \cdot A_{dc}}{h_{dc}}, \text{ or } Q_{dc} = k \cdot Q_{w1}, \text{ where } k \text{ is subject to optimization.}
\]
Following the same procedure as in the echelon 2 case, the average echelon 1 inventory cost per item per unit time is as follows:

\[
\text{Avg Echelon 1 Cost per unit} = \frac{1}{\rho} \left( \sqrt{2 \cdot K_{DC} \cdot \rho \cdot h_{DC} + h_{DC} \cdot z \cdot \sigma \cdot \sqrt{L_{DC} + L_{WII}}} \right) A_{DC}^{-1/2} + c_{DC}
\]

Note that if the echelon 2 lead time is stochastic then the variance of the aggregate demand within a time period of \( L_{DC} + L_{WII} \) is:

\[
\sigma^2 \cdot A_{DC} \cdot \left( L_{DC} + \mu_{w} \right) + \left( \rho \cdot A_{DC} \right)^2 \cdot \left( \sigma_{w} \right)^2
\]

and thus, the safety stock for the DC is:

\[
z_{DC} \cdot \sqrt{\sigma^2 \cdot A_{DC} \cdot \left( L_{DC} + \mu_{w} \right) + \left( \rho \cdot A_{DC} \right)^2 \cdot \left( \sigma_{w} \right)^2}.
\]

The average echelon 1 cost per unit in this case is:

\[
\sqrt{2 \cdot K_{DC} \cdot \rho \cdot A_{DC} \cdot h_{DC} + h_{DC} \cdot SS_{DC} + c_{DC} \cdot \rho \cdot A_{DC}}
\]

\[
\frac{\rho \cdot A_{DC}}{\rho \cdot A_{DC}}
\]

\[
\text{Avg Echelon 1 Cost per unit} = \frac{\sqrt{2 \cdot K_{DC} \cdot \rho \cdot A_{DC} \cdot h_{DC} + h_{DC} \cdot SS_{DC} + c_{DC} \cdot \rho \cdot A_{DC}}}{\rho \cdot A_{DC}}
\]

4.1.2 Periodic Review Approach

Assume that both echelons apply the \((R, S)\) policy or identically \((R, s, S)\) where \(s = S - 1\) periodic review policy. Let \(R_{WII}\) and \(R_{DC}\) be the reorder periods for the warehouses and distribution centers respectively, and \(S_{WII}\) and \(S_{DC}\) be the corresponding order-up-to levels.

Schwarz (1973) proves that in the deterministic demand case single-cycle policies are optimal for one central stock and \(N\) identical local-stock systems. Schwarz also points out that single cycle policies are near optimal in the more general case. This implies that the reorder period of the distribution center \(R_{DC}\) is an integer multiple of the reorder period of the warehouses \(R_{WII}\), i.e. \(R_{DC} = k \cdot R_{WII}\), where in the deterministic demand case with no backlogs or loss of sales \(k\) is an integer that satisfies:

\[
k \cdot (k - 1) \leq \frac{K_{DC}}{N \cdot K_{WII}} \cdot \frac{h_{WII}}{h_{DC}} \leq k \cdot (k + 1)
\]

77
Although this formula is derived for the deterministic demand case, the insights it provides can be applied to the stochastic demand case. The formula suggests that among the major factors that determine the optimum value of \( k \) are the ratios of inventory holding costs and the fixed ordering costs for local and central facilities.

If the ordering cost of the central facility is higher compared to the fixed order costs of the local facilities, then this provides an incentive for central facilities to order less frequently, which implies ordering in higher quantities and thus holding higher levels of average inventories (i.e. centralizing the inventory). On the other hand, if the inventory holding stock at the local stocks is much less compared to the holding cost at the central facilities, the formula drives \( k \) to be smaller, which implies that the central facility orders more frequently and in smaller quantities which decreases the level of centralized inventory (i.e. decentralizing the inventory).

In a stochastic setting, in addition to all of these factors, one would expect \( k \) to depend on by how much the uncertainty reduces if inventory is centralized, which in turn depends on the uncertainty in the demand process \( (\sigma/\rho) \) and the number of warehouses per distribution center.

In the two-echelon distribution network with stochastic demand, we will assume that a single cycle-policy is characterized by the fact that the central-stock locations, i.e. the distribution centers receive a delivery when the local-stocks, i.e. the warehouses place a replenishment order. In this way, the distribution centers ship immediately when they receive stock, obviating the need to hold it for an extra period of time. Let’s consider such an order cycle and assume that the reorder period for the distribution center is given by the relation \( R_{DC} = k \cdot R_{WF} \).

For multi-level inventory control it is possible to define the stock policy either with respect to “installation stock” or “echelon stock”. When ordering decisions are based exclusively on the installation stock—the inventory position at that installation only, the control is often decentralized since such inventory policies do not require information about the stock levels at other installations. Axsater and Rosling [3] argue that for certain policies, such as \((S, Q)\) rule, the cost effectiveness of the installation stock policy may be limited due to the lack of information about the entire system. To account for such information, in such systems echelon stock should be used rather than installation stock. Echelon inventory position is the sum of the installation stocks at an installation and all of its downstream installations. Axsater and Rosling point out that neither echelon nor
installation stock policies can completely characterize the state of a multi-stage inventory system, and thus, only in special cases the optimal policies are based on either of these concepts. For the setting of two-stage distribution network, we will try to see how these costs behave comparatively to each other.

4.1.2.1 Echelon Stock Approach

As mentioned earlier, for the two-echelon distribution network we assume that both echelons apply the \((R, S)\) policy and that \(R_{DC} = k \cdot R_{WH}\), where the reorder period of the DCs is an integer multiple of the reorder period of the warehouses. The optimal value of \(k\) depends on the inventory cost function and some factors such as uncertainty in the system, number of warehouses per distribution center and the ratios of fixed order costs and inventory holding costs at Echelon 1 and Echelon 2.

Both echelons have a pre-specified service level which determines the safety factors \(z_{WH}\) and \(z_{DC}\) for the warehouses and distribution centers respectively. The choice of service levels at different levels in the system has a significant impact on the costs of the logistics system. In addition, different combinations of the two service levels can lead to identical system service levels. Service level represents the basic trade-off in the inventory system: cost versus increased customer satisfaction. Let \(h_{WH}\) and \(h_{DC}\) be the installation stock holding costs per unit per unit time. Then, the echelon holding costs are \(h'_{DC} = h_{DC}\) and \(h'_{WH} = h_{WH} - h_{DC}\) respectively.

Assume that the distribution center is able to meet the Echelon 2 demand in the first \(k-1\) Echelon 2 order cycles with high certainty (the choice of safety stock and the corresponding stock-out probabilities are defined in the section where Echelon 1 inventory is analyzed). However, the distribution center may run out of stock in the \(k'\)th Echelon 2 replenishment cycle with probability \(1 - \alpha\). Therefore, we can assume that the lead time is stochastic only in the \(k'\)th Echelon 2 order cycle and constant in the rest of the \(k\) cycles. Figure 4-4 illustrates how Echelon 2 inventory behaves if we assume that cycle \(k\) has a stochastic lead time with an expected value of \(\mu_L\) and standard deviation of \(\sigma_L\). If the DC has sufficient stock, the lead time is \(L_2\), whereas if the DC does not have sufficient
stock then the lead time is \( R_{WH} + L_2 \), since the DC ships its next order after \( R_{WH} \) time units.

\[
\text{Leadtime} = \begin{cases} 
L_2 & \text{with prob. } \alpha \\
R_{WH} + L_2 & \text{with prob. } 1 - \alpha 
\end{cases}, \quad \alpha \text{ is the service level of the DC for the last cycle.}
\]

\[
E(\text{Leadtime}) = \mu_L = \alpha \cdot L_2 + (1 - \alpha) \cdot (R_{WH} + L_2) = L_2 + (1 - \alpha) \cdot R_{WH}
\]

\[
E(\text{Leadtime}^2) = \alpha \cdot L_2^2 + (1 - \alpha) \cdot (R_{WH} + L_2)^2
\]

\[
Var(\text{Leadtime}) = E(\text{Leadtime}^2) - E(\text{Leadtime})^2
\]

\[
\sigma_L^2 = \alpha \cdot L_2^2 + (1 - \alpha) \cdot (R_{WH} + L_2)^2 - \left[ \alpha \cdot L_2 + (1 - \alpha) \cdot (R_{WH} + L_2) \right]^2 = \alpha \cdot (1 - \alpha) \cdot R_{WH}
\]

It is interesting to observe that the variance for the lead time at the warehouses is dependent on the reorder period and has a concave shape with respect to the service level at the distribution center (see Figure 4-3) with the maximum variance occurring at the service level of 0.5.

![Figure 4-3: Variance of Lead Time for the Warehouses vs Service Level at the Distribution Center](image)

From the previous discussion about the \((R, S)\) model, we know that both the order-up-to levels and the safety stocks increase when uncertainty is introduced into lead time.
We have shown that the safety stock for constant lead time \((SS_{CLT})\) and stochastic lead time \((SS_{SLT})\) are as follows:

\[
SS_{CLT} = z \cdot \sqrt{\sigma^2 \cdot A_{WH} \cdot (R_{WH} + L_{WH})}
\]

\[
SS_{SLT} = z \cdot \sqrt{\sigma^2 \cdot A_{WH} \cdot (R_{WH} + \mu_L) + \rho^2 \cdot A_{WH} \cdot \sigma_L^2}
\]

The corresponding order-up-to levels are:

\[
S_{SLT} = \mu_D \cdot (\mu_L + R) + SS_{SLT} \quad \text{Order-up-to level for stochastic lead time}
\]

\[
S_{CLT} = \mu_D \cdot (L + R) + SS_{CLT} \quad \text{Order-up-to level for constant lead time}
\]

If we make the analysis based on expected lead time, the inventory at a warehouse looks as in Figure 4-4. Uncertainty in cycle \(k\) requires that a higher order is placed to increase the safety stock and thus disturbs the pattern. From Figure 4-4:

\[
I_b = S_{SLT} - \mu_D \cdot \mu_L \quad \text{is the inventory level at point B, when order is received in cycle } k.
\]

\[
I_b - SS_{SLT} = S_{SLT} - \mu_D \cdot \mu_L - SS_{SLT} = \mu_D \cdot (\mu_L + R_{WH}) + SS_{SLT} - \mu_D \cdot \mu_L - SS_{SLT} = \mu_D \cdot R_{WH}
\]

\[
I_c = S_{CLT} - \mu_D \cdot L \quad \text{is the inventory position at point C, when stock is received in cycle 1.}
\]

\[
I_c - SS_{CLT} = S_{CLT} - \mu_D \cdot L - SS_{CLT} = \mu_D \cdot (L + R_{WH}) + SS_{CLT} - \mu_D \cdot L - SS_{CLT} = \mu_D \cdot R_{WH}
\]

\[
I_A = SS_{CLT} - \mu_D \cdot (\mu_L - L) \quad \text{is the inventory position before stock is received in cycle } k.
\]

\[
SS_{CLT} - I_A = \mu_D \cdot (\mu_L - L) = \mu_D \cdot (1 - \alpha) \cdot R_{WH}
\]

Note that the average order placed for cycle \(k\) is \(\mu_D \cdot R_{WH} + (S_{SLT} - S_{CLT})\), which is equal to \(\mu_D \cdot R_{WH} + [SS_{SLT} - SS_{CLT} + \mu_D \cdot (\mu_L - L)]\).

