THREE INVENTORY MODELS

FOR NON-TRADITIONAL SUPPLY CHAINS

by

BENJAMIN VAUGHN NEVE

CHARLES P. SCHMIDT, COMMITTEE CHAIR

DAVID M. MILLER
BURCU B. KESKIN
ALEXANDER E. ELLINGER
SUBHABRATA CHAKRABORTI
CHARLES R. SOX

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Information Systems,
Statistics, and Management Science
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2011

ABSTRACT

This work considers three different non-traditional supply chain structures with similar demand and replenishment parameters, and similar solution techniques.

In the first article, we develop an inventory model that addresses inventory rationing based on customer priority. We use the framework of a multi-echelon inventory system to describe the physics of a critical level policy. To extend from previous research, we allow multiple demand classes while minimizing a cost objective. We assume a continuous-review, base stock replenishment policy and allow for full backordering. Simulation is used to estimate total expected cost, applying variance reduction to reduce sampling error. First differences are estimated using a Perturbation Analysis unique to inventory rationing literature, heuristics are used to minimize costs.

In the second article, we consider a "stockless" hospital supply chain with inaccurate inventory records. The model presented here is conditional on the level of accuracy in a particular hospital department, or point-of-use (POU). Similar to previous research on inventory inaccuracy, we consider both actual net inventory and recorded inventory in deriving the performance measures. The resultant model is a periodic-review, cost minimization inventory model with full backordering that is centered at the POU. Similar to the previous article, we assume a base stock ordering policy, but in addition to choosing the optimal order-up-to level, we seek the optimal frequency of inventory counts to reconcile inaccurate records. We present both a service level model and a shortage cost model under this framework.

In the final article, we consider a hybrid hospital supply chain with both regular and emergency ordering when inventory records are inaccurate. The resultant model is an extension from the previous article where there are opportunities for both regular replenishments and emergency replenishments. We seek an optimal solution to an approximate cost model, and then we compare the results to a simulation-optimization approach.

DEDICATION

This dissertation is dedicated to my family first: Annalie, Jamison, and Natalie - and to anyone else who encouraged me along the way.

LIST OF ABBREVIATIONS AND SYMBOLS

a_k	=	Truncation error value in simulated replication k
B_i	=	Total backorders at stage i or at the end of period i
$B_{i,i}$	=	External backorders at stage i
$B_{i,i-1}$	=	Internal backorders at stage i
b_i	=	Backorder cost per unit in stage i
C	=	Total cost or expected total cost, a function of different decision variables
C_k	=	Total cost generated by simulated replication k
$\bar{C}$	=	Expected or average total cost, or average daily cost
c_b	=	Backorder cost per unit
c_e	=	Emergency ordering cost per unit
c_h	=	Holding cost per unit
c_p	=	Shortage penalty cost, per unit
D	=	Total demand in a particular period or stage
DCS	=	Abbreviation for "demand class inventory system"
D^L	=	Lead time demand, a Poisson random variable
DLA	=	Abbreviation for the Defense Logistics Agency
Δ_j	=	Signifies a change in value over some variable with subscript j

E_i	=	Emergency base stock level in shift $i = 1, 2$
EO_{ij}	=	Emergency order quantity in shift i of day j
FR_i	=	Fill rate in period i
FR_{min}	=	Minimum fill rate allowed as a constraint
g_X	=	Probability function for some distribution X
G_X	=	Cumulative probability function for some distribution X
G_X^0	=	Complimentary cumulative probability function for some distribution X
G_X^1	=	First order loss function for some distribution X
h	=	Inventory holding cost per unit, per unit time
I	=	End of period actual net inventory, or on-hand inventory for some stage
$I^{'}$	=	Recorded inventory at the beginning of a period
$\underline{I}^{'}$	=	Recorded inventory level at the beginning of a period
$\bar{I}^{'}$	=	Recorded inventory level at the end of a period
IN	=	Net inventory for a particular stage
IPA	=	Abbreviation for "infinitesimal perturbation analysis"
IP	=	Inventory position for a particular stage
j	=	A widely used indexing variable, often represents days or stages
k	=	Fixed cost to count and reconcile inventory records, also used for indexing
L	=	Replenishment lead time
λ	=	Poisson demand parameter, usually for a distinct period or stage
M	=	Partition point for minimum truncation error
m	=	Index of the highest stage where net inventory is positive

μ = Poisson demand parameter, usually a convolution of several demand rates

N = Number of demand classes or number of days in a counting cycle

n = Number of simulation replications

NI = Actual net inventory at the beginning of a particular period

p = Probability that demand is successfully recorded

PA = Abbreviation for "perturbation analysis"

p_i = Probability of an internal demand at a particular stage i

POU = Abbreviation for "point of use"

q_i = Probability of an external demand at stage i

RD = Amount of recorded demand in a particular period

RO_j = Regular order quantity in day j

S = Base stock level or standard deviation

SF = Inventory shortfall for a given period

s_i = Reserve stock level for stage i, or base stock level for stage $i = N$

SSS = Abbreviation for "serial stage inventory system"

$t_{i,1-\alpha/2}$ = The t-distribution value for i degrees of freedom and $\alpha/2$ right tail probability

UD = Amount of unrecorded demand in a particular period

US = Number of units short in a period

X = Represents a general random variable or distribution

x_i = Value of the last Bernoulli trial at stage i

y_{ij} = The product of the last Bernoulli trials for stages i through j

$z_{(1+P^*)/2}$ = The $(1 + P^*)/2$ fractional point of the unit normal distribution

ACKNOWLEDGEMENTS

I first acknowledge God's hand in all of my educational endeavors and especially in the completion of this dissertation. Also, I am grateful to my advisor and my committee for all of their encouragement and insight. I am greatly indebted to the Alabama Productivity Center for the industry experience and for providing the doorway to research that is relevant to both industry and academia. Lastly, I acknowledge the University of Alabama and the department of Information Systems, Statistics, and Management Science for the opportunity to pursue this research and my degree.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This work is focused on inventory management for non-traditional supply chain structures. By non-traditional we mean that the supply chain structures are industry-specific and are different from most textbook supply chains. One of the main obstacles in dealing with industry-specific supply chain structures, is that each industry has unique challenges which are difficult to address analytically. Specifically, we look at three general cases, which are as follows: 1) customer demand is not fulfilled on a first-come, first-served basis; 2) customer demand is not accurately recorded; and 3) inventory is replenished by more than one method. Under these unique structures, we use some common elements to portray, model, and minimize the overall inventory costs for each situation.

The first common assumption for the models presented here is that random customer demand follows a Poisson distribution. This allows for more straightforward derivations of the performance measures. The Poisson distribution is especially helpful as we consider multiple demand classes in the first paper and both recorded and unrecorded demand in both the second and third papers.

We have also assumed, across all three papers, that replenishment occurs according to a base stock inventory policy. This $(S-1, S)$ policy is a special case of an (r, Q) policy, and it is more applicable when there is not a fixed setup cost. In our case, we have assumed no fixed ordering cost in any model, and so, naturally, we have used a base stock policy. However, we do not claim that the base stock policy is necessarily the optimal policy under the conditions considered here. But, as our research expands on earlier works which also assumed base stock replenishment systems, we find the models here contribute to the body of models that utilize such policies effectively. Most importantly, a portion of this research is based on actual industry operations where base stock policies are used, therefore, we also mimic what is done in practice.

Another common element is the treatment of stock shortages. We assume full backordering in all three papers, with either a shortage penalty cost or a fill rate constraint. With this assumption of full backordering, it should also be noted that in each of our models, we are minimizing total costs rather than minimizing inventory levels. In light of this last assumption, one could consider this research as evidence for the useful application of base stock replenishment under Poisson demand when modeling a unique inventory structure with full backordering.

Lastly, our methods of solving the models presented here are also similar. We employ computer simulation to estimate costs and performance measures in articles 1 and 3. We also use similar forms of sample path analysis in papers 1 and 3, where paper 1 uses perturbation analysis and paper 3 uses marginal analysis. Our approach in paper 2 is different, in that we were able to approach the model analytically due to a decreased complexity. These solution approaches are applied both through computer simulation and analytical derivation.

Through a review of the literature, as found in each of the papers, we have found this research to be novel. In paper 1 our application of perturbation analysis is unique to the inventory rationing literature. In paper 2, the hospital inventory system we present is a new structure not found in the modeling literature before, nor is the analytical solution approach. Furthermore, paper 3 expands on the unique inventory system in paper 2 by allowing for emergency ordering. This last paper is unique to two separate streams of research and acts as a bridge between them. While the solution approach in paper 3 is based on previous research, the added assumption of inventory inaccuracy requires some novel modifications of the procedure which are shown to be successfully implemented.

The results of our efforts include minimizing solutions for each of the models. Though not always optimal, our solution approaches find optimal solutions most of the time and are very close to optimal otherwise. Furthermore, from our results we have gained new insight into these non-traditional supply chains that will allow for the advancement of research and the tactical improvement of the related industries.

1: INVENTORY RATIONING MODEL FOR MULTIPLE DEMAND CLASSES WITH BACKORDERING

In most industry situations, inventory is given to customers on a first-come, first-served basis. The main body of inventory-related literature thus focuses on this situation. However, there are cases where inventory is rationed to customers based on any number of reasons, including backorder costs, service level commitments, supply contract values, and so forth.

Here, we develop an inventory model that addresses inventory rationing based on customer priority derived by backorder costs. Similar to recent literature, we use the framework of a multi-echelon inventory system to describe the physics of a critical level policy. However, instead of considering fill rates as in the previous work, we seek to minimize a cost objective. Under a similar framework other research has considered 2 demand-classes, and we extend to multiple demand classes. We assume a continuous-review, base stock replenishment policy and allow for full backordering.

Due to the complexity of the model, we will be assuming Poisson demand and a base-stock replenishment system. In solving the inventory model, we utilize simulation to estimate total expected cost and apply variance reduction to reduce sampling error in computing cost. First differences are estimated using a form of Perturbation Analysis unique to inventory rationing literature, and we use the first differences to minimize costs using heuristics.

1.1 INTRODUCTION

Many of the inventory policies discussed in the literature and used in practice assume that all customers are equally important and thus, demand is met on a first-come, first-served basis. In practice, however, an inventory manager might have a set of customers who receive preferential treatment. Reasons to differentiate customers might include personal relationships, differing backorder costs, contractual agreements, differing dollar volumes, or other cost and revenue-related issues. When customers are differentiated for any reason, a more involved inventory policy may be needed to meet the different customer needs while minimizing inventory-related costs.

One method from the literature is to choose a “critical level” for on-hand inventory, also called a threshold level. Low priority customer demand is backordered when inventory levels fall to the critical level, while high priority customer demand is consistently filled as long as on-hand stock remains. Once inventory is replenished sufficiently to fill backorders and increase on-hand inventory beyond the critical level, orders from low priority customers are again filled as usual. This method of inventory control has been called a threshold inventory rationing policy (or critical level rationing policy) in the literature, and we adhere to the same terminology in this research as we model a critical level rationing policy.

In industry, there are many cases where inventory rationing might be usefully applied. Deshpande et al. (2003) cite the case where the same spare part for an equipment component is needed by different branches of the military but is stored in a single location. Suppose that for a particular component the Air Force requires an 85% fill rate, and the Navy requires a 95% fill rate. The Defense Logistics Agency (DLA) might decide to round-up the fill rate to hold enough components in inventory to meet the higher fill rate of 95%. They might also physically separate the inventory with separate replenishment policies for each “pile” of parts. One pile would be large enough to meet the Air Force fill rate, while the other pile would be sized to meet the Navy fill rate. In either case, the inventory held may be more than necessary to meet both fill rates while minimizing costs.

Deshpande et al. (2003) show that with a critical level policy, DLA might be able to reduce inventory-related costs while still meeting the different service expectations. They assign priorities to each branch of the military (i.e. classify Navy demand as high priority, and Air Force demand as low-priority). Then, by setting a critical inventory level at which to backorder low priority demand, the procedures would be in place to ensure Navy's higher fill rate would be met while inventory costs are reduced.

Another example comes from industry, where customers are prioritized based on differing shortage costs. Some suppliers to auto manufacturers incur steep fines for delayed shipments to certain customers, and other manufacturers incur similar backorder costs with different contractual customers. Similarly, when a firm seeks to segment customers based on a price for higher service, critical rationing of inventory is again applicable. Or, rationing my apply to a firm that uses the same inventory stockpile to fill demand on two fronts, for instance, Internet sales in addition to retail sales.

The premise of this paper is to develop a multiple demand-class inventory model, where customers are prioritized base on backorder costs and inventory is rationed to the different customer classes using a critical level rationing policy.

This paper is organized with section 1.2 covering the related literature and defining some of the basic elements of demand class inventory models. Section 1.3 introduces the demand class model, and we also re-frame the model as a serial stage inventory system. Section 1.4 defines the model notation and equations needed to compute performance measures, which are then used to define a novel objective cost function. In Section 1.5 we discuss a simulation approach that relies on the steady state distributions of the performance measures. We also present our unique application of perturbation analysis for an inventory model with integer-valued decision variables in Section 1.6. In Section 1.7 we provide a preliminary analysis of currently partial results, as the computational portion of this research is ongoing. We conclude the paper in Section 1.8 by presenting some of the insights and future research related to the work presented here. Throughout the body of this paper, we refer to the Appendix found at the end of this document.

1.2 LITERATURE

In the literature there have been several contributions on multiple demand class inventory models, and we cite some of the research more similar to our own. Below, we categorize the related research and discusses the contribution of our work.

Critical Level Rationing Policy - As explained in the introduction, we assume a critical level rationing policy, which defines the way we ration inventory for the different classes of customers. Most of the literature assumes a static critical level policy, while some consider dynamic rationing policies where critical levels may change over time. Here we are developing a static critical level rationing policy.

Number of Demand Classes - This refers to how many groups, or classes, of customers are defined. These customer classes are prioritized so that demand from each class can be met immediately or backorded, depending on the assigned priority. Models from the literature define either 2 customer classes or the model is generalized with N customer classes. Here, we develop a model of N demand classes and analyze a set of test cases for a three, four, and five-demand class formulation.

Treatment of Shortages - Shortages are treated as either backorders or as lost sales. The penalty cost associated with the shortage is often how customer classes are prioritized. In this paper we assume shortages are backordered, and we assign priorities based on backorder costs.

Formulation Objective - Though not always optimized, models typically are formulated around an objective to minimize inventory-related costs, or to minimize inventory levels. When the objective is to minimize inventory levels, the model includes constraints that enforce the desired fill rates for each class of customers. In a few cases, cost is minimized with fill rate constraints, and in other cases the model may be formulated only to analyze expected fill rates. In this paper, we are attempting to minimize the inventory-related costs, and we do not specify any constraints.

Replenishment Policy - This refers to the way inventory is replenished from the supplier. The replenishment policies considered in multiple demand class literature are the same as seen in other inventory literature. Common policies include the reorder point, order quantity policy (r, Q), the one-for-one base-stock policy $(S-1, S)$, or the reorder point, order-up-to level policy (s, S). While the replenishment policy parameters are often included as decision variables in the model, sometimes the replenishment policy is only discussed and the parameters predetermined. In this paper we analyze a base stock replenishment policy with order-up-to level S, and we include both the order-up-to level, S, and the rationing policy parameters as decision variables in the model.

Model Tools - Initially, multiple demand class models used dynamic programming to formulate and solve the model, such as Veinott (1965), Topkis (1968). Kleign and Dekker (1998), Moon and Kang (1998) use simulation to estimate the performance measures and the effects of the different model parameters. Continuous-time models from such papers as Dekker et al. (2002), Deshpande et al. (2003) and Arslan et al. (2007) require an analytical approach, finding the distributions of the performance measures based on some or all of the unknown parameters. After the relationships are defined using an analytical approach, algorithms are developed in an attempt to optimize the system. Some later papers such as Vicil and Jackson (2006) and Teunter and Haneveld (2008) use Markov Chains to describe the inventory system, and then they seek to optimize the system based on steady-state probabilities.

