Inventory Optimization in a Retail Multi-Echelon Environment

by

Rintiya Arkaresvimun

B.Eng. Industrial Engineering Chulalongkorn University, 2002

Submitted to the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Logistics

at the

Massachusetts Institute of Technology

June 2008

© **2008** Rintiya Arkaresvimun All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this document in whole or in part.

Signature of Author **.......................** Master of Engineering in Logistics Program, Engineering Systems Division May 6, **2008**

Certified by...........

Dr. Larry Lapide Director, Demand Management, MIT Center for Transportation and Logistics $\frac{1}{\sqrt{1}}$ / Thesis Supervisor

A ccepted **by...............................** Prof. Yossi Sheffi

Professor, Engineering Systems Division Professor, Civil and Environmental Engineering Department Director, Center for Transportation and Logistics Director, Engineering Systems Division

RCHNVES

Inventory Optimization in a Retail Multi-Echelon Environment

By

Rintiya Arkaresvimun Submitted to the Engineering Systems Division on May 9, 2008 in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Logistics

ABSTRACT

The objective of the study is to find an optimal inventory distribution in a retail three-echelon environment, consisting of a supplier, a **DC,** and stores. An inventory model is built **by** replicating the echelons' periodic, order-up-to-level policies with all echelons' transactions integrated. Network carrying cost is set as an objective function, while the store target service level and the store's minimum order-up-to-levels are set as constraints. **A** heuristic approach, that combines the optimization and simulation methods, is used to find the optimal inventory distribution. The results show that the optimal network carrying cost can be achieved **by** having low inventory and low service level at the **DC.** In addition, the impact of the echelons' deviations from the optimal policies as well as the impact of the upstream echelon's service disruptions on the other echelons confirms the interrelation between the echelons in the network. The analyses also illustrate that high target service level can be accomplished **by** keeping high inventory at the stores and the **DC.**

Thesis Supervisor: Dr.Larry Lapide Title: Director, Demand Management, MIT Center for Transportation and Logistics

ACKNOWLEDGEMENT

I would like to thank Dr.Larry Lapide for his guidance and support on my thesis. His expertise in supply chain consulting really contributes to the success of my thesis.

I would like to thank Dr.Chris Caplice for giving me helpful advice on the inventory simulation. I have also used a lot of knowledge that he taught in the logistics system class in my thesis.

I appreciated the cooperation and support from the people in my thesis sponsor company. I have learned a lot about the supply chain practices in the retail business. Their recommendations also helped improve the quality of my thesis.

I would like to give special thanks for all my classmates who have been very supportive. Their comments on my thesis and presentation were extremely valuable. I will miss the time that we have worked really hard together in the MLOG lab.

Lastly, I would like to thank my family for their continuous support and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Retailers are concerned with product availability on the shelf. Customers are now becoming less willing to wait for a product to be available and continued unacceptable levels of out-of-stocks may result in the loss of customers. This fact is supported **by** a study from **AC** Neilson which indicates that 20 percent of out-of-stock situations result in store switching and lost sales (Vuyk,C., **2003).** It is a logistics task to respond to changes in customers' behavior. However, logistics elements are significantly expensive. Focusing on customer service can confine the drive for operational efficiency and cost reduction. For example, companies may hold too much inventory in warehouses to serve unexpected demand. Therefore, it is necessary to carry out the logistics task effectively and efficiently through the right allocation of resources in the supply chain.

This study focuses on finding the right allocation of inventory in a retail multiechelon network given a desired service level. In this context, the "right" allocation means an optimal inventory level at each echelon that enables minimal total network inventory carrying cost, while still allowing the retailers to achieve a target service level.

The research is based on a case study of RetailCo, a leading pharmacy and convenience store chain in the United States, and SupplierCo, a big manufacturer of privatelabel products. The supply chain network in the study consists of **3** echelons: SupplierCo's warehouse, RetailCo's distribution center **(DC),** and RetailCo's stores. **A** simulation model replicating each echelon's inventory policy is built and decisions on inventory level are made **by** minimizing network inventory carrying cost. Compromising on stores' service level and

shelf availability are unacceptable to RetailCo, and are thus set as constraints in the model. The model is developed using a multi-echelon optimization approach, in which every echelon's transactions are integrated. This approach allows us to analyze the impact of one echelon's inventory policy and service level on those of the other echelons. It also enables us to determine the impact of service disruptions of the upstream echelons on the downstream echelons. Three stock keeping units (SKU) with different average daily demand are selected to measure the impact of sales volume on the inventory level and inventory distribution in the network. Some attributes such as delivery frequency and lead time are also set as parameters in the model, enabling us to measure the effect of changes in these parameters on the optimal inventory distribution.

1.1 RetailCo's Supply Chain Network Overview

RetailCo is a leading pharmacy and convenience store chain having 14 DCs supplying products to approximately 6,200 stores across the US. SupplierCo is a large private-label manufacturer of pharmaceuticals and nutritional products supplying a number of retailers in the US. In this study, only one DC and all one hundred stores that it supplies are selected. The scope of the study starts from SupplierCo's warehouse and ends at stores as presented in Figure 1.

Figure 1 : RetailCo's supply chain network in the model

1.2 SupplierCo's Current Practice

SupplierCo sells private-label products to a number of retailers. Products sold to different retailers are identical but become unique once they are tagged with the retailers' brands. Private-label products usually require long manufacturing lead time, especially those that SupplierCo procures from second-tier manufacturers. Lead time of the products supplied to RetailCo ranges from 14 to 84 days. In addition, most items need to be quarantined after production to make sure that they meet the FDA's requirements. This quality inspection period can range from 2 to 10 days depending on the type of product. Failure of the products to pass the quality inspection means scrapping of the entire production lot.

SupplierCo manages finished good inventory in three forms: component, non-labeled product, and labeled product. Component inventory is in the form of non-packaged items. Non-labeled product inventory is in the form of packaged items without customer labels, while labeled product inventory consists of items that are tagged with customers' unique labels. Labeled product inventory is replenished from non-labeled product inventory and non-

labeled product inventory is filled from component inventory, which must be produced at least at a minimum production quantity. This study only focuses on labeled product inventory that is unique to RetailCo since other forms of inventory are shared among SupplierCo's customers.

For labeled product items sold to RetailCo, SupplierCo employs an order-up-to-level inventory policy with daily review. An order is generated once the inventory falls below the order-up-to-level with the ordered quantity in multiple of case pack quantity.

The current order-up-to-level is 12 weeks of supply which is based on an inventory agreement between SupplierCo and RetailCo. For seasonal items, inventory is built in advance before periods of high demand and the order-up-to-level can be much higher than 12 weeks.

SupplierCo works with RetailCo on a Collaborative, Planning, Forecasting and Replenishment (CPFR) program, in which it has access to RetailCo's DC inventory management system. The orders are created automatically by the system, and are reviewed by SupplierCo on a fixed schedule basis. The schedule to review the order can be different for each DC. Inventory is reserved on the same day that SupplierCo receives the order. If inventory is insufficient to fulfill all the incoming orders from RetailCo's DCs, each order is filled with the same proportion of available inventory to total incoming order. However, the order can still be filled if it has not yet been delivered.

Shipments are also delivered on a fixed schedule basis and the schedule for each DC can be different. Figure 2 shows an example of a fixed schedule appointment agreed upon between the DC and the stores. It is noted that the delivery lead times can be different if orders are made on different days in a week. The days highlighted in blue represent the ordering days and those highlighted in grey represent the delivery days.

Figure 2 : An example of fixed schedule appointment agreed upon between DC and store

1.3 DC's Current Practice

The DC uses a periodic review, order-up-to-level inventory policy. An order-up-tolevel is generated by the inventory management system by considering safety stock, transportation lead time, and ordering cycle. The ordering cycle is the greater of two ordering cycles or review period alternatives. One is the ordering cycle agreed upon between RetailCo and SupplierCo. The other is the economic ordering cycle. Orders are automatically generated by the system up to the target quantity or order-up-to-level quantity. SupplierCo has access to the system and retrieves the orders on a fixed schedule basis. Shipments from SupplierCo also arrive at the DC on a fixed schedule, which is normally longer than the actual transportation lead time.

Orders from the stores are received on a fixed schedule basis and the schedule can be different by store. The inventory is checked and reserved for the orders at the end of the day that the DC gets the order. There is no exact allocation rule in the case of insufficient inventory.