Thus, inventory holding cost over \(k\) cycles for an echelon 2 facility is given by:

\[
SS_{CLT} + SS_{SLT} - \mu_D \cdot (\mu_L - L) \cdot (\mu_L - L) \cdot h_{WH} + \frac{SS_{SLT} + S_{SLT} - \mu_D \cdot \mu_L}{2} \cdot (R_{WH} - (\mu_L - L)) \cdot h_{WH} + \\
\frac{SS_{CLT} + S_{CLT} - \mu_D \cdot L}{2} \cdot R_{WH} \cdot (k - 1) \cdot h_{WH}
\]
Dividing the inventory holding cost by $k \cdot R$, we get the average inventory holding cost for a warehouse as follows:

\[
\frac{SS_{CLT} + SS_{LT} - \mu_D \cdot (\mu_L - L)}{2} \cdot \frac{(\mu_L - L)}{k \cdot R} \cdot h_{WH}' + \frac{SS_{SLT} + SS_{LT} - \mu_D \cdot \mu_L \cdot (R - (\mu_L - L))}{k \cdot R} \cdot h_{WH}' + \frac{SS_{CLT} + SS_{LT} - \mu_D \cdot L}{2} \cdot \frac{(k - 1)}{k} \cdot h_{WH}' =
\]

\[
\left[ SS_{CLT} - \frac{\mu_D \cdot (\mu_L - L)}{2} \cdot \frac{(\mu_L - L)}{k \cdot R} \cdot h_{WH}' \right] + \left[ SS_{SLT} + \mu_D \cdot \frac{(R - (\mu_L - L))}{k \cdot R} \cdot h_{WH}' + \frac{SS_{CLT} + \mu_D \cdot R}{2} \cdot \frac{(k - 1)}{k} \cdot h_{WH}' =
\]

\[
\left\{ \left( \frac{1}{k} \cdot SS_{SLT} + \frac{k - 1}{k} \cdot SS_{CLT} \right) + \mu_D \cdot \frac{R - (\mu_L - L)}{k \cdot R} \cdot \left( SS_{SLT} - SS_{CLT} + \mu_D \cdot (\mu_L - L + R) \right) \right\} \cdot h_{WH}' =
\]

\[
\left\{ \left( \frac{1}{k} \cdot SS_{SLT} + \frac{k - 1}{k} \cdot SS_{CLT} \right) + \mu_D \cdot \frac{(1-\alpha)}{k} \cdot \left( SS_{SLT} - SS_{CLT} + \mu_D \cdot (2-\alpha) \cdot R \right) \right\} \cdot h_{WH}'
\]
If we assume that \( \mu_L - L = (1 - \alpha) \cdot R \) is small compared to \( k \cdot R \), or \( 1 - \alpha \) small compared to \( k \), the cost reduces to \( \left\{ \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} \right\} + \frac{\mu_D \cdot R}{2} \cdot h'_{WII} \), which is the holding cost if we use a unique safety stock over all cycles by using an average of \( SS_{CLT} \) and \( SS_{SLT} \), where the weights are \( (k-1)/k \) and \( 1/k \) respectively. (\( 1 - \alpha \) is the probability of a stock-out at the DC during the \( k' \)th cycle).

To simplify the rest of the analysis, let’s assume that a unique safety stock is kept over all \( k \) cycles. This approach slightly overestimates the actual inventory holding cost at the warehouses, but gives a simple approach to gain insight when we think in terms of echelon stock. So, from now on we can approximate the average behavior of Echelon 2 stock by \( k \) identical cycles such that:

\[
SS_{WII} = \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT}
\]

for one facility.

\[
Q_{WII} = \mu_D \cdot R_{WII}
\]

Thus, average inventory holding and ordering cost for one facility is as follows:

\[
\frac{K_{WII}}{R} + \left\{ \left( \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} \right) + \frac{\mu_D \cdot R}{2} \cdot h'_{WII} \right\} \cdot \frac{c_{WII} \cdot \mu_D \cdot R}{R}
\]

Dividing by the demand rate, we get the average inventory cost per unit as:

\[
\frac{K_{WII}}{R \cdot \mu_D} + \left\{ \left( \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} \right) + \frac{\mu_D \cdot R}{2} \right\} \cdot h'_{WII} + c_{WII} = \]

\[
\frac{K_{WII}}{R \cdot \rho \cdot A_{WII}} + \left\{ \left( \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} \right) + \frac{\rho \cdot A_{WII} \cdot R}{2} \right\} \cdot h'_{WII} + c_{WII}
\]

Once we have identified Echelon 2 inventory cost, it remains to find the Echelon 1 inventory cost. The system below illustrates the behavior of Echelon 1 and Echelon 2 on-hand stock levels. For simplicity, let’s assume that the single-cycle policy has \( k = 2 \), i.e. \( R_{DC} = 2 \cdot R_{WII} \). Whenever the distribution center receives an order the warehouses place an order so that the distribution center can immediately ship the order received. Warehouses place orders at time \( t \), \( t + R_{WII} \), \( t + 2R_{WII} \), \( t + 3R_{WII} \), ... , \( t + 9R_{WII} \). These points in time also correspond to the shipment times for the distribution center if there is available stock on
hand. An interval of \( L_{\text{WH}} = L_2 \) time units elapses till the warehouses receive the stock that they order. The distribution center, on the other hand, places orders at time points \( t - L_1 \), \( t + R_{\text{DC}} - L_1 \), \( t + 2R_{\text{DC}} - L_1 \), \( t + 3R_{\text{DC}} - L_1 \), and \( t + 4R_{\text{DC}} - L_1 \). \( L_{\text{DC}} = L_1 \) time units elapse till the distribution center receives the stock ordered. This means that the distribution center receives stock at \( t \), \( t + R_{\text{DC}} \), \( t + 2R_{\text{DC}} \), \( t + 3R_{\text{DC}} \), and \( t + 4R_{\text{DC}} \) (as noted earlier these points also correspond to Echelon 2 replenishment instants).

As it can be seen from Figure 4-5, in the case when the distribution center has stock, the time required to ship Echelon 2 order is constant and equal to \( L_{\text{WH}} = L_2 \). On the other hand, during the \( k \)'th Echelon 2 replenishment cycle the distribution center may run out of stock. In this case, the distribution center cannot ship stock immediately. Let’s assume that when the warehouses order at \( t + R_{\text{WH}} \), the distribution center is out of stock. The distribution center reviews the Echelon 1 inventory position at the time instant \( t + R_{\text{DC}} - L_1 \), places an order, receives and ships the order to the warehouses at \( t + R_{\text{DC}} \). The warehouses receive the stock at \( t + R_{\text{DC}} + L_2 = t + 2 \cdot R_{\text{WH}} + L_2 \). Thus, in case of a stock-out at the distribution center (which happens with probability \( 1 - \alpha \)) the Echelon 2 lead-time is \( R_{\text{WH}} + L_2 \).
Figure 4-5: Expected Behavior of Echelon 2 Aggregate Stock Level and Echelon 1 Stock Level 
\((k=2)\)

As mentioned before, each warehouse has enough safety-stock to protect it against deviations from expected demand in \(R_{\text{wII}} + L_2\) units of time. The amount ordered by a warehouse is equal to the expected demand within a market area of \(A_{\text{wII}}\) and a time interval of length \(R_{\text{wII}}\). The demand within a time interval of \(R_{\text{wII}}\) units is normally distributed with a mean of \(R_{\text{wII}} \cdot \rho \cdot A_{\text{wII}}\) and a variance of \(R_{\text{wII}} \cdot \sigma^2 = R_{\text{wII}} \cdot \sigma^2 \cdot A_{\text{wII}}\).

The aggregate order of all warehouses for one Echelon 2 replenishment cycle is a superposition of normally distributed processes, and therefore, is itself normally distributed with mean \(R_{\text{wII}} \cdot \rho \cdot A_{\text{DC}}\) and a variance of \(R_{\text{wII}} \cdot \sigma^2 \cdot A_{\text{DC}}\). This implies that the order placed by a distribution center should cover the average aggregate order for \(k\) Echelon 2 replenishment orders. Due to normality, the aggregate order for \(k\) Echelon 2 orders is
Normally distributed with mean \( k \cdot R_{wh} \cdot \rho \cdot A_{dc} \) and a variance of \( k \cdot R_{wh} \cdot \sigma^2 \cdot A_{dc} \). Let \( S_{dc} \) be the order-up-to level for the distribution center. Once Echelon 1 inventory position is raised to \( S_{dc} \) when the DC places an order, no other stock is received by the DC within the next \( R_{dc} + L_{dc} \) time units. Therefore, Echelon 1 order-up-to level \( S_{dc} \) is composed of two components: the average depletion of Echelon 1 inventory within a time interval of \( R_{dc} + L_{dc} \) units and the safety stock needed to guard against the variation of Echelon 1 demand during this interval.

\[
S_{dc} = \text{Average Depletion of Echelon 1 Inventory during } (t, t + R_{dc} + L_{dc}) + SS_{dc}
\]

\[
S_{dc} = (R_{dc} + L_{dc}) \cdot \rho \cdot A_{dc} + SS_{dc}
\]

\[
E(OHS^*)_{dc} = \text{Average Order of a DC} + SS_{dc} = R_{dc} \cdot \rho \cdot A_{dc} + SS_{dc}
\]

\[
E(OHS^*)_{dc} = SS_{dc}
\]

As mentioned above the average order of a DC covers the aggregate demand of warehouses within \( k \) Echelon 2 replenishment orders and is equal to \( k \cdot R_{wh} \cdot \rho \cdot A_{dc} \) or \( R_{dc} \cdot \rho \cdot A_{dc} \). The safety stock of a distribution center is the amount of stock that is kept to protect the system against the deviations of aggregate warehouse demands within \( R_{dc} + L_{dc} = k \cdot R_{wh} + L_1 \) units of time. Since the lead time from the supplier to the DC is constant, the variance of the aggregate demand during a period of length \( R_{dc} + L_{dc} \) is given by \((R_{dc} + L_{dc}) \cdot \sigma^2 \cdot A_{dc} \). Thus, the safety stock for the distribution center is:

\[
SS_{dc} = z_{dc} \cdot \sqrt{(R_{dc} + L_{dc}) \cdot \sigma^2 \cdot A_{dc}} = z_{dc} \cdot \sigma \cdot \sqrt{(R_{dc} + L_{dc}) \cdot A_{dc}}
\]

Let \( \delta = 1 - \Phi^{-1}(z_{dc}) \) be the stock-out probability over a period of \( R_{dc} + L_1 \). This occurs if the demand during this period exceeds \( S_{dc} = (R_{dc} + L_{dc}) \cdot \rho \cdot A_{dc} + SS_{dc} \). In this time interval, demand is Normally distributed with mean and standard deviation of \((R_{dc} + L_{dc}) \cdot \rho \cdot A_{dc} \cdot \sigma \cdot \sqrt{A_{dc} \cdot (R_{dc} + L_{dc})})\). Similarly, let \( 1 - \alpha \) be the stock-out probability during the \( k \)'th cycle. To have a stock-out at the end of the \( i \)'th cycle, the cumulative demand during \( L_{dc} + i \cdot R_{wh} \) should exceed the order-up to level \( S_{dc} \). Numerical calculations show that the stock-out probability for the first \( k - 1 \) cycles is
practically zero and that \( \delta = 1 - \Phi^{-1}(z_{DC}) \) and \( 1 - \alpha \) are equal to each other. The following figure represents the stock-out probability for each cycle in a system with the following parameters: \( R_{WH} = 0.01 \), \( \rho = 8 \), \( \sigma = 1 \), \( A_{DC} = 100 \), \( z_{DC} = 0.85 \), and \( L_{DC} = 0.1 \).