The main motivation for the structure of our model is from Arslan et al. (2007) who consider a single product, continuous review, (r, Q) policy, for N demand classes. It has nearly all the assumptions of our model, but does not model a cost objective as we do. Rather, it seeks to minimize the expected inventory level subject to fill rate constraints. We use their method of structuring the inventory model as a serial stage inventory system that represents the multiple demand class system, and then based on the serial-stage model structure, we similarly derive the performance measures. Arslan et al. (2007) use a single pass algorithm to come to a near-optimal solution, which is different from the perturbation analysis approach that we employ.

More specifically, we use the steady state distributions of the performance measures to generate random instances of the total cost using a computer simulation. Then, we employ a unique approach to perturbation analysis where we estimate first differences over each decision variable based on the sample path. Finally, we present a heuristic to minimize total expected cost, and then report on the preliminary analysis of its efficiency and accuracy.

Aside from Arslan et al. (2007), this is only the second example in the literature to use a serial-stage inventory system as the framework to model a multiple demand class inventory system. Thus, this research represents the first cost-objective model under the serial-stage framework. To our knowledge, this is also the first example of using a modified form of perturbation analysis to estimate first differences in an inventory model setting. We discuss this element of the research further in Section 1.6.

Table 1.1: Related Literature Summary

Literature	*Rationing Policy*	*No. of Classes*	*Shortages*	*Model Objective*	*Ordering Policy*	*Model Tools*	*Special*
Veinott (1965)	*None	N	Partial Backordering	Min cost	Periodic Base-stock	Dynamic Programming	Introduced critical levels
Topkis (1968)	Static / Dynamic	N	Partial Backordering	Min cost	Periodic Base-stock	Dynamic Programming	Argued / proved critical levels opt.
Kaplan (1969)	Dynamic	2	Backorders	Min cost	None	Dynamic Programming	First numerical results
Nahmias and Demmy (1981)	Static	2	Backorders	Fill rate analysis	(r,Q)	Analytical / Algorithmic	First continuous-time model
Kleign and Dekker (1998)	Static	N	Backorders	Fill rate analysis	(S-1,S)	Simulation	Numerically verified fill rates
Moon and Kang (1998)	Static	N	Backorders	Min cost	(r,Q)	Analytical / Simulation	Deterministic Demand
Dekker et al. (2002)	Static	N	Lost Sales	Min cost w/fill rate constraints	(S-1,S)	Analytical / Algorithmic	First to use fill rates as constraint
Deshpande et al. (2003)	Static	2	Backorders	Min cost	(r,Q)	Analytical / Algorithmic	Studied different clearing policies
Vicil and Jackson (2006)	Static	N	Backorders	Min inv. w/fill rate const.	(S-1,S)	Markov Chain	Proved optimal solution
Arslan et al. (2007)	Static	N	Backorders	Min inv. w/fill rate const.	(r,Q)	Analytical / Algorithmic	Used Serial Stage framework
Teunter and Haneveld (2008)	Dynamic	2	Lost Sales	Min cost	None	Analytical	First to consider rationing times
Neve and Schmidt	Static	N	Backorders	Min cost	Base-stock (S-1, S)	Analytical / Algorithmic	Use IPA to estimate differences

1.3 GENERAL FRAMEWORK

We develop the framework of our model in the same manner as Arslan et al. (2007), describing first a general multiple demand class system (DCS) and then equating it to a serial-stage inventory system (SSS). In our demand class system, we focus on a finished goods warehouse that stocks a single product in anticipation of future demand.

1.3.1 DCS Framework

The warehouse is assumed to meet the demands of its customers based on customer priority, so we need a way to segregate customers into sets, or classes. Since we allow full backordering, and since it is generally assumed that backorder costs outweigh holding costs, we will use differing backorder costs as the defining parameter. We divide the customer base into N demand classes based on backorder costs, where b_i is the backorder cost for customers in demand class i, and $b_1 > b_2 > \cdots > b_{N-1} > b_N \geq 0$. As implied by the descending order of the backorder costs, we prioritize the demand classes with class 1 having the highest priority, and class N having the lowest priority. Simply put - the greater the customer backorder cost, the higher priority the customer is assigned.

We assume, similar to other related literature, that customer demand follows a Poisson process where one item is ordered at a time and demand from different customer classes are independent. We define $\lambda_i > 0$ to be the Poisson demand rate for demand class i, and the aggregate, or total, demand rate is the sum of the individual class demand rates: $\lambda = \sum_{i=1}^{N} \lambda_i$.

As demand occurs, the warehouse decides whether to fill an order, or to backorder it, based on current on-hand inventory and based on which class the demand comes from. We define a critical level rationing policy by setting rationing levels c_i for classes $i = 1, 2, \cdots, N-1$. We decide to fill a demand from class i only if current on-hand inventory is above the critical level of the higher priority class, c_{i-1}. If we decide to immediately fill an order, then the total on-hand inventory is decreased. We backorder class i demand once on-hand inventory reaches or falls below the critical level, c_{i-1}. We can say, generally, that c_i units of stock are reserved for customers from classes 1 through i.

An example of a three-class DCS is shown in Figure 1.1, where the warehouse is represented by its inventory "meter". The critical rationing levels, marked c_1 and c_2 on the meter, are arbitrarily set at 5 and 11 in the example, respectively - and because we are assuming a static critical level policy, these values do not change once they are chosen. On-hand inventory at the warehouse is represented by the height of the shaded region, with on-hand inventory level at current time t given by $I(t)$. Since $I(t) \leq c_2$, any orders from demand class 3 would be backordered. Similarly, demand at time t from class 2 will be filled since $I(t) > c_1$, and demand at time t from class 1 will be filled since $I(t) > 0$. The critical rationing levels work in the same manner for the general case.

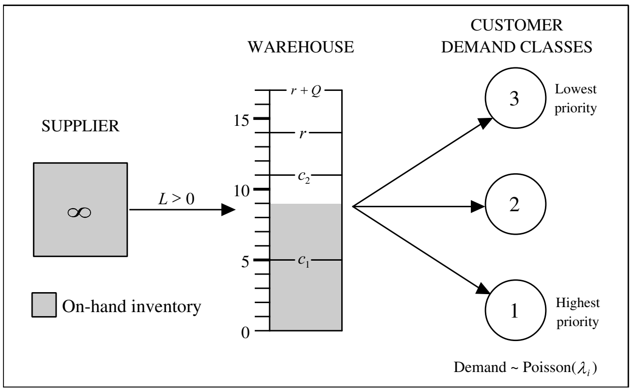


Figure 1.1: DCS Inventory System

As stated previously, we are using a static critical level rationing policy, so whenever on-hand inventory is equal to c_{i-1} or below, any demands from class i will be backordered. The total unfilled class i backorders at time t is given by $B_{i,i}(t)$, which we call the external backorders. The backorder cost, b_i, for these class i external backorders is incurred over time with units given in dollars per units-backordered per unit-time. The double i, i subscript will help later to distinguish *external* backorders from *internal* backorders.

We assume the outside supplier has sufficient capacity to meet all demand from the warehouse, which is why, in Figure 1.1, the supplier is shown with infinite on-hand inventory. Also, there is a deterministic lead-time, $L > 0$, after placing an order for replenishment with the supplier. Initially we model a continuous review, one-for-one, base-stock replenishment policy $(S-1, S)$. When operating under no, or a sufficiently small, fixed ordering cost, we assume base-stock replenishment. However, under a large positive fixed order cost system, an (r, Q) or (r, S) policy would be preferred. Figure 1.1 is demonstrating the (r, Q) case, where $r+Q$ is the maximum inventory.

We base our continuous review replenishment policy on the inventory position, $IP(t)$, which includes on-hand and on-order inventory. Under an (r, Q) policy we would order Q units when the inventory position reaches the reorder point, r. Under the one-for-one base-stock policy, we place an order with our supplier each time we see a customer demand.

When replenishment arrives from the supplier we need a way to intelligently choose whether to clear outstanding backorders or to refill inventory (or a little of both) - especially when the incoming replenishment is not sufficient to clear all backorders and refill all the 'gaps' in inventory. We assume that we allocate the replenishment between unfilled backorders *and* inventory gaps on a first-come, first-served (FCFS) basis as was done by Arslan et al. (2007). This approach is basically the threshold clearing mechanism as defined by Deshpande et al. (2003), which was shown to perform better than using replenishment to first clear backorders based on FCFS, and then using whatever remained for inventory replenishment. The method assumed here also allows us to derive the performance measures more directly.

We define class i shortfall to be the total inventory gap for class i, plus the total external backorders for classes $1, \cdots, i$. We denote the shortfalls at time t, $SF_i(t)$, for class $i = 1, \cdots, N-1$ with the following relation:

$$SF_i(t) = (c_i - I(t))^+ + \sum_{j=1}^{i} B_{j,j}(t) \tag{1.1}$$

The first term could be called the inventory gap, or the number of units needed to bring on-hand inventory back up to the critical level, c_i, for class i. The second term is the total external backorders for class i and all higher priority classes at time t. For class N, the shortfall only includes the external backorders as no stock is reserved for the lowest-priority demand class, as given below:

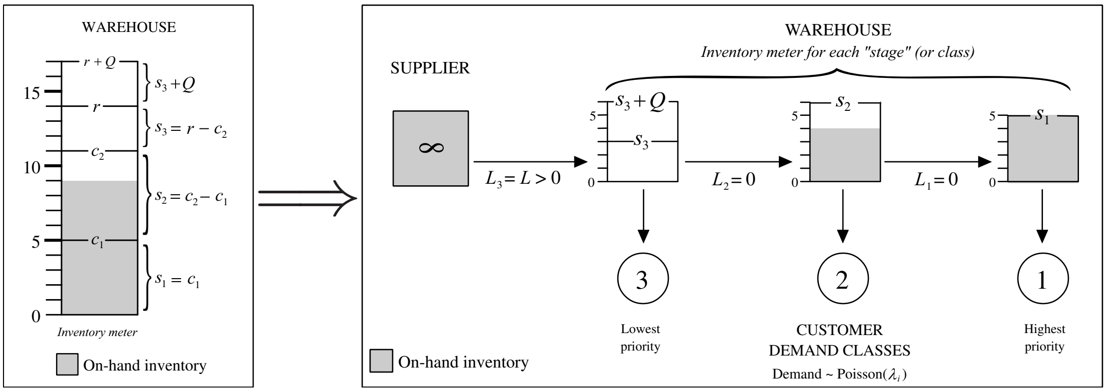
$$SF_N(t) = \sum_{j=1}^{N} B_{j,j}(t) \tag{1.2}$$

Notice that, for each class i, shortfalls increase either when a backorder occurs or when on-hand inventory drops further below the critical level. So, to further explain the backorder clearing mechanism, whenever a demand causes shortfall to increase for any of the classes, the occurrence is added to an ordered list of demand-generated shortfalls. We assume that a separate ordered list of shortfall events is created for each class of customers. Arslan et al. (2007) present an algorithm, iterating over the demand classes, that basically allocates the incoming replenishment to the shortfall occurrences as recorded in the ordered list. This process helps to balance between clearing backorders and refilling inventory, and it ensures that a particular class cannot have both customer backorders and sufficient on-hand inventory simultaneously. For an in-depth explanation of the backorder clearing algorithm, see Arslan et al. (2007).

We now follow the lead of Arslan et al. (2007) and use the framework of a serial-stage system to model the demand class system. Taking this approach helps to describe the relationships between the different demand classes, and to develop a more direct objective cost function.

1.3.2 SSS Framework

To restructure the DCS model as a SSS model we need to describe the transformation of both the rationing and the replenishment policy parameters, as well as the derivation of performance measures. Customers are assigned priorities and grouped into customer classes in the same way described for the DCS, and all the associated costs remain the same. However, instead of thinking of the warehouse having a single stockpile of inventory, we imagine that each customer demand class is assigned an inventory stage with its own inventory. The customer demand distributions remain unchanged for each class, but now, where each class is represented by an inventory stage, there is also the consideration of "downstream" demand which we call *internal* demand.



(a) Reserve stock levels

(b) Example with 3 classes

Figure 1.2: Transforming a DCS to a SSS

We first transform the critical levels, c_i, and call the transformed variables *reserve stock levels*, denoted s_i as shown in the SSS model in Figure 1.2b. Looking at Figures 1.2a and 1.2b, we see the same DCS example from Figure 1.1 re-modeled as a SSS with reserve stock levels instead of critical levels. In Figure 1.2a it shows that the reserve stock level for stage 1 is $s_1 = c_1$, and we have the general reserve stock substitution for stages $i = 2, \cdots, N-1$ given by $s_i = c_i - c_{i-1}$.

This implies the following relationship between reserve stock levels and critical levels:

$$c_i = \sum_{j=1}^{i} s_i \text{ ; for } i = 1, \cdots, N-1 \tag{1.3}$$

At stage N there is no critical level, so we modify the warehouse replenishment policy by defining the reorder point for *stage* N, given by $s_N = r - c_{N-1}$, though we assume the same ordering quantity, Q. However, in this paper we are assuming a base stock replenishment policy from the external supplier with base stock level S, which is a special case of the (r, Q) policy when $Q = 1$ and $r = S - 1$. Therefore, we can say that the base stock level in the SSS case is $s_N + 1$, which gives a reorder point of s_N and an order quantity of $Q = 1$.

The structure of the SSS model implies that each stage i is replenished by stage $i+1$ while meeting class i external customer demand *and* stage $i-1$ internal demand. The SSS model assumes that the reserve stock levels, s_i at *internal* stages $i = 1, \cdots, N-1$, act as the base stock levels for each of the stage-specific replenishment policies. Then, we assume continuous review, one-for-one base-stock replenishment policies at each internal stage within the system. To accurately mimic the performance under the DCS framework, we set the lead time between the internal stages to $L_i = 0, i = 1, \cdots, N-1$. Lead time to stage N is given by $L_N = L > 0$, since external replenishment from the supplier is unaffected by the transformation to a SSS.

Mathematically, we treat the system as if it were a multi-echelon, or serial-stage, inventory system - the main difference being that external demand is allowed at all stages of the system. We call the inventory at stage i the local inventory, given by $I_i(t)$. Under the DCS we only considered $I(t)$, which is the total on-hand inventory, and here it has the same meaning and is the sum of the local on-hand inventories for each class. Note in the SSS example in Figure 1.2b that $I_3(t) = 0; I_2(t) = 4$; and $I_1(t) = 5$. The total on hand inventory for the SSS would be the same as for the DCS in Figure 1.1, that is $I(t) = \sum_{i=1}^{N} I_i(t) = 9$.

Now that we have introduced some of the elements of the SSS representation of the DCS, we can discuss how rationing occurs. In the DCS, we relied on the level of total on-hand inventory coupled with the critical level, c_i. Here, the rationing policy is handled by the structure of the SSS, utilizing the information about local stage i inventory, $I_i(t)$, and reserve stock level, s_i. Basically, if there is available stock at stage i, then class i customers will be served; if there is no stock available at stage i, then class i customer demand will be backordered. The amount reserved for each customer demand class i is based on the reserve stock level, s_i.

When class i demand cannot be met, due to lack of on-hand inventory at stage i, the backorder is called an *external* backorder. For any stage i, external backorders, $B_{i,i}(t)$, have already been defined in our discussion of the DCS model, and they are no different in the SSS case. They are the total, unsatisfied class i customer demand at time t, and we incur a cost b_i for each unit backordered.

When internal demand cannot be met - that is, the demand from a downstream stage, it creates what is called an *internal* backorder. Internal backorders were also introduced with the DCS model as in Equation 1.1, but they were called class i shortfall. In the SSS case, we define $B_{i,i-1}(t)$ to be the internal backorders at stage i, where the subscript is used as follows:

$B_{i,i-1}(t)$ is the number of replenishment requests from stage $i-1$ that have yet to be filled by stage i by time t. Thus, the number of internal backorders at stage i is equivalent to the class $i-1$ shortfall, giving the following relationship:

$$B_{i,i-1}(t) = S_{i-1}(t) = (c_{i-1} - I(t))^{+} + \sum_{j=1}^{i-1} B_{j,j}(t) \tag{1.4}$$

Arslan et al. (2007) showed how the SSS formulation approach operates identically to the DCS formulation. We assume the results from the previous work are valid and move forward to the development of the SSS-based mathematical model to minimize an objective cost function.

