

IBM Research Report

Empirical Methods for Two-Echelon Inventory Management with Service Level Constraints Based on Simulation-Regression

Lin Li

Sabre Holdings
3150 Sabre Drive
Southlake, TX 76092
USA

Karthik Sourirajan, Kaan Katircioglu

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
USA



Research Division

Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

EMPIRICAL METHODS FOR TWO-ECHELON INVENTORY MANAGEMENT WITH SERVICE LEVEL CONSTRAINTS BASED ON SIMULATION-REGRESSION

Lin Li*

Sabre Holdings
3150 Sabre Drive
Southlake, TX 76092, USA

Karthik Sourirajan
Kaan Katircioglu

IBM TJ Watson Research Center
1101 Kitchawan Road, Route 134
Yorktown Heights, NY 10598, USA

*The research was performed when the author was a post-doc at IBM TJ Watson Research Center, Yorktown Heights, NY

ABSTRACT

We present a simulation-regression based method for obtaining inventory policies for a two-echelon distribution system with service level constraints. Our motivation comes from a wholesale distribution system in the consumer products industry with thousands of products that have different cost, demand, and lead time characteristics. We need to obtain good inventory policies quickly so that supply chain managers can run and analyze multiple scenarios effectively in reasonable amount of time. While simulation-based optimization approaches can be used, the time required to solve the inventory problem for a large number of products is prohibitive. On the other hand, available quick approximations are not guaranteed to provide satisfactory solutions. Our approach involves sampling the universe of products with different problem parameters, obtaining their optimal inventory policies via simulation-based optimization and then using regression methods to characterize the inventory policy for similar products. We show that our method obtains near-optimal policies and is quite robust.

1 INTRODUCTION

There are numerous types of inventory problems that arise from the differences in their supply chain structures, product characteristics and business objectives that drive the way companies want to manage their inventory policies. Over many decades, this led to a fairly high level of proliferation of models and publications in the inventory control literature. In this paper, we will explain a two-echelon problem we faced in a client project.

Our client is a major distributor in the United States. They are a wholesale buyer of over 40,000 different products from more than 2,000 manufacturers and vendors. They do not have any manufacturing operations. They sell the products to various different types of customers. Every day, they process a few million customer order transactions and deliver products to over 25,000 locations. They have a relatively simple distribution network that consists of two echelons (see Figure 1). In the first echelon, there are two central distribution centers and in the second there are over 25 local warehouses or distribution centers (LDCs). We consider the case where vendors ship their products to the CDCs, which in turn distribute the supply to all LDCs that have demand for the products. The CDCs are not merely distribution centers that do cross docking, but they actually hold inventory with two purposes: to fulfill supply orders from LDCs and to fulfill orders from customers in their territories. Vendors fulfill supply orders from the CDCs with random lead times. The CDCs replenish inventory in the LDCs also with random lead times. In addition, order arrivals and order quantities from customers to the CDCs and the LDCs are random. Therefore, the CDCs and the LDCs must hold safety stock in order to meet customer demand with high probability (typically above 95%). Since inventory costs, and supply lead times are different in different sites, the optimal inventory policies in the CDCs and the LDCs are also different.

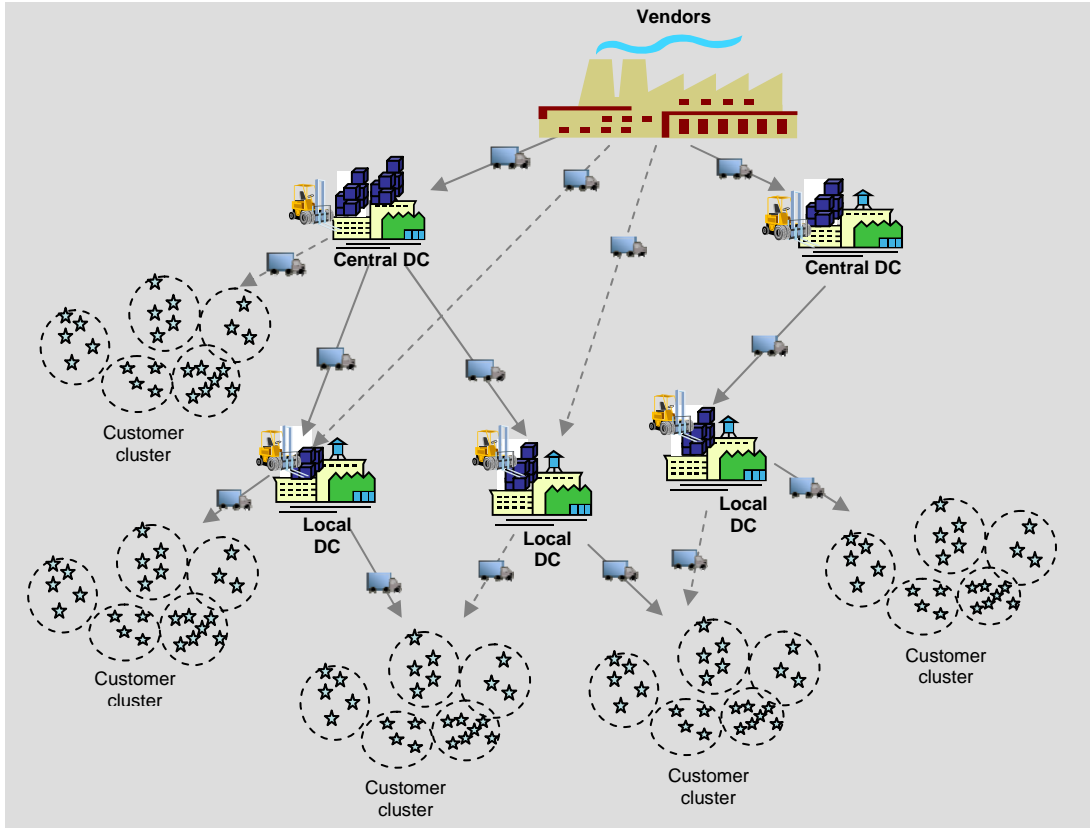


Figure 1: Two-echelon distribution network

The client's objective is to find the inventory policies at the CDCs and the LDCs that achieve their service level requirements for each product while minimizing the sum of their replenishment order costs and inventory holding costs.