![Graph showing stock-out probabilities for different cycles.]

Figure 4-6: Stock-out Probabilities for \( k = 7 \) and \( k = 15 \)

\[ (\delta = 1 - \Phi^{-1}(z_{DC}) = 1 - \alpha = 0.1977) \]

To find the Echelon 1 stock, we need to add the safety stock kept at Echelon 2 facilities to the safety stock kept by the central distribution center. Therefore, the safety stock for Echelon 1 is:

\[
SS_{i} = z_{DC} \cdot \sigma \cdot \sqrt{(R_{DC} + L_{DC}) \cdot A_{DC} + n_{WH} \cdot SS_{WH}}, \text{ where}
\]

\[
SS_{WH} = \frac{1}{k} \cdot SS_{SLT} + \frac{(k - 1)}{k} \cdot SS_{CLT} \text{ as before.}
\]

This represents the safety stock that must be present in the system to account for the variations in the Echelon 1 stock depletion. In other words, \( SS_{i} \) is stock either at the distribution center and warehouses, or in-transit from the supplier to Echelon 1 or from Echelon 1 to Echelon 2.

Now, the average Echelon 1 inventory cost can be computed (note that \( h'_{DC} \) and \( h'_{WH} \) are the echelon inventory holding costs).

Echelon 1 Cost per Order = \( K_{DC} + h'_{DC} \cdot (SS_{i} + \frac{Q_{i}}{2}) \cdot R_{DC} + c_{DC} \cdot Q_{i} \), where \( Q_{i} \) is the echelon 1 order quantity given by

\[
Q_{i} = k \cdot R_{WH} \cdot \rho \cdot A_{DC} = R_{DC} \cdot \rho \cdot A_{DC}.
\]
Cost per Unit Time = \( \frac{K_{\text{DC}} + h'_{\text{DC}} \cdot (SS_1 + \frac{Q_1}{2}) \cdot R_{\text{DC}} + c_{\text{DC}} \cdot Q_1}{R_{\text{DC}}} \)

\[ \frac{K_{\text{DC}}}{R_{\text{DC}}} + h'_{\text{DC}} \cdot (SS_1 + \frac{Q_1}{2}) + c_{\text{DC}} \cdot \rho \cdot A_{\text{DC}} \cdot \]

Average Echelon 1 Cost per Unit = \( \frac{K_{\text{DC}}}{R_{\text{DC}} \cdot \rho \cdot A_{\text{DC}}} + \frac{h'_{\text{DC}}}{\rho \cdot A_{\text{DC}}} \cdot (SS_1 + \frac{Q_1}{2}) + c_{\text{DC}} \),

where \( SS_1 = z_{\text{DC}} \cdot \sigma \cdot \sqrt{(R_{\text{DC}} + L_{\text{DC}}) \cdot A_{\text{DC}}} + n_{\text{wh}} \cdot SS_{\text{WH}} \).

4.1.2.2 Installation Stock Approach

When we consider the installation stock rather than the echelon stock, the analysis for Echelon 2 facilities remains the same since the echelon stock and the installation stock are identical for the warehouses. However, the distribution center policy computation is affected due to the fact that the installation stock level at the distribution centers is not identical to the echelon stock level. Note that in the echelon stock approach, the holding cost at the warehouses is defined as the difference between the holding costs of the warehouse and the distribution center in contrast to the case in the installation stock approach, where the warehouses charge their real cost.

Let the reorder period for the DCs be an integer multiple \((k)\) of the warehouses, and define each order placed at intervals of \( R_{\text{wh}} = R \) as an Echelon 2 order (referring to the aggregate order placed by the warehouses). Each Echelon 2 order is normally distributed with a mean of \( R_{\text{wh}} \cdot \rho \cdot A_{\text{DC}} \) and a variance of \( R_{\text{wh}} \cdot \sigma^2 \cdot A_{\text{DC}} \).

This implies that the order placed by a distribution center, which we will refer to as Echelon 1 order, should cover the average aggregate order for \( k \) Echelon 2 replenishment orders and the amount of inventory depletion during a time interval of \( L_{\text{DC}} \) units. As mentioned earlier one Echelon order has an expected value of \( R_{\text{wh}} \cdot \rho \cdot A_{\text{DC}} \); thus, on \( k \) such cycles the expected value is \( k \cdot R_{\text{wh}} \cdot \rho \cdot A_{\text{DC}} \).
Let $z_{DC}$ be the safety factor for the distribution centers. This corresponds to a particular service level, i.e. $\Phi(z_{DC})$, for the length of the period between the time instances when the DC places an order and receives the stock. The stock-out probability during this interval is given by $1 - \Phi(z_{DC})$, where $\Phi(\cdot)$ is the Gaussian cumulative distribution function. Once a DC places an order and raises its inventory position to the order-up-to level, no other stock reaches the DC within the next $L_{DC} + R_{DC}$ time units. This implies that the safety stock held at the DCs should cover the uncertainty in the DC market area within a time period of $L_{DC} + R_{DC}$. Thus, DC safety stock should be:

$$SS_{DC} = z_{DC} \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot \sigma^2 \cdot A_{DC}} = z_{DC} \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}}$$

Similarly, the order-up-to point at the DCs should cover the demand in the DC market area within a time period of $L_{DC} + R_{DC}$ and provide safety stock against possible stock-outs during this period (for simplicity we will refer to these $k$ cycles as an Echelon 1 cycle).

As in the installation stock case, given that the safety stock level is $z_{DC} \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}}$, the probability of stocking out during the first $k-1$ cycles is practically zero. Let $1 - \alpha$ be the probability of a stock-out situation during the $k$th Echelon 2 order cycle. $1 - \alpha$ and $1 - \Phi(z_{DC})$ are identical as before.

When we consider the installation stock:

$$S_{DC} = (k \cdot R_{WH} + L_{DC}) \cdot \rho \cdot A_{DC} + SS_{DC} = k \cdot R_{WH} \cdot \rho \cdot A_{DC} + z_{DC} \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}}$$

The expected behavior of the Echelon 1 inventory can be seen in Figure 4-7. This figure shows that the inventory holding cost at a DC over an Echelon 1 cycle (i.e. $k \cdot R_{WH}$) is $(1 + 2 + \cdots + k - 1) \cdot R_{WH} \cdot Q_1 / k + R_{DC} \cdot SS_{DC}$. Thus, inventory holding and ordering cost over an Echelon 1 cycle is:

$$K_{DC} + h_{DC} \left[ \frac{k \cdot (k-1)}{2} \cdot (R_{WH} \cdot \rho \cdot A_{DC}) \cdot R_{WH} + k \cdot R_{WH} \cdot SS_{DC} \right] + c_{DC} \cdot k \cdot R_{WH} \cdot \rho \cdot A_{DC}$$
Dividing the inventory cost over one Echelon 1 cycle by $R_{DC} = k \cdot R_{WH}$, we get the average inventory holding and ordering cost at the DCs as:

$$\frac{K_{DC}}{k \cdot R_{WH}} + h_{DC} \cdot \left[ \frac{(k-1)}{2} \cdot (R_{WH} \cdot \rho \cdot A_{DC}) + SS_{DC} \right] + c_{DC} \cdot \rho \cdot A_{DC}.$$  

Since $SS_{DC} = z_{DC} \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot \sigma^2 \cdot A_{DC}} = z \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}}$, the average inventory holding and ordering cost at a distribution center is:

$$\frac{K_{DC}}{k \cdot R_{WH}} + h_{DC} \cdot \left[ \frac{(k-1)}{2} \cdot (R_{WH} \cdot \rho \cdot A_{DC}) + z \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}} \right] + c_{DC} \cdot \rho \cdot A_{DC}.$$  

Thus, the average cost per unit is:

$$\frac{K_{DC}}{k \cdot R_{WH} \cdot \rho \cdot A_{DC}} + \frac{h_{DC}}{\rho \cdot A_{DC}} \cdot \left[ \frac{(k-1)}{2} \cdot (R_{WH} \cdot \rho \cdot A_{DC}) + z \cdot \sigma \cdot \sqrt{(k \cdot R_{WH} + L_{DC}) \cdot A_{DC}} \right] + c_{DC}$$
4.2 Total Logistics Cost

Per unit inventory cost for the 1 DC – N warehouse system turns out to be the same under echelon and installation stock approach because of the way the echelon costs are allocated and the definition of the replenishment policy for the installation stock. If we use per unit costs under echelon approach, the average total logistics cost (TLC) for the 1 DC – N warehouse system is the sum of per unit inventory, transportation and facility operating costs:

\[
\text{TLC} = \frac{K_{WH}}{R_{WH} \cdot \rho \cdot A_{WH}} + \left( \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} \right) + \frac{R_{WH} \cdot A_{WH}}{2} \cdot \frac{h_{WH}'}{\rho \cdot A_{WH}} + c_{WH} + \\
\frac{K_{DC}}{k \cdot R_{WH} \cdot \rho \cdot A_{DC}} + \frac{h_{DC}'}{\rho \cdot A_{DC}} \cdot (SS_{1} + \frac{Q_{1}}{2}) + c_{DC} + \\
\Gamma_{WH} \cdot \sqrt{\frac{2}{3}} \cdot A_{WH}^{1/2} + \Gamma_{DC} \cdot \sqrt{\frac{A_{WH}}{A_{DC}}} \cdot \left( \frac{A_{DC}}{A_{WH}} - 1 \right) \cdot \frac{\sqrt{2} \cdot A_{WH}}{3} + \\
\frac{F_{WH} \cdot \xi_{WH}^{\eta_{WH}}}{\rho \cdot A_{WH}} + V_{WH} + \frac{F_{DC} \cdot \xi_{DC}^{\eta_{DC}}}{\rho \cdot A_{DC}} + V_{DC}, \text{ where}
\]

\[
SS_{1} = \rho_{DC} \cdot \sigma_{\sqrt{(R_{DC} + L_{DC}) \cdot A_{DC} + n_{WH} \cdot SS_{WH}}},
\]