1.4 MODEL FORMULATION

Now that we have framed the DCS as a SSS, we again refer to Arslan et al. (2007) in deriving the needed performance measures. First, we present the objective cost function which is new to the literature, second, we present the equations and notation for the performance measures, and finally, we present the steady-state distributions of each.

1.4.1 Cost-Based Objective Function

We begin by introducing a cost objective that is based on a multiple demand class inventory model with backorders, stochastic demand, and a continuous review base-stock inventory policy. In its present form, the cost function given in Equation 1.5 looks little different than familiar objective functions in the literature, such as in Zipkin (2000). The main difference is the addition of external backorders, $B_{i,i}$, at each stage and each with a unique cost, b_i, rather than a single external backorder term at the final stage of the SSS.

Our total system cost for N demand classes under a SSS framework is given by:

$$C(s_1, s_2, \cdots, s_N) = \sum_{i=1}^{N} (h \cdot I_i + b_i \cdot B_{i,i}) \tag{1.5}$$

In the above equation, h is the inventory holding cost in dollars per unit - per unit time, and b_i is the cost in dollars per unit backordered at stage i, per unit time. I_i and $B_{i,i}$ represent the steady-state forms of stage i on-hand inventory and external backorders, respectively. As such, the objective function is a function of random variables, I_i and $B_{i,i}$, which means that any specific calculation of the objective function will be a single observation from the distribution of total cost. We therefore require that the objective be an expected value, rather than a single observation, and we also use notation referencing the decision variables. The total *expected* cost is given below, and will serve as the objective function of our model.

$$E\left[C(s_1, s_2, \cdots, s_N)\right] = \bar{C}(s_1, s_2, ..., s_N) = E\left[\sum_{i=1}^{N} (h \cdot I_i + b_i \cdot B_{i,i})\right] \tag{1.6}$$

Because the expected cost is a function of the reserve stock levels (which is demonstrated below), we will minimize the expected total cost by choosing the best reserve stock level s_i for each stage i. To derive the necessary steady state distributions and show their dependence on the reserve stock levels, we need to define additional notation.

1.4.2 SSS Performance Measures

The SSS structure we use for our model is the same as used in Arslan et al. (2007), so we develop the performance measures using the same approach. It should be noted that this approach for developing the inventory and backorder performance measures for a multi-echelon inventory model was developed before by Graves (1985).

For $i = 1, 2, \cdots, N$:

$D_i(s,t) =$ total external demand at stage i during the time interval $(s,t]$ - we assume $t > s$. Because we assume that demand follows a Poisson distribution for each demand class, the distribution of the random variable $D_i(s,t)$ is Poisson with mean $(t-s)\lambda_i$.

$B_i(t) =$ total number of unfilled internal and external backorders at stage i at time t, given by

$$B_i(t) = B_{i,i-1}(t) + B_{i,i}(t) \tag{1.7}$$

$IN_i(t) =$ net inventory level for stage i at time t

$IP_i(t) =$ inventory position for stage i at time t, which is the net inventory plus on-order inventory.

Net inventory is a key performance measure, as it is used to derive the on-hand inventory and backorders. Net inventory for stage i after a lead time, $L_i > 0$, is given by:

$$IN_i(t+L_i) = IP_i(t) - B_{i+1,i}(t) - D_i(t,t+L_i) - \sum_{j=1}^{i-1} D_j(t,t+L_i) \tag{1.8}$$

The first term, which is the inventory position at time t, represents on-hand inventory at time t plus the inventory that is on-order at time t - all of which becomes available on-hand inventory by time $t + L_i$, at the latest. The second term is the internal backorders unfilled by stage $i+1$ at time t (meaning, the amount ordered from stage $i+1$ by stage i that has yet to be filled at time t), which by definition cannot be filled during the replenishment lead time. The third term simply represents the stage i external demand during the lead time, directly from class i customers. The last term represents the internal demand from all downstream (higher priority) stages, as we are assuming a one-for-one, base-stock replenishment within the system. To shorten the expression we simply combine the third and fourth terms into one summation:

$$IN_i(t + L_i) = IP_i(t) - B_{i+1,i}(t) - \sum_{j=1}^{i} D_j(t, t + L_i) \tag{1.9}$$

We use $IN_i(t)$ to describe the on-hand inventory and total backorders at stage i at some time t. The following relationships will be used in our calculations:

$$I_i(t) = [IN_i(t)]^+ \tag{1.10}$$

$$B_i(t) = [-IN_i(t)]^+ \tag{1.11}$$

For a given stage $i, i = 1, \cdots, N$, Equation 1.10 shows how the positive part of net inventory is the on-hand inventory at time t, and Equation 1.11 shows how the negative part of net inventory is the total backorders at time t. You will note that we have not yet defined external backorders, $B_{i,i}(t)$, in relation to net inventory. This requires a sequential derivation process that will be described later on. This is important, since backorder cost is applied to external backorders only, and not to total backorders as giving in Equation 1.11.

1.4.3 Steady-State Equations

Now, we want to rewrite our equations in their steady-state form, in much the same manner as Arslan et al. (2007). Doing so allows us to compute the total expected cost, which the previous research did not attempt. At stage N, Equation 1.9 is reduced to the steady state form:

$$IN_N = IP_N - D^L \tag{1.12}$$

Equation 1.12 gives the steady-state form of net inventory at stage N. The backorder term from Equation 1.9 was dropped due to our assumption that the external supplier always has sufficient capacity to meet our replenishment requests. The notation for the steady-state lead time demand has been modified using the following definition:

$$D^L = \sum_{i=1}^{N} D_i^L \tag{1.13}$$

Where D_i^L represents the random variable of total external class i demand during lead time $L_N = L > 0$. Since class i lead time demand is a Poisson process with rate $\lambda_i L$, we can say that the distribution of total lead time demand at stage N is Poisson with mean $(\lambda_1+\lambda_2+\cdots+\lambda_N)L$. For notational purposes, we define $\lambda = \sum_{i=1}^{N} \lambda_i$. We can then say that:

$$D^L \sim \text{ Poisson}(\lambda L) \tag{1.14}$$

The other term in Equation 1.12 is the steady-state inventory position at stage N, which can be shown to vary uniformly between $s_N + 1$ and $s_N + Q$. Note that stage N has the same assumptions used in an (r, Q) single-stage model from Zipkin (2000), which verifies this relationship. We assume, for this paper, a one-for-one base-stock policy with the external supplier, where $Q = 1$, so the inventory position in steady state stays constant at the base-stock level, $s_N + 1$, giving steady-state net inventory at stage N, where s_N has no restrictions on sign:

$$IN_N = s_N + 1 - D^L \tag{1.15}$$

For stages $i = 1, \cdots, N-1$ we can write the steady state form for net inventory thus:

$$IN_i = s_i - B_{i+1,i}; i = 1, 2, \cdots, N-1 \tag{1.16}$$

Equation 1.16 is derived from Equation 1.9, where the demand term is dropped since lead time for each stage $i = 1, \cdots, N-1$ is $L_i = 0$. Under one-for-one base-stock replenishment between stages, the inventory position, IP_i, remains constant over time and equals the reserve stock level, s_i. The random variable in Equation 1.16 is the stage $i+1$ internal backorders, $B_{i+1,i}$, and its distribution will be discussed shortly.

With these steady-state equations defined, we can now find the distribution of net inventory for any stage i, given the associated reserve stock level s_i and given the distribution of the applicable random variable. Later, we will treat each s_i as decision variables and define a policy, $\pi = (s_1, s_2, \cdots, s_N)$, which will affect the probability distributions of the SSS performance measures and the total expected inventory and backorder costs.

Now that we have the steady-state form of net inventory, we can write the steady-state forms of on-hand inventory and backorders for stages i, $i = 1, 2, \cdots, N$:

$$I_i = [IN_i]^+ \tag{1.17}$$

$$B_i = [-IN_i]^+ \tag{1.18}$$

$$B_i = B_{i,i-1} + B_{i,i} \tag{1.19}$$

Equations 1.17, 1.18, and 1.19 are simply the steady-state form of Equations 1.10, 1.11, and 1.7, respectively. However, we define the internal backorders at stage 1 to be zero, as there are no downstream stages below stage 1, so $B_{1,0} = 0$. Thus, total stage 1 backorders is simply equal to stage 1 external backorders, that is:

$$B_1 = B_{1,1} \tag{1.20}$$

Once we have the steady-state distribution for total stage i backorders, B_i, we must condition on values of B_i to obtain the distribution of either external back orders, $B_{i,i}$, or internal backorders, $B_{i,i-1}$. Again using an approach from Arslan et al. (2007), we use a binomial distribution to describe the conditional distributions of both internal and external backorders, where we condition on the value of total backorders.

$$\Pr\left(B_{i,i-1} = x \mid B_i = n\right) = \begin{pmatrix} n \\ x \end{pmatrix} p_i^x (1 - p_i^x)^{n-x} \tag{1.21}$$

$$p_i = \frac{\sum_{j=1}^{i-1} \lambda_j}{\sum_{j=1}^{i} \lambda_j} \tag{1.22}$$

$$\Pr\left(B_{i,i} = x \mid B_i = n\right) = \begin{pmatrix} n \\ x \end{pmatrix} q_i^x (1 - q_i^x)^{n-x} \tag{1.23}$$

$$q_i = \frac{\lambda_i}{\sum_{j=1}^{i} \lambda_j} \tag{1.24}$$

Equation 1.21 gives the conditional distribution for internal stage i backorders, and Equation 1.23 gives the conditional distribution for stage i external backorders. The probability of an internal backorder at stage i (once stage i inventory falls to zero) is given by Equation 1.22. Similarly, Equation 1.24 is the probability of an external backorder at stage i when the current on-hand inventory is zero.

Like Arslan et al. (2007), we argue the conditional distribution is binomial because stage i internal backorders are generated randomly at the Poisson rate of internal demand (the sum of all higher-priority class demand rates), and external backorders are generated randomly according to the stage i Poisson demand rate. So, if we know the total backorders, we can use the binomial distribution, as each arriving backorder is either internal or external, with probability p_i and q_i, respectively.

Now that we have defined the relationships necessary to compute the performance measures, we use a sequential process to derive the steady-state performance measure distributions for all stages of the SSS. As previously stated, we assume continuous review, one-for-one base-stock replenishment through the external supplier. The following steps of the derivation process are similar to those in Arslan et al. (2007), though we focus on the base-stock replenishment model which is a special case.

First: Use Equation 1.15 to derive the steady-state distribution for IN_N given: s_N, the Poisson demand rates λ_i from all stages $i = 1, 2, \cdots, N$, and the deterministic lead time to the supplier $L > 0$.

Second: Use the stage N net inventory distribution to develop the distribution for total stage N backorders, B_N, using Equation 1.18.

Third: Using Equation 1.21, condition on B_N to derive the steady state distribution of stage N internal backorders.

Fourth: Use Equation 1.16 and the given value for s_{N-1} to derive the steady-state distribution of net inventory for stage $N - 1$.

After the fourth step, the sequential derivation process starts over again at the second step, but at the next lower stage. Continue in this manner, iterating down through all the stages until we have found the steady-state distributions for the performance measures for every stage of the SSS. At stage 1, however, we derive the steady state distribution of external backorders by simply using the negative part of net inventory, as required in Equation 1.20.

1.5 COST OBJECTIVE

Now that we have a way to derive the steady state distributions for the performance measures, we present the resultant objective function for N demand classes in expanded form, and introduce two methods to estimate its value. The first is a basic simulation approach, followed by a similar approach that utilizes variance reduction.

1.5.1 Expanding the Objective Function

From equation 1.6 we can expand to the general equation as follows:

$$\begin{aligned}\bar{C}(s_1, s_2, ..., s_N) &= h \cdot s_1 - h \cdot p_2 \cdot G^1_{B_{3,2}}(s_2) + (h + b_1)G^1_{B_{2,1}}(s_1) \\ &+ \sum_{j=2}^{N-2} \Big(h \cdot s_j - h \cdot p_{j+1} \cdot G^1_{B_{j+2,j+1}}(s_{j+1}) + (h + b_j - b_j p_j)G^1_{B_{j+1,j}}(s_j) \Big) \\ &+ h \cdot s_{N-1} - h \cdot p_N \cdot G^1_{D^L}(s_N + 1) + (h + b_{N-1} - b_{N-1}p_{N-1})G^1_{B_{N,N-1}}(s_{N-1}) \\ &+ h \cdot (s_N + 1) - h \cdot E\left[D^L\right] + (h + b_N - b_N p_N)G^1_{D^L}(s_N + 1) \qquad (1.25)\end{aligned}$$

The derivation of this expanded form of the cost objective for N demand classes can be found in the appendix. Note that Equation 1.25 contains several first-order loss functions, denoted $G^1_X(x)$. These first-order loss functions are based either on the distribution of lead time demand ($X \sim D^L$), or one of the distributions for internal backorders $X \sim B_{j+1,j}$. Also, these first-order loss functions are defined for all real x. The result is a cost function that is a non-separable, non-linear function with integer-valued decision variables and discrete randomness.

And while the cost objective appears to be separable, note that many of the first order loss functions are based on distributions of internal backorders. Recall the approach used to derive these distributions. Due to the repeated need to condition on the value of total backorders at each stage, these first order loss functions are actually dependent on all decision variables from all higher stages. As a result, these loss functions are quite complex and are impossible to separate into forms that are only dependent on single decision variables. In addition to being non-separable, we were unable to derive a closed form expression of the objective function for more than two demand classes. Therefore, we require a method to estimate or approximate the total expected cost.

Also, because we are only interested in integer-valued arguments, we need a way to repeatedly compute the differences over each of the decision variables, $s_1, s_2, \ldots, s_N$, to apply heuristics and minimize the objective value. After several attempts, we were unable to derive the exact differences necessary for minimization for more than 2 demand classes. Other researchers, such as Moon and Kang (1998), faced similar issues when attempting to obtain analytical solutions for cost-objective inventory rationing models for more than two classes. Thus, we will use simulation and a unique form of Perturbation Analysis (PA) to estimate both the total expected cost and the differences necessary for optimization.

For the remainder of this section, we discuss the use of simulation to approximate the total expected cost for a given set of decision variables. Due to apparent sampling error in the estimation, we attempt to reduce the variation using a stratified sampling approach.

In the next section, we will apply PA along the sample paths generated during each simulation run. Taking this approach allows us to estimate first differences over each decision variable. It also provides a means to apply optimization heuristics that will be discussed later.

1.5.2 Basic Simulation Approach

Due to the complex nature of the analytically derived function for total *expected* cost, as given in Equation 1.25, we use a simulation approach to generate random terms from the steady state distribution of *total* cost, as given in Equation 1.26 below.

$$C(s_1, s_2, \cdots, s_N) = h(I_1 + I_2 + \cdots + I_N) + b_1 B_{1,1} + b_2 B_{2,2} + \cdots + b_N B_{N,N} \tag{1.26}$$

Each of the terms on the right hand side of Equation 1.26 can be generated sequentially using their steady state distributions. As such, the total cost function can be viewed as the steady state distribution of total cost. Based on the structure of the model, we develop a simulation that can generate one instance of the total cost $C_k(s_1, s_2, \cdots, s_N)$ from its steady state distribution, where C_k is the cost generated by the k^{th} simulation replication.