The objective of our project was to help the client analyze the impact of any supply chain actions on the operating costs and working capital. We developed a web-based solution to be used by supply chain managers for scenario analysis and decision-support. Because of the high number of products that they have to manage in all locations, we had to develop solution methods that can quickly evaluate the impacts of supply chain actions. Since calculating exact optimal solutions can take considerable amount of time, we developed a simulation-regression approach to evaluate the different scenarios quickly. Our approach involves sampling the universe of products with different problem parameters, obtaining their optimal inventory policies via simulation-based optimization and then using regression methods to characterize the inventory policy for similar products.

The rest of this paper is organized as follows. We present a brief review of existing literature in the next section, followed by the problem definition and formulation in section 3. Then, we explain the solution approach we used to obtain near-optimal solutions quickly in section 4. We present results of numerical experiments in section 5 and finally conclude with a summary discussion of lessons learned and potential future research in section 6.

2 LITERATURE REVIEW

Although the inventory control literature is very rich, in the case of multi-echelon inventory problems with service constraints and random lead times, there is much fewer work. Random lead times are always

challenging even for single echelon problems since the technique of using the inventory position to derive long run average cost and service is not very useful. When we complicate this further with multiple locations and two levels in distribution, it becomes much harder to formulate and solve the problem using known techniques. Hence, researchers resorted to heuristic techniques or approximations to address such problems. However, we know from experience that in overwhelming majority of inventory problems, the cost function is fairly flat around the optimal policy and hence it is relatively easier to find good policies using heuristics or approximations.

Ehrhardt (1979) gave one of the earliest useful and practical solution through a power approximation approach for a single item, single location problem with linear inventory holding and backlogging costs when replenishment lead times are fixed and inventory is reviewed periodically. The paper gives simple approximations to calculate near optimal (s,S) inventory policies where s is the reorder level and S is the order up-to level. Under this policy, when inventory position drops down to s , a replenishment order is placed at an amount that brings the inventory position to S . Inventory position is defined as inventory on hand plus replenishment orders in the pipeline minus backlogged demand. The paper used test results from 288 problems to show that the approximation gives policies that yield costs within 1% of the optimal value. Ehrhardt and Mosier (1984) revise this power approximation to better handle low demand variability and scaled demand units.

Based on this power approximation approach, Schneider and Ringuest (1990) came up with analytical approximations for calculating these policies for the same problem, but with service constraints. They defined service as the long run average backlog relative to the average demand. The approximation requires the knowledge of the mean and the variance of demand. Computational tests show that the actual service is within 1% of the target. Sridhar and Fu (1998) study the same inventory problem where replenishment lead times are random and order arrivals can cross each other (known as “order crossing”). They show a simulation-based procedure that can be used to calculate near-optimal inventory policies. They assume the objective is to minimize long run average cost with service level constraint where service is defined by fill rate (i.e. fraction of demand satisfied immediately with inventory on hand). Their method gives policies with costs within 5% of the optimal for the problems they tested. A good literature review of previous work that uses simulation approach can also be found in Sridhar and Fu (1998).

Boyaci and Gallego (2001) study a serial system where there are multiple stages of production or distribution activities in a series and there is a fill rate service constraint at the final stage. They develop an optimal solution and a computationally efficient heuristic where they restrict the number of stages that can hold inventory to two. A prior work done by Gallego and Zipkin shows that this restriction has very little impact on the cost. Simchi-Levi and Zhao (2005) also study a single product, multi-stage problem with service constraints. Their problem is more general in that it has a tree structure (i.e. each location can have multiple suppliers and each supplier can supply to multiple locations). They assume supply lead times are stochastic and demand follows a Poisson process. All unmet demand is backlogged. They derive recursive equations for calculating the amount of delay replenishment orders will experience. Based on this, they provide approximations and algorithms to calculate base-stock policies in all locations.

There are many works that study two-echelon inventory problems without service constraints. Readers interested in studying these problems can look at Schwarz et al (1985), Axsater (1990, 1993, 1997, 1998, 2003), Chen and Zheng (1997), Cachon (2001).

Amongst the work that address two echelon problems with service constraints, the most relevant one to this paper is Schneider et al. (1995) which extends Ehrhardt’s power approximation to a two-echelon problem with service constraint. They assume a single warehouse with a “probability of no stock out” service constraint that supplies multiple stores, each of which has a fill rate constraint. All locations use (s,S) policies. The warehouse has a fixed replenishment lead time from the outside supplier and the stores have stochastic replenishment lead times from the warehouse. They provide an iterative procedure for adjusting lead time distribution the stores experience due to possible shortages at the warehouse. They provide numerical study results that show the accuracy of the approximation.