\[
SS_{CLT} = \sigma_{\sqrt{\rho \cdot A_{WH} \cdot (R_{WH} + L_{WH})}}^2,
\]

\[
SS_{SLT} = \sigma_{\sqrt{\rho \cdot A_{WH} \cdot (R_{WH} + L_{WH}) + \mu_{L}^2 \cdot A_{WH}^{1/2} \cdot \sigma_{L}^{2}}},
\]

\[
\mu_{L} = \alpha \cdot L_{1} + (1-\alpha) \cdot (R_{WH} + L_{2}) = L_{2} + (1-\alpha) \cdot R_{WH},
\]

\[
\sigma_{L}^{2} = \alpha \cdot L_{1}^{2} + (1-\alpha) \cdot (R_{WH} + L_{2})^{2} - [\alpha \cdot L_{1} + (1-\alpha) \cdot (R_{WH} + L_{2})]^{2} = \alpha \cdot (1-\alpha) \cdot R_{WH},
\]

\[
Q_{1} = Q_{DC} = k \cdot R_{WH} \cdot \rho \cdot A_{DC} = R_{DC} \cdot \rho \cdot A_{DC},
\]

\[
n_{DC} = M / A_{DC} \text{ and } n_{WH} = A_{DC} / A_{WH}.
\]

\[
\xi_{WH} = SS_{WH} + Q_{WH}, \text{ where } SS_{WH} = \frac{SS_{SLT}}{k} + \frac{(k-1)}{k} \cdot SS_{CLT} \text{ and } Q_{WH} = \rho \cdot A_{WH} \cdot R_{WH},
\]

\[
\xi_{DC} = SS_{DC} + Q_{DC}, \text{ where } SS_{DC} = \rho_{DC} \cdot \sigma_{\sqrt{(R_{DC} + L_{DC}) \cdot A_{DC}}},
\]

In calculating the facility capacity, we assume that the capacity for the storage facility is determined by the maximum on-hand inventory. Note that, we consider the case only with one item. If there are more items, one also needs to consider the fact that all items are not going to be at their maximum inventory levels simultaneously. Hence, taking
the total maximum inventory in that case will be an over-estimation for the facility capacity.

Assuming that management has decided on the service level to be provided to the end customer, then at the design stage the above cost involves the unknowns of:

- Echelon market areas \( A_{WH} \) and \( A_{DC} \),
- The service level of the distribution center \( \alpha_{DC} \), and thus \( z_{DC} \),
- The number of Echelon 2 cycles in one Echelon 1 cycle \( (k) \)
- The review period at the warehouses \( R_{WH} \).

As mentioned under the section periodic inventory review, as a heuristic \( R_{WH} \) can be taken such that the order quantity is close to the economic order quantity. In this case the order period turns out to be:

\[
R_{WH} = \frac{\sqrt{2 \cdot K_{WH}}}{h_{WH} \cdot \rho \cdot A_{WH}}.
\]

### 4.2.1 Single-Echelon Distribution System vs. Two-Echelon Distribution System

An important issue in multi-echelon distribution system design is the choice of number of echelons in the system. In this study, we focus on the cost components in a two-echelon system. Another question to be investigated is under what settings the choice of a two-echelon distribution system is preferable over a single-echelon system.

For the constant demand case, Webster and Robinson [62] isolate key determinants of two-echelon superiority. According to their analysis, where they do not consider inventory costs, the superiority of a two-echelon or single-echelon system depends on the ratios of echelon transportation costs over echelon fixed facility costs. They define echelon cost rates for fixed facility costs as \( f_{WH} = F_{WH} \) and \( f_{DC} = F_{DC} - F_{WH} \). In a similar fashion, echelon transportation costs per unit are defined as \( \gamma_{DC} = \Gamma_{DC} \) and \( \gamma_{WH} = \Gamma_{WH} - \Gamma_{DC} \). In these equations \( \Gamma_{DC} \) is the cost of shipping one unit from a DC to the warehouses, and \( \Gamma_{WH} \) is the cost of shipping one unit from a warehouse to the end customer.

A major result that Webster and Robinson find is that if \( \gamma_{WH} / f_{WH} \leq \gamma_{DC} / f_{DC} \), then a single echelon network is optimal. Alternatively, if a two-echelon network is optimal,
then $\gamma_{WH} / f_{WH} > \gamma_{DC} / f_{DC}$. They interpret the ratio $\gamma_{WH} / f_{WH}$ as an Echelon 2 bang-for-buck ratio indicating the per mile transportation savings rate over the cost rate for reducing Echelon 2 distance (each time a warehouse is opened, the average echelon distance that a unit must travel at a rate of $\gamma_{WH}$ is reduced). In a similar fashion, they interpret $\gamma_{DC} / f_{DC}$ as an Echelon 1 bang-for-buck ratio indicating the per-mile transportation savings rate over the cost rate for reducing Echelon 1 distance. If the Echelon 2 bang-for-buck ratio is not more than the Echelon 1 bang-for-buck ratio, then the number of Echelon 2 facilities will not be more than the number of Echelon 1 facilities in the optimal solution, which implies that there is no economic reason to open warehouses.

However, a two-echelon network is not optimal whenever $\gamma_{WH} / f_{WH} > \gamma_{DC} / f_{DC}$ due to two-echelon network inefficiency. Factors contributing to inefficiency include a double handling of the item as it moves through a distribution center and a warehouse, the maintenance of inventory at two levels in the network, an increase in average shipping distance to the customer as shipments travel over a less direct route through a warehouse, and the effect of splitting a long route into two smaller routes when economies of distance are present ($\beta < 1$). Webster and Robinson state that the larger the inefficiency, the larger the difference between $\gamma_{WH} / f_{WH}$ and $\gamma_{DC} / f_{DC}$ before a two-echelon network is optimal. The significance of two-echelon network inefficiency increases with economies-of-scale parameter for the facilities and the difference between variable facility operating costs (it is assumed that for the DC the variable facility cost is greater). On the other hand, inefficiency decreases with $\beta$, i.e. the economies-of-distance parameter.

The single-echelon system is composed of identical facilities that serve identical markets as in the GOMA model. The single-echelon GOMA model with stochastic demand under periodic review has been analyzed by Duran Murrieta [22]. For the single-echelon model, the system is composed of identical facilities that serve their own independent markets that are identical in shape and area. To make the systems comparable in terms of cost, the assumptions regarding the inventory policy and system parameters have to be similar. We will assume that the lead time from the supplier to these facilities is fixed, as was the case for the supplier-DC lead time in the two-echelon case. Again, for compatibility the inventory will be controlled under a periodic review policy.

From the analysis done for the periodic stock policies, the inventory cost over a replenishment cycle of length $R$ is given by:
\[ K + h \cdot (SS + Q/2) \cdot R + c \cdot Q, \text{ where} \]
\[ Q = \mu_d \cdot R = \rho \cdot A \cdot R \text{ and } SS = z \cdot \sigma \cdot \sqrt{A \cdot (R + L)}. \]

Thus, average inventory cost per unit can be computed by finding the cost per unit time and then dividing by the demand rate per unit time. Therefore, the average inventory cost holding and ordering per unit for the single-echelon is:
\[
\frac{K}{R \cdot \rho \cdot A} + \frac{h}{\rho \cdot A} \left( SS + \frac{\rho \cdot A \cdot R}{2} \right) + c.
\]

\( z \) in the above safety stock is the safety factor that determines the service level provided to the end customer. Thus, for the single-echelon case it should be identical with the safety factor at the warehouses under a two-echelon setting. \( L \) is the lead time between the supplier and the storage facility. As mentioned earlier, a practical approach is to choose \( R \) such that the order quantity is approximately equal to the economic order quantity.

In a single-echelon setting we will assume that the market areas are diamond shaped and the distance metric is rectilinear as in the two-echelon case. It is possible to define other market shapes and distance norms as in Erlenkotter [24]. The definition of transportation and facility operating cost per unit remain the same as before. In terms of transportation, the single-echelon model behaves like the warehouses under a two-echelon setting since they face spatially uniform demand over their market areas. Incorporating all cost components, the total logistics cost per unit for the single-echelon turns out to be:

\[
TLC(A) = \frac{K}{R \cdot \rho \cdot A} + \frac{h}{\rho \cdot A} \left( SS(A) + \frac{\rho \cdot A \cdot R}{2} \right) + c + \Gamma \cdot \frac{\sqrt{2}}{3} \cdot A^{1/2} + \frac{F \cdot \xi(A)^8}{\rho \cdot A} + V.
\]

The definition of the parameters in the above equation remains as in the two-echelon case. Remember that \( SS(A) = z \cdot \sigma \cdot \sqrt{A \cdot (R + L)} \) is the safety stock kept by the facility and \( \xi(A) = SS + Q = z \cdot \sigma \cdot \sqrt{A \cdot (R + L)} + \rho \cdot A \cdot R \) represents the maximum amount of inventory held by the storage facility.

Let \( M \) be the total market area and \( A^* \) the optimum market area for one facility, and assume that the lead time \( L \) is independent of the number of facilities, or alternatively the market area of each facility. Thus, the optimum number of facilities in the single-echelon case is \( M/A^* \) (rounded up or down). For instance, consider the case where the
reorder period is chosen such that the order quantity is approximately equal to the economic order quantity:

\[ R = \frac{Q^*}{\mu_D} = \sqrt{\frac{2 \cdot K \cdot \mu_D}{h}} \cdot \frac{1}{\mu_D} = \sqrt{\frac{2 \cdot K}{h \cdot \mu_D}} = \sqrt{\frac{2 \cdot K}{h \cdot \rho \cdot A}}. \]

Then the total logistics cost in terms of the market area reduces to:

\[ TLC(A) = \lambda_{\text{inv}} \cdot A^{-1/2} + \lambda_{\text{transp}} \cdot A^{1/2} + \lambda_{\text{facil}} \cdot A^{\theta/2 - 1} = \]

\[ \left( \frac{h \cdot z \cdot \sigma}{\rho} \cdot \sqrt{R + L} + \frac{2 \cdot h \cdot K}{\rho} \right) \cdot A^{-1/2} + \left( \Gamma \cdot \frac{\sqrt{2}}{3} \right) \cdot A^{1/2} + \frac{F}{\rho} \left( z \cdot \sigma \cdot \sqrt{R + L} + \frac{2 \cdot K \cdot \rho}{h} \right)^{\theta} \cdot A^{\theta/2 - 1} \]

The optimal market area is given by the value \( A^* \) where the marginal increase in per unit costs of transportation is equal to the marginal savings per unit due to economies of scale in inventory and facility costs. In other words:

\[ \left. \frac{\partial TLC(A)}{\partial A} \right|_{A=A^*} = 0 \]

\[-\frac{1}{2} \cdot \lambda_{\text{inv}} \cdot (A^*)^{-3/2} + \frac{1}{2} \cdot \lambda_{\text{transp}} \cdot (A^*)^{-1/2} + \left( \frac{\theta}{2} - 1 \right) \cdot \lambda_{\text{facil}} \cdot (A^*)^{\theta/2 - 2} = 0 \]