1.4 Stores' Current Practice

Stores manage inventory using a periodic review, order-up-to level policy. Each store has a fixed ordering and receiving schedule. Normally the schedules are set by the DC to balance limited transportation capacity and workload at the DC. Frequency of delivery for

each store can also be different. Stores with high volume usually have higher frequency of delivery. The order-up-to level is calculated daily by the store inventory management system based on the sum of forecasted sales and safety stock. Sales forecast is computed using a moving average approach, while safety stock is computed from volatilities of forecast errors during review period plus lead time. The period between ordering day and receiving day for each store can be different based on the fixed schedule that the DC agreed on with stores. Service level is calculated weekly at a SKU-echelon level by dividing the number of stores experiencing out-of-stock at the end of the week by total number of stores. Target service level is set and compared against actual service level.

CHAPTER 2

LITERATURE REVIEWS

The study focuses on optimizing the inventory of low demand items in a retail multiechelon environment. The model is built using a multi-echelon optimization to simulate and find an optimal inventory distribution in the network. The multi-echelon optimization is preferable and gives more practical and optimal solution than the single-echelon optimization. This chapter compares 2 methods, describing drawbacks of single-echelon optimization, and the benefits of multi-echelon optimization. In single echelon optimization, it is assumed that an upper echelon can offer unlimited supply. In reality, there may be service disruption from equipment breakdown or scarcity of raw material supply, resulting in failure to offer **100%** service level. Failure of the upstream echelon to serve demand of lower echelon may result in declining service level at the lower echelon. This chapter describes the relationship between an echelon's service level and the other echelons' service level.

There are several methods which are used in multi-echelon optimization. Each method is described and the drawbacks from using each method are also mentioned in this chapter. In this study, we build the model **by** replicating RetailCo's and SupplierCo's current inventory policy. **All** echelons use order-up-to-level with different review period. This chapter provides better understanding of how periodic review, order-up-to-level policy works and when it is used.

2.1 Single Echelon VS Multi-Echelon Optimization

A report from Evant Inc. (2003) indicates the problems of single-echelon optimization to achieve true network inventory optimization. The problems are caused by unawareness of the impact of replenishment strategies applied to one echelon on the other echelons. Singleechelon optimization can bring about excess safety stock and suboptimal inventory allocation. Customer service can fail despite excess inventory in the network due to inventory misallocation. Stock out at the customer-facing locations can occur in spite of more than acceptable service level between echelons.

Hausman and Erkip (1994) describe 2 approaches used for developing inventory policies: Independent single-echelon and multi-echelon inventory control. In the first approach, each echelon is responsible for its own stocking policy, regardless of the others' policies. Once the lower echelons determine their policies, the results from their policies are combined and used as demand for the upper echelon. The upper echelon then develops its single-echelon inventory policy using its own performance objective.

In the multi-echelon approach, all inventory control parameters are determined simultaneously by considering the interrelationship between echelons. The performance goal, such as fill rate or service level, at each echelon is related to those at the others. The multiechelon policies always create results that dominate single-echelon policies when there is no managerial or organizational issue involved. These managerial and organizational issues are independent performance evaluation, job satisfaction, and motivation. Considerable savings of multi-echelon policies over single-echelon policies are shown in Figure 3.

Figure 3: Total inventory investment versus expected emergency backorder days/year/ warehouse (Source : Hausman and Erkip (1994))

Silver, Pyke, and Peterson (1998) mention 3 serious flaws when using single-echelon inventory approach in a multi-echelon situation. First, in single-echelon approach, it is assumed that the upper echelons have enough stock to fulfill the lower echelons' demand. This is usually not true in practice since upper echelons cannot have infinite supply. Second, the approach ignores the cost implications of one echelon's inventory policy on the other echelons in the network. Third, the approach fails to reduce the bullwhip effect. Even if the demand of the end-item has low variability, the orders placed further up the network can become larger and less frequent, thus creating higher variability. Under the single-echelon

approach, the upper echelons might end up carrying a large safety stock to protect against the infrequent demands.

2.2 Relationship of echelons' service levels in multi-echelon environment

In a multi-echelon environment, the service level offered by one echelon can have a direct impact on the service level perceived by the end customers. The number of out-ofstock occasions experienced at the warehouse affects the number of shortages at the retailers. Diks, De Kok, and Lagodimos (1996) discuss that one of the main challenges of cost-efficient and effective supply chain management is to determine the target service level at each echelon so that the network target service level can be achieved at minimum cost. The authors also explain that service measures are needed as a means to obtain direct information on the physical performance of the supply chain because shortage cost is usually difficult to obtain in real world. The authors distinguish between 2 different types of performance measures: internal and external performance measures. Internal performance measures are related to the service provided to internal customer, while the latter are those received by external customer at the customer-facing points. The authors also mention extensive numerical experiments and simulations, which reveal that cost optimal policies under service level constraint are mostly achieved by low stocks at intermediate stages.

2.3 Method used in multi-echelon optimization

According to Evant Inc. (2003), there are 3 approaches commonly used to solve multi-echelon inventory problems: Sequential Approach, Distribution Resources Planning (DRP), and Multi-echelon Optimization.

The sequential approach splits a multi-echelon environment into individual echelons and uses a single-echelon approach to optimize each echelon separately. Product demand at each echelon is considered independent from demand at other echelons. Demand can be calculated either **by** using historical orders from lower echelon or **by** passing up the customer demand to the upper echelon. This approach can lead to lack of visibility up and down the demand chain. Demand at the upper level can be distorted from the bullwhip effect because each echelon develops demand forecast separately. The approach also ignores the impact of one echelon's changes in replenishment strategies on other echelons. Figure 4 shows the modeling of sequential approach.

Figure 4 : Sequential approach

Evant Inc. (2003) describes DRP approach as an extension to the Material Requirement Planning (MRP) approach in production planning. Demand for the product in the upper echelon is dependent on demand in the lower echelon. In this approach, net requirements at lower echelon are calculated from forecasted end-customer demand, safety stock, and inventory status; then it is offset by the lead time from the upper echelon to the lower echelon. The sum of the time-phased net requirement from all points in the lower echelon is then passed up to the upper echelon to replenish itself. The major drawback is that

this approach does not determine safety stock. Instead, the safety stock decision is generally made by an unscientific approach, leading to excess inventory. Lack of correlation between echelon's safety stocks also makes it impractical to optimize network inventory. In addition, inventory cost is barely considered in this approach; therefore cost trade-off has to be considered manually. This approach also does not offer network visibility and true network optimization.

The multi-echelon approach determines the inventory policy for each echelon at the network level in a single optimization model. The objective is to minimize total network inventory, while meeting the end-customer's service level. Primary demands are used to drive forecasts in all echelons, thus reducing the bullwhip effect from passed-up demand. Decision on replenishment at each echelon takes into account demand and lead time variations of not only the immediate upper echelon but also all the upper echelons. Since this method uses a single model to find the optimal result, it offers synchronized order strategies, leading to the most optimal result, compared to the first two approaches. Figure 5 shows the modeling of the multi-echelon approach.

Figure 5: Multi-echelon approach

2.4 Periodic review, order-up-to-level (R,S) System

Silver, Pyke, and Peterson (1998) explain that a periodic review, order-up-to-level policy is commonly used when multiple items are ordered from the same supplier or when resource sharing is required. In this system, an order is created every R time units to lift the inventory position up to S. Figure 6 graphically shows the typical inventory behavior of the system.

Figure 6 : (R,S) System

The review period (R) is often determined by the external factors such as frequency of delivery. The order-up-to-level (S) is calculated based on the expected demand over the review interval (R) plus a replenishment lead time (L) and safety stock.

(R,S) system offers significant saving of coordination effort required to manage the orders and shipments of multiple items; however this system might result in higher carrying costs than continuous review systems.

CHAPTER 3

METHODOLOGY

This chapter describes the development of the inventory model and the method used to find the network optimal inventory distribution in a retail three-echelon environment, comprising of a supplier's warehouse, a retailer's DC, and retail stores. To achieve this end, a combination of simulation and optimization approach is used.