The problem we address in this paper is based on the operational environment of our client and their objectives and has a number of differences from the literature. We assume both the CDC and the LDCs experience stochastic supply lead times; customer service levels at the LDCs are defined as the probability of satisfying demand at least partially; unfilled demand at the LDCs are lost whereas the CDCs can backlog unsatisfied replenishment orders.

3 PROBLEM DEFINITION

We consider the two-echelon supply chain outlined in section 1. Although there is more than one CDC, for a given product only one of the CDC stocks inventory and provides supply to the LDCs. For this reason, we will assume that there is one CDC and multiple LDCs in our model. Daily demand for each product is independent, identically distributed Normal random variable. Each location carries inventory for the products with demand at that location. Each location has a daily review of inventory and can place supply orders based on the review. All locations use (s, S) inventory policies (i.e., when inventory position drops down to s , a supply order is placed at an amount that brings the inventory position to S .) The LDCs place the supply order to the CDC and the CDC places the supply order to the vendor. Unfilled customer demands are lost in both the CDC and the LDCs whereas the unfilled LDC supply orders are backlogged at the CDC. Therefore, inventory position is defined as inventory on hand plus supply orders in the pipeline minus the backlogged demand. Backlogged demand is always zero for LDCs since they lose demand when they are short of inventory. Every replenishment order has a fixed order cost, and it arrives after a lead time, also assumed to be an independent and identically distributed Normal random variable.

Since the CDC also serves customer demand, we create a pseudo-LDC at the CDC that serves the customers of CDC. The pseudo-LDC faces zero lead time from the CDC for its orders. The equivalent of the pseudo-LDC in practice is the area at the CDC that functions as a mini-LDC and is used to serve customer demand. This way, we maintain the two-echelon structure of the problem where the CDC only satisfies demands from the LDCs (including the pseudo-LDC) and the LDCs satisfy customer demands. The sequence of events in a daily cycle for each product is as follows:

1. LDCs observe their customer demands
2. LDCs receive supply scheduled to arrive that day.
3. LDCs fulfill customer demand as much as possible with inventory on hand.
4. LDCs place their supply orders on the CDC.
5. CDC observes their total demand (sum of customer demand at CDC and LDC supply orders)
6. CDC receives supply scheduled to arrive that day
7. CDC satisfies its total demand as much as possible from on-hand inventory and backlogs any LDC order that cannot be satisfied.
8. CDC places its supply order on the vendor.
9. CDC and LDCs account for lost sales, backlogs, order costs, and cost of inventory on hand.

We will use the following notation where a period typically represents a day.

Indices

V : vendor

i : index representing CDC and LDCs ($i = 0$ indicates CDC, $i = 1, \dots, n$ indicates LDCs, including the pseudo - LDC)

J_i : number of customers served by location i

j : index representing customers for location i ($j = 1, \dots, J_i$)

t : index representing time period

Random variables

$O_{ij}(t)$: order quantity in period t from customer j in location i

$D_i(t)$: demand in period t in location i

$L_i(t)$: supply replenishment lead time for an order placed in period t in location i

Inventory policy-based variables

$B_i(t)$: backlog at the end of period t at location i

$I_i(t)$: inventory on hand at the end of period t at location i

$IP_i(t)$: inventory position the end of period t at location i just before placing a replenishment order

$Q_i(t)$: size of replenishment order placed at the end of period t by location i

$Q_i^+(t)$: sum of replenishment orders in the pipeline that arrive in period t at location i

$Q_i^-(t)$: sum of replenishment orders in the pipeline that arrive after period t at location i

Unit costs

h_i : unit inventory holding cost at location i

a_i : order cost for each replenishment order placed by location i

We assume that $L_i(t)$ is a stationary normal random variable, i.e., the distribution of $L_i(t)$ does not vary with t but every location i potentially has different replenishment lead time distributions. We present the derivation of quantities at the LDCs and the CDC next. While we model the random variables and their effects accurately in our simulation setting, we present the random variables as deterministic in the equations below for ease of understanding.

In order to formulate $Q_i^+(t)$, the orders that arrive at location i in period t , we define a indicator variable $\theta_{it}^+(k) = \begin{cases} 1, k + L_i(k) = t \\ 0, \text{otherwise} \end{cases}$. We have $Q_i^+(t) = \sum_{k=0}^{t-1} \theta_{it}^+(k) Q_i(k)$.

In order to formulate $Q_i^-(t)$, the orders that arrive at location i after period t , we define a indicator variable $\theta_{it}^-(k) = \begin{cases} 1, k + L_i(k) > t \\ 0, \text{otherwise} \end{cases}$. We have $Q_i^-(t) = \sum_{k=0}^{t-1} \theta_{it}^-(k) Q_i(k)$.

For all LDCs $i=1, \dots, n$, we have the derivation of quantities as follows.

The total demand in period t at the LDCs is given by:

$$D_i(t) = \sum_{j=1}^{J_i} O_{ij}(t), i=1, \dots, n$$

Inventory on hand at the LDCs at the end of period t is given by:

$$I_i(t) = (I_i(t-1) + Q_i^+(t) - D_i(t))^+, i=1, \dots, n.$$

At the end of each period t , each LDC places orders on the CDC at an amount $Q_i(t)$. All orders placed must bring the inventory position up to S_i . That is,

$$Q_i(t) = (S_i - IP_i(t))^+, i=1, \dots, n$$

For the CDC, the total demand in period t is the sum of replenishment orders from LDCs and can be written as:

$$D_0(t) = \sum_{i=1}^n Q_i(t)$$

Therefore, the inventory on hand at the CDC at the end of period t is

$$I_0(t) = (I_0(t-1) + Q_0^+(t) - B_0(t-1) - D_0(t))^+$$

where, $B_0(t)$ represents the backlogged replenishment orders from LDCs and is given by

$$B_0(t) = (B_0(t-1) + D_0(t) - I_0(t-1) - Q_0^+(t))^+$$

The inventory position at the CDC before placing an order in period t can be written as:

$$IP_0(t) = I_0(t) + Q_0^-(t) - B_0(t).$$

Replenishment orders are similar to LDCs and can be written as:

$$Q_0(t) = (S_0 - IP_0(t))^+$$

$$\text{Now, let } \varepsilon_i(t) = \begin{cases} 1, & \text{if } Q_i(t) > 0 \\ 0, & \text{otherwise} \end{cases}$$