The general form of the total logistics cost function with respect to the market area \( A \) is:

\[ TLC(A) = \frac{\alpha_1}{A} + \alpha_2 \cdot \sqrt{A} + \alpha_3 \cdot \sqrt{A} \cdot \alpha_4 \cdot \left( \alpha_5 \cdot \sqrt{A} + \alpha_6 \cdot A \right)^{\theta}, \text{ where} \]

\[ \alpha_1 = \frac{K}{R \cdot \rho}, \quad \alpha_2 = \frac{h \cdot z \cdot \sigma \cdot \sqrt{R + L}}{\rho}, \quad \alpha_3 = \Gamma \cdot \frac{\sqrt{2}}{3}, \quad \alpha_4 = \frac{F}{\rho}, \quad \alpha_5 = z \cdot \sigma \cdot \sqrt{R + L}, \quad \alpha_6 = \rho \cdot R \]

Although this function may be neither convex nor concave over the range of \( A \), a unique local optimum exists. This local optimum is also a global optimum (the constants in the above equation are non-negative). For instance, consider the following graph for the case where \( \alpha_1 = 1, \quad \alpha_2 = 100, \quad \alpha_3 = 100, \quad \alpha_4 = 2, \quad \alpha_5 = 3, \quad \alpha_6 = 1, \quad \theta = 0.4. \)
Figure 4-8: Total Logistics Cost vs. Market Area in a Single-Echelon System

For the single-echelon model, the general behavior of the logistics cost components and the optimum total logistics cost with respect to the number of distribution facilities in the system is as follows:

Figure 4-9: Logistics Cost Components vs. Number of Facilities in a Single-Echelon System (Numerical results for the case where $M = 1000$, $K = 1$, $R = 0.02$, $\rho = 10$, $\sigma = 50$, $h = 0.2$, $L = 0.2$, $\alpha = 0.95$, $\theta = 0.8$, $c = 4$, $\Gamma = 0.2$, $F = 10$, $V = 2$, Optimum Number of Facilities $N^* = 6$)
Equivalently, the cost components can be viewed with respect to the market area. The roles of the cost components are reversed since an increase in market area corresponds to a decrease in the number of facilities. As market area increases the transportation cost per unit increases, whereas the facility and the inventory costs per unit decrease.

Figure 4-10: Logistics Cost Components vs. Number of Facilities in a Single-Echelon System

Let $N^*$ be the optimum number of facilities and $TLC^*$ be the optimum total logistics cost. For the single-echelon system, the following observations can be made with respect to the changes in system parameters:

- Decreasing the fixed order cost $K$ increases $N^*$ and decreases $TLC^*$.
- Decreasing the uncertainty in demand ($\sigma$) increases $N^*$ and decreases $TLC^*$.
- Increasing $\rho$ (demand intensity) increases $N^*$ and decreases $TLC^*$.
- If we decrease both $\rho$ and $\sigma$ such that the same coefficient of variation is maintained, $N^*$ decreases and $TLC^*$ increases.
- Increasing the holding cost decreases $N^*$ and increases $TLC^*$.
- Increasing the lead time $L$ decreases $N^*$ and increases $TLC^*$.
- Decreasing the service level $\alpha$ increases $N^*$ and decreases $TLC^*$.
- Decreasing $\theta$, i.e. increasing economies-of-scale results in lower $TLC^*$. $N^*$ may decrease or remain the same.
- Changing the variable cost per order $c$ or the variable facility cost $V$ does not affect $N^*$, whereas $TLC$ increases if they are increased and decreases otherwise.
- Increasing transportation cost per unit $F$ results in a higher value of $N^*$ and increases $TLC^*$.
- Increasing the fixed facility cost per unit $F$ decreases $N^*$ and increases $TLC^*$. On the other hand, decreasing $F$ results in higher $N^*$ and lower $TLC^*$.
- Increasing the review period $R$ reduces $N^*$ and increases $TLC^*$.

### 4.2.2 Total Logistics Cost in the Two-Echelon Network

Understanding the insights provided by the single-echelon model is important for the analysis of the two-echelon case. Remember that for the two-echelon system the total logistics cost is given by:

$$
TLC = \frac{K_{WH}}{R_{WH} \cdot \rho \cdot A_{WH}} + \left( \frac{1}{k} \cdot SS_{SLT} + \frac{(k-1)}{k} \cdot SS_{CLT} + \frac{\rho \cdot A_{WH} \cdot R_{WH}}{2} \right) \cdot \frac{h'_{WH}}{\rho \cdot A_{WH}} + c_{WH} + \frac{K_{DC}}{k \cdot R_{WH} \cdot \rho \cdot A_{DC}} + \frac{h'_{DC}}{\rho \cdot A_{DC}} \cdot (SS_{1} + \frac{Q_{1}}{2}) + c_{DC} + \frac{1_{WH} \cdot \sqrt{\frac{2}{3}}}{\sqrt{A_{WH}}} + \frac{1_{DC} \cdot \sqrt{\frac{2}{A_{WH}}} \cdot \frac{\sqrt{A_{WH}}}{A_{DC} - 1} \cdot \frac{\sqrt{2 \cdot A_{WH}}}{3} +}{\frac{F_{WH} \cdot \varphi_{WH}}{\rho \cdot A_{WH}} + V_{WH} \cdot \frac{F_{DC} \cdot \varphi_{DC}}{\rho \cdot A_{DC}} + V_{DC},}
$$

where

$$
SS_{1} = z_{DC} \cdot \sigma \cdot \frac{\sqrt{(R_{DC} + L_{DC}) \cdot A_{DC}} + n_{WH} \cdot SS_{WH}}{\sqrt{(R_{WH} + L_{WH})}}
$$

$$
SS_{CLT} = z \cdot \sqrt{\sigma ^{2} \cdot A_{WH} \cdot (R_{WH} + L_{WH})}
$$

$$
SS_{SLT} = z \cdot \sqrt{\sigma ^{2} \cdot A_{WH} \cdot (R_{WH} + \mu_{L}) + \rho \cdot A_{WH} ^{2} \cdot \sigma_{L} ^{2}}
$$

$$
\mu_{L} = \alpha \cdot L_{2} + (1 - \alpha) \cdot (R_{WH} + L_{2}) = L_{2} + (1 - \alpha) \cdot R_{WH},
$$

$$
\sigma_{L} ^{2} = \alpha \cdot L_{2} ^{2} + (1 - \alpha) \cdot (R_{WH} + L_{2}) ^{2} - [\alpha \cdot L_{2} + (1 - \alpha) \cdot (R_{WH} + L_{2})] ^{2} = \alpha \cdot (1 - \alpha) \cdot R_{WH},
$$

$$
Q_{1} = Q_{DC} = k \cdot R_{WH} \cdot \rho \cdot A_{DC} = R_{DC} \cdot \rho \cdot A_{DC},
$$

$$
n_{DC} = M / A_{DC} \text{ and } n_{WH} = A_{DC} / A_{WH},
$$

$$
\varphi_{WH} = SS_{WH} + Q_{WH}, \text{ where } SS_{WH} = \frac{SS_{SLT}}{k} + \frac{(k-1)}{k} \cdot SS_{CLT} \text{ and } Q_{WH} = \rho \cdot A_{WH} \cdot R_{WH},
$$

$$
\varphi_{DC} = SS_{DC} + Q_{DC}, \text{ where } SS_{DC} = z_{DC} \cdot \sigma \cdot \sqrt{(R_{DC} + L_{DC}) \cdot A_{DC}}.
$$
As mentioned earlier, echelon market areas $A_{WH}$ and $A_{DC}$, and the number of Echelon 2 cycles in one Echelon 1 cycle-i.e. $k$ the reorder period multiple-are all subject to optimization at the design stage. This is because the choice of optimum $k$ depends on the corresponding market areas as well as other system parameters.

**Logistics Costs with Respect to $k$ in a Two-echelon Network with Given Market Areas**

To investigate how the system parameters in a two-echelon system affect the optimum value of $k$, let’s consider the case where the market areas and all other parameters are known for the warehouses and distribution centers.

Let’s assume that the reference case has the following system parameters:

$k_{WH}=2$, $k_{DC}=8$, $R_{WH}=0.02$, $\alpha_{WH}=0.95$, $\alpha_{DC}=0.8$, $I_{WH}=0.06$, $L_{DC}=0.01$,

$h'_{WH}=0.01$, $h'_{DC}=1$, $\rho=8$, $\sigma=8$, $F_{WH}=5$, $F_{DC}=10$, $\Gamma_{DC}=0.1$, $\Gamma_{WH}=0.6$, $\theta_{WH}=0.8$,

$\theta_{DC}=0.8$, $c_{WH}=1$, $c_{DC}=2$, $V_{WH}=10$, $V_{DC}=20$, $A_{DC}=200$, $A_{WH}=10$.

For this choice of parameters, the optimum value of $k$ is 2 and the optimum logistics cost is 6.81. The functional forms of the logistics cost components and the total logistic cost function are as follows:
From the above figures, it can be seen that the total logistics cost is convex in $k$ and the inventory cost per unit at a warehouse is monotonically decreasing in $k$. This can be intuitively explained by the fact that as $k$ increases the safety stock for the warehouse approaches the safety stock value when the lead times are constant, i.e. the higher the $k$ value, the closer is the warehouse lead time to being deterministic. Thus, this results in a decrease in the safety stock held at the warehouse implying that the capacity requirements for the warehouse decreases as well, which explains the similar behavior of facility costs per unit at the warehouse. On the other hand, the inventory cost at the DC shows a convex pattern with a local optimum. Moderate increases in $k$ result in the aggregation of the uncertainty and thus may result in lower inventory cost. After $k$ reaches a certain point, the total stock requirements for the DC increase so much that it is not profitable to increase $k$ anymore. As $k$ increases the DC has to hold stock over a longer period. Therefore, the expected max inventory that a DC holds increases, which in turn increases the capacity requirement at the DC. As a result, a higher $k$ value implies a higher fixed facility cost for the DC.

It can be observed from the figures that the optimum $k$ value given by the total inventory cost per unit is higher than the optimum $k$ value given by the total logistics cost function. This occurs due to the effect of the facility cost since the transportation cost per unit is fixed. Therefore, decreasing the effect of the facility cost can reduce the gap
between these $k$ values. Choosing the economies of scale factor for the facilities to be zero also closes this gap since the facility cost per unit becomes flat, thus, making the capacity decision unimportant in terms of cost. Increasing economies of scale—i.e. decreasing $\theta$—results in a higher $k$ value, implying that inventory in the system is pushed to the DC if there are greater economies of scale.

Some other observations that can be made regarding the impact of other cost parameters on $k$ include the fact that increasing the holding cost at the warehouses increases $k$ and pushes the inventory to the distribution center. On the other hand, increasing the holding cost at the DC decreases the optimum $k$ value such that the DC holds less stock now and the inventory is pushed to the warehouses. An increase in the fixed order cost for the DC, however, results in a bigger $k$ value, whereas an increase in the fixed order cost $K_{WH}$ does not affect the $k$ value much for this cost setting. The following figure represents how some of the cost components behave when the ordering cost for the DC is increased 10 times. In this case, the optimum $k$ value increases to 6 and the optimum logistics cost goes up to 7.40.