Section 3.1 identifies the SKUs and the store segments used in the analysis. Based on the selected SKUs and stores, section 3.2 summarizes the types of data collected and how they are used in the model. The demand data is tested in Section 3.3 to confirm that it fits a Poisson distribution. Random daily demand is then generated and used as inputs in the model. Section 3.4 explains the development of the inventory model and lists all the equations used to calculate the values in the model. A heuristic approach is then developed and used to find the optimal network carrying cost.

3.1 Scope of Analysis

RetailCo is concerned with finding the optimal network inventory distribution of their private-label products, manufactured by SupplierCo. SKUs under the private-label category are ranked by average daily demand and divided into 3 groups: 1) High volume, 2) Medium volume, and 3) Low volume. SKUs with volume lower than the $25th$ percentile are categorized as "Low volume" SKUs; those with volume between the $25th$ and $75th$ percentile are "Medium volume" SKUs; those with volume higher than the $75th$ percentile are "High

volume" SKUs. One SKU is randomly selected from each group to represent its entire group. List of selected SKUs are presented in Table 1.

Table 1 : List of selected SKUs

To simplify the model, only one SupplierCo's warehouse and one RetailCo's DC are selected. Due to limitations of the Excel spreadsheet, we decrease the number of the decision variables in the model by aggregating all stores into segments.

All one hundred stores supplied by the selected DC are grouped into 12 segments and one store is randomly selected from each segment to represent its segment in the model. From this method, we make an assumption that all stores in the same segment have identical demand for the same product item. Segmentation is performed based on characteristics and criteria provided by RetailCo. These characteristics include 1) store size in square feet, 2) dollar sales volume, and 3) frequency of delivery from the DC to the stores. Table 2 shows the characteristics and criteria used to segment the stores.

Table 2 **:** Characteristics and criteria for store segmentation

From the above criteria, twelve store segments are obtained. A single store is then randomly selected from each segment to act as a proxy for its segment in the model. Table 3 shows the twelve store-segments and the number of stores in each segment.

	Store Size in Square Feet			
Dollar Sales Frequency Volume of Delivery Volume	Very Small	Small	Normal	Large
Low		25		
Mid		30	15	
High				

Table 3 : Store-segments and the number of stores in segments

3.2 Data Collection

Six-week daily point-of-sales (POS) data and one year worth of weekly demand data of the selected SKUs at the representative stores were extracted from RetailCo's store frontend system. These sets of data are used to generate random daily demand data, which is described further in section 3.4.

Other attributes of the selected SKUs are also collected to be set as parameters in the model. These attributes include the SKUs' minimum presentation quantities at each store, case pack quantity required by SupplierCo, manufacturing and quality inspection lead times at SupplierCo, and unit cost at RetailCo. Minimum presentation quantity is defined as the quantity of the products facing on the shelf or the minimum quantity that should always be on the shelf. Cost at SupplierCo is confidential and cannot be obtained; therefore, estimated value is used.

The current order-up-to-levels at the stores, the DC, and the supplier were also collected. These policies are later input in the model to estimate the current network

performance in term of service level and average inventory level. The performance of the current inventory policies is then compared to that of the optimal policies obtained from the model. The causes of differences in the results are analyzed and areas for improvement are recommended for the current inventory policies.

3.3 Random Daily Demand Generation

The collected six-week daily demand data shows that on average, stores' daily demand ranges from 0 to 1 unit for "Low" and "Medium" sales volume SKUs and 0 to 4 units for "High" volume SKUs with 0 being most frequently observed. It is noted that the daily demand at the retail store is very low, resembling a Poisson distribution, which is usually used to characterize low demand items.

A Chi-square test is used to test whether the demand of the selected SKUs from a sample store fits the Poisson distribution. P-values obtained from the test with high, medium, and low sales volume SKUs are 0.494, 0.847, and 0.943, respectively. These high p-values denote the appropriateness of using a Poisson distribution to characterize the daily demand of the SKU at the stores. However, it is important to note that due to the low number of records, the result of the Chi-square test may not be reliable. To ascertain that the daily demand resembles the Poisson distribution, observed frequency of sales is plotted against expected frequency under the Poisson distribution. The resulting graphs of the demand frequency of the high, medium, and low sales volume SKUs presented in Figure 7, Figure 8, and Figure 9 further confirm that a Poisson distribution can be used to characterize daily demand at the stores.

Figure 7 : Observed frequency of sales versus actual frequency of sales of high sales volume SKU at sample store

Figure 8 . Observed frequency of sales versus actual frequency of sales of medium sales volume SKU at sample store

Figure 9 : Observed frequency of sales versus actual frequency of sales of low sales volume SKU at sample store

To generate random daily demands of the selected SKUs using a Poisson distribution function, we need actual average daily demand at the representative stores. First, average weekly demand is calculated from one-year weekly demand data extracted from store frontend systems. Then, the average daily demand is obtained by dividing the average weekly demands by 7.

Two-year of random daily demand of the representative stores is then generated using a Poisson distribution function with the average daily demand calculated above. To represent the demand of the entire segment, the randomly generated demand of each representative store is multiplied by the number of stores in its segment. The assumption that the stores in the same segment have identical demand is used. It is noted that by using this method to scale-up the segment demand, the demand variability of the store-echelon may be overstated as illustrated in the following equations.

= Estimated demand of entire segment **j** on day i yji

= Number of stores in segment **j** n_i

$$
y_{ji} = n_j * x_{ji}
$$

$$
E(y_i) = n_i * E(x_i)
$$

Therefore $V(y_i) = n_i^2 * V(x_i)$

From the above equations, the demand variability of the segment is n^2 times the demand variability of the representative store. In reality, when stores' demands are aggregated to the DC, the segment variability may be less than the variability obtained from this method.

The number of stores included in the model is constrained by the model capacity. Including all stores in the model may overload the model due to too many decision variables as well as too much stores' transactional data. The selected scaling-up method seems to be the simplest and the best alternative thus far. Figure 10 summarizes the steps used to generate the random daily demand data.

Figure 10 **:** Steps used to generate random daily demand data

3.4 Model Development

The inventory model is divided into **3** sections: 1) stores, 2) DC, and 3) supplier. The current inventory policies of RetailCo and SupplierCo are replicated in the model by using the actual review periods and delivery schedules. The only values that are changed are the order-up-to-levels at each echelon, which are set as decision variables in the model. A heuristic approach that combines simulation and optimization is used in the model to find the optimal inventory distribution that offers minimal network carrying cost. Factors that are considered important in a retail environment such as high target service level at retail stores and minimum presentation quantities are set as constraints in the model to ensure that the optimal result from the model is applicable to the actual retail environment. Figure 11 shows the structure of the inventory model.

Figure 11 : Structure of the inventory model

3.4.1 Stores Section

The stores section is divided into 12 segments according to the store segmentation mentioned in Section 3.1. Two-year random daily demands of the store-segments obtained from Section 3.3 are used as inputs in the store level. Store level parameters include minimum presentation quantities of selected SKU, and ordering and receiving schedules at representative stores. Store-segments' order-up-to-levels are set as decision variables in the model. The following equations are used to calculate a store-segment's ending inventory, and ordered quantity. It is assumed that minimum presentation quantities and ordering and receiving schedules of the stores in the same segment are the same. It is also assumed that inventory is reviewed and order is created in the beginning of the day on a fixed schedule basis.

Let
$$
j
$$
 = Rank of store-segment by the representative stores' historical average
daily sales; where $j = 1, 2, 3, 4, 5, \dots, n$ (n = Total number of store-segments)

i = The **ith** day; where i = 1, 2, 3, 4, 5......, N (N = Total number of days in operation)

Ending Inventory $_{ji}$ = MAX (Ending Inventory $_{ji(i-1)}$ + Received Quantity $_{ji}$ - Demand $_{ji}$, 0)

$$
k_{jx}
$$
 = The xth scheduled ordering day of store j in a week; where k ε {1, 2, 3,
..., 7} (1 = Sunday, 2 = Monday, ..., 7 = Saturday)

Order Quantity $_{ji}$ = MAX(Store Order - up - to - level_j - Ending Inventory $_{j(i-1)}$,0) ; when MOD(i, 7) = k_{ix}

Store-segment's received quantity is equal to the delivered quantity at the DC.