Here when $\varepsilon_i(t)$ is 1 it indicates that at the end of period t , location i places a replenishment order. The total cost in period t is given by

$$C(t) = \sum_{i=0}^n (h_i I_i(t) + a_i \varepsilon_i(t))$$

Since we assume demand is stationary, we are interested in minimizing long run average cost subject to a minimum long run average service level, α . In order to calculate the service level, we will define the following indicator variables:

$$\delta_i(t) = \begin{cases} 1, & \text{if } D_i(t) > 0 \text{ \& } I_i(t-1) + Q_i^+(t) > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$o_i(t) = \begin{cases} 1, & \text{if } D_i(t) > 0 \\ 0, & \text{otherwise} \end{cases}$$

Here, $\delta_i(t)$ is 1 if there is a demand and we satisfy at least one unit of demand. This captures the unique service level the way our client wants to define. Then, we can formulate the problem we want to solve as follows:

$$\begin{aligned} & \text{Minimize } \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T C(t) \\ & \text{s.t. } \frac{\sum_{t=1}^T \delta_i(t)}{\sum_{t=1}^T o_i(t)} \geq \alpha, \forall i = 1, \Lambda, n \end{aligned}$$

The left hand side of the constraint is the long run average service level achieved at the LDCs. Note that the replenishment lot size decision in period t impacts inventory and service in subsequent periods. Therefore, an analytical solution to this problem is not easy to calculate. One can attempt to use a dynamic programming (DP) formulation. However, the state space is large in dimension and any known techniques to solve the underlying DP will not be very fruitful. In the next section, we will present alternative practical ways for calculating good inventory policies.

4 SOLUTION APPROACH

The problem described in section 3 is complex and intractable. The objective of our project was to develop a tool used by managers to do scenario analyses quickly and therefore we need to develop solution

methods that can yield good solutions quickly. While simulation-optimization approaches can be used, the time required to solve the inventory problem for a large number of products is prohibitive. On the other hand, available quick approximations are not guaranteed to provide satisfactory solutions. We use a simulation-regression approach in this paper that involves sampling the universe of products with different problem parameters, obtaining their optimal inventory policies via simulation-optimization and then using regression methods to characterize the inventory policy for similar products. In order to evaluate the effectiveness of our method, we develop a heuristic that solves the problem approximately and compare the performance of the heuristic and our simulation-regression approach to the optimal solution obtained using simulation-optimization.

In all our solution approaches, we set the difference between the reorder point (s_i) and the order up to levels (S_i) to be the economic order quantity as follows:

$$(S_i - s_i) = \sqrt{2 * a_i * E(D_i(t)) / h_i}$$

Here $E(D_i(t))$ is the expected value of demand in period t in location i . In order to meet the minimum lot size restrictions for each product faced by our client, we use either $(S_i - s_i)$ or the minimum lot size whichever is larger.

Since we reformulated the problem to turn it into a two-echelon system where the first echelon serves only the orders from the second echelon, we need to calculate the optimal service level at the CDC for the orders from the LDCs. This would then implicitly determine the inventory policy at the CDC. We use the service level decision at the CDC as the anchor for our solution approaches outlined in the following sections. We will first present the heuristic we developed followed by a discussion of the simulation-based optimization approach. We then discuss the regression model we developed to calculate the optimal service levels at the CDCs.

4.1 Heuristic for the two-echelon inventory optimization problem

The motivation for our heuristic comes from similar approaches used for two-echelon inventory systems. For each product, we set the service level requirements at the LDC, α_i ($i = 1, \dots, n$) to be equal to α , the service level requirement of the product. Let α_0 be the service level at the CDC for the orders from the LDCs. The lead times LDCs face need to be adjusted to account for inventory shortages the CDC may experience. We compute the adjusted lead time faced by the LDCs as follows:

$$\bar{L}_i(t) = L_i(t) + (1 - \alpha_0)L_0(t), \forall i = 1, \dots, n$$

The above equation implies that the lead time seen by an LDC will be the lead time from the CDC to the LDC when the CDC has inventory in stock and will be sum of the lead time from the CDC to the LDC and the lead time from the vendor to the CDC when the CDC is out of stock. This is an overestimation of the lead time for the LDCs since $L_0(t)$ is the maximum possible lead time when there is a shortage.

We assume outside supplier never delays orders. We then model the distribution of the demand during lead time at the LDCs based on their daily demand distribution and their adjusted lead time distribution. We model the distribution of the demand during lead time at the CDC using the total demand distribution across all LDCs and the lead time distribution at the CDC.