![Graphs showing Total Logistics Cost vs k, Total Inventory Cost vs k, Inventory Cost at WH vs k, Inventory Cost at DC vs k](image-url)

Figure 4-12: Inventory and Total Logistics Cost per unit when $K_{DC}$ is increased

Increasing the uncertainty in demand, i.e. $\sigma$, results in a decrease in $k$. Higher demand intensity decreases the optimum $k$ value as well, which pushes the inventory to the warehouses since a higher certainty in demand obviates the need to aggregate the inventory
at the DC. Both the demand intensity and the standard deviation have an impact on the degree of uncertainty (i.e. the coefficient of variation $\sigma/\rho$).

If the lead time from the supplier to the distribution centers $L_{dc}$ is increased, the optimum $k$ value increases. The optimum $k$ value is not very sensitive to the lead time from the distribution center to the warehouses, i.e. $L_{wh}$. Note that for this case, the system configuration is given as the market areas of both echelons are known. This result implies that for a given configuration, the lead time from the supplier to the distribution system plays a greater role in determining the deployment of inventory between echelons than the internal transfer time $L_{wh}$ from the DC to the warehouses.

When the warehouse serves a bigger number of warehouses, cost optimization results in a higher optimum $k$ value. This observation implies that in case of higher number of warehouses served by a DC, aggregating the inventory at a central stock may help reduce the logistics costs. Another observation relating to the increase of the number of warehouses is that the smaller the demand intensity the more sensitive $k$ is to the increases in the number of warehouses. In other words, for smaller demand intensity (keeping coefficient of variation constant) the system requires a smaller number of warehouses to reach a particular $k$ value. This might be explained by the fact that lower demand intensity implies higher risk.

Having a bigger review period $R$ at the warehouses results in a lower optimum $k$ value. A higher $R$ value increases the safety stock requirements at the facilities. A lower $k$ value in this case counteracts an additional increase of the safety stock at the DC.

Decreasing the fixed facility cost at the DC or increasing the fixed facility cost at the warehouse increases the optimum value of $k$ (see the graph below). For instance, if the fixed cost for the DC is decreased to 1 then the optimum $k$ turns out to be 4 and the optimum logistics cost goes down to 6.58.
Changing the transportation costs per unit, however, has no impact on optimum $k$ since the corresponding market areas are fixed. Therefore, for the case where we know the market areas, or alternatively the configuration of the distribution system, the optimum value of $k$ is determined by the interplay between the facility and inventory costs per unit. It should be emphasized that an increase in $k$ affects the safety stock at the DC and at the WH in opposite ways. For the warehouses a higher $k$ value lowers the safety stock, whereas for the DC it increases the safety stock. Thinking in terms of safety stock, the competing factors are the increase in the DC inventory cost and the decrease in the inventory holding cost at a warehouse times the number of warehouses.

**Logistics Costs in a Two-echelon Network with Given Market Areas for the DC**

As a further step to gaining insight about the two-echelon system, it is possible to fix the number of distribution centers, alternatively Echelon 1 market area, and observe how the logistic cost and optimum $k$ value evolves with respect to the changes in Echelon 2 market areas. Since the Echelon 1 market area is fixed, it is expected that the cost components for the warehouses behave as in the single-echelon system.

The following case considers a system with the following parameters:
$K_{WH} = 2$, $K_{DC} = 8$, $R_{WH} = 0.01$, $\alpha_{WH} = 0.95$, $\alpha_{DC} = 0.8$, $L_{WH} = 0.06$, $L_{DC} = 0.01$, $h'_{WH} = 5$, $h'_{DC} = 1$, $\rho = 10$, $\sigma = 0.1$, $F_{WH} = 5$, $F_{DC} = 8$, $\Gamma_{DC} = 0.1$, $\Gamma_{WH} = 0.6$, $\theta_{WH} = 0.8$, $\theta_{DC} = 0.8$, $c_{WH} = 1$, $c_{DC} = 2$, $V_{WH} = 3$, $V_{DC} = 2.5$, $A_{DC} = 200$.

For this choice of parameters optimum number of warehouses is 7, optimum $k$ value is 4 and the optimum logistics cost is 11.62.
From the above figures it can be observed that as the number of Echelon 2 facilities increases, the transportation cost per unit for the warehouses decreases due to the fact that they serve a smaller area and thus need lower capacity—just as in the single-echelon case. Inventory cost for the warehouses, on the other hand, increases because the consolidation effect is lost as the market area becomes smaller. Since facility cost is a function of inventory and market area, the facility cost for the warehouses increases at a faster rate than the inventory cost.

Transportation cost per unit for the Echelon 1 facility increases since the distribution center serves more Echelon 1 facilities as $N$ increases. The increase in inventory cost per unit at the DC can be explained by the fact that Echelon 1 safety stock is increased as the number of warehouses is increased (inventory cost at DC represents the Echelon 1 inventory cost). Facility cost on the other hand remains the same since the optimum $k$ value remains the same, which fixes the maximum expected inventory and thus the capacity at the DC.

Another result consistent with the single-echelon case is the fact that the total logistics cost function has a local minimum—which is also the global minimum—with respect to the number of Echelon 2 facilities. Therefore, there exists a unique optimum choice for the number of warehouses. In addition, note that for this case where Echelon 1 market area is fixed, optimum $k$ value is quite insensitive to the number of warehouses served by a DC. This may suggest that Echelon 1 market area has a greater impact on the optimum $k$ value rather than the Echelon 2 market area. The competing factors are facility cost of the warehouses and inventory cost versus transportation cost.

As in the previous case, it is possible to check for the sensitivity of the results with respect to the other system parameters. For instance, increasing the standard deviation of demand results in fewer warehouses and a lower $k$ value, and increases the optimum total logistics cost value. For instance, choosing $\sigma$ to be 10 rather than 0.1 decreases the optimum number of warehouses to 5, decreases the optimum $k$ value to 3 and increases the optimum total logistics cost to 12.26 (see the graphs below). An interesting observation for this case is the fact that when the number of warehouses in the system is increased the optimum $k$ value increases, whereas the overall optimum $k$ value over the range of $n_{WH} = N$ decreases. Note that for this case, a change in the optimum $k$ value also reflects
itself in the inventory and facility cost per unit at the DC since the capacity requirements change with respect to $k$.

![Graphs showing cost components vs. number of WH]

Figure 4-15: Logistics Cost Components When Echelon 1 Areas are fixed and the uncertainty in demand, i.e. $\sigma$, is increased

Remember that when both echelon market areas were fixed increasing $\sigma$ resulted in a lower optimum $k$ value and increasing the number of warehouses resulted in a higher $k$ value, which is consistent with the result here.
Increasing the lead time from the DC to the warehouses, decreases the optimum number of warehouses whereas the optimum $k$ value seems to be unaffected by this change, which was also observed when both echelon areas are fixed. It turns out that increasing the lead time from the supplier to the DC increases the optimum value of $k$ as observed before; however, the optimum number of warehouses is not very sensitive to the change in $L_{DC}$. Another observation is that when the demand density is low, sensitivity to the change in $L_{DC}$ is higher.

Both the optimum number of warehouses and the optimum $k$ value is sensitive to a change in the review period at the Echelon 2 facilities. A decrease in both of these decision variables is observed and the optimal total logistics cost increases as $R$ is increased. For instance, increasing the review period from 0.01 to 2 results in an optimal number of 6 warehouses (which was 7) and an optimal value of 2 for $k$ (which was 4). An interesting observation for the total logistics cost curve is the fact that in this case it turns out to be neither convex nor concave but still the local optimum in $N$ is also global optimum. This result was also observed for the single-echelon case (see Figure 4-16).

![Figure 4-16: Logistics Cost Components When Echelon 1 Areas are fixed and the Review Period $R$ is increased](image)

Finally, let's investigate the sensitivity with respect to fixed order costs and inventory holding costs. When the DC has a higher fixed order cost $K_{DC}$, the optimum
value of \( k \) and the total optimum logistics cost increase. On the other hand, the optimum number of warehouses \( N^* \) is quite insensitive to the change in \( K_{DC} \). In other words, the configuration of the system remains the same but the inventory is pushed to the DC. Increasing the fixed order cost of the warehouses \( K_{WH} \), on the other hand, changes the configuration of the system by decreasing the optimum number of warehouses \( N^* \). Optimum \( k \) value seems to be quite insensitive to the change in \( K_{WH} \), which implies that how the inventory is placed among echelons depends more on \( K_{DC} \) rather than \( K_{WH} \) when the Echelon 1 market area is fixed and the Echelon 2 market area is to be optimized.

Changing the inventory holding cost at Echelon 1 or Echelon 2 has an impact on both the market segmentation, i.e. system configuration, and the deployment of inventory in the system. Increasing the holding cost at the warehouses, for instance, reduces \( N^* \) and increases the optimum \( k \) value, which implies that the inventory is centralized at the DC and the market area of the Echelon 2 facilities is increased to benefit from the inventory consolidation effect. On the contrary, increasing the holding cost at the DC decreases \( N^* \) and \( k \), thus pushing the inventory to the warehouses which now have smaller market areas. Although the change in holding costs affect \( k \) in opposite ways, they both decrease \( N^* \) to make use of the consolidation effect at the Echelon 2 facilities. When the holding costs are increased, at certain points the behavior of \( k \) with respect to the number of warehouses in the system may follow a pattern opposite to the one when \( \sigma \) is increased. The optimum \( k \) value decreases as the number of warehouses is increased for high order cost values (see Figure 4-17).
Figure 4-17: Logistics Cost Components When Echelon 1 Areas are fixed and the order cost for the warehouses is increased to 100.

Logistics Costs in a Two-echelon Network with Given Market Areas for the WH

It is also possible to consider the reverse of the situation where the market areas of the Echelon 1 facilities are fixed. The market areas of the warehouses can be taken as fixed, and the optimization can be carried out with respect to $k$ and the distribution center market
areas. This case may be applicable when there exists already a single-echelon system composed of warehouses and a second echelon is to be added to the distribution system.

Let's again use the same parameter values as in the previous case and choose $A_{wh} = 5$.

$K_{wh} = 2$, $K_{dc} = 8$, $R_{wh} = 0.01$, $\alpha_{wh} = 0.95$, $\alpha_{dc} = 0.8$, $L_{wh} = 0.06$, $L_{dc} = 0.01$, $h'_{wh} = 5$, $h'_{dc} = 1$, $\rho = 10$, $\sigma = 0.1$, $F_{wh} = 5$, $F_{dc} = 8$, $\Gamma_{dc} = 0.1$, $\Gamma_{wh} = 0.6$, $\theta_{wh} = 0.8$, $\theta_{dc} = 0.8$, $c_{wh} = 1$, $c_{dc} = 2$, $V_{wh} = 3$, $V_{dc} = 2.5$.

For these values of system parameters the optimum number of warehouses in an Echelon 1 market area turns out to be 17, optimum $k$ value is 5 and the total logistics cost is 14.06. See the figures below for the functional behavior of the logistics cost components.