Received Quantity $\hat{I}_{ji} = DC$ Delivered Quantity \hat{I}_{ji}

In this model, we can calculate service level either by looking at the day that the store experiences an out-of-stock or the quantity of lost sales because we know the exact demand that we cannot fulfill. However, in reality, it is very difficult to measure the lost sales or unfulfilled demand. Therefore, in this study, we will calculate service level by using the first method. The service levels of the store-segments that have high average daily demand are given higher weight than those with lower average daily demand in the calculation of the store-echelon service level. The equations to calculate service levels are presented below.

Service Level. **=** Number of days with out -of -stock Number of days in operation

 \sum (Average Daily Demand_j × Service Level_j) Store -Echelon Service Level = **j='** $\sum_{j=1}^{\infty}$ Average Daily Demand_j

Average Daily Demand_j =
$$
\frac{\sum_{i=1}^{N} Demon_{ij}}{N}
$$

Average inventory level is calculated from the store's ending inventory.

Average Inventory_j =
$$
\frac{\sum_{i=1}^{N} \text{Ending Inventory}_i}{N}
$$

Store - Echelon Average Inventory $=$ $\sum_{j=1}$ Average Inventory

Maintaining store inventory levels above store-SKU minimum presentation quantities is considered **highly** important for retailers. Therefore, these numbers are set as minimum order-up-to-level constraints for store-segments.

Minimum Order - up - to - level_j = Minimum Presentation Quantity_i * Number of Stores in Segment_i

Figure 12 shows a diagram representing the model structure for the store level.

Figure 12 : Structure of inventory model: Store section

3.4.2 Distribution Center Section

Orders generated from stores are used as input demand for the **DC.** The **DC** order-upto-level is set as a decision variable. It is assumed that the **DC** sets the priority of the stores based on the historical average daily sales. **A** store with higher daily sales is given a higher priority in the case that the **DC** does not have enough inventory to serve all stores.

Figure 13 shows an example of the DC's inventory allocation. Assume that the DC has 20 units to be allocated and there are 4 stores that order in the same day, accounting for 26 units. Ranking by historical average daily sales, we have store-segment 2, 1, 4, and 3 in a sequence. Inventory available to allocate to store-segment 2 is 20 units; therefore storesegment 2 gets all it wants which is 10 units. After allocating to store-segment 2, inventory available to store-segment 1 is 10 units (20-10); therefore store-segment 1 also gets its full ordered quantity which is 6 units. After allocating to store-segment 2 and **1,** inventory available for store-segment 4 is 4 units; therefore the store-segment gets 4 units even though it orders 6 units. Store-segment 3 gets nothing.

Figure 13 : Example of DC's inventory allocation

Two types of inventories are defined in this section: 1) non-reserved inventory, and 2) actual inventory. Non-reserved inventory is the inventory that has not been allocated to any order, while actual inventory is the inventory that is physically available, but not all is able to be allocated since some portions has already been reserved for the past orders. An availableto-promise quantity (ATP) is one form of non-reserved inventory and is the inventory that is available to be allocated.

The equations below show the calculations in the DC section.

Total Demand From Store_i = \sum Order Quantity \sum_{ji} **j=1**

DC Total Reserved Quantity_i = MIN (Total Demand From Store_i, DC ATP_i)

DC NonReserved Ending Inventory $_i = MAX(DC$ NonReserved Ending Inventory $_{i,j}$ $+DC$ Received Quantity. - DC Total Reserved Quantity_i,0)

DC Actual Ending Inventory $_i = MAX($ DC Actual Ending Inventory $_{i-1}$ + DC Received Quantity_i - DC Total Delivered Quantity_i,0)

It is assumed that the inventory at the DC is reserved at the end of the day that it receives orders from the stores. Therefore, the ATP quantity also includes received quantity from the supplier in the same day.

DC ATP $_i = DC$ NonReserved Ending Inventory $_{i+1} + DC$ Received Quantity_i

As explained above, the DC allocates inventory to the stores based on their priority. Therefore, the inventory available to allocate to the stores is the remaining quantity after allocating to stores with higher priority.

j-1 DC Available Inventory $_{ii} = MAX(DCATP_i - \sum CrefC}$ and U_i , 0) **j=1**

Reserved Quantity μ = MIN(Order Quantity μ , DC Available Inventory μ)

Delivered Quantity to a store is equal to reserved quantity for that store but the delivery day is lagged from the day that the inventory is reserved by DC replenishment lead time. DC replenishment lead time (L_1) is the number of days between the DC receiving the order and the truck from DC arriving the stores. DC replenishment lead time for different

ordering days in a week can be different depending on the fixed schedule appointment between the stores and the DC. In this model, it is assumed that inventory in-transit belongs to the upstream echelon, which is the DC in this case. Therefore, the inventory during transportation to the stores is still held by the DC.

DC Delivered Quantity $\mathbf{p} =$ Reserved Quantity $\mathbf{p}(\mathbf{i} \cdot \mathbf{L}_1)$; Where L_1 = DC Replenishment Lead Time (To Stores)

Orders to the supplier are created on a fixed schedule basis and only when the beginning inventory plus pending inventory falls below the order-up-to-level. The order quantity needs to be in a multiple of case pack since the supplier only delivers in a case pack quantity.

 k_{DC} = The DC's scheduled ordering day in a week; where k ϵ {1, 2, 3, ..., 7} $(1 =$ Sunday, $2 =$ Monday, ..., $7 =$ Saturday)

 DC Order Quantity_i = CEILING(MAX(DC Order - up - to - level - DC NonReserved Ending Inventory $_{i-1}$ – DC Pending Order_i,0) ,Case Pack Quantity) ; when MOD(i, 7) = k_{DC}

DC's pending order on any day i is the cumulative sum of all DC's order quantities from day 0 to day i-I less the cumulative sum of all DC's received quantity from day 0 to day $i-1$.

DC Pending Order_i =
$$
\sum_{i=0}^{i-1}
$$
 DC Order Quantity_i - $\sum_{i=0}^{i-1}$ DC Receiver0.

Received quantity at the DC is equal to the delivered quantity determined by the supplier.

DC Received Quantity **i=** Supplier Delivered Quantity ⁱ

Both the DC's service level to each store and the overall service level are determined **by** total fulfilled demand, as follows :

Service level<sub>DC-8
fore_j =
$$
\frac{\sum_{i=1}^{N} DC\text{Delivered Quantity}_{ji}}{\sum_{i=1}^{N}\text{Order Quantity}_{ji}}
$$</sub>

DC Service Level =
$$
\frac{\sum_{i=1}^{N} \sum_{j=1}^{n} DC\text{ Delivered quantity}_{ji}}{\sum_{i=1}^{N} \sum_{j=1}^{n} Order\text{ Quantity}_{ji}}
$$

DC's average inventory is determined by averaging the actual daily ending inventory.

DC Average Inventory =
$$
\frac{\sum_{i=1}^{N} DC \text{Actual Ending Inventory}_{i}}{N}
$$

Figure 14 shows the model structure at the DC level.

Figure 14 : Structure of inventory model: DC section

3.4.3 Supplier Section

An order-up-to-level inventory policy with daily review is set at the supplier. Inventory is checked every day and orders are created if the inventory level falls below the order-up-to-level. Orders from the DC are used as input demand and the supplier's order-upto-level is set as a decision variable to calculate other values including ending inventory, orders to production, and delivered quantity to the DC. The supplier section in this model covers all processes from production to finished goods inventory of labeled-products specific to RetailCo. Even though SupplierCo delivers its products to many of the RetailCo's DCs, it is assumed that inventory shown in this model is only for serving the selected DC. In reality, the inventory level at SupplierCo can be much higher than the result obtained from the model.

Just as in the DC section, two types of inventory are defined in supplier section since the inventory is reserved earlier than the actual delivery day. Therefore, the actual inventory and non-reserved inventory may be different and should be taken into account when allocating inventory or calculating average inventory.

The equations below show the calculations in the supplier's section.

Supplier NonReserved Ending Inventory_i = MAX(Supplier NonReserved Ending Inventory_{i-1} + Produced Quantity, - Supplier Reserved Quantity, 0)

Supplier Actual Ending Inventory_i = MAX(Supplier Actual Ending Inventory_{i-1} + Produced Quantity, - Supplier Delivered Quantity, 0)

It is assumed that inventory at the supplier is reserved at the end of the day that it receives orders from the DC. Therefore, ATP quantity also includes replenished inventory from the production on the same day.