The steps of the simulation are based on the serial-stage structure and on the demand assumptions previously defined for the model. After n simulation replications, we average the total cost values, C_k, $k = 1, \ldots, n$, to give the estimated total expected cost, $\bar{C}(n)$, as shown in the following equation:

$$\bar{C}(n) = \sum_{k=1}^{n} \frac{C_k(s_1, s_2, \ldots, s_N)}{n} \tag{1.27}$$

The simulation requires the following parameters and decision variables before running:

h : inventory holding cost per unit, per unit time

b_i : backorder cost per unit backordered externally at stage i, per unit time

L : lead time from the external supplier

λ_i : class-specific Poisson demand rate per unit time for stage i

s_i : reserve stock level (decision variable)

N : the number of demand classes in the simulation

An alternate form for Equation 1.26 is given when we replace on hand inventory, I_i, with the positive part of net inventory, IN_i^+. The modified form of the equation is given below, and will reduce the number of variables needed during simulation.

$$C(s_1, s_2, \cdots, s_N) = h(IN_1{}^+ + IN_2{}^+ + \cdots + IN_N{}^+) + b_1 B_{1,1} + b_2 B_{2,2} + \cdots + b_N B_{N,N} \tag{1.28}$$

Therefore, during each simulation run the model will require the following values to compute cost:

$IN_i{}^+$: the positive part of net inventory at every stage i

$B_{i,i}$: the external backorders at every stage i

To compute the above values, the simulation generates values from several different probability distributions. Lead time demand is generated from a Poisson distribution (see Equation 1.14), and the N external backorder distributions are generated using N different Binomial distributions (see Equation 1.23). Applying Equations 1.15, 1.16, 1.18, 1.19, and 1.20, we can compute an instance of total cost.

Essentially, one run of the simulation is generating a random value for total cost. As suggested previously, we will average the results of several simulation runs to generate total *expected* cost, which is the objective we seek to minimize. The basic simulation algorithm follows (note the similarities between the algorithm and the method of deriving the steady-state performance measures discussed previously).

Basic Simulation Algorithm:

Step 0 Generate a value for total demand, D^L, from the Poisson(λL) distribution.
Compute $IN_N = s_N + 1 - D^L$ and set the total cost $C = h \cdot IN_N^+$.
Set stage $i = N$.

Step 1 Compute $B_i = [-IN_i]^+$.

Step 2 Generate a value for internal demand, $B_{i,i-1}$, from a Binomial(p_i, B_i) distribution

Step 3 Compute $B_{i,i} = B_i - B_{i,i-1}$ and add the value of $b_i \cdot B_{i,i}$ to the total cost.
Compute $IN_{i-1} = s_{i-1} - B_{i,i-1}$ and add the value of $h \cdot IN_{i-1}^+$ to the total cost.

Step 4 Set $i = i - 1$. If $i = 1$ go to Step 5, otherwise go to Step 1.

Step 5 Add the value of $b_1 \cdot IN_1^-$ to the total cost.

Step 6 Return the total cost, $C(s_1, s_2, \cdots, s_N)$.

The above simulation steps generate one replication from the steady-state distribution of total cost for a particular set of reserve stock levels, $s_1, \cdots, s_N$.

To estimate the expected total cost for a given set of decision variables requires that we replicate the simulated cost n_0 times and then compute the average cost over the n_0 replications. Initially, we run the simulation for an arbitrary number of replications, say for $n_0 = 30$ replications. From this initial run, we compute the sample average, $\bar{C}(n_0)$, and the sample variance, $S^2(n_0)$. The equations are given below:

$$\bar{C}(n_0) = \sum_{k=1}^{n_0} \frac{C_k}{n_0} \tag{1.29}$$

$$S^2(n_0) = \sum_{k=1}^{n_0} \frac{(C_k - \bar{C}(n_0))^2}{n_0 - 1} \tag{1.30}$$

By analyzing the replications of total cost, C_k, for the initial simulation run of size n_0, we choose a sample size, $n \geq n_0 + 1$, to ensure an estimate error of $\epsilon = 0.075$ at a 95% confidence level (see Law and Kelton (2005)). To do this, we increment n until the following condition holds:

$$n = \min_i \left\{ i > n_0 : t_{i-1,1-\alpha/2} \cdot \sqrt{\frac{S^2(n_0)}{i}} \leq \epsilon \right\} \tag{1.31}$$

After the initial run of n_0 replications, we will require an additional $n - n_0$ replications. Note that we use an error of $\epsilon = 0.075$ as a balance between accuracy and computational processing effort.

Since the dispersion of the simulation results may depend on the parameters and decision variables, we perform the above sample size selection process for each run of the simulation. This should ensure, with 95% confidence, that the estimated total expected cost is within an error of $\epsilon \leq 0.075$ for the possible decision variable combinations for a given set of cost and demand parameters.

The full simulation flow chart is given in the Figure 1.3, as it demonstrates the sample size selection process coupled with the simulation algorithm.

When we were initially exploring the validity of the cost estimate, there appeared to be some significant sampling error. Therefore, to validate our initial simulation approach and reduce sampling error, we present the following improved simulation approach.

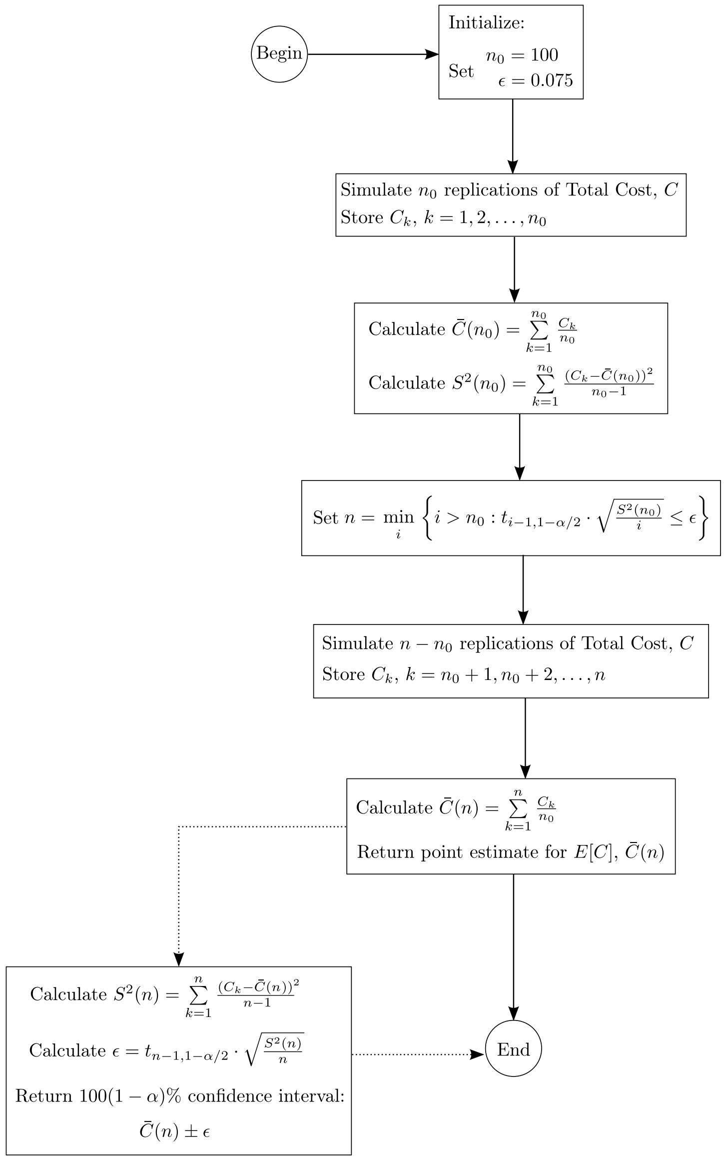


Figure 1.3: Basic Simulation Flowchart: Computing expected total cost

1.5.3 Simulation with Variance Reduction

In an attempt to decrease the effect of sampling error, and potentially increase the speed of computing the total expected cost, we will condition on the value of total lead time demand, D^L, as given below:

$$\bar{C} = \sum_{i=0}^{\infty} E[C|D^L = i] Pr(D^L = i) \tag{1.32}$$

Estimating the above infinite sum of conditional expectations requires that we define a new set of partitions of the possible values of D^L. We will truncate the resultant sum while minimizing the truncation error. We define the partitions as follows:

$$\begin{aligned} \text{Partition 0:} &\quad D^L \leq s_N + 1 \\ \text{Partitions } k,\ k = s_N + 2, \ldots, M\text{:} &\quad D^L = k \\ \text{Partition } M+1\text{:} &\quad D^L > M \end{aligned}$$

Computing the expected cost for the different partitions will require different approaches. When D^L is between zero and $s_N + 1$, for instance, the derivation of the performance measures yields no backorders, leading to a cost equation with no first order loss functions. In this situation, we can easily compute the total expected cost for Partition 0, denoted $\bar{C}_0$, as given below:

$$\bar{C}_0 = h(s_1 + s_2 + \cdots + s_N + 1) Pr(D^L \leq s_N + 1) - h \sum_{i=1}^{s_N+1} i Pr(D^L = i) \tag{1.33}$$

For the remaining partitions, when $D^L > s_N + 1$, the total expected cost will have to be computed using the sum of the sample averages from each partition k, where $k > s_N + 1$. We use $\bar{C}_k$, to denote the sample average for a particular partition, k. Assuming that for partition k we simulate n_k observations of total cost to compute the average, and that C_{kl} is the l^{th} replication of total cost from partition k, we have the following:

$$\bar{C}_k = Pr(D^L = k) \cdot \sum_{l=1}^{n_k} \frac{C_{kl}}{n_k} \tag{1.34}$$

As stated above, we must minimize the truncation error and sampling error. To minimize truncation error, we basically stop the conditioning process at demand value $k = M$, after the magnitude of expected cost falls below an error value, $0 < \epsilon << 1$.

For sampling error, we use an approach similar to that used in our basic simulation approach, but it is applied to each partition, k. This means, that we use an arbitrary initial sample size for all partitions, $n_0 = 30$. Then, through the course of the simulation for partition k, we define n_k to be the total number of replications required to meet an estimate error value at a specified level of confidence over all replications.

For more specific information regarding our approach to finding n_k, know that we employ a method of variance reduction for conditional expectations based on Karian and Dudewicz (1991) and Minh (1989) - see Appendix A.5 for a more detailed description of the process. The complete simulation algorithm for this simulation approach is demonstrated using a flow chart in Figure 1.4.

The estimated cost is given by:

$$\bar{C} = \bar{C_0} + \sum_{k=s_N+2}^{M} \bar{C_k} \tag{1.35}$$

1.5.4 Brief results of variance reduction technique

It turns out that the variance reduction technique was found to be more accurate in computing total expected cost, as hoped. In addition, the variance has been reduced significantly with the coefficient of variation found to be less than half that of the original simulation approach. It would be interesting to pursue a more in-depth analysis of the two simulation approaches, but that is left to future research.

With a way to compute cost at a specified level of error, we move now to the optimization approach. We will discuss the idea of perturbation analysis (PA), where we will estimate first differences after each simulation run.

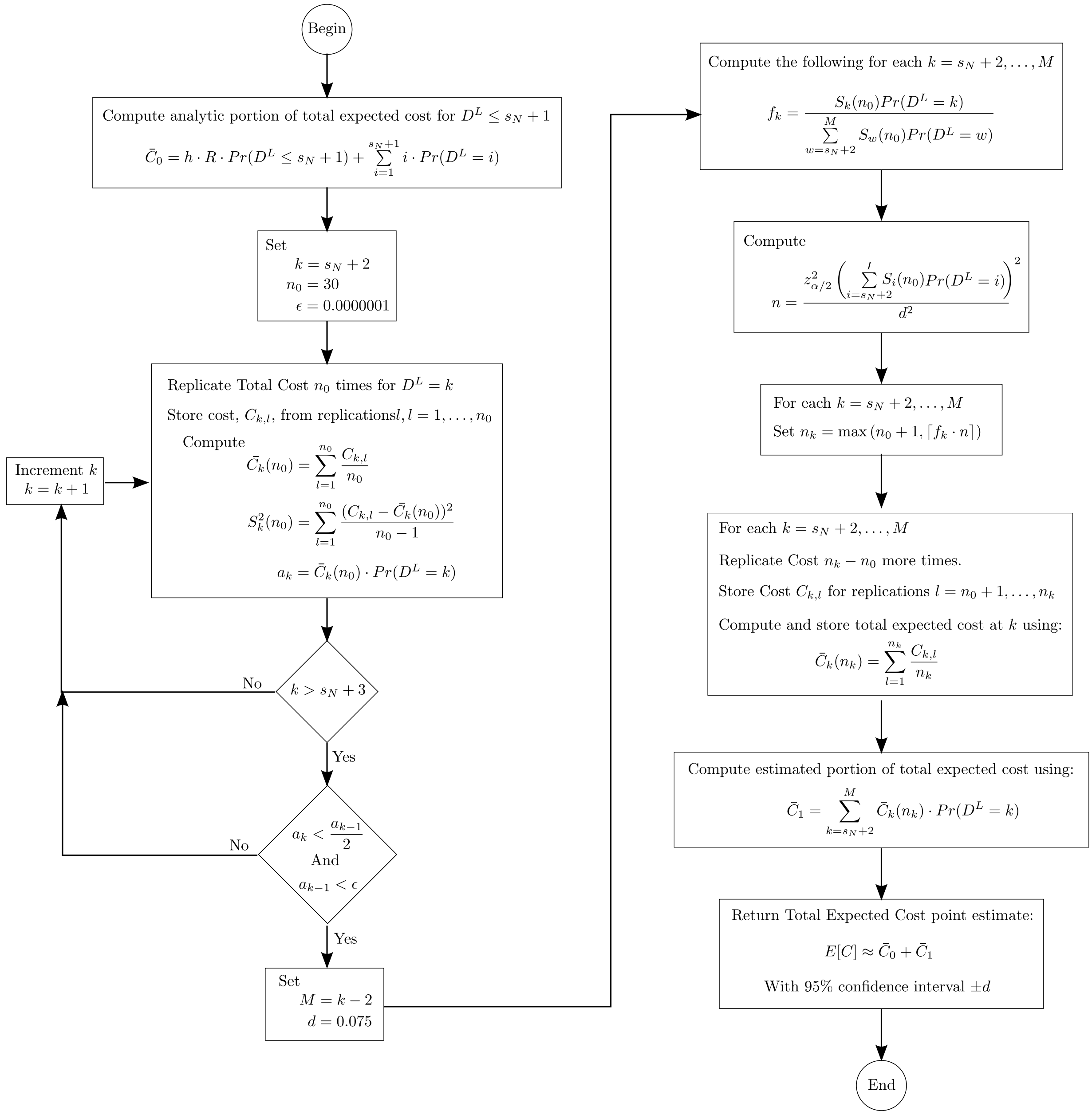


Figure 1.4: Simulation Flowchart for Variance Reduction: Computing expected total cost

1.6 DIFFERENCE ESTIMATION

Normally, to minimize total expected cost we would compute the first differences, or the derivatives, over the decision variables and then apply some form of gradient-search heuristic to minimize the objective. Due to the difficulty of analytically deriving the first differences for our model, we will use a form of PA that allows us to estimate first differences.

We define first differences over each of the decision variables, $s_1, s_2, \ldots, s_N$, as:

$$\Delta_j C = C(s_j + 1) - C(s_j) \tag{1.36}$$

where all other variables, s_i, $i = 1, \ldots, N$ and $i \neq j$, are left unchanged.

The cost, C, represents the total inventory cost generated over a single sample path. As we employ simulation to implement PA, we consider a single sample path to be the state of the system after a single simulation replication. Similar to the approach taken to compute cost, we will estimate the first differences by averaging them over several simulation replications.

However, we need not simulate separately for each decision variable, as the PA technique allows for first-difference estimation over all decision variables in a single simulation replication. We now discuss the mechanics of sample path perturbation analysis, tailoring the approach to the unique structure of the demand class inventory system.

1.6.1 Perturbation Analysis

To begin PA, the chosen decision variables are perturbed and then their effect on cost is traced through the course of one instance, or sample path, as generated by one simulation replication. In other words, we basically track how *small changes* to the decision variables change the cost of the system, thus estimating the derivatives or differences. When using Infinitesimal PA in general, a *small* change is an infinitesimal change given by a small $\nu > 0$, where ν is a real number. However, in our inventory system the decision variables are integers, net inventory is an integer, backorders are given by an integer, and the probability distributions are discrete.