Note that in this case since the warehouse market areas are fixed, the transportation cost per unit at the warehouses is constant. On the other hand, the transportation cost per unit at the DC is monotonically increasing in the number of warehouses as expected.

As more warehouses are added to the market area of the DC, or alternatively as the Echelon 1 market area increases, the optimum $k$ values decrease. A lower $k$ value for the warehouses implies that the lead time between the DC and the warehouse is certain for fewer cycles, thus the safety stock is increased. This increases the maximum expected inventory level, and thus, the facility cost at the warehouses. The steps in the cost curves are at the points where the optimum $k$ value changes.

Inventory cost per unit at the DC decreases as the distribution center serves a larger market area (inventory consolidation effect). The facility costs per unit at the distribution center decreases since the area of the DC grows and the maximum inventory level, i.e. capacity requirements drop, as more warehouses are added to the market area of the DC.

The total logistics cost still has a local optimum with respect to the number of warehouses which is also a global optimum for the logistics cost.
Figure 4-18: Logistics Cost Components vs. $N$ (number of warehouses) when Echelon 2 Areas $A_{wh}$ are fixed

Increasing the uncertainty in demand (i.e., $\sigma$) for this case where the Echelon 2 market areas are fixed results in a decrease in the optimum number of warehouses in an Echelon 1 market area ($N^*$) and in a lower optimum $k$ value, which is in accord with the previous observations.
Changing the review period, however, produces functional behaviors that are different than in the above figure. For example, increasing $R$ to 1, increases $N^*$ to 94 and decreases the optimal $k$ value to 2. For big $R$ values, as the number of warehouses are increased the optimal $k$ value increases as opposed to the situation in the base case (see Figure 4-19).

Figure 4-19: Logistics Cost Components vs. $N$ (when $A_{rt}$ is fixed and $R$ is increased)

Increasing the certainty in demand, i.e. increasing the intensity factor $\rho$, results in a 1 DC – $N$ warehouse system with fewer warehouses and a lower optimum $k$ value.
Increasing the intensity factor to 30 gives a total logistics cost curve that is neither convex nor concave; however, as in the previous cases, still a local optimum which is also a global optimum in $N$ exists.

In this system, $N^*$ and the optimal $k$ value are highly insensitive to the changes in $L_{WH}$. Increasing $L_{DC}$ increases $N^*$; however, $k$ is not very sensitive to $L_{DC}$ also. If $L_{DC}$ is increased high enough, then optimum $k$ value finally decreases accompanied by the increase in $N^*$.

Increasing the holding cost for the warehouses reduces $N^*$ and increases the optimum $k$ value, thus centralizing the inventory at the DC. On the other hand, increasing the holding cost at the DC has an opposite effect, i.e. $N^*$ increases and optimum $k$ value decreases, which implies that inventory is pushed to the warehouses.

As in the previous cases, neither $N$ nor $k$ is very sensitive to changes in the fixed warehouse ordering cost $K_{WH}$. However, if it is really increased to a very high level then the optimum $k$ value increases placing the inventory at the DC and $N^*$ increases as well. The system configuration and the placement of inventory is more sensitive to changes in $K_{DC}$ than in $K_{WH}$. Increasing $K_{DC}$ increases both $N^*$ and the optimum $k$.

The effects of changing the fixed facility costs and per unit transportation costs are as expected. For instance, if the fixed facility cost per unit for the warehouses $F_{WH}$ is increased, then $N^*$ decreases and optimum $k$ increases creating a more centralized system. On the other hand, if $F_{DC}$ is increased the opposite occurs and $N^*$ increases and $k^*$ decreases, which results in a more decentralized system.

Increasing the transportation cost per unit at the DC has the same effect as increasing the fixed facility costs per unit at the warehouses: $N^*$ decreases and $k^*$ increases. As expected, since the warehouse market areas are fixed changing the transportation cost per unit for the warehouses has no impact on $N^*$ and $k^*$.

A final observation relates to the changes in the service level of the distribution centers. If the service level at the Echelon 1 facilities is decreased, $N^*$ decreases and $k^*$ increases, which means that when uncertainty at the upstream of the warehouses is high, it is more cost-effective to centralize the inventory.

The previous cases show that the total logistics cost function in the two-echelon system is not necessarily convex over the decision variable space. Note that in practice there exist budget constraints that provide an upper bound on the number of distribution
centers and warehouses that can be opened. For a given market area, one can find the optimum system configuration and the optimum deployment of inventory within the two-echelon distribution network by exhausting all possible combinations of fixed number of warehouses and distribution centers, finding the corresponding optimum k value for each and then choosing the minimum cost configuration. However, this may result in a very huge number of enumerations.

Another approach is fixing k first and optimizing with respect to $A_{WH}$ and $A_{DC}$. The space for k is one dimensional and most probably requires considerably fewer enumerations. For the case where k is fixed the total logistics cost function assumes the following analytical form (assuming $\theta = 0$, i.e. there exists extreme economies of scale in the facility operating costs):

$$TLC(A_{WH}, A_{DC}) = \frac{\beta_1}{A_{WH}} + \beta_2 \cdot \sqrt[\frac{\beta_3}{A_{WH}}} + \frac{\beta_5}{\sqrt{A_{WH}}} + \beta_6 \cdot \sqrt{A_{WH}} + \frac{\beta_7}{A_{DC}} + \beta_8 \cdot \frac{1}{\sqrt{A_{DC}}} + \beta_9 \cdot \frac{A_{WH}}{\sqrt{A_{DC}}}$$

where

$$0 < A_{WH} < A_{DC} \quad \text{and} \quad 0 < A_{DC} < M,$$

and M is the total market area.

From the single-echelon case studied before, we already know that the functional form

$$TLC(A) = \frac{\alpha_1}{A} + \frac{\alpha_2}{\sqrt{A}} + \alpha_3 \cdot \sqrt{A} + \alpha_4 \cdot \left(\frac{\alpha_5 \cdot \sqrt{A} + \alpha_6 \cdot A}{A}\right)^\theta$$

is neither convex nor concave over the range of A. Therefore, the above cost function is not guaranteed to be convex over the space of $A_{WH}$ and $A_{DC}$.

To get some insight about the behavior of the optimum system configuration and stock allocation, optimum values for the number of Echelon 1 and Echelon 2 facilities as well as the optimum k value were found by considering all possible combinations. In the following example, the upper bounds for the number of distribution centers, number of warehouses and k value were set respectively to 30, 90 and 30. The choice of the parameter values are summarized in the following table. The first entry in the table corresponds to the optimum system configuration in the base case.
Table 4-1: Sensitivity Analysis with Respect to Parameters in a Two-Echelon Network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Base Case</th>
<th>New Value</th>
<th>N-dc</th>
<th>N-wh</th>
<th>k</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>12.96</td>
</tr>
<tr>
<td>$K_{WH}$</td>
<td>1</td>
<td>100</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>18.24</td>
</tr>
<tr>
<td>$K_{DC}$</td>
<td>40</td>
<td>4,000</td>
<td>1</td>
<td>49</td>
<td>16</td>
<td>17.78</td>
</tr>
<tr>
<td>$R_{WH}$</td>
<td>0.01</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>289.42</td>
</tr>
<tr>
<td>$L_{WH}$</td>
<td>0.06</td>
<td>6000</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>17.00</td>
</tr>
<tr>
<td>$L_{DC}$</td>
<td>0.01</td>
<td>1000</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>13.08</td>
</tr>
<tr>
<td>$h_{WH}$</td>
<td>10</td>
<td>1,000</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>20.15</td>
</tr>
<tr>
<td>$h_{DC}$</td>
<td>1</td>
<td>1,000</td>
<td>1</td>
<td>44</td>
<td>1</td>
<td>24.18</td>
</tr>
<tr>
<td>$\rho$</td>
<td>10</td>
<td>1,000</td>
<td>26</td>
<td>27</td>
<td>1</td>
<td>10.08</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>22.10</td>
</tr>
<tr>
<td>$F_{WH}$</td>
<td>50</td>
<td>5,000</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>40.53</td>
</tr>
<tr>
<td>$F_{DC}$</td>
<td>80</td>
<td>8,000</td>
<td>1</td>
<td>48</td>
<td>1</td>
<td>44.84</td>
</tr>
<tr>
<td>$\Gamma_{WH}$</td>
<td>0.6</td>
<td>600</td>
<td>30</td>
<td>90</td>
<td>6</td>
<td>214.23</td>
</tr>
<tr>
<td>$\Gamma_{DC}$</td>
<td>0.1</td>
<td>100</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>14.42</td>
</tr>
<tr>
<td>$\theta_{WH}$</td>
<td>0.8</td>
<td>0.1</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>12.68</td>
</tr>
<tr>
<td>$\theta_{DC}$</td>
<td>0.8</td>
<td>0.1</td>
<td>11</td>
<td>4</td>
<td>29</td>
<td>11.60</td>
</tr>
<tr>
<td>$V_{WH}$</td>
<td>3</td>
<td>30</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>39.96</td>
</tr>
<tr>
<td>$V_{DC}$</td>
<td>2.5</td>
<td>25</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>35.46</td>
</tr>
<tr>
<td>$c_{WH}$</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>21.96</td>
</tr>
<tr>
<td>$c_{DC}$</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>30.96</td>
</tr>
<tr>
<td>$\alpha_{WH}$</td>
<td>0.95</td>
<td>0.65</td>
<td>2</td>
<td>25</td>
<td>2</td>
<td>12.84</td>
</tr>
<tr>
<td>$\alpha_{DC}$</td>
<td>0.85</td>
<td>0.55</td>
<td>2</td>
<td>24</td>
<td>2</td>
<td>13.01</td>
</tr>
</tbody>
</table>

As expected, it is possible to observe that the parameters $V_{DC}$, $V_{WH}$, $c_{DC}$ and $c_{WH}$ affect only the value of the total logistics cost and do not have any influence on the optimum market area segmentation or stock allocation. It turns out that the system is also quite insensitive to the value of $L_{DC}$, which was assumed to be constant. However, increasing the lead time from the supplier to the DCs results in a higher logistics cost.

Increasing the fixed ordering cost for the warehouses ($K_{WH}$) decreases the number of warehouses within an Echelon 1 market area. Increasing $K_{DC}$, on the other hand, pushes the inventory to the first echelon and increase the number of warehouses per DC.

From the table, it can be observed that factors that push the inventory to the distribution centers include increasing the holding cost at the warehouses ($h_{WH}$), increasing the fixed warehouse facility cost ($F_{WH}$), increasing the transportation costs per unit and
decreasing $\theta_{DC}$ (i.e. increasing economies of scale for the distribution centers). On the other hand, increases in $R_{w1}$, $h_{DC}$, $\rho$, $\sigma$, and $F_{DC}$ decrease the optimum $k$ value and thus push the inventory to the warehouses.

The table also shows that an increase in $K_{DC}$, $h_{DC}$, and $F_{DC}$ results in an increase in the number of warehouses and a decrease in the number of distribution centers. Interestingly, as the demand intensity increases and the standard deviation is kept constant optimum market areas of both echelons become smaller, in other words, number of facilities at both echelons increases. On the other hand, increasing the standard deviation of demand, results in fewer facilities at both echelons.