Supplier Reserved Quantity $_i = MIN (DC Order Quantity_i, Supplementary F)$

Supplier ATP $_i$ = Supplier NonReserved Ending Inventory $_{i+1}$ + Produced Quantity is

Delivered quantity to the DC is equal to the reserved quantity. Shipments are delivered to the DC on a fixed schedule basis. The delivery day is lagged from the day that the inventory is reserved by the supplier replenishment lead time (L_2) , the time between the supplier receiving the orders and the trucks from the supplier arriving the DC. Again, it is assumed that inventory in-transit belongs to the upstream echelon, which is the supplier in this case.

Supplier Delivered Quantity_i = Supplier Reserved Quantity_{(i-Lo}) ; Where L_2 = Supplier Replenishment Lead Time (To DC)

The minimum production order quantity or minimum order quantity to the second-tier supplier are not considered here due to the fact that SupplierCo can sell the remaining inventory to other retailers. Therefore, minimum production order quantity or minimum order quantity should not be set as an ordering constraint in this model. However, the replenished
quantity still needs to be in multiple of case pack. The pending orders, the past orders that have not been fulfilled, are taken into account when calculating the next order quantity, as follows **:**

Supplier Order Quantity ⁱ= CEILING(MAX(Supplier Order - up - to - level -Supplier NonReserved Ending Inventory,. - Supplier Pending Orderi ,O) ,Case Pack Quantity)

Supplier's pending order on any day i is the cumulative sum of all supplier's production order quantities from day 0 to day i-1 less the cumulative sum of all supplier's produced quantity from day 0 to day $i - 1$.

Supplier Pending Order_i =
$$
\sum_{i=0}^{i-1}
$$
Supplier Order Quantity – $\sum_{i=0}^{i-1}$ Product Quantity

Production replenishment lead time (L_3) is the period between orders being created to inventory being replenished. This lead time includes production lead time and quality inspection lead time. It is taken into account when calculating the day that production replenishes the inventory at the warehouse.

Produced Quantity_i = Supplier Order Quantity_{(i-L3}) ; Where L_3 = Production Replenishment Lead Time

The supplier's service level is determined by its ability to fulfill the ordered quantity from the DC.

Supplier Service Level =
$$
\frac{\sum_{i=1}^{N} \text{Supplier Delivered Quantity}_{i}}{\sum_{i=1}^{N} \text{DC Order Quantity}_{i}}
$$

As in the other echelons, supplier's average inventory is calculated from its actual ending inventory.

Figure 15 shows the model structure of the supplier section.

Figure 15 : Structure of the inventory model: Supplier section

3.4.4 Objective Function

The objective function is to minimize network inventory carrying cost. Due to the confidentiality of the cost data between the parties involved in the analysis, we set some assumptions to enable us to calculate the network inventory carrying cost. First, we assume that the inventory carrying cost per unit value is the same for all echelons. Second, we assume that the unit cost of the inventory at the supplier is 60% of that at the retailer. It is important to remember that this unit value is an arbitrary number only for the purpose of the

analysis. The model allows a user to change the parameters in the case that more accurate data can be obtained.

The functions to calculate network average inventory and carrying cost are given below.

Let $m =$ Number of echelons in the network

Network Average Inventory = $\sum_{i=1}^{m}$ Average Inventory_m

Network Carrying Cost

 $=$ \sum (Average Inventory_m \times Carrying Cost Per Unit Value_m \times Unit Cost Value_m) i=1

The objective function and constraints are given below:

Order - up - to - level_j \geq Minimum Order - up - to - level_j

Decision variables include order-up-to-levels at store-segments, the DC, and the supplier. Since there are too many decision variables for a spreadsheet based algorithm to solve, a heuristic approach is used to minimize the time to run the model.

First, we set the order-up-to-levels at all echelons such that they can offer 100% service level. Then, the process is divided into 2 sub-processes: 1) store-echelon optimization and 2) supplier and DC optimization.

In the first sub-process, store-segments' order-up-to-levels are set as decision variables, while the network carrying cost is set as the objective function. Then, an iterative approach is used to find the optimal order-up-to-levels at the store level. The Excel solver is run by varying one store-segment's order-up-to-level at a time until all the stores have been run and the process loops until the network's carrying cost converges. It is noted that using this method, we give the highest priority to the stores.

In the second sub-process, after we get the stores' optimal order-up-to-levels, we list all the possible combinations of DC's order-up-to-level and supplier's order-up-to-levels that satisfy the store-segment service level constraint. The combination that gives the lowest network carrying cost is selected as the optimal answer. To reduce the number of possible solutions, we set the ranges for both the supplier's and the DC's order-up-to-levels. One end of the ranges are the lowest order-up-to-levels that allow the supplier or the DC to achieve 100% service level given 100% service levels initially set at the other echelons. The other end of the ranges are the order-up-to-levels at the supplier or the DC that still allow the storeechelon's service level to achieve store target service level given 100% service level at the other echelons. Figure 16 presents the diagram showing the objective function, constraints, and decision variables in the two sub-processes.

Figure 16: Objective function, decision variables, and constraint in the inventory optimization

CHAPTER 4

RESULTS AND ANALYSIS

This chapter is divided into two sections: 1) Results and 2) Sensitivity Analyses. Section 4.1 presents the analysis of the optimal network inventory distributions and the optimal combination of the echelons' service levels obtained from the inventory model. The echelons' average inventory levels and carrying costs under the optimal inventory policies are compared to those under the current inventory policies employed **by** RetailCo and SupplierCo. The causes of the differences in the inventory policies are analyzed and the opportunities for performance improvement are identified.

In the literature review section, we discuss the interrelationship that exists between the echelons, especially the echelons that are juxtaposed against each other. In Section 4.2, sensitivity analyses are performed to assess how an echelon's actions can affect the other echelons. The first analysis estimates how an echelon's deviation from the optimal inventory policies impacts the other echelons in the network. The second analysis estimates the effect of changes in the network target service level on the echelons' average inventory levels and distribution, as well as the echelons' service levels. As mentioned Chapter **1,** the product category that we select in the study requires a long manufacturing lead time and strict quality inspection. Product failure is usually found in the final stage, or the quality inspection process. Therefore, the possibility of service disruptions at the supplier can be a threat to downstream echelon. The third analysis quantifies the impact of the supplier's service disruption on the downstream echelons given different probabilities of service disruption. One of the most popular initiatives to reduce the network carrying cost is to increase the delivery frequency

between the echelons. The final analysis is performed to assess the inventory carrying cost savings from the increase in delivery frequency.

4.1 Results

This section presents the optimal result obtained from running the model at the store target service level of 97.5%. Three SKUs with different average daily demand are selected. For each of them, five sets of two-year random daily demand data are run in the model and the optimal results are averaged and presented in Figure 17 and Figure 18.

The graph in Figure 17 shows that maintaining low average inventory level at the intermediate echelon or the DC offers optimal network inventory carrying cost, regardless of the average daily sales demand. On average, the average inventory at the DC should be maintained at around 20% of the total network inventory. There is no exact trend in the inventory distribution at the supplier and the stores. For SKUs with higher average daily demand, the supplier needs to maintain higher average inventory level than the stores. The stores maintain a higher portion of inventory than the supplier when the average daily demand is low.

It is noted that the constraints set at the store level may possibly play major roles in the inventory distribution. High store target service level and the minimum order-up-to-levels set for merchandising purposes may prevent the stores from further decreasing their inventory level. As a result, the stores may maintain more inventory than what is needed to respond to customer demand and demand variability. Excess safety stock and the sporadic nature of demand at the stores reduce the necessity of the upstream echelons to maintain high safety stock, thus resulting in the low average inventory level.

Manufacturing and inspection lead time may also play major roles in building up the inventory level at the supplier. If the lead time is long, the supplier needs to build up inventory to account for the demand during the lead time and can result in higher inventory level than what is needed to respond to demand.

Figure 17 : Optimal inventory distribution for SKUs with different average sales unit volume

The graph in Figure 18 shows the same trend as Figure 17. Among all echelons, the DC can offer the lowest service level while still allowing the stores to achieve target service level. The lower the average daily demand, the lower the DC's service level can be. The supplier's service level can also be further reduced for SKUs with lower average daily demand. However, the range to which it can go is much lower than the DC's service level as can be seen from the graph that the supplier still needs to maintain a service level higher than **90%.**

It is noted that constraints may possibly play a major role in dictating the storeechelon's service level. Since the stores need to maintain their service levels higher than target service level and order-up-to-levels higher than minimum order-up-to-levels to satisfy merchandising purposes, the average inventory level that the stores maintain can be higher than what is really needed to serve the expected demand, thus resulting in high echelon service levels. Again manufacturing and inspection lead time may also play major roles in increasing the service level offered by the supplier.