We then set the reorder points at all the DCs as follows:

$$s_i = \mu_i^{LD} + z_{\alpha_i} \sigma_i^{LD}, \forall i = 0, \dots, n$$

where μ_i^{LD} is the mean demand over lead time at location i , σ_i^{LD} is the standard deviation of demand over lead time at location i and z_{α_i} is the inverse of the standard normal distribution evaluated at α_i . This reorder level is very commonly used because of its simplicity and its optimality for some simple class of inventory problems. We then compute the average daily total cost as:

$$TC = \sum_{i=0}^n \left(h_i \left(s_i + \frac{(S_i - s_i)}{2} \right) + a_i \frac{\mu_i}{(S_i - s_i)} \right)$$

where μ_i is the mean daily demand at location i . Using this approximation, we then search for the value of α_0 that gives the lowest total cost. We follow the procedure above for all products to achieve the service level at the CDC and the inventory policies at the CDC and the LDCs.

We outline our simulation-based optimization framework in the next section followed by a discussion of the regression model we developed to determine the optimal service level at the CDC as a function of supply chain parameters.

4.2 Simulation-optimization for the two-echelon inventory optimization problem

In order to develop a simulation-optimization framework, we first developed a discrete event simulation model of the two echelon supply chain problem outlined in section 3. Simulation models to analyze supply chain performance have been used in literature (see Sridhar and Fu 1998 for a review). Our simulation model had some unique differences due to the client's operational conditions. Since we model a pseudo-LDC, we had to adjust the simulation model to account for the fact that if the pseudo-LDC ran out of inventory, but the CDC has inventory in stock, it was possible to get inventory from the CDC to satisfy the customer demand at the pseudo-LDC immediately. In order to do this, we modify the sequence of events outlined in section 3 by introducing an additional event right after step 7 that allows receiving orders at the LDC later in the day. This way the pseudo-LDC can receive orders placed on the same day and thus capture the impact of getting emergency inventory from the CDC. We also needed to make sure that the portion of the pseudo-LDC orders that are placed to satisfy customer demand on the same day are not backlogged at the CDC since any customer demand that is not satisfied by the end of the day is lost. The second unique aspect of our simulation model is the computation of the service level statistics. As described earlier, our client uses a service level definition that is different from what is used in literature and we had to account for that. We use the standard normal random variable generators to generate the demands and the lead times. To evaluate the supply chain cost and performance at each setting, we run the simulation model for 200 periods. We collect statistics starting from period 31 to avoid initialization effects. We also controlled the seed of the random variable generator to expose the policies we simulated to the same random conditions. This reduces the chance of picking poorer policies as optimal.

Once we had the simulation model set up, we created a simulation-based optimization framework that had two levels. In the inner level, for a given (s,S) policy at the CDC, we use bisection search to find the optimal (s,S) policies at the LDCs that minimize the total cost while achieving the desired service levels. Since we fix the $(S - s)$ value based on the economic order quantities, the average order costs are fixed. Therefore, we only need to search for the lowest value of s at the LDCs that achieves the desired service levels. In the outer loop, we find the value of s at the CDC that minimizes the total cost using a cost-based bisection-type search.

Notice that each time we change the value of s at the CDC, we need to run a bisection search for the LDCs to find their optimal s values. Thus, we run nested searches with the evaluation of each search point requiring a simulation run. Using such an approach in our tool to find the optimal inventory policies would not be practical since we have 50,000 products and managers would like to try different scenarios quickly before making a decision. In order to provide reasonable solutions quickly, we use regression methods to model the optimal service level at the CDC as a function of the supply chain parameters. We then use this regression model to drive the identification of inventory policies. We outline the regression model in the next section.

4.3 Regression Model

We can simplify the computation of optimal CDC service level α_0^* by approximating it using a regression model. The reason we choose to approximate service level instead of reorder point s_0 is due to the fact that the service level is a more normalized measure than the reorder point.

$$\log \frac{\alpha_0^r}{1-\alpha_0^r} = \beta_0 + \beta_1 \log(N_{LDC}) + \beta_2 \log(Q) + \beta_3 \log(\mu_D) + \beta_4 \log(\sigma_D) + \beta_5 \log(\mu_L) \quad (1)$$

$$+ \beta_6 \log(\sigma_L) + \beta_7 \log(h) + \beta_8 \log(\alpha_F)$$

where N_{LDC} is the number of LDCs in the two echelon network, Q is the average order quantity at the CDC given by $(S-s)$, (μ_D, σ_D) is the mean and standard deviation of the daily demand at the CDC, (μ_L, σ_L) is the mean and standard deviation of the supply lead time faced by the CDC, h is the unit holding cost and α_F is the target service level at the LDC.

In the next section, we outline the method we use to convert the estimate of the optimal service level at the CDC into inventory policies that can be implemented in practice and also provide a discussion on the computational experiments that we use to test the performance of our solution approaches.

5 COMPUTATIONAL RESULTS

In this section, we analyze the performance of inventory policies obtained using the solution approaches we presented in section 4. We use a simulation-based optimization method (section 4.2) to benchmark the solutions obtained from our two approximation methods for computing the CDC service level. The first approximation method uses a heuristic search based on an approximate cost function (section 4.1) and the second approach uses a regression model of the supply chain parameters (section 4.3). Once we find the CDC service level, we need to derive accurate inventory policies that supply chain managers can implement. Deriving accurate inventory policies based on a given CDC service level requires sophisticated approximation techniques. In this paper, we experiment with a bisection search based method and a simple approximation as an initial attempt towards finding good approximations. Overall, we analyze the following five different methods in our numerical experiments:

- Simulation-based optimization (SBO): This approach uses a bisection search based on the expected costs generated via Monte Carlo simulation to find the optimal CDC service level and inventory policies. The results from this method are used as the benchmarks to evaluate the performances of the other methods.
- Heuristic-based method (HBM): We obtain the CDC service level, $\hat{\alpha}_0^h$, using the heuristic given in section 4.1. We then use a simulation-based bisection search to find the optimal s values at the CDC and the LDCs that meet the service level requirements at the LDCs with $\hat{\alpha}_0^h$ as the service level at the CDC.
- Regression-based method (RBM): We obtain the CDC service level, $\hat{\alpha}_0^r$, using the regression model given in section 4.3. We then use a simulation-based bisection search to find the optimal s values at the CDC and the LDCs that meet the service level requirements at the LDCs with $\hat{\alpha}_0^r$ as the service level at the CDC.