The configuration of the system seems to be more sensitive to the service level of the warehouses rather than the service level of the distribution centers.
Chapter 5

Conclusions and Further Research Opportunities

In this research we analyzed a two-echelon distribution network design, where the first echelon is composed of distribution centers and the second echelon is made up of warehouses. The effects of market segmentation and inventory deployment were investigated in terms of their effect on the total system logistics cost that incorporates inventory holding and ordering costs, transportation costs and facility operating costs. The optimum system configuration and stock allocation is thus found by considering these cost components simultaneously rather than making sequential optimization for each separate cost component. The GOMA Model developed by Erlenkotter [24] was extended to incorporate inventory costs and stochasticity of demand and lead time faced by the warehouses. In addition, a two-echelon distribution network was analyzed rather than a single-echelon system as in the GOMA mode, and sensitivity of the total cost function with respect to different system parameters was investigated.

Although for the single-echelon case it is relatively easy to come up with an optimum solution, for the two-echelon networks it is hard to come up with an analytic solution for the optimum number of facilities in each echelon and the optimum $k$ value that determines how stock is allocated between these two-echelons. We analyze how some system parameters such as the intensity and uncertainty of demand, lead times, reorder periods, holding and ordering costs, transportation and facility costs affect the system configuration and the inventory deployment within a single and two-echelon logistics network.

In our analysis, we considered a single-item distribution system, and lateral shipments between the market areas of the distribution centers were not allowed. As a further research, it is possible to extend the single-item network to a multi-item network
and see how this affects the market segmentation and inventory deployment within the market area. The constraint of no lateral shipments may be relaxed as well but this would result in a highly complicated analysis. In our analysis, the demand distribution was assumed to be normal and the market area shapes for all facilities were taken to be identical. One possible extension is to consider the problem under a different demand structure and possibly allow for non-identical market areas.

Another interesting future research includes the analysis of the service levels at the warehouses and distribution centers, and the identification of different combinations for these levels that lead to identical system service levels.
Appendix

I. The Economic Order Quantity (EOQ) Model

The classical Economic Lot Size Model provides a framework where the simple trade-offs between ordering and storage costs can be seen. This model considers a facility that faces a constant demand for a single item and places orders from another in the distribution network which has sufficient or unlimited quantity of the product. The assumptions of this model are:

- Demand is constant at a rate of $D$ items per unit time.
- Order quantities are fixed at $Q$ items per order.
- A fixed cost of $K$ is incurred each time an order is placed.
- A linear inventory holding cost $h$ is accrued for every unit held in inventory per unit time.
- Initial inventory and the lead time is zero.
- The planning horizon is infinite.

![Diagram](image)

Figure A-1: Inventory versus time in the EOQ Model

An optimal policy satisfies the “Zero Inventory Ordering Property”, which states that every order is received exactly when the inventory level drops to zero. Average total cost per unit of time is $KD/Q + hQ/2$. The optimal order quantity which is referred to as the Economic Order Quantity (EOQ) is given by:
Cost per cycle = $K + h \cdot \frac{Q}{2} \cdot T + c \cdot Q$, where $T$ is the length of a cycle ($T = Q/D$).

Dividing by $T$ we obtain cost per unit time $C(Q)$ as a function of $Q$.

$$C(Q) = \frac{K \cdot D}{Q} + h \cdot \frac{Q}{2} + c \cdot Q$$

Taking the derivative of $C(Q)$ and setting the result equal to zero yields the first-order condition $\frac{dC(Q)}{dQ} = \frac{h}{2} = \frac{K \cdot D}{Q^2} = 0$. The second-order condition $\frac{d^2C(Q)}{dQ^2} = -\frac{2 \cdot K \cdot D}{Q^3}$ shows that the second derivative is positive for any positive $Q$, and hence, $C(Q)$ is a convex function in $Q$, which implies that the first order conditions are sufficient to find the optimum $Q$. Solving the first-order conditions yields: $Q^* = \sqrt{\frac{2KD}{h}}$.

This quantity is the economic order quantity (EOQ), or the economic lot size. A key point in the EOQ model is that the total ordering and holding costs are relatively stable around the economic order quantity. For this reason, a firm is better served by ordering a convenient lot size close to the economic order quantity rather than the precise EOQ.

II. Expected Shortages per Cycle (Continuous Review)

Given that the lead-time demand $Y$ has a normal distribution with mean $\mu_Y$ and standard deviation $\sigma_Y$, and $s = \mu_Y + z \cdot \sigma_Y = \mu_Y + SS$, where $z$ is the safety factor, the expected backorders per replenishment order are as follows:

$$E[\text{Backorders per replenishment}] = \int_{y=s}^{\infty} (y-s) \cdot f_Y(y) \cdot dy = \int_{y=\mu_Y+SS}^{\infty} (y-\mu_Y-SS) \cdot f_Y(y) \cdot dy.$$

Since demand is normally distributed with mean $\mu_Y$ and standard deviation $\sigma_Y$, the random variable $U = (Y - \mu_Y) / \sigma_Y$ has a standard normal distribution. Make the substitution $u = (y - \mu_Y) / \sigma_Y$. This implies that $dy = du \cdot \sigma_Y$. Thus,

$$\int_{y=\mu_Y+SS}^{\infty} (y-\mu_Y-SS) \cdot f_Y(y) \cdot dy = \int_{u=SS/\sigma_Y}^{\infty} (u \cdot \sigma_Y - SS) \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_Y} \cdot \exp\left(-\frac{u^2}{2}\right) \cdot \sigma_Y \cdot du.$$
\[
= \sigma_y \int_{u=z}^{\infty} \left( u - \frac{SS}{\sigma_y} \right) \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp \left( -\frac{u^2}{2} \right) \cdot du = \sigma_y \int_{u=z}^{\infty} (u - z) \cdot \phi(u) \cdot du,
\]

where \( \phi(u) \) is the probability density function for a standard normal random variable. The integral \( \int_{u=z}^{\infty} (u - z) \cdot \phi(u) \cdot du \) is called "the partial expectation function", denoted by \( L(z) \).

Therefore, the expected backorders per order cycle are \( \sigma_y \cdot L(z) \).

III. Expected Distribution Center to Warehouse Distance

![Two-echelon Configuration Diagram](image)

Figure A-2: Two-echelon Configuration

(DC is denoted by a dark circle, and WHs denoted by a light circle)

A diamond defines the set of points equidistant from a facility and is the optimal market area shape for the rectilinear norm [35]. Consider the two-echelon configuration above. This represents one independent unit within the total market area \( M \). Remember that warehouses serve a spatially-uniform demand within their market areas. Given centered
facilities and constant, identical market areas for the warehouses, the average Echelon 2
distance, i.e. the average distance between a warehouse and the customers it serves is
equivalent to the average distance expression in the single-echelon GOMA model:

\[
\text{Average Echelon 2 distance} = ED_2 = \sqrt{2 \cdot A_{wh}} / 3.
\]

Average distance from a DC to the warehouses within the Echelon 1 market area
depends on the distance between adjacent facilities and the value of \( n_{wh} = \frac{A_{dc}}{A_{wh}} \), i.e. the number of warehouses. Let’s assume that there exists a warehouse co-located with the
DC, otherwise to find \( n_{wh} \), all we need to do is to subtract one from \( A_{dc} / A_{wh} \). This, however, does not affect the average distance calculation.

Given a warehouse market area of \( A_{wh} \), the distance between adjacent facilities is
\( \sqrt{2 \cdot A_{wh}} \) or equivalently \( 3 \cdot ED_2 \). Webster and Robinson state that configurations that are
efficient with respect to echelon 1 distances have diamond shaped market areas. For a
diamond shaped market area, \( A_{dc} / A_{wh} \) takes the values of 1, 9, 25, 49, ..., or from the
figure:

\[
A_{dc} / A_{wh} = 1 + 8 \cdot \sum_{i=1}^{N} i.
\]

\( N \) represents the number of warehouse markets between the DC and the outermost
warehouse in the Echelon 1 market area. Let’s refer to warehouses equidistant from the DC
as a layer. Thus, the first layer has 8 warehouses, the second layer has 16, the third layer
has 24, and finally the \( N \)th layer has \( 8 \cdot N \) warehouses. Thus, the total number of
warehouses is:

\[
1 + 8 + 2 \cdot 8 + 3 \cdot 8 + ... + N \cdot 8 = 1 + 8 \cdot \sum_{i=1}^{N} i = 1 + 4 \cdot N \cdot (N + 1), \quad \text{and}
\]

\[
N = \frac{\sqrt{A_{dc} / A_{wh}} - 1}{2}.
\]

The average distance from a DC to the WHs is:

\[
\frac{1}{(A_{dc} / A_{wh})} \cdot \left\{ 0 \cdot ED_2 + 8 \cdot 3ED_2 + 16 \cdot 6ED_2 + 24 \cdot 9ED_2 + ... + (8 \cdot N) \cdot (3 \cdot N)ED_2 \right\} = \\
\frac{1}{(A_{dc} / A_{wh})} \cdot 24ED_2 \cdot \left( 1 + 2^2 + 3^2 + ... + N^2 \right) = \\
\frac{1}{(A_{dc} / A_{wh})} \cdot 24ED_2 \cdot \sum_{i=1}^{N} i^2 = 
\]

122
\[
\frac{1}{(A_{DC} / A_{WH})} \cdot 24ED_2 \cdot \frac{N \cdot (N+1) \cdot (2N+1)}{6}.
\]

Plugging in \( N = \frac{\sqrt{A_{DC} / A_{WH}} - 1}{2} \) and \( ED_2 = \sqrt{2 \cdot A_{WH}} / 3 \):

\[
\frac{1}{(A_{DC} / A_{WH})} \cdot 24ED_2 \cdot \frac{N \cdot (N+1) \cdot (2N+1)}{6} =
\]

\[
\frac{4}{(A_{DC} / A_{WH})} \cdot \frac{\sqrt{2 \cdot A_{WH}}}{3} \cdot \frac{\sqrt{A_{DC} / A_{WH}} - 1}{2} \cdot \frac{\sqrt{A_{DC} / A_{WH}} + 1}{2} \cdot \sqrt{A_{DC} / A_{WH}} =
\]

\[
\frac{(A_{DC} / A_{WH} - 1)}{\left(\frac{\sqrt{A_{DC} / A_{WH}}}{3}\right)} = (A_{DC} / A_{WH} - 1) \cdot \left(\frac{A_{WH}}{A_{DC}}\right)^{1/2} \cdot \sqrt{2 \cdot A_{WH}} / 3.
\]

Therefore, average DC-WH distance is given by \((A_{DC} / A_{WH} - 1) \cdot \left(\frac{A_{WH}}{A_{DC}}\right)^{1/2} \cdot \sqrt{2 \cdot A_{WH}} / 3\). This value is exact for a diamond shaped market areas and can be used as an approximation otherwise [62].
Bibliography


[22] Duran Murrieta, R., Effects of Supply Chain Strategy in Distribution Networks, MIT