Therefore, in our case we will look at a $\nu = 1$ and change, or perturb, each reserve stock level independently by adding a $\nu = 1$ to each $s_j, j = 1, 2, \cdots, N$. Perturbing each of the reserve stock levels creates a sequence of changes to the state of the system following a single simulation replication. Tracking the affect on the cost function allows us to compute an estimated first difference for each perturbed decision variable. This approach is different than a classic IPA approach which, instead, assumes continuous variables. These continuous variables are usually perturbed infinitesimally to estimate the derivatives with respect to each of the variables.

The body of inventory modeling literature for multiple demand classes does not have any examples where PA has been applied. Thus, our integer PA approach is new to this stream of research.

So, we will now use integer PA to estimate the first differences which will, in turn, direct the optimization heuristic to be presented later. First, we will introduce some new terminology and notation associated with the model, followed by the algorithm used to implement PA for our model. The PA algorithm will begin at the end of each simulation replication and will use the explicit values generated during the run to compute each estimate. The following cost difference equation, based on Equation 1.28, demonstrates the differences we want to estimate.

$$\Delta_j C = h\left[\Delta_j IN_1{}^+ + \cdots + \Delta_j IN_N{}^+\right] + b_1 \Delta_j B_{1,1} + \cdots + b_N \Delta_j B_{N,N} \text{ ; for } j = 1, \cdots, N \quad (1.37)$$

Note from Equation 1.28 that the total cost is dependent on all of the reserve stock levels, s_j, $j = 1, \cdots, N$, but we simplify the notation here for ease of use. As stated previously, we define $\Delta_j C$ to mean the first difference $C(s_j+1) - C(s_j)$, where all s_i, $i \neq j$, are held constant. The meanings of $\Delta_j IN_i$, $\Delta_j B_{i,i}$, $\Delta_j B_{i,i-1}$ and $\Delta_j B_i$ follow the same logic. To simplify Equation 1.38 we introduce the following theorem:

Theorem 1.1 *$\Delta_j IN_i = 0$, $\Delta_j B_i = 0$, $\Delta_j B_{i,i-1} = 0$, and $\Delta_j B_{i,i} = 0$ for $j < i$*

The proof of the theorem can be found in the Appendix, though it is a straightforward concept. Basically, the theorem demonstrates that perturbing some reserve stock level s_j, at some stage j, $j \geq 1$, will not affect the performance measures at higher stages, $i > j$ (i.e.

lower priority stages). This structure allows us to modify Equation 1.38 to only address those decision variables, s_j, $j = 1, \ldots, N$, that affect the performance measures for stages $i = 1, \ldots, j$, as follows:

$$\Delta_j C = h\left[\Delta_j IN_1{}^{+} + \cdots + \Delta_j IN_j{}^{+}\right] + b_1 \Delta_j B_{1,1} + \cdots + b_j \Delta_j B_{j,j} \text{ ; for } j = 1, \cdots, N \quad (1.38)$$

In its current state, the above equation still requires that we compute $\Delta_j IN_i$ and $\Delta_j B_{i,i}$ for all $i = 1, \cdots, j$. We have found that this is unnecessary and wastes computer processing capacity. Recall that we assumed that our integer PA step would occur at the end of each simulation run, so we would already have information regarding all the performance measures at every stage of the inventory system. We will use this static information to further simplify the difference function given in Equation 1.38. First, we know that the following is true:

THEOREM 1.2 $\Delta_i IN_i = 1$, *for all* $i = 1, \cdots, N$.

So, from Theorem 1.2 we see that perturbing s_j at some stage j directly affects net inventory at stage j by an increase of one unit. However, this increase will not necessarily affect the costs specific to stage j. Taking a broader look, we find that perturbing s_j for any $j \geq 1$ could possibly effect the static value of cost at any single stage stage $i \leq j$. And in fact, each perturbation will only affect the cost specific to exactly one stage $i \leq j$, as will be shown.

The effect that such a perturbation will have depends on the static values generated during the simulation run. More specifically, we only need to know whether net inventory is positive, $IN_i < 0$, or negative, $IN_i \geq 0$, for each $i \leq j$. We therefore introduce the key theorems for these two possibilities, or cases, that rely on the value of the simulation-generated net inventory at each stage; theorem proofs are found in the Appendix.

Case 1: Negative stage i net inventory, $IN_i < 0$:

THEOREM 1.3 *(a) If* $IN_i < 0$ *for some* i, $2 \leq i < N$, *then* $\Delta_j B_{i,i} = [-1 + x_i] \cdot x_{i+1} \cdots x_j$, *for all* $j = i+1, \cdots, N$. *(b) If* $IN_1 < 0$, *then* $\Delta_j B_{1,1} = x_2 \cdot x_3 \cdots x_j$, *for* $j = 2, \cdots, N$.

THEOREM 1.4 *For any i, $2 \leq i \leq N$, if $IN_i < 0$, then $\Delta_j IN_{i-1} = x_i \cdot x_{i+1} \cdots x_j$, for $j = i, \cdots, N$.*

To give a short description of the x_k, we use Bernoulli trials to simulate the conditional distributions given in Equations 1.21 and 1.23. We define x_i to be the final Bernoulli trial generated at a particular stage i, where $x_i = 1$ for internal backorders and $x_i = 0$ for external backorders. Since the value of this random variable is static at the end of a simulation replication, we can use it to determine the effect of perturbing the decision variables.

For example, in the case of negative net inventory, $IN_i < 0$, increasing a reserve stock level means decreasing the number of total backorders, B_i. The method used to simulate the system requires that a reduction in total backorders means a reduction in the number of Bernoulli trials required to compute the conditional distributions for internal and external backorders. The value of x_i determines whether the reduction in total backorders means reduced internal or reduced external backorders. See Observation A.2 for more details.

Looking at what the above theorems tell us, we see that if $IN_i < 0$, then we can directly compute the change in performance measures using the result of the last Bernoulli trials x_k, where $i \leq k \leq j$. Since these performance measures tie directly to cost, we have a way to estimate cost differences when $IN_i < 0$.

Case 2: Non-negative stage i net inventory, $IN_i \geq 0$:

THEOREM 1.5 *For $2 \leq i \leq N$, if $IN_i \geq 0$, then $\Delta_j IN_{i-1} = 0$, $j = i, \cdots, N$.*

THEOREM 1.6 *For i, $1 \leq i \leq N$, if $IN_i \geq 0$, then $\Delta_j B_k = 0$, $\Delta_j B_{k,k-1} = 0$, $\Delta_j B_{k,k} = 0$, for $k \leq i$, and $j \geq k$.*

THEOREM 1.7 *For some i, $2 \leq i \leq N$, and any $j \geq i$, if $\Delta_j IN_i = 0$, then $\Delta_j IN_k = 0$, for $k < i$.*

The above theorems show that for some stage, i, where net inventory is nonnegative, the perturbation of variable s_j, for some $j \geq i$, will not affect the performance measures at stages $1, \cdots, i-1$. Since the system performance measures for downstream stages are unaffected, there is no further affect on total cost. We define stage m to be the highest stage where net inventory is nonnegative. The structure of the inventory system requires that $IN_1 \geq 0; \cdots; IN_m \geq 0$ and $IN_{m+1} < 0; \cdots; IN_N < 0$.

For $j \geq m$, Theorems 1.4 and 1.3 give the relations for net inventory difference $\Delta_j IN_i$ and external backorder difference $\Delta_j B_{i,i}$ for any stage i, $m+1 \leq i \leq j$. Note also, that on-hand inventory, IN_i^+, is consistently zero and is unchanged since $IN_i < 0$ for stages $i = m+1, \cdots, j$.

At stage m, we have that on-hand inventory equals net inventory, that is, $\Delta_j IN_m^+ = \Delta_j IN_m = x_{m+1} \cdots x_j$, by Theorem 1.4. This is because net inventory at higher stages is negative, $IN_{m+1} < 0$, and, for stages m and lower, net inventory is non-negative, $IN_m \geq 0$. Also we have that $\Delta_j B_{m,m} = 0$ by Theorem 1.6.

If $j = m$, then of course $\Delta_m IN_m^+ = \Delta_m IN_m = 1$ by Theorem 1.2. By Theorems 1.6 and A.12 we know there is no further effect from perturbing s_j, so $\Delta_j IN_i^+ = \Delta_j IN_i = 0$ and $\Delta_j B_{i,i} = 0$ for stages i, where $i = 1, \cdots, m-1$ when $j \geq m$.

If $j < m$, then we know $\Delta_j IN_j^+ = \Delta_j IN_j = 1$, and by the same logic as above, Theorems 1.6 and A.12 show that there is no further effect since $IN_1 \geq 0; \cdots; IN_j \geq 0$. Again, we have that $\Delta_j IN_i^+ = \Delta_j IN_i = 0$ for stages i, where $i = 1, \cdots, j-1$ and $\Delta_j B_{i,i} = 0$ for stages i, where $i = 1, \cdots, j$ when $j < m$.

Based on the above logic, we can now simplify the derivation of the cost difference using the fact that there is a highest stage m where $IN_m \geq 0$.

$$\begin{aligned}
\text{For } j > m: \quad \Delta_j C &= h\Delta_j IN_m^+ + b_{m+1}\Delta_j B_{m+1,m+1} + \cdots + b_j \Delta_j B_{j,j} \\
&= h(x_{m+1} \cdots x_j) + (b_{m+1}(-1 + x_{m+1})x_{m+2} \cdots x_j) + \cdots + (b_j(-1 + x_j)) \\
\text{For } j \leq m: \quad \Delta_j C &= h\Delta_j IN_j^+ = h
\end{aligned}$$

Using the above approach will allow us to estimate the first difference appropriate to our model and with some computational efficiency. We now present an algorithm for integer PA, that will estimate the cost differences for each simulation replication.

1.6.2 Algorithm for PA

We will apply PA to each replication generated during the computation and simulation of the cost objective as described in Section 1.5. This will require that, in addition to computing cost, the simulation will keep track of net inventory, internal and external backorders, the highest stage, m, where net inventory is positive, and the last Bernoulli trial, x_k, that is found each time the internal and external backorders are generated. Assuming that this is done, we present an algorithm that utilizes the structure of the sample path, as described above, to estimate the first differences over the decision variables, $s_1, s_2, \ldots, s_N$.

We define $y_{i,j}$ to be the product of the last Bernoulli trials $x_i, x_{i+1}, \ldots, x_j$, as shown below:

$$y_{i,j} = \prod_{w=i}^{j} x_w \tag{1.39}$$

To begin the algorithm we assume that a single cost replication has just completed, where all we need to know is the costs, b_i and h, the highest stage with positive net inventory, m, and the last Bernoulli trials, x_i, for stages $i = m+1, \ldots, N$. The basic steps are outlined by the flow diagram in Figure 1.5.

Applying this algorithm at the end of each replication and then averaging the estimated differences over a sufficient number of replications will provide estimates for the expected first differences over each decision variable. However, as we employ a variance reduction technique in the simulation for cost, our PA approach must be modified accordingly.

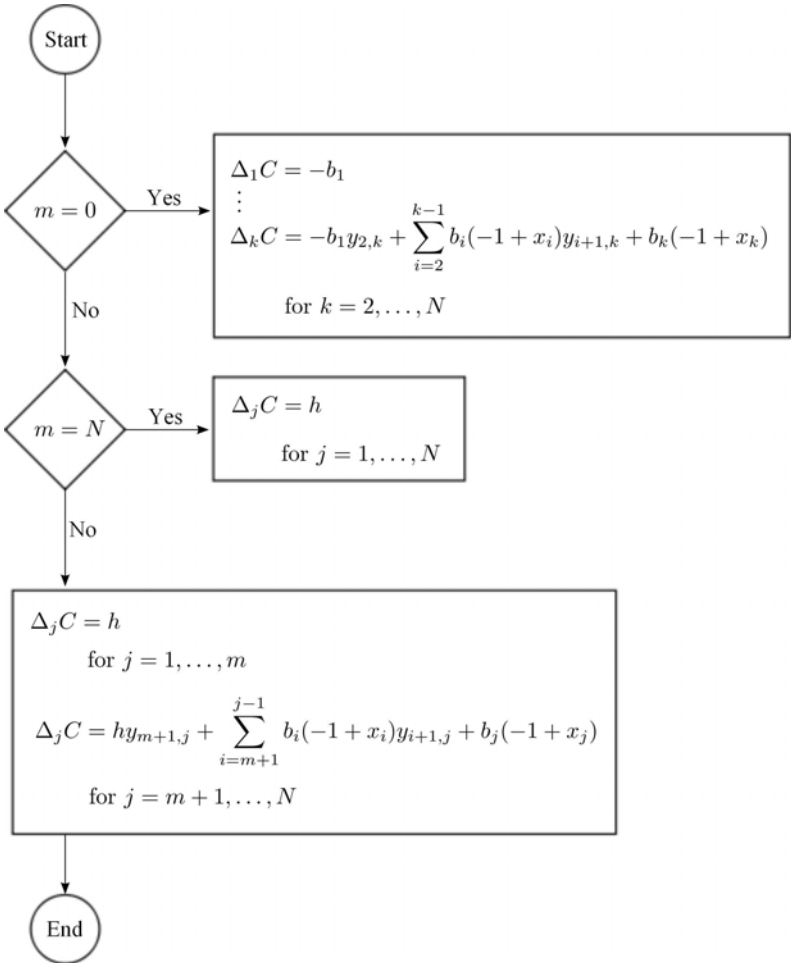


Figure 1.5: IPA Algorithm Outline: applied at the end of each simulation replication

1.6.3 Simulating PA

In the simulation for estimating cost we condition on the value of lead time demand, D^L, to reduce sampling error. Below, we review the simulation for cost as we discuss the approach used in estimating the first differences.

Remember that a portion of the expected cost, $\bar{C}_0$, was computed directly when $D^L \leq s_N+1$, as the net inventory at each stage was nonnegative, $IN_i \geq 0$. Clearly, when inventory is nonnegative, perturbing the decision variables will only increase holding costs, leading to a method of direct computation of first differences under these conditions, given by $\Delta_j \bar{C}_0 = h$.

Furthermore, for $D^L > s_N + 1$ the expected cost could not be computed directly, so we utilized simulation to compute the remaining $\bar{C}_k$ for $D^L = k$, for $k = s_N + 2, \ldots, M$, where M is the truncation point of the infinite sum. We will follow the same simulation approach (in fact, we embed the PA simulation within the cost simulation) as we generate the estimates for the expected first differences for each value of $D^L = k$, given by $\hat{\Delta}_j \bar{C}_k$.

Here, we define $\hat{\Delta}_j \bar{C}_k$ to be the average of the differences generated over n_k replications, where n_k is found during the cost simulation as described previously. To compute these averages, we store the sample path cost differences from each replication, l, denoted $\Delta_j C_{kl}$.

Then, as with cost, we will average the cost difference estimates over the n_k replications for each $D^L = k \leq M$, to find the following:

$$\hat{\Delta}_j \bar{C}_k = \sum_{l=1}^{n_k} \frac{\Delta_j C_{kl}}{n_k} \tag{1.40}$$

Finally, after applying the PA algorithm at the end of each simulation replication, we can estimate $\Delta_j C$. This is estimated using the estimated expected value of the first differences computed directly for $D^L \leq s_N + 1$, as given by $\Delta_j C_0 \cdot \Pr(D^L \leq s_N + 1)$, added to the estimated expected values of the first differences from the simulation for $s_N + 1 < D^L \leq M$, as given by $\sum_{k=s_N+2}^{M} \hat{\Delta}_j \bar{C}_k \cdot \Pr(D^L = k)$. This idea is expressed in the following equation:

$$\Delta_j C \approx \Delta_j \bar{C}_0 \Pr(D^L \leq s_N + 1) + \sum_{k=s_N+2}^{M} \hat{\Delta}_j \bar{C}_k \Pr(D^L = k) \tag{1.41}$$

Having successfully estimated the first differences over each of the decision variables, we move now to our preliminary numerical analysis. We also introduce a heuristic approach for minimizing total expected cost based on the estimated first differences.