Figure 18 : Optimal echelons' service level for SKUs with different average sales volume

Further analysis is conducted to measure how the current inventory policies perform compared to the optimal inventory policies. The current inventory policies or order-up-tolevels are input in the model with the same five sets of random daily demand that are used to find the optimal inventory policies. The average inventory levels under the current policies are then compared to those of the optimal inventory policies. The comparison of the echelons' average inventory levels from both policies is shown in Figure **19** and the order-upto-levels in units and days of supply under the current and optimal inventory policies are show in table 4.

Both policies result in the same trend of inventory distribution, in which the average inventory level at the **DC** is the lowest compared to those of the other echelons. However, the average inventory levels and the order-up-to-levels under both inventory policies are very different, especially at the supplier and the stores.

Figure **19** : Current vs optimal average inventory on-hand

Echelon	Order-Up-To-Level					
	Current		Optimal			
	Units	Days of Supply	Units	Weeks of Supply		
Supplier	319	84	125	35		
DC	85	20		22		
Retail Stores	248	59	157	33		

Table 4: Comparison of current vs optimal order-up-to-level

One of the possible explanations for this large difference is the flaws in the underlying assumptions used in the model. We assume that all stores in the same segment have identical daily demands as well as minimum presentation quantities, while these assumptions may not be true in reality.

Another possible explanation is the difference in the optimization approaches used in the model versus in reality. In the model, we use multi-echelon inventory optimization, while single-echelon inventory optimization is used in reality. The model is built by integrating the transactions of all echelons and the optimal inventory policies are obtained from a single performance objective function set at the network level. In reality, a single-echelon approach is used. Each echelon has its own performance objective function and does not consider the interrelationship between itself and the other echelons. In reality, the DC and the supplier also have their own target service levels, which were agreed upon between the involved parties. In contrast, target service levels of the upstream echelons are not set as constraints to determine order-up-to-levels in the model. Furthermore, inventory policy at the supplier is overwritten manually without considering the optimality of the policies by setting the same order-up-tolevels for all SKUs. This combination of factors results in suboptimal inventory policies and leading to higher network inventory level.

In summary, the differences in the echelons' average inventory can be explained by the plausible flaws in the assumptions and the use of different optimization approaches, as well as human intervention to set up the inventory policies. The analysis also shows that the benefit of a multi-echelon optimization outweighs that of a single-echelon optimization and that the optimal inventory policies can be obtained by having low inventory and service level at the intermediate echelon. Since different SKUs have different demands and lead times, setting a uniform inventory policy across all SKUs leads to suboptimal inventory carrying cost, and therefore these differences should be taken into account to achieve the optimal cost. However, dealing with all SKUs can be a daunting task, therefore it is recommended that SKUs be segmented by demand and lead time, and a uniform inventory policy be used for all SKUs in the same segment.

4.2 Sensitivity Analysis

Analyses are performed to assess how sensitive the optimal results are to the echelons' deviations from the optimal inventory policies. The first analysis assesses the impact of the DC's and the supplier's variations from the optimal inventory policies on the other echelons. The second analysis is performed to see changes of the optimal inventory policies when the store target service level is altered. The third analysis focuses on the assessment of how service disruptions at the suppliers affect the other echelons. The last analysis is performed to quantify the inventory carrying cost savings obtained from a change in delivery frequency from the DC to the stores.

4.2.1 Impact of the Echelons' Deviations from the Optimal Inventory Policies

The objective of this analysis is to estimate the impact of the DC's deviation from the optimal inventory policy on the echelons' average inventory levels and service levels. Only one set of demand data is used in the analysis. The DC's order-up-to-level is varied, while those of the other echelons are held at the optimal point when store target service level is 97.5%. The echelons' average inventory levels and service levels are estimated and presented in Figure 20 and Figure 21 respectively. The red dotted line shows the DC's order-up-to-level under the network optimal policies.

From Figure 20, the DC's changes in the order-up-to-level seem to have more impact on the average inventory level of the store-echelon than the supplier. Decreasing the DC's order-up-to-level below the optimal point reduces the network average inventory level, especially at the DC and the stores. On the other hand, increasing the DC's order-up-to-level

above the optimal point increases the network average inventory level, but only at its own echelon. Very marginal impact on inventory level is realized at the other echelons.

Figure 20 : Impact of changes in DC's order-up-to-level on the echelons' average inventory levels

From Figure 21, decreasing the order-up-to-level below the optimal point, even though it offers savings on the inventory carrying cost, compromises the stores' service levels. The DC's service level significantly drops, resulting in a decrease in the store-echelon's service level. Some stores' service levels also drop below the target service level. Interestingly, a higher service level at the **DC** only marginally increases the store-echelon's service level. Since the **DC** may order more from the supplier while the supplier's average inventory level remains constant, the supplier's service level slightly drops.

Figure 21 : Impact of changes in DC's order-up-to-level on the echelons' service level

We can imply from the analysis that maintaining the order-up-to-level at **70** units or slightly above the network optimal policies seems to be the best option for the DC. Beyond an order-up-to-level of 80 units, the additional inventory carrying cost associated with increasing the DC's order-up-to-level beyond the optimal point outweighs the marginal benefit from improving the store-echelon's service level. The savings from decreasing the order-up-to-level also does not justify the decrease in the stores' service level.

Another analysis is done to estimate the impact of the supplier's deviation from the network optimal inventory policies under store target service level of 97.5%. The results of the impact on the echelons' average inventory levels and service levels are shown in Figure 22 and Figure 23 respectively.

From Figure 22, decreasing the supplier's order-up-to-level below the network optimal inventory policies significantly decreases its average inventory level and moderately decreases the other echelons' average inventory levels. On the other hand, increasing supplier's order-up-to-level beyond the optimal point considerably increases its average

inventory level, but seems to have little or no effect on the other echelons' average inventory levels.

Figure 22: Impact of changes in supplier's order-up-to-level on the echelons' average inventory levels

Figure 23 shows that decreasing the supplier's order-up-to-level slightly decreases its service level. However, a minor reduction in the supplier's service level creates substantial impact on the DC's service level, which finally leads to lower-than-target service levels at the stores. On the other hand, increasing supplier's order-up-to-level increases DC's service level up to 85% when order-up-to-level is around 130; beyond which, DC's service level remains constant. Supplier's and stores' service levels remain almost constant with increasing supplier order-up-to-level.

Figure 23 : Impact of changes in supplier's order-up-to-level on echelon's service levels

A possible explanation is that the increase in inventory at the supplier does not alter the quantity that the DC and the stores order from their upper echelons; therefore it does not impact the average inventory level at the other echelons. The result from this analysis shows that there is very little or no benefit for the supplier to increase the order-up-to-level significantly beyond the network optimal policies. On the other hand, the potential lost sales from the decreased service level at the store-echelon due to a decrease in the supplier's orderup-to-level below the optimal point may offset the savings from lower inventory carrying cost.

With the assumption that there is no service disruption in the network and all echelons are committed to a common goal of minimizing network inventory, these sensitivity analyses shows that maintaining the order-up-to-level at the network optimal policies seem to be the best possible option for the whole network. The analyses also show that an upper echelon's service level is only positively correlated to the lower echelon's service level up to a certain service level. Moving beyond the point gives no significant benefit to either party.

The analyses also shows that an echelon's service and inventory level is highly dependent on the other echelons', therefore setting up the performance objective without considering this interrelationship between the echelons can result in unnecessary inventory in the network as well as service failure despite enough inventory being held in the network.

4.2.2 Impact of Changes in the Store Target Service Level

The objective of this analysis is to measure the impact of changes in the store target service level on the echelons' average inventory levels and service levels. Five sets of demand data are run to find the optimal policies at different store target service levels and the results at each service level are averaged. The impact on the echelons' average inventory level and on the echelons' service levels are shown in Figure 24 and Figure 25 respectively.

From Figure 24, an increase in the store target service level can significantly increase the network average inventory level, especially at the stores and the DC. Supplier inventory level, however, remains almost the same with increasing store target service level.