- Heuristic with reorder point approximation (HRPA): We obtain the CDC service level, $\hat{\alpha}_0^h$, using the heuristic given in section 4.1. We then approximate the reorder point as $\hat{s}_0^c = \mu_0^{LD} + z_h \sigma_0^{LD}$ where $(\mu_0^{LD}, \sigma_0^{LD})$ is the mean and standard deviation of the CDC lead time demand, and $z_h = \Phi^{-1}(\hat{\alpha}_0^h)$ with $\Phi^{-1}(\cdot)$ being the inverse function of standard normal distribution.
- Regression with reorder point approximation (RRPA): We obtain the CDC service level, $\hat{\alpha}_0^r$, using the regression model given in section 4.3. We then approximate the reorder point as $\hat{s}_0^r = \mu_0^{LD} + z_r \sigma_0^{LD}$ where $(\mu_0^{LD}, \sigma_0^{LD})$ is the mean and standard deviation of the CDC lead time demand, and $z_r = \Phi^{-1}(\hat{\alpha}_0^r)$ with $\Phi^{-1}(\cdot)$ being the inverse function of standard normal distribution.

In our numerical experiments, we first calibrate our regression model using a well defined training data set. We then analyze the performances of the policies based on a test data set. Our training data set consists of 200 items whose parameters span over a wide spectrum:

- The number of LDCs can be between 1 and 25
- The item cost ranges from \$1 to \$12,000
- The coefficient of variation (CV) of demand varies from 0.1 to 16
- The CV of lead time ranges from 0.14 to 1.61
- The service level requirements at the LDCs take four different values $\{0.90, 0.95, 0.97, 0.99\}$

Altogether, we had 800 scenarios (200 items times analyzed for 4 service level settings). We applied simulation-based optimization on these 800 scenarios to get the optimal CDC service levels. We then ran a regression analysis to approximate optimal CDC service level based on the item and supply chain characteristics. We obtained the following coefficients for the regression model in our experiments:

$\beta_0 = 0.43$, $\beta_1 = 0.15$, $\beta_2 = 1.04$, $\beta_3 = -0.76$, $\beta_4 = -0.43$, $\beta_5 = -1.24$, $\beta_6 = -0.01$, $\beta_7 = 0.29$ and $\beta_8 = 1.88$. The adjusted R^2 value was 0.66.

Next, we tested the performance of the five different methods (SBO, HBM, RBM, HRPA and RRPA) we mentioned above using a new (but representative) set of items. Let c_i be the expected total cost for method i .

We compute the optimality gap ε_i as $\varepsilon_i = \frac{c_i - c_{SBO}}{c_{SBO}} \times 100\%$.

Table 1 summarizes the performances of different methods in terms of the cost metrics. For the four different LDC service level requirements, we show the expected total cost c_i along with the 95% confidence interval and the optimality gap for the five methods. We observe in Table 1 that HBM outperforms RBM when LDC service level requirement is high while RBM outperforms HBM when the LDC service level requirement is lower. Overall, both HBM and RBM have good performances with optimality gaps of less than 6%. This indicates that the CDC service levels we computed using the heuristic and the regression model are good approximations to the optimal levels. We also observe from Table 1 that the 95% confidence intervals are narrow enough to indicate that the optimality gaps are statistically significant. The naive reorder point approximations used by HRPA and RRPA results in a much higher expected total cost and optimality gaps increase as the LDC service level requirements go down. This indicates that we need to explore more sophisticated curve-fitting methods to find good solutions for problems with low service levels. For the purpose of our project, since the client has very high service level targets, these methods had sufficient performances.

Table 1: Performance of different methods

LDC SLA	Method	Expected Total Cost c_i	95% CI c_i^{low}	95% CI c_i^{high}	Optimality Gap ε_i (%)
0.99	SBO	\$4,037,779	\$4,027,146	\$4,048,411	-
	HBM	\$4,147,235	\$4,136,307	\$4,158,163	2.71%
	RBM	\$4,195,333	\$4,184,317	\$4,206,349	3.90%
	HRPA	\$4,690,856	\$4,679,563	\$4,702,150	16.17%
	RRPA	\$4,832,409	\$4,821,070	\$4,843,749	19.68%
0.97	SBO	\$3,244,001	\$3,233,667	\$3,254,335	-
	HBM	\$3,399,855	\$3,389,873	\$3,409,837	4.81%
	RBM	\$3,414,831	\$3,404,193	\$3,425,469	5.27%
	HRPA	\$4,409,174	\$4,397,882	\$4,420,466	35.92%
	RRPA	\$4,505,716	\$4,494,546	\$4,516,886	38.89%
0.95	SBO	\$2,886,264	\$2,876,897	\$2,895,631	-
	HBM	\$3,035,197	\$3,025,639	\$3,044,755	5.29%
	RBM	\$3,024,720	\$3,014,628	\$3,034,813	4.93%
	HRPA	\$4,252,750	\$4,241,519	\$4,263,981	47.34%
	RRPA	\$4,329,884	\$4,318,768	\$4,341,000	50.02%
0.90	SBO	\$2,420,562	\$2,411,668	\$2,429,456	-
	HBM	\$2,485,791	\$2,477,007	\$2,494,575	2.69%
	RBM	\$2,468,658	\$2,459,639	\$2,477,677	1.99%
	HRPA	\$4,006,672	\$3,995,556	\$4,017,787	65.53%
	RRPA	\$4,063,193	\$4,052,139	\$4,074,247	67.86%