1.7 PRELIMINARY ANALYSIS AND ATTEMPTED OPTIMIZATION HEURISTIC

This section represents the portion of this article that is currently in process. However, we do offer some very basic preliminary analysis and an attempt at optimizing the system described in this paper.

1.7.1 Preliminary Analysis

While a full analysis has not been completed, we have so far found that, on the average, our utilization of PA successfully predicts changes in cost due to perturbations of the decision variables within an error of 0.18%.

Also, the variance reduction technique that we employ generates expected cost estimates with very little error at the 95% confidence level as compared to the basic simulation approach. Recall that the simulation was designed to estimate cost within an error of at most 0.075 at the 95% confidence level. Currently, averaging the estimated expected cost over several simulation *runs* nets a 95% confidence interval for cost with an error of just 0.17% with an actual error of around 0.005. This improved accuracy is partially based on our setting the initial simulation run length $n_k = 30$, which, for many $D^L = k$, is more than sufficient to guarantee a specified level of error.

Clearly, more analysis is required to have a better understanding of how the simulation performs, but for the purpose of demonstration for the dissertation, we move on to discuss the optimization heuristic used to minimize the total estimated expected cost of the system.

1.7.2 Optimization

Having a method to estimate first differences for the total expected cost, we now seek to minimize the cost of the demand class system using heuristics. Having first difference estimates allows us to apply several different types of heuristics, the most common is a form of the steepest descent heuristic. However, as we are interested in discrete decision variables, we instead look at the largest change in cost over discrete variable changes. We call this discrete version of the descent heuristic the deepest descent heuristic, and an important element of this approach is the starting solution.

Although it isn't formal yet, our preliminary analysis and other numerical trials suggest that their are both local and global optima. Thus, applying the deepest descent heuristic from different starting points can lead to different minimizing solutions.

The most successful starting point we have found in applying the deepest descent heuristic is when we start at the solution $s_N = \lfloor L \cdot \sum \lambda_i \rfloor$, and $s_i = 0$, for $i = 1, \ldots, N-1$. While it bears further exploration both analytically and numerically, starting at this point and applying a deepest descent heuristic based on the estimated first differences appears to lead to the global minimum.

As stated previously, currently this section is currently in process, and will be continually updated.

1.8 CONCLUSION

In addressing an inventory rationing model, we have presented a multiple demand class system with backordering. We have structured the model based on a serial stage inventory system, as done in other literature. However, we provide the first example of a cost-objective model under this structure.

As experienced by other researchers, we were unable to develop a closed-form, analytical approach to the multi-class inventory rationing model under a cost objective and backordering. Instead, we employed simulation and a form of perturbation analysis (PA) to estimate the total expected cost and the first difference, an approach that has not been employed previously in the literature to address this type of model.

Based on our preliminary analyses, though incomplete, it appears that the PA approach is successful in determining viable estimates for the first differences over each of the decision variables. Using these difference values and ensuring a good starting point, we have been able to find optimal solutions to the simulation model in a limited study.

1.8.1 Future Work

Clearly, the analyses are incomplete, and will be a focus in the immediate future. Based on the analyses, further development of the optimization heuristic may also be beneficial.

Also, there is some promise in developing an easy-to implement spreadsheet version of the model that could find good solutions at a reasonable rate using solver or Visual Basic. This new approach could use the expected values of the Binomial distributions given in Equations 1.21 and 1.23, instead of simulating the entire system. The approach would still require conditioning on values of the lead time demand, D^L, but instead of sample path analysis you could think of it as steady-state path analysis.

Now that the base-stock replenishment case is complete, it would be interesting to continue the research to include the case of a fixed ordering cost - allowing for an (r, Q) replenishment policy when economically beneficial.

Note that our simulation computes the total expected cost using the performance measures needed to also compute the fill rate. It would be both interesting and potentially simple to recreate, or modify, the service level problem from Arslan et al. (2007) that was the main motivation for this paper. We could then compare our model results to the previous work on the service level problem.

Lastly, it is our desire that this type of model can be utilized in industry. A study could be completed regarding the prevalence of contemporary inventory rationing systems currently in use. If in existence, it would be rewarding to research how our model may apply to the industry case. It is likely that this and other inventory rationing models lend insight to industries related to inventory management, such as revenue management and yield management.

REFERENCES FOR ARTICLE 1

Arslan, H., S. C. Graves, T. A. Roemer. 2007. A single-product inventory moel for multiple demand classes. *Management Science* **53**(9) 15.

Dekker, R., R. M. Hill, M. J. Kleijn, R. H. Teunter. 2002. On the (s - 1, s) lost sales inventory model with priority demand classes. *Naval Research Logistics* **49**(6) 539–610.

Deshpande, V., M. A. Cohen, K. Donohue. 2003. A threshold inventory rationing policy for service-differentiated demand classes. *Management Science* **49**(6) 21.

Graves, S. C. 1985. A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science* **31**(10).

Kaplan, A. 1969. Stock rationing. *Management Science* **15**(5) 8.

Karian, Z. A., E. J. Dudewicz. 1991. *Modern Statistical, Systems, and GPSS Simulation: The First Course*. W.H. Freeman and Company, New York.

Kleign, M. J., R. Dekker. 1998. An overview of inventory systems with several demand classes. Tech. rep., Erasmus University.

Law, A. M., W. D. Kelton. 2005. *Simulation Modeling and Analysis*. McGraw-Hill, New York.

Minh, D. L. 1989. A variant of the conditional expectation variance reduction technique and its application to the simulation of the gi/g/1 queues. *Management Science* **25**(11).

Moon, I., S. Kang. 1998. Rationing policies for some inventory systems. *The Journal of the Operational Research Society* **49**(5) 10.

Nahmias, S., W. S. Demmy. 1981. Operating characteristics of an inventory system with rationing. *Management Science* **27**(11) 10.

Teunter, R. H., W. K. K. Haneveld. 2008. Dynamic inventory rationing strategies for inventory systems with two demand classes, poisson demand and backordering. *European Journal of Operational Research* **190** 23.

Topkis, D. M. 1968. Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes. *Management Science* **15**(3) 17.

Veinott, A. F. 1965. Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. *Operations Research* **13**(5) 18.

Vicil, O., P. Jackson. 2006. An exact optimal solution to a threshold inventory rationing model for multiple priority demand classes. Tech. rep., Cornell University.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.

A APPENDIX

A.1 Definitions and Observations

These first definitions, observations and theorems are either described in the text, or are extensions of what was shown in Arslan et al. (2007).

Definition The probability that the next demand at stage i is internal is given by $p_i = \frac{\sum_{k=1}^{i-1} \lambda_k}{\sum_{k=1}^{i} \lambda_k}$, where $i = 2, 3, \cdots, N$. The probability that the next demand at stage i is external is given by $q_i = 1 - p_i = \frac{\lambda_i}{\sum_{k=1}^{i} \lambda_k}$, where $i = 2, 3, \cdots, N$. For stage 1, there is only external demand.

Observation A.1 $B_1 = B_{1,1} = [-IN_1]^+$

Proof

We know that $B_i = B_{i,i-1} + B_{i,i}$ for all N stages. But stage 1 only serves external demand, and we define $B_{1,0} \equiv 0$. So $B_1 = 0 + B_{1,1}$, where $B_1 = [-IN_1]^+ \Longrightarrow B_1 = B_{1,1} = [-IN_1]^+$ ■

Theorem A.1 *Internal backorders $B_{i,i-1}$, $i = 2, \cdots, N$, are randomly distributed according to a Binomial(p_i,B_i).*

Proof

This was shown to be true by Arslan et al. (2007), and we assume the validity of the original work. ■

A.2 Model-Related Proofs

Based on the structure of the serial stage system, the following theorems and observations are given to support the computation, simulation, and analysis that are performed in this paper.

Observation A.2 *(Theorem 1.3 from text) Internal Backorders at some stage $i = 2, \cdots, N$, given by $B_{i,i-1}$, can be generated using the sum of a Binomial(p_i, $B_i - 1$) and a Bernoulli trial with probability p_i, which we write as x_i. This relationship is written as $B_{i,i-1} \sim$ Binomial$(p_i, B_i - 1) + x_i$. Also, $B_i = 0 \Longrightarrow B_{i,i-1} = 0$ and $B_i = 1 \Longrightarrow B_{i,i-1} = x_i$for all $i = 2, \cdots, N$.*

Proof

By Theorem A.1 which was shown by Arslan et al. (2007), we know that $B_{i,i-1} \sim \text{Binomial}(p_i, B_i)$. By definition, the Binomial(p, n) distribution is identically distributed to the sum of n Bernoulli trials.

We use this property to separate the Binomial$(p_i,\ B_i)$ into B_i Bernoulli trials, and then combine the first $B_i - 1$ Bernoulli trials into a Binomial$(p_i,\ B_i - 1)$ random variable, which is summed with a single Bernoulli random variable with probability p_i. We define x_i to be the Bernoulli random variable, where

$$x_i = \begin{cases} 1\,; & \text{successful trial (internal demand)} \\ 0\,; & \text{otherwise (external demand)} \end{cases}$$

We can then write $B_{i,i-1} \sim \text{Binomial}(p_i, B_i - 1) + x_i$, and we will use this relationship for our calculations. Here, x_i will represent the *last* Bernoulli trial in computing the internal backorders at stage $i = 2, \cdots, N$. As such, if $B_i = 1$ then we define $B_{i,i-1} = x_i$. Also, $B_i = 0 \Longrightarrow B_{i,i-1} = 0$, because naturally there can be no internal backorders since $B_i \equiv B_{i,i-1} + B_{i,i}$.
■

Theorem A.2 *IN_i, B_i, $B_{i,i-1}$, and $B_{i,i}$ depend only on the decision variables $s_i, s_{i+1}, \ldots, s_N$ and can be written in function notation (i.e. $IN_i(s_i, s_{i+1}, \cdots, s_N)$).*

Proof

For $i = N$, we have that $IN_N = s_N - D^L$, with D^L independent of the decision variables. So, IN_N is dependent only on s_N, written as $IN_N(s_N)$. Similarly, $B_N = [-IN_N(s_N)]^+$, which shows B_N is only dependent on s_N, written as $B_N(s_N)$. We also have from Observation A.2 that $B_{N,N-1} = \text{Binomial}(p_N, B_N(s_N) - 1) + x_N$, which shows that $B_{N,N-1}$ also depends only on s_N through its dependence on B_N. Finally, since $B_{N,N} = B_N(s_N) - B_{N,N-1}(s_N)$, we see that external backorders at stage $i = N$ are also only dependent on s_N.

By induction, assume theorem is true for $i = N, N-1, \cdots, k+1$

For $i = k$ we have that $IN_k = s_k - B_{k+1,k}$, where B_{k+1} is only dependent on the decision variables $s_{k+1}, s_{k+2}, \cdots, s_N$ by the induction assumption. Therefore, $IN_k = s_k - B_{k+1,k}(s_{k+1}, \cdots, s_N)$ is only dependent on $s_k, s_{k+1}, \cdots, s_N$. Also, $B_k = [-IN_k(s_k, \cdots, s_N)]^+$, so it is only dependent on $s_k, s_{k+1}, \cdots, s_N$. Internal backorders are given by $B_{k,k-1} = \text{Binomial}(p_k, B_k(s_k, \cdots, s_N) - 1) + x_k$, which shows that $B_{k,k-1}$ depends only on $s_k, \cdots, s_N$. Finally, since $B_{k,k} = B_k(s_k, \cdots, s_N) - B_{k,k-1}(s_k, \cdots, s_N)$, it is obvious that external backorders at stage k are dependent on the decision variables $s_k, \cdots, s_N$.

$\therefore$ For some stage $i = 1, \cdots, N$, IN_i, B_i, $B_{i,i-1}$, and $B_{i,i}$ depend on the decision variables $s_i, s_{i+1}, \cdots, s_N$. ■

THEOREM A.3 *$IN_i < 0 \Longrightarrow IN_j < 0, B_j > 0$ for $j = i, \cdots, N$, where $i = 1, \cdots, N$.*

PROOF (By induction)

$IN_i < 0 \Longrightarrow IN_i < 0$

Assume $IN_i < 0, IN_{i+1} < 0, \cdots, IN_j < 0$ for $j < N$.

$$\begin{aligned} IN_j < 0 &\iff s_j - B_{j+1,j} < 0 \\ &\iff B_{j+1,j} > s_j \geq 0 \\ &\iff B_{j+1} > 0 \\ &\iff IN_{j+1} < 0 \end{aligned}$$ ■

THEOREM A.4 *$IN_i \geq 0 \Longrightarrow IN_j = s_j$, for $j = 1, \cdots, i-1$, and $i = 2, \cdots, N$*

PROOF (by induction)

For $j = i - 1$ we have that $IN_{i-1} = s_{i-1} - B_{i,i-1}$ for $i = 2, \cdots, N$. We also know that $B_i = 0$ since $IN_i \geq 0$ and $B_i = [-IN_i]^+$. By Observation A.2 this means that $B_{i,i-1} = 0$. Therefore $IN_{i-1} = s_{i-1}$.

Assume theorem is true for $j = i-1, i-2, \cdots, k+1$, then for $j = k$ we have $IN_k = s_k - B_{k+1,k}$. Since $IN_{k+1} \geq 0$ by the induction assumption, we know $B_{k+1} = 0 \Longrightarrow B_{k+1,k} = 0$. Therefore, $IN_k = s_k$.

Thus, knowing that $IN_i \geq 0$ for some stage $i = 2, \cdots, N$, we know that for stages $j = 1, \cdots, i-1$ $IN_j = s_j$. ■

A.3 IPA-Related Proofs

To set up our solution procedure using IPA, we want to derive some useful properties of the model that will reduce the computation time of the simulation and add to our understanding of the model structure. Therefore, in the following theorems and observations, we are assuming that the random variables are replaced with static numerical values. The actual quantities are not important, only that they are static representations of random variables - though we do not change notation. As we commence with IPA and this replacement is made, we can begin to look at how small changes in the decision variables affect the costs and other values of the system in a particular instance.

For example, we have net inventory at stage N given by $IN_N = s_N - D^L$. In such a case, the random variable for lead time demand, $D^L \sim \text{Poisson}(\lambda L)$, is assumed to be a static value. That is, during IPA we treat D^L as a static, realized instance of lead time demand. In such a case, if we were to increase s_N by one unit, then net inventory, would be directly increased by one unit. So, because we have assumed that D^L represents a static numerical value, IN_N becomes a static numerical value that is directly dependent on the value of s_N. The definitions, observations, and theorems that follow adhere to this same assumption for all random variables associated with the model.

Definition $\Delta_j X(s_j) = X(s_j+1) - X(s_j)$, where $X(s_j)$ represents some performance measure that depends, at least, on variable s_j, where s_j is the reserve stock level for stage j.

Theorem A.5 *(Theorem 1.1 from text)* $\Delta_j IN_i = 0$, $\Delta_j B_i = 0$, $\Delta_j B_{i,i-1} = 0$, *and* $\Delta_j B_{i,i} = 0$ $j < i$

Proof

We define $\Delta_j IN_i$ to be the change in stage i net inventory given that reserve stock level s_j is increased by one unit. We define $\Delta_j B_i$, $\Delta_j B_{i,i-1}$, and $\Delta_j B_{i,i}$ similarly.