Figure 25 shows that increasing inventory at the store-echelon may only cause a marginal improvement in the store-echelon's service level. On the other hand, increasing inventory at the DC results in a proportionally bigger improvement in the DC's service level.

Figure 25 : Impact of changes in the stores' service level on the echelons' service levels

The analysis implies that in order to keep the service level high, it is important that the inventory be shifted to the downstream echelons, which are the stores and the DC. Maintaining high service level at the stores and the DC seems to make their service levels less sensitive to the supplier's service level. This can be seen from the graph that the supplier's service level drops slightly when the other echelons' service levels increase. However, the retailer needs to balance the trade-off between the expected decrease in lost sales at the stores and the significant increase in inventory carrying cost in order to make a better decision.

4.2.3 Impact of Supplier's Service Disruption

The objective of this analysis is to estimate the impact of supplier's service disruptions on the echelons' service levels. The products that we selected require a strict quality check to ensure that they comply with FDA regulations. After production, the finished goods need to be quarantined for 2 to 10 days for quality inspections. If the products fail the inspections, the entire production lot needs to be discarded. Therefore, the impact of quality failure can be substantial. This analysis is designed to estimate the impact of service disruptions due to quality failure.

The model is modified to include the probability of service disruptions at the supplier. For each day, a random number between 0 and 1 is generated. We assume that a service disruption occurs if the random number is less than the specified probability. If there is a service disruption, the produced quantity, if any, is set to 0 and the pending order after the day that the service disruption occurs is cancelled.

The base case is when all echelons maintain the order-up-to-levels at the network optimal policies given 97.5% store target service level and the probability of the supplier's service disruption is zero. Fifty sets of random probabilities are tested and the echelons' service levels are recorded. Since the probability of the supplier's service disruption is unknown, different probabilities, including 1%, 2.5%, 5%, 7.5%, and 10%, are tested and compared. The echelons' service levels from fifty sets of data are averaged and shown in Figure 26. The percentage change in the echelons' service levels from the optimal service levels (base case) are shown in Table 5.

Figure 26 : Impact of supplier's service disruption on the echelons' service levels

Changes in	Probability of the Supplier's Service Disruption					
Service Level	1.00%	2.50%	5.00%	7.50%	10.00%	
Supplier	0.48%	1.69%	3.81%	5.43%	7.83%	
IDC	12.08%	29.19%	68.63%	76.92%	87.59%	
Stores	1.60%	8.75%	26.32%	37.92%	52.20%	

Table 5 : Impact of the supplier's service disruption on the echelons' service levels

The data in Table 5 shows that the supplier's service disruptions have least impact on its own service level. However, the slight drop in the supplier's service level has a significant impact on the DC's service level, which finally results in a decrease in the store-echelon's service level. Figure 26 shows that with only 1% probability of the supplier's service disruption, the store-echelon service level has already dropped below the target service level of 97.5%.

This analysis implies that the optimal order-up-to-levels at the stores and the DC may not be sustainable if there is even a slight chance of supplier's service disruption. Therefore, higher inventory level than what is recommended from the model may need to be held at one

or more echelons to protect against service disruptions. The amount of inventory added to the network depends on the probability of service disruption and more complex calculation may be required to determine the amount to be held. The calculation should also take into account the cost of lost sales at the retail stores in order to balance the increase in carrying cost.

4.2.4 Impact of Change in Delivery Frequency on Carrying Cost

The objective of this analysis is to measure how much savings can be gained when the DC increases the delivery frequency to the stores. The base scenario is the current situation in which 18 stores receive twice-a-week deliveries and 80 stores receive once-a-week deliveries from the DC. The new scenario is 100% twice-a-week delivery frequency to all stores. Five sets of demand data are tested under both scenarios and the echelons' average inventory levels and service levels are averaged. The results are shown in Figure 27.

With more frequent delivery, the overall network inventory level decreases by 12%. The highest inventory cost savings are at the stores and the DC where average inventory level is reduced by 15% and 14% respectively, while the savings at the supplier are only 7%.

With more frequent deliveries, the stores need less safety stock, thus reducing the stores' order-up-to-levels and average inventory levels. At the DC level, orders from the stores are received more frequently with smaller volume, resulting in a decrease in the DC's average inventory level even though the total demand from the stores is not much different from that in the base scenario.

Figure 28 graphically explains how the average inventory at the DC can be reduced with more frequent delivery to the stores. The average inventory at the DC in scenario 1-1, in which the DC delivers to the stores once a week on day 4 and the DC receives the shipment from the supplier on day 6, is 12 units. In scenario 1-2, when the DC delivers to the stores twice a week on day 1 and day 4 and receives the shipment from the suppliers on the same day as in scenario 1-1, the average inventory at the DC drops to 10 units. The comparison of inventory in Scenario 2-1 and 2-2 follows the same trend.

Figure 28 **:** Examples of average inventory calculation under once-a-week delivery and twice-a-week delivery

When the inventory level at the DC drops more frequently, the chance that the inventory level is below the order-up-to-level when the inventory is checked is higher, thus leading to more frequent orders with smaller volume to the supplier. Referring to the examples in Figure 28 again; if the DC's ordering day is on day 3 instead of day 6, the DC will order on day 3 under scenario 1-2 and 2-2, in which the delivery frequency to the stores is twice-a-week, while it will not order under scenario 1-1 and 2-1.

With more frequent orders from the DC, the supplier's average inventory level reduces. However, the reduction in the average inventory level at the supplier is smaller than that at the DC since the actual delivery schedule does not change.

However, there is a limit to which the inventory cost savings can be realized at the stores because of the minimum inventory constraints set by the stores' presentation quantities. In addition we also need to consider the increase in transportation cost, which can be doubled, and the increase in ordering costs, as well as handling costs, at the stores. Economy of scale may not be achieved with more frequent delivery since the orders from the stores can be very small. Further research which includes the other relevant costs need to be done for more accurate decision making.

CHAPTER 5

REVIEWS AND CONCLUSION

This chapter summarizes the work that has been done in this thesis and the results obtained from the study. The chapter consists of three sections. The first section summarizes the research questions and reviews the methods used to obtain the results. The second section provides a complete summary of the results obtained from the study. The last section provides the recommendation to improve the model and results.

5.1 Research Questions and Methods

The study focuses on finding the optimal inventory distribution in a retail threeechelon environment, consisting of a supplier, a retailer's DC, and stores. The objective is to find the minimum network inventory carrying cost, while still allowing the retailer to achieve high target customer service level. The study is based on the case study of RetailCo, a leading pharmacy and convenience store chain in the US and SupplierCo, a big manufacturer of private-label products. The overview of the network and the current practices are described in detail in Chapter 1.

In this study, a multi-echelon inventory simulation and optimization model is used to find the optimal inventory distribution since it offers significant benefit over a single-echelon optimization, which is currently used in the network. Chapter 2 describes benefits of multiechelon optimization and compares it to single-echelon optimization. An inventory model is built in a simple Excel spreadsheet by replicating the current periodic, order-up-to-level policies at the RetailCo and SupplierCo with all echelon transactions integrated. The network carrying cost is minimized by varying the echelons' order-up-to-levels while constraining store target service level and store's minimum order-up-to-levels. A heuristic approach is used to find the optimal inventory policies due to a significant number of variables and large amount of transactional data. The detailed equations used in the model and the steps to find the optimal results are described in Chapter 3.

5.2 Result Summary

The results are discussed in Chapter 4. The results show that optimal network carrying cost can be achieved **by** having low inventory level and service level at the intermediate echelon, or the **DC** in this study. The comparison of the optimal and current cost performance confirms that the benefit of a multi-echelon optimization outweighs that obtained from a single-echelon optimization, given that only the carrying cost is considered. In addition, the performance analysis recommends that **SKU** be segmented **by** demand and lead time and an inventory policy be set **by SKU** segment to achieve lower inventory carrying cost. The sensitivity analyses strengthen the past research finding that there is an interrelationship between the echelons' inventory policies and service levels, especially between the juxtaposed echelons. This is confirmed **by** the estimation of the impact of the echelons' inventory policies on the other echelons' average inventory levels and service levels. Slight deviations from the optimal inventory policies can result in significant customer service failure at the customer-facing echelons. The study also shows that the echelons' service levels are not always correlated to the juxtaposed echelons. Increasing the service level at an upstream echelon does not always create a benefit for downstream echelons. Therefore, consideration of the echelons' interrelationship is very important when setting up the inventory policies to ensure that the echelons' policies work well together and reduce the possibility of excess safety stock and service failure.