In HBM and RBM, we search for the inventory policies that satisfy the service level requirements at the LDCs. However, in HRPA and RRPA, we use an approximation to set the inventory policies. In order to analyze the impact of the reorder point approximation on LDC service levels in HRPA and RRPA, we collected service level statistics for each item-location pair. Let $\bar{\alpha}_F$ be the daily demand weighted average service levels at the LDCs and ζ_F be the daily demand weighted standard deviation of the service levels at the LDCs across all item-location pairs. Let n_F denote the number of item-location pairs for which the service level at the LDCs falls below the desired target. We have a total of 1506 item-location pairs in our experiments. Table 2 shows the performance of HRPA and RRPA in terms $\bar{\alpha}_F$, ζ_F and n_F (in number and percentage).

Table 2: LDC service level using reorder point approximation

LDC SLA	HRPA			RRPA		
	$\bar{\alpha}_F$	ζ_F	n_F	$\bar{\alpha}_F$	ζ_F	n_F
0.99	0.99423	0.0076	78 (5%)	0.99585	0.0071	72 (5%)
0.97	0.99051	0.0103	93 (6%)	0.99269	0.0098	86 (6%)
0.95	0.98753	0.0124	108 (7%)	0.99022	0.0119	104 (7%)
0.90	0.98161	0.0161	122 (8%)	0.98498	0.016	118 (8%)

We can see from Table 2 that both HRPA and RRPA exceed the average service level requirements at the LDCs in all cases. The service levels achieved by HRPA and RRPA fall below targets only for a very small portion of item-location pairs. Since our client measures the service level requirements at an aggre-

gate level across all item-location pairs, we conclude that the reorder point approximations over-achieve targets.

Table 3 presents the average CDC service levels obtained using SBO, RBM and HBM. We can see from Table 3 that while RBM results in a lower CDC service level compared to SBO, while HBM results in a higher CDC service level.

Table 3: Average CDC service level

LDC SLA	SBO	RBM	HBM
0.99	0.7312	0.5724	0.7568
0.97	0.6568	0.5638	0.7295
0.95	0.6336	0.5549	0.7083
0.90	0.6128	0.5317	0.6718

Figure 2 shows an example of a typical expected total cost curve as a function of the CDC service level. We want to emphasize that the cost function for most of items is very flat around the optimal CDC service level. Therefore, small inaccuracies in predicting the optimal CDC service levels should not cause costs to be significantly higher than the optimals.

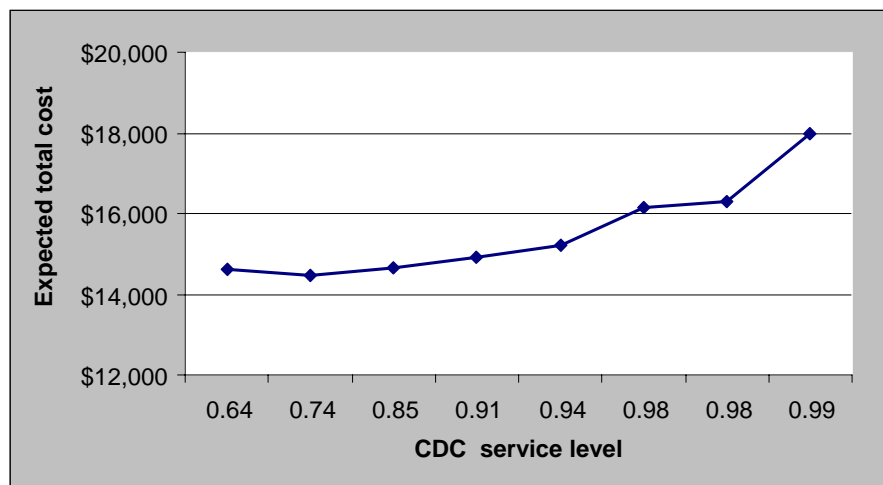


Figure 2: Expected total cost at different CDC service levels

6 SUMMARY

In this paper, we presented solution approaches to find approximate inventory policies for a two echelon inventory system with stochastic demand, stochastic lead times, fixed order costs and customer service level requirements. The anchor for our solution approaches was identifying the optimal service level at the CDC. We then used the service level at the CDC to derive the inventory policies at the CDC and the LDCs. We developed two methods for calculating the service level at the CDC: a heuristic that searches for the optimal CDC service level based on an approximate cost function and a simulation-regression based method that uses a regression model to find the optimal CDC service level as a function of supply chain parameters.

We find that both methods give CDC service levels that provide excellent cost performance compared to simulation-based optimization method. Since the heuristic method required a search to obtain the CDC service levels, the time required to execute the heuristic for 50,000 products can make it prohibitive for

use in a tool that requires short run times. We find that the simulation-regression based method is able to provide solutions that are comparable to or better than the heuristic. The time required to compute the CDC service level using a regression model is also very minimal. Thus, it serves as the best alternative to use for computing the CDC service level in our tool.