By Theorem A.2 we know that IN_i, B_i, $B_{i,i-1}$, and $B_{i,i}$ each depend on variables $s_i, s_{i+1}, \cdots s_N$. Since we assume that $j < i$, we know that IN_i, B_i, $B_{i,i-1}$, and $B_{i,i}$ do not depend on s_j. Therefore, $\Delta_j IN_i = 0$, $\Delta_j B_i = 0$, $\Delta_j B_{i,i-1} = 0$, and $\Delta_j B_{i,i} = 0$, for $j < i$. ■

Theorem A.6 *(Theorem 1.2 from text) $\Delta_i IN_i = 1$, for all $i = 1, \cdots, N$.*

Proof

We know $IN_i = s_i - B_{i+1,i}$, for $i = 1, 2, \cdots, N-1$. By Theorem A.2 we know $B_{i+1,i}$ is dependent only on $s_{i+1}, s_{i+2}, \cdots, s_N$, for $i = 1, 2, \cdots, N-1$. We can then write $\Delta_i IN_i = IN_i(s_i + 1) - IN_i(s_i) = (s_i + 1 - B_{i+1,i}) - (s_i - B_{i+1,i}) = 1$. So for $i = 1, \cdots, N-1$, we have shown that $\Delta_i IN_i = 1$.

For $i = N$, we have that $IN_i = s_i - D^L$. By definition, we know that D^L, assumed to be a static instance of the random variable, is independent of the reserve stock levels, s_i, for all $i = 1, 2, \cdots, N$. We can then write $\Delta_N IN_N = (s_N + 1 - D^L) - (s_N - D^L) = 1$. So for $i = N$ we have shown that $\Delta_i IN_i = 1$. Therefore, $\Delta_i IN_i = 1$, for all $i = 1, \cdots, N$. ■

Theorem A.7 *For some $i = 1, \cdots, N$, if $IN_i < 0$, then $\Delta_j B_j = -1$, for any $j = i, \cdots, N$.*

Proof

Assume for some $i = 1, \cdots, N$ that $IN_i < 0$. Then by Theorem A.3 for $j = i, \cdots, N-1$, $IN_j < 0$, and $B_j > 0$. By definition, $B_j = B_{j+1,j} - s_j > 0$. We then apply the perturbation Δ_j and get $\Delta_j B_j = \Delta_j B_{j+1,j} - 1 = -1$.

For $j = N$ we know, by Theorem A.3 that $IN_N < 0$, and $B_N > 0$. So $B_N = D^L - s_N > 0$, where D^L is a static value as previously discussed. We apply Δ_N and get $\Delta_N B_N = \Delta_N D^L - 1 = -1$.

Therefore, when $IN_i < 0$ for $i = 1, \cdots, N$, we have shown that $\Delta_j B_j = -1$ for $j = i, \cdots, N$. ■

Theorem A.8 *For some $i = 2, \cdots, N$ if $IN_i < 0$, then (a) $\Delta_j B_{j,j-1} = -x_j$ and (b) $\Delta_j B_{j,j} = -1 + x_j$ for all $j = i, \cdots, N$.*

Proof

(a) By Theorem A.2 we know $B_{j,j-1} \sim \text{Binomial}(p_j, B_j - 1) + x_j$ for $j = 2, \cdots, N$, where x_j is defined as the last Bernoulli trial. Because of IPA, we assume that the random variables are replaced by static numeric values, and we can rewrite $B_{j,j-1}(s_j) = \text{Binomial}(p_j, B_j - 1) + x_j$. By

Theorem A.2 we have included function notation showing dependence on (at least) the variable s_j. So x_j represents one static instance of the last Bernoulli trial used to compute internal backorders, and the Binomial$(p_j, B_j - 1)$ term represents the numeric value of the first $B_j - 1$ Bernoulli trials in one instance.

By Theorem A.7 we know that $IN_i < 0 \implies \Delta_j B_j = -1$, for any stage $i = 1, \cdots, N$ and $j = i, \cdots, N$. This means that if stage i net inventory is negative, then for any stage $j = i, \cdots, N$, increasing s_j by one unit means total stage j backorders will be reduced by one unit.

Reducing total backorders by one unit reduces the number of Bernoulli trials used to generate internal backorders, and we remove the last Bernoulli trial, x_j. This is a direct result of increasing s_j, so $B_{j,j-1}(s_j + 1) = \text{Binomial}(p_j, B_j - 1)$.

Finally, we apply the Δ_j and get $\Delta_j B_{j,j-1} = B_{j,j-1}(s_j + 1) - B_{j,j-1}(s_j) = \text{Binomial}(p_j, B_j - 1) - (\text{Binomial}(p_j, B_j - 1) + x_j) = -x_j$, for all $j = i, \cdots, N$. Thus, we have shown that $IN_i < 0 \implies \Delta_j B_{j,j-1} = -x_j$, for $i = 2, \cdots, N$ and $j = i, \cdots, N$.

(b) By Theorem A.7 we know for $i = 1, \cdots, N$ that $IN_i < 0 \implies \Delta_j B_j = -1$ for $j = i, \cdots, N$. Also, by (a) we know $\Delta_j B_{j,j-1} = -x_j$ for $j = i, \cdots, N$. So we can apply Δ_j to the relation $B_{j,j} = B_j - B_{j,j-1}$, giving $\Delta_j B_{j,j} = \Delta_j B_j - \Delta_j B_{j,j-1} = -1 + x_j$. Thus we have also shown that $IN_i < 0 \implies \Delta_j B_{j,j} = -1 + x_j$, for $i = 2, \cdots, N$ and $j = i, \cdots, N$. ■

Observation A.3 *For any i, $2 \leq i \leq N$, if $\Delta_j B_i = -1$ for some j, $i \leq j \leq N$, then $\Delta_j B_{i,i-1} = -x_i$.*

Proof

Observation A.2 shows that for $i = 2, \cdots, N$, $B_{i,i-1} = \text{Binomial}(p_i, B_i - 1) + x_i$, where x_i is the last of B_i Bernoulli trials used to compute internal backorders $B_{i,i-1}$. Therefore, if total backorders, B_i, are reduced by one unit (speaking of the static values), then internal backorders, $B_{i,i-1}$, are reduced by the value of x_i, which is either 1 or 0 by definition.

Saying it more formally, if we assume for some $j = i, \cdots, N$ that $\Delta_j B_i = -1$. As stated above, this one-unit decrease in total backorders at stage i will result in a reduction of internal backorders at stage i in the amount of x_i. Because this reduction was assumed to be a result of increasing s_j, we can write that $\Delta_j B_i = -1 \Longrightarrow \Delta_j B_{i,i-1} = -x_i$. This holds for all $i = 2, \cdots, N$, and for any $j = i, \cdots, N$. ■

Theorem A.9 *For all $j = 2, \cdots, N$, if $IN_i < 0$ for some $i < j$, then (a) $\Delta_j B_i = -x_{i+1} \cdots x_j$ and (b) $\Delta_j B_{i,i-1} = -x_i \cdots x_j$*

Proof (By induction on i)

Let j be any stage where $j = 2, \cdots, N$, and initialize induction by showing true for $i = j-1$. So, suppose $IN_{j-1} < 0$.

(a) From Theorem A.3 we have that $IN_{j-1} < 0 \Longrightarrow B_{j-1} > 0 \Longrightarrow B_{j-1} = B_{j,j-1} - s_{j-1}$. We also have from Theorem A.3 that $IN_{j-1} < 0 \Longrightarrow IN_j < 0$. From Theorem A.8 and by applying the Δ_j we get $\Delta_j B_{j-1} = \Delta_j B_{j,j-1} = -x_j$. Which shows that $IN_{j-1} < 0 \Longrightarrow \Delta_j B_{j-1} = -x_j$.

(b) From (a) we know $\Delta_j B_{j-1} = -x_j$. We know that x_j is either 1 or 0, so we have that

$$\Delta_j B_{j-1} = \begin{cases} -1 & ; \text{if } x_j = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

From Observations A.2 and A.3 we have that

$$\Delta_j B_{j-1,j-2} = \begin{cases} -x_{j-1} & ; \text{if } \Delta_j B_{j-1} = -1 \\ 0 & ; \text{otherwise} \end{cases}$$

This is equivalent to saying that $\Delta_j B_{j-1,j-2} = -x_{j-1} \cdot x_j$. So, $IN_{j-1} < 0 \Longrightarrow \Delta_j B_{j-1,j-2} = -x_{j-1} \cdot x_j$.

Assume by induction that for all j, $2 \leq j \leq N$, if $IN_i < 0$ for some i, $k+1 \leq i < j$, then let $\Delta_j B_i = -x_{i+1} \cdots x_j$ and $\Delta_j B_{i,i-1} = -x_i \cdots x_j$. We need to show that (a) and (b) hold for $i = k$ if $IN_k < 0$. So, suppose that $IN_k < 0$.

(a) We know that $IN_k < 0 \Longrightarrow B_k > 0 \Longrightarrow B_k = B_{k+1,k} - s_k$ and $IN_{k+1} < 0$. We then apply the Δ_j and get $\Delta_j B_k = \Delta_j B_{k+1,k} = -x_{k+1} \cdot x_{k+1} \cdots x_j$ by the induction assumption.

(b)

From above, we have that $IN_k < 0 \Longrightarrow \Delta_j B_k = -x_{k+1} \cdot x_{k+1} \cdots x_j$, and so

$$\Delta_j B_k = \begin{cases} -1 & ; \text{if } x_{k+1} = x_{k+2} = \cdots x_j = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

Again, from Observations A.2 and A.3 we have that

$$\Delta_j B_{k,k-1} = \begin{cases} -x_k & ; \text{if } \Delta_j B_k = -1 \\ 0 & ; \text{otherwise} \end{cases}$$

Using the same logic as before, we can say this is equivalent to $\Delta_j B_{k,k-1} = -x_k \cdot x_{k+1} \cdots x_j$.

Thus, we have shown both (a) and (b) for $i = k$ and the induction is complete. ∎

Theorem A.10 *(a) If $IN_i < 0$ for some i, $2 \leq i < N$, then $\Delta_j B_{i,i} = [-1 + x_i] \cdot x_{i+1} \cdots x_j$, for all $j = i+1, \cdots, N$. (b) If $IN_1 < 0$, then $\Delta_j B_{1,1} = x_2 \cdot x_3 \cdots x_j$, for $j = 2, \cdots, N$.*

Proof

For $i = 2, \cdots, N-1$ we know $B_{i,i} = B_i - B_{i,i-1} \Longrightarrow \Delta_j B_{i,i} = \Delta_j B_i - \Delta_j B_{i,i-1}$ for $j = i+1, \cdots N$. From Theorem A.9, if $IN_i < 0$, then $\Delta_j B_{i,i} = (-x_{i+1} \cdot x_{i+2} \cdots x_j) - (-x_i \cdot x_{i+1} \cdots x_j)$ After factoring we get $\Delta_j B_{i,i} = [-1 + x_i] \cdot x_{i+1} \cdots x_j$.

When $i = 1$ we know that $B_{1,1} = B_1$. Therefore, $\Delta_j B_{1,1} = \Delta_j B_1 = -x_2 \cdot x_3 \cdots x_j$. ∎

Theorem A.11 *(Theorem 1.4 from text) For any i, $2 \leq i \leq N$, if $IN_i < 0$, then $\Delta_j IN_{i-1} = x_i \cdot x_{i+1} \cdots x_j$, for $j = i, \cdots, N$.*

Proof

We know for all $i = 2, \cdots, N$ that $IN_{i-1} = s_{i-1} - B_{i,i-1}$. We apply the Δ_j and get $\Delta_j IN_{i-1} = -\Delta_j B_{i,i-1}$. We know from Theorem A.9 that if $IN_i < 0$, then $\Delta_j B_{i,i-1} = -x_i \cdot x_{i+1} \cdots x_j \Longrightarrow \Delta_j IN_{i-1} = x_i \cdot x_{i+1} \cdots x_j$. ∎

Theorem A.12 *(Theorem 1.5 from text) For $2 \leq i \leq N$, if $IN_i \geq 0$, then $\Delta_j IN_{i-1} = 0$, $j = i, \cdots, N$.*

Proof

Suppose $IN_i < 0$ for some $i = 2, \cdots, N$. By Theorem A.4 we know that $IN_k = s_k$ for all $k = 1, \cdots, i-1$, so $IN_{i-1} = s_{i-1}$. We apply Δ_j and get $\Delta_j IN_{i-1} = \Delta_j s_{i-1} = 0$, for $j = i, \cdots, N$. ■

Theorem A.13 *$\Delta_j IN_i = 0$ or 1 for all $j, i = 1, \cdots, N$.*

Proof

For all $i = 2, \cdots, N$, where $j < i$ we know $\Delta_j IN_i = 0$ from Theorem A.5.

For all $i = 1, \cdots, N$, where $j = i$ we know $\Delta_j IN_i = 1$ from Theorem A.6.

So for $j \leq i$, $i = 1, \cdots, N$ we have that $\Delta_j IN_i = 0$ or 1.

For $i = 1, \cdots, N-1$, and $j > i$ we have two cases to consider: either $IN_i < 0$ or $IN_i \geq 0$.

Case $IN_i < 0$

By Theorem A.3 we know $IN_k < 0$, for all $k \geq i$. By Theorem A.11, if $IN_{i+1} < 0$ for some $i = 1, \cdots, N-1$, then $\Delta_j IN_i = x_{i+1} \cdot x_{i+2} \cdots x_j$ for $j > i$. And since for all i, x_i is either 0 or 1, we have that $\Delta_j IN_i = 0$ or 1.

Case $IN_i \geq 0$

Suppose $IN_i \geq 0$ for i, $2 \leq i \leq N$. Then, $\Delta_j IN_{i-1} = 0$ for $j = i, \cdots, N$ By Theorem A.12. This shows that $\Delta_j IN_i = 0$ for $j > i$ when $IN_i \geq 0$.

$\therefore$ we have shown for all $j, i = 1, \cdots, N$ that $\Delta_j IN_i = 0$ or 1. This makes sense in our IPA approach, as we are analyzing the affect of increasing s_j by one unit at some stage j. Intuitively, this one-unit increase in the base-stock level should not decrease net inventory, nor should it increase net inventory more than one unit at a particular stage. ■

Theorem A.14 *(Theorem 1.6 from text) For i, $1 \leq i \leq N$, if $IN_i \geq 0$, then $\Delta_j B_k = 0$, $\Delta_j B_{k,k-1} = 0$, $\Delta_j B_{k,k} = 0$, for $k \leq i$, and $j \geq k$.*

Proof

$$
\begin{aligned}
IN_i \geq 0 \quad &\Longrightarrow IN_k = s_k \geq 0 \text{ for } k = 1, \cdots, i-1 \text{ by Theorem A.4} \\
&\Longrightarrow IN_k \geq 0 \text{ for } k = 1, \cdots, i \\
&\Longrightarrow B_k = 0 \text{ for } k = 1, \cdots, i \text{ since } B_k = [-IN_k]^+.
\end{aligned}
$$

$\Longrightarrow \Delta_j B_k = 0$, $j = k, \cdots, N$ since $\Delta_j IN_k \geq 0$ for all j, k by Theorem A.13.

$\Longrightarrow \Delta_j B_{k,k-1} = 0$ since $B_k = 0 \Longrightarrow B_{k,k-1} = 0$ for all k by Observation A.2.

$\Longrightarrow \Delta_j B_{k,k} = 0$ because $B_{k,k} = B_k - B_{k,k-1}$ for all k. ■

THEOREM A.15 *(Theorem 1.7 from text) For some i, $2 \leq i \leq N$, and any $j \geq i$, if $\Delta_j IN_i = 0$, then $\Delta_j IN_k = 0$, for $k < i$.*

PROOF

Suppose for some i, $2 \leq i \leq N$, and any $j \geq i$, that $\Delta_j IN_i = 0$. This means that changing s_j has no effect on IN_i, therefore it has no effect on B_i because $B_i = [-IN_i]^+$. The structure of the problem requires that since B_i hasn't changed, neither does $B_{i,i-1}$ or $B_{i,i}$. This implies that $\Delta_j B_i = \Delta_j B_{i,i-1} = \Delta_j B_{i,i} = 0$. Finally, since $IN_{i-1} = s_{i-1} - B_{i,i-1}$, then $\Delta_j IN_{i-1} = 0$.