Setting up the store target service level is an important task for the retailers. The study shows that an increase in the store target service level has more impact on the stores and the DC. Thus, the decision to set up a higher store target service level should consider the overall increase in the inventory levels in different echelons and balance the trade-off between increased inventory carrying costs and decreased lost sales.

Furthermore, the decision on the inventory policies should also consider the probability of service disruption at the upstream echelons. The study shows that with the optimal inventory policies, only a slight probability of service disruption at the supplier can result in a significant service failure at the downstream echelons. Therefore, additional inventory beyond the optimal inventory policies may be required to protect against service disruption.

Additional study to quantify the benefit of an increase in the delivery frequency between the DC and the stores is conducted. Significant savings are realized at the DC and the stores due to reduced safety stocks associated with more frequent deliveries. Smaller benefit is realized at the supplier due to more frequent orders from the DC. Detailed discussion on the results is presented in Chapter 4.

The study provides the reader insights into the benefits that can be obtained from the collaboration between the echelons in the network to create the optimal multi-echelon inventory policies. Collaborative planning between the parties to set up a single performance objective and integration of the information to provide demand visibility up the supply chain is required to achieve the true optimal inventory policies. Because of the complexity of the problems, sophisticated technology may be needed.

5.3 Future Research

Because of the complexity of the problems, a number of assumptions are made in the model. First the stores are segmented into groups and only the representative stores are included in the model to represent their segments. It is assumed that daily demand of the stores in the same segments is identical. This is not true in reality. To make the results more accurate, the model should be able to reflect the original demand from all stores. However, in a retail environment, the number of stores can go up to several thousands, making it impossible to find the tools that can manage this amount of data. Therefore, we should change focus from finding the tool to finding the appropriate segmentation method. Thus, more detailed study on segmentation is required.

To find the optimal result, we use a heuristic approach by splitting the optimization into two sub-processes. The store-echelon is optimized first and then all combinations of the order-up-to-levels at the DC and the supplier are enumerated. The combination that gives the lowest average inventory is chosen as the optimal result. Using this method, the stores are given higher priority than other echelons. Being able to optimize the inventory by simultaneously varying all echelons' order-up-to-levels can potentially provide a more accurate result.

The study also shows the impact of the upstream echelon's service disruptions on the other echelons' service levels given different probabilities of service disruption. Additional safety stock may be required to protect against service failure. Future study is recommended to find the way to incorporate the probability of service disruption into the safety stock calculation.

This study only focuses on the inventory carrying cost and neglects ordering cost, transportation cost, and lost sales. Future research may want to consider all other relevant costs to make the solution more practical.

 $\mathcal{A}^{\mathcal{A}}$

BIBLIOGRAPHY

Axsiter, **S.,** Zhang, W. **(1999), A** joint replenishment policy for multi-echelon inventory control. *International Journal of Production Economics, 59(1-3),* 243-250.

Diks, E.B., De Kok, **A.G.,** Lagodimos, **A.G. (1995).** Multi-echelon systems: **A** service measure perspective. *European Journal of Operational Research 95 (1996),* 241-263

Feeney, G.J., Sherbrooke, **C.C.** (1966). The **(S-1,S)** inventory policy under compound Poisson demand. *Management Science, 12(5),* 91-411.

Forsberg, R. (1995). Optimization of order-up-to **S** policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research 81(1),* 143-153.

Hausman, W.H., Erkip, **N.K.** (1994). Multi-Echelon vs. Single-Echelon Inventory Control Policies for Low-Demand Items. *Management Science, 40(5),* **597-602.**

Lee, C.B. **(2003).** *Multi-Echelon Inventory Optimization.* Evant Inc.

Masters, **J.M. (1993).** Determination of near optimal stock levels for multi-echelon distribution inventories. *Journal of Business Logistics, 14(2),* **165-195.**

Nozick, L.K., Turnquist, M.A. **(1999). A** two-echelon inventory allocation and **DC** location analysis. *Logistics and Transportation Review, 37(6),* 425- 441.

Silver, E.A., Pyke, **D.** F., **&** Peterson, R. **(1998).** *Inventory management and production planning and scheduling.* New York. John Wiley **&** Sons.

Vuyk, **C. (2003).** Out-of-stocks : **A** nightmare for retailer and supplier. *Beverage World,* 122(2), 555

Wilson, B.L., Kingdon, M.&Reeve, T. (1995), Best practices in retailing: quick hits in store level merchandise and inventory management, Chain Store Age.

APPENDIX A: Simulation and Optimization Method

Al Optimization Sub-Process 1 : Objective Function and Constraints in the Optimization of Stores' Order-up-to-levels

order-up-to-level at a time until all the stores have been
run and the process loops until the network's carrying cost converges

Remark: Before running the solver, the initial value of the decision variables must be set. The initial value of the echelon' order-up-to-level is the lowest number that enables all echelons' service levels to achieve 100% service level.

A2 Optimization Sub-Process 1: Iterative Approach to Find the Optimal Store's Orderup-to-levels

The macro is written to iteratively run the stores' order-up-to-levels. The solver continues running until the network carrying cost converges. The script is shown below.

Sub InventoryOptimiation()

```
Range("J6").FormulaR1C1 = "10000000"
Do Until Range("J7").Value < 1
Range("I6").Copy
Range("J6").Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks
  :=False, Transpose:=False
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$763"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$764"
SolverSolve userFinish:=True
SolverOk SetCell:="$l$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$765"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$766"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$767"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$768"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$769"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="0", ByChange:="$H$770"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$771"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$772"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$773"
SolverSolve userFinish:=True
SolverOk SetCell:="$I$6", MaxMinVal:=2, ValueOf:="O", ByChange:="$H$774"
SolverSolve userFinish:=True
```
Loop

A3 Optimization Sub-Process 2 **:** Iterative Approach to Find the Optimal Order-up-tolevels of the **DC** and the supplier

After we obtain the optimal stores' order-up-to-levels, we **innumerate the order-up-to-levels** of the **DC** and the supplier and list all possible combinations that still enable the stores to meet the **store target service level constraint. The combination** that offers **minimum network carrying cost is selected** as the optimal result. **The ranges of the possible order-up-to-levels** are set to **minimize the time to run the model.**

A3 Optimization Sub-Process 2 : Iterative Approach to Find the Optimal Order-up-tolevels of the DC and the supplier (Continued)

VB Script is written to automate the innumeration of the combinations of the DC's order-upto-level and the supplier's order-up-to-level. The store target service level constraints are in bold text.

Dim DC_Count As Integer Dim Supplier_Count As Integer

Supplier $Count = Range("G13")$. Value DC_Count = Range("S13").Value

For Supplier_Count = Range("G13").Value To Range("F13").Value $Range("F14").Value = Supplementary$ For DC Count = $Range("S13")$.Value To Range("R13").Value $Range("R14")$.Value = DC Count **If Range("I763").Value >= Range("D3").Value And Range("I764").Value >= Range("D3").Value And Range("I765").Value >= Range("D3").Value And Range("I766").Value >= Range("D3").Value And Range("I767").Value >= Range("D3").Value And Range("I768").Value >= Range("D3").Value And Range("I769").Value >= Range("D3").Value And Range("I770").Value >= Range("D3").Value And Range("I771").Value >= Range("D3").Value And Range("I772").Value >= Range("D3").Value And Range("I773").Value >= Range("D3").Value And Range("I774").Value >= Range("D3").Value Then** Range("N5:Z5").Select Selection.Copy Sheets("Result").Select Range("A65536").End(xlUp).Offset(1).Select Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks :=False, Transpose:=False Application.CutCopyMode = False ActiveWorkbook.Save Sheets("Simulation").Select End If Next DC_Count Next Supplier_Count

End Sub

APPENDIX B : Optimal Results And Sensitivity Analyses

B1 : Optimal Inventory Policies at 97.5 % Store Target Service Level

B2 : Optimal Inventory Policies at Different Store Target Service Levels

B3 **: Optimal Inventory Policies under Current Delivery Frequency and 100% Twice-a-week Delivery Frequency to the Stores**