We find that the approximations we use to estimate the optimal inventory policies at the warehouses based on the CDC service levels can be improved. While the approximations over-achieve the service level requirements at the LDCs, their cost performance is not as good as the other methods. This will be the focus of our future research in which we will explore other models to estimate optimal inventory policies.

REFERENCES

- Axsater, Sven. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research* 38 (1) 64-69.
- Axsater, Sven. 1993. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* 41 (4) 777-785.
- Axsater, Sven. 1997. Simple evaluation of echelon stock (R,Q) policies for two-level inventory systems. *IIE Transactions* 29 661-669.
- Axsater, Sven. 1998. Evaluation of installation stock based (R,Q)-policies for two-level inventory systems with poisson demand. *Operations Research* 46 (3) S135-S145.
- Axsater, Sven. 2002. Heuristic Methods for centralized control of one-warehouse, N-retailer inventory systems. *M&SOM* 4 (1) 75-97.
- Axsater, Sven. 2003. Approximate optimization of a two-level distribution inventory system. *Int'l J. of Prod. Econ.* 81-82 545-553.
- Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2000. *Discrete-event system simulation*. 3rd ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- Bashyam, S., Fu, M.C. Optimization of (s,S) inventory systems with random lead times and a service level constraint. *Management Science* 44 (12) S243-S256.
- Boyaci, T., Gallego, G. 2001. Serial production / distribution systems under service constraints. *M&SOM* 3 (1) 43-50.
- Cachon, P. Gerard. 2001. Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. *Operations Research* 49 (1) 79-98.
- Chen, F., Zheng, Y. 1997. One-warehouse multiretailer systems with centralized stock information. *Operations Research* 45 (2) 275-287.
- Cheng, R. C. H. 1994. Selecting input models. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 184-191. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Chien, C. 1989. Small sample theory for steady state confidence intervals. Technical Report No. 37, Department of Operations Research, Stanford University, Stanford, California.
- Ehrhardt, R. 1979. The power approximation for computing (s,S) inventory policies. *Management Science* 25 (8) 777-786.
- Ehrhardt, R., and Mosier, C. 1984. A revision of the power approximation for computing (s,S) policies. *Management Science* 30 (5) 618-622.
- Gupta, S. S., K. Nagel, and S. Panchapakesan. 1973. On the order statistics from equally correlated normal random variables. *Biometrika* 60:403-413.
- Hammersley, J. M., and D. C. Handscomb. 1964. *Monte Carlo methods*. London: Methuen.
- Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis*. 3rd ed. New York: McGraw-Hill, Inc.

- Mugglenet 2005. Interview with J. K. Rowling. Available via <http://www.mugglenet.com/jkrinterview.shtml> [accessed January 3, 2008].
- Schneider, H., Rinks, D.B., Kelle, Peter. 1995. Power approximation for a two-echelon inventory system using service levels. *POM* 4 (4) 381-400.
- Schwarz, L.B., Deurmeyer, B.L., and Badinelli, R.D. 1985. Fill rate optimization in a one-warehouse N-identical retailer distribution System. *Management Science* 31 (4) 488-498.
- Schneider, H. and Ringuest J.L. 1990. Power approximation for computing (s,S) policies using service level. *Management Science* 36 (7) 822-834.
- Schruben, L. W. 1979. Designing correlation induction strategies for simulation experiments. In *Current issues in computer simulation*, ed. N. R. Adam and A. Dogramaci, 235–256. New York: Academic Press.
- Simchi-Levi, D., Zhao, Y. 2005. Safety stock positioning in supply chains with stochastic lead times. *M&SOM* 7 (4) 295-318.
- Steiger, N. M. 1999. *Improved batching for confidence interval construction in steady-state simulation*. Ph.D. thesis, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available via www.lib.ncsu.edu/etd/public/etd-19231992992670/etd.pdf [accessed January 3, 2008].
- The University of Chicago Press. 2003. *The Chicago manual of style*. 15th ed. Chicago: The University of Chicago Press.

AUTHOR BIOGRAPHIES

LIN LI is currently a Senior Operations Research Analyst at Sabre Holdings, Southlake, TX. She was a postdoctoral researcher in the Business Analytics & Mathematical Sciences Department at IBM's TJ Watson Research Center. She received Ph.D degree in Operations Research from Columbia University. She Her email address is ll2053@gmail.com.

KARTHIK SOURIRAJAN is a Research Staff Member in the Mathematical Sciences Department at IBM's TJ Watson Research Center. He received a B.E.(Hons.) in Mechanical Engineering at Birla Institute of Technology and Science, Pilani, India followed by M.S. and Ph.D. degrees in Industrial Engineering from Purdue University, West Lafayette, Indiana. His research interests include the application of operations research techniques to solving real world problems generally related to supply chains such as forecasting, inventory management, pricing, integrated facility location and logistics. His email address is ksourira@us.ibm.com.

KAAN KATIRCIOGLU is a senior research scientist at IBM Research. He has more than fifteen years of experience in the field of Operations Research, Management Science and Logistics. He received his B.Sc. in Industrial Engineering M.Sc. in Statistics, and Ph.D. in Management Science / Transportation & Logistics. He joined IBM's T.J. Watson Research Center in 1996 as a research staff member. He has worked on over 30 projects related to supply chain management for IBM and its customers. He has over 35 scientific article publications, book chapters, conference presentations, and appeared in many events as an invited speaker or panelist. He has patents and pending patent applications for his work. He is a member of INFORMS, DAS and IEEE. His email address is kaan@us.ibm.com.