The same argument follows with $\Delta_j IN_{i-1} = 0 \Longrightarrow \Delta_j IN_{i-2} = 0$, continuing until $\Delta_j IN_1 = 0$. At this last iteration, we also know that $\Delta_j B_i = \Delta_j B_{i,i} = 0$ using the same argument as above. ■

A.4 Expanding the Cost Objective for Computation

Basic form of the expected total cost objective is given by Equation 1.6 from the text, and the derivation to a general expanded form of the equation follows.

$$\bar{C}(s_1, s_2, ..., s_N) = E\left[\sum_{j=1}^{N} (hI_j + b_j B_{j,j})\right] \tag{A.1}$$

From the basic form of the cost objective in 1, we expand first by bringing the expectation through the expression and then substitute using the definition of external backorders $B_{j,j} = B_j - B_{j,j-1}$.

$$\bar{C}(s_1, s_2, ..., s_N) = \sum_{j=1}^{N} (hE[I_j] + b_j E[B_j - B_{j,j-1}]) \tag{A.2}$$

Again, we bring the expectation through. We use the definitions of $I_j = (IN_j)^+$ and

$B_j = (-IN_j)^+$ for substitution and we get:

$$\bar{C}(s_1, s_2, ..., s_N) = \sum_{j=1}^{N} \left(h \cdot E\left[(IN_j)^+\right] + b_j \cdot E\left[(-IN_j)^+\right] - b_j \cdot E\left[B_{j,j-1}\right]\right) \tag{A.3}$$

From 3 we pull out the last term of the summation, as the definitions for net inventory are different for class N than for higher priority classes. The definitions are $(-IN_j)^+ = (B_{j+1,j} - s_j)^+$ and $(IN_j)^+ = (s_j - B_{j+1,j})^+$ for $j = 1, 2, \cdots, N-1$. For class N we have $(-IN_N)^+ = (D^L - (s_N + 1))^+$ and $(IN_N)^+ = ((s_N + 1) - D^L)^+$. The substitution gives the following:

$$\begin{aligned}\bar{C}(s_1, s_2, ..., s_N) = &\sum_{j=1}^{N-1} \left(h \cdot E\left[(s_j - B_{j+1,j})^+\right] + b_j \cdot E\left[(B_{j+1,j} - s_j)^+\right] - b_j \cdot E\left[B_{j,j-1}\right]\right) \\ &+ h \cdot E\left[\left((s_N + 1) - D^L\right)^+\right] + b_N \cdot E\left[\left(D^L - (s_N + 1)\right)^+\right] - b_N \cdot E\left[B_{N,N-1}\right]\end{aligned} \tag{A.4}$$

In this next step, we use the definition $y - X = (y - X)^+ - (X - y)^+$ which gives the general substitution $(y - X)^+ = (y - X) + (X - y)^+$. After the substitution is made, we combine terms and get the following:

$$\begin{aligned}\bar{C}(s_1, s_2, ..., s_N) = &\sum_{j=1}^{N-1} \left(h \cdot E\left[(s_j - B_{j+1,j})\right] + (h + b_j)E\left[(B_{j+1,j} - s_j)^+\right] - b_j \cdot E\left[B_{j,j-1}\right]\right) \\ &+ h \cdot E\left[((s_N + 1) - D^L)\right] + (h + b_N)E\left[\left(D^L - (s_N + 1)\right)^+\right] - b_N \cdot E\left[B_{N,N-1}\right]\end{aligned} \tag{A.5}$$

Here we carry the expectation through where possible, and we use the definition of first-order loss functions to replace $E\left[(X - y)^+\right]$ with $G_X^1(y)$. Here, X is distributed according to lead time demand, $X \sim D^L$, or according to internal backorders, $X \sim B_{i,i-1}$, $i = 2, \dots, N$.

$$\bar{C}(s_1, s_2, ..., s_N) = \sum_{j=1}^{N-1} \Big(h \cdot s_j - h \cdot E\left[B_{j+1,j}\right] + (h + b_j) G^1_{B_{j+1,j}}(s_j) - b_j \cdot E\left[B_{j,j-1}\right] \Big)$$
$$+ h \cdot (s_N + 1) - h \cdot E\left[D^L\right] + (h + b_N) G^1_{D^L}(s_N + 1) - b_N \cdot E\left[B_{N,N-1}\right] \quad \text{(A.6)}$$

We pull out the first and last terms in the remaining summation, as class 1 and class $N - 1$ are special cases of the summation. For class 1, the highest priority class, the term $-b_1 \cdot E\left[B_{1,0}\right] = 0$ so we remove it. For class $N - 1$, the eventual substitution for $E\left[B_{N,N-1}\right]$ is different than for classes $1, 2, \cdots, N - 2$.

$$\bar{C}(s_1, s_2, ..., s_N) = h \cdot s_1 - h \cdot E\left[B_{2,1}\right] + (h + b_1) G^1_{B_{2,1}}(s_1)$$
$$+ \sum_{j=2}^{N-2} \Big(h \cdot s_j - h \cdot E\left[B_{j+1,j}\right] + (h + b_j) G^1_{B_{j+1,j}}(s_j) - b_j \cdot E\left[B_{j,j-1}\right] \Big)$$
$$+ h \cdot s_{N-1} - h \cdot E\left[B_{N,N-1}\right] + (h + b_{N-1}) G^1_{B_{N,N-1}}(s_{N-1}) - b_{N-1} \cdot E\left[B_{N-1,N-2}\right]$$
$$+ h \cdot (s_N + 1) - h \cdot E\left[D^L\right] + (h + b_N) G^1_{D^L}(s_N + 1) - b_N \cdot E\left[B_{N,N-1}\right] \quad \text{(A.7)}$$

In equation 7 we have the expectation $E\left[B_{j+1,j}\right]$. We use the general definition of expectations to begin our transformation:

$$E\left[B_{j+1,j}\right] = E\left[E\left[B_{j+1,j} | B_{j+1}\right]\right]$$

The distribution of the conditional probability is binomial, given by

$$Pr\left[B_{j+1,j} = j | B_{j+1} = n\right] = \begin{pmatrix} n \\ j \end{pmatrix} p^j_{j+1} (1 - p_{j+1})^{n-j}$$

and has the expected value of $p_{j+1} \cdot n$. This gives

$$E\left[E\left[B_{j+1,j} | B_{j+1}\right]\right] = E\left[p_{j+1} \cdot B_{j+1}\right] = p_{j+1} E\left[B_{j+1}\right]$$

The expected total backorders is given by

$$E\left[B_{j+1}\right] = E\left[(-IN_{j+1})^{+}\right] = E\left[(B_{j+2,j+1} - s_{j+1})^{+}\right] = G^{1}_{B_{j+2,j+1}}(s_{j+1})$$

So the substitution will be $E\left[B_{j+1,j}\right] = p_{j+1} \cdot G^{1}_{B_{j+2,j+1}}(s_{j+1})$, for $j = 1, 2, \cdots, N-2$. For class $N-1$, the logic is similar but the substitution is $E\left[B_{N,N-1}\right] = p_N \cdot G^{1}_{D^L}(s_N + 1)$.

Also, $E\left[D^L\right] = \lambda L$, from the definition of the Poisson distribution which we have assumed.

Finally, we have the full expansion of the cost objective for N demand classes, given by:

$$\begin{aligned}\bar{C}(s_1, s_2, ..., s_N) &= h \cdot s_1 - h \cdot p_2 \cdot G^{1}_{B_{3,2}}(s_2) + (h + b_1)G^{1}_{B_{2,1}}(s_1) \\ &+ \sum_{j=2}^{N-2} \Big(h \cdot s_j - h \cdot p_{j+1} \cdot G^{1}_{B_{j+2,j+1}}(s_{j+1}) + (h + b_j - b_j p_j)G^{1}_{B_{j+1,j}}(s_j)\Big) \\ &+ h \cdot s_{N-1} - h \cdot p_N \cdot G^{1}_{D^L}(s_N + 1) + (h + b_{N-1} - b_{N-1}p_{N-1})G^{1}_{B_{N,N-1}}(s_{N-1}) \\ &+ h \cdot (s_N + 1) - h \cdot \lambda L + (h + b_N - b_N p_N)G^{1}_{D^L}(s_N + 1) \end{aligned} \tag{A.8}$$

A.5 Conditional Expectation Variance Reduction

This is a short note to show how to estimate the expected mean total cost to a given accuracy with the fewest number of observations (see both Karian and Dudewicz (1991) and Minh (1989)). Let $A \subseteq \{0, 1, \cdots\}$ be an arbitrary subset of the sample space of D^L. Let C be a random variable equal to the total cost incurred corresponding to a sampled value of D^L and the related internal and external back orders, given a set of fixed values for the base stock levels. By C_A denote the random variable conditional on $D^L \in A$, that is C_A is generated as follows: (1) sample D^L in A by choosing $D^L = a$ with probability $\Pr(a)/\Pr(A)$ (2) determine what the related internal and external back orders are (3) compute the total cost. Suppose $A = \{a\}$ and define $\overline{C}_a = E[C_A]$ and $\sigma_a^2 = V[C_A]$. We now show how to calculate the mean and variance of C_A in terms of these fundamental means and variances.

$$
\begin{aligned}
\overline{C}_A &= \sum_{a\in A} E[C|D^L = a]\frac{\Pr(a)}{\Pr(A)} \\
&= \sum_{a\in A} \overline{C}_a \frac{\Pr(a)}{\Pr(A)} \qquad (1) \\
V[C_A] &= E[(C_A - \overline{C}_A)^2] \\
&= \sum_{a\in A} E[(C_A - \overline{C}_A)^2|D^L = a]\frac{\Pr(a)}{\Pr(A)} \\
&= \sum_{a\in A} E[(C_a - \overline{C}_a + \overline{C}_a - \overline{C}_A)^2]\frac{\Pr(a)}{\Pr(A)} \\
&= \sum_{a\in A} [\sigma_a^2 + (\overline{C}_a - \overline{C}_A)^2]\frac{\Pr(a)}{\Pr(A)} \\
&= \sum_{a\in A} [\sigma_a^2 + \overline{C}_a^2 - 2\overline{C}_a\overline{C}_A + \overline{C}_A^2]\frac{\Pr(a)}{\Pr(A)} \\
&= \sum_{a\in A} [\sigma_a^2 + \overline{C}_a^2]\frac{\Pr(a)}{\Pr(A)} - \overline{C}_A^2 \qquad (2)
\end{aligned}
$$

Next we turn to how to estimate the mean total cost. We consider a partition of the sample space $\bigcup_{i=0}^m B_i = \{0, 1, \cdots\}$ where $B_i \cap B_j = \emptyset$ for $i \neq j$. Initially, we'll assume that $s_N + 1 \geq 0$ and we'll take $B_i = \{i\}$ for $i = 0, \cdots, s_N + 1$. Then

$$
\begin{aligned}
E[C] &= \sum_{i=0}^{m} E[C|B_i]\Pr(B_i) \\
&= \sum_{i=0}^{s_N+1} (h(s_N + 1 - i) + hs_{N-1} + \cdots + hs_1)\Pr(B_i) + \sum_{i=s_N+2}^{m} \overline{C}_{B_i}\Pr(B_i)
\end{aligned}
$$

Noting that $R = 1 + s_N + \cdots + s_1$, this simplifies to

$$
E[C] = hR\Pr(D^L \leq s_N + 1) - h\sum_{i=1}^{s_N+1} i\Pr(D^L = i) + \sum_{i=s_N+2}^{m} \overline{C}_{B_i}\Pr(B_i) \qquad (3)
$$

Since it appears difficult to calculate $\overline{C}_{B_i}$ directly, we'll replace it by a sample average and thus come up with an estimator of $E[C]$. Let n_i be the number of observations of $\overline{C}_{B_i}$ and let $C_{B_i}^j$

be the jth such observation. Then our estimator is

$$hR\Pr(D^L \leq s_N + 1) - h\sum_{i=1}^{s_N+1} i\Pr(D^L = i) + \sum_{i=s_N+2}^{m} \Pr(B_i)\sum_{j=1}^{n_i} \frac{C_{B_i}^j}{n_i} \tag{4}$$

The variance of this estimator is

$$\begin{aligned} V\left[\sum_{i=s_N+2}^{m} \Pr(B_i)\sum_{j=1}^{n_i} \frac{C_{B_i}^j}{n_i}\right] &= \sum_{i=s_N+2}^{m} \Pr(B_i)^2 V\left[\sum_{j=1}^{n_i} \frac{C_{B_i}^j}{n_i}\right] \\ &= \sum_{i=s_N+2}^{m} \frac{\Pr(B_i)^2}{n_i} V[C_{B_i}] \end{aligned} \tag{5}$$

The question now arises as to how best to partition the set $\{s_N + 1, s_N + 2, \cdots\}$. Let B represent a generic member of a partition of this set and suppose we split B into B^+ and B^- where $B^+ \cup B^- = B$ and $B^+ \cap B^- = \emptyset$. Let n be the number of times we sample C_B and suppose we sample C_{B^+} and C_{B^-} n^+ and n^- times, respectively, where $n^+ + n^- = n$. We show below that under certain conditions the contribution of the two split sets to the variance is less than the contribution of the original set, i.e., that

$$\frac{\Pr(B)^2}{n} V[C_B] - \frac{\Pr(B^+)^2}{n^+} V[C_{B^+}] - \frac{\Pr(B^-)^2}{n^-} V[C_{B^-}] > 0 \tag{6}$$

First, let us choose n^+ and n^- so that the sum of the two variance contributions is minimized. Formally, solve the following nonlinear program:

$$\begin{aligned} \min \quad & \tfrac{\Pr(B^+)^2}{n^+} V[C_{B^+}] + \tfrac{\Pr(B^-)^2}{n^-} V[C_{B^-}] \\ s.t. \quad & n^+ + n^- = n \end{aligned}$$

We'll treat the variables as continuous so that we can use the method of Lagrange multipliers. The Langrangian is

$$\mathcal{L} = \frac{\Pr(B^+)^2}{n^+} V[C_{B^+}] + \frac{\Pr(B^-)^2}{n^-} V[C_{B^-}] - \lambda(n - n^+ - n^-)$$

and so

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial n^+} &= -\frac{\Pr(B^+)^2}{{n^+}^2} V[C_{B^+}] + \lambda = 0 \\
\frac{\partial \mathcal{L}}{\partial n^-} &= -\frac{\Pr(B^+)^2}{{n^-}^2} V[C_{B^+ -}] + \lambda = 0 \\
n^+ &= \frac{1}{\sqrt{\lambda}} \Pr(B^+) \sqrt{V[C_{B^+}]} \\
n^- &= \frac{1}{\sqrt{\lambda}} \Pr(B^-) \sqrt{V[C_{B^-}]}
\end{aligned}
$$

Now use the condition $n^+ + n^- = n$ to solve for $1/\sqrt{\lambda}$:

$$
\frac{1}{\sqrt{\lambda}} = \frac{n}{\Pr(B^+)\sqrt{V[C_{B^+}]} + \Pr(B^-)\sqrt{V[C_{B^-}]}}
$$

and so

$$
\begin{aligned}
n^+ &= \frac{\Pr(B^+)\sqrt{V[C_{B^+}]}}{\Pr(B^+)\sqrt{V[C_{B^+}]} + \Pr(B^-)\sqrt{V[C_{B^-}]}} n \\
n^- &= \frac{\Pr(B^-)\sqrt{V[C_{B^-}]}}{\Pr(B^+)\sqrt{V[C_{B^+}]} + \Pr(B^-)\sqrt{V[C_{B^-}]}} n
\end{aligned}
$$

The value of the objective function at this solution is

$$
\frac{(\Pr(B^+)\sqrt{V[C_{B^+}]} + \Pr(B^-)\sqrt{V[C_{B^-}]})^2}{n}
$$

Equation (6) is now equivalent to

$$
\frac{\Pr(B)^2 V[C_B] - (\Pr(B^+)\sqrt{V[C_{B^+}]} + \Pr(B^-)\sqrt{V[C_{B^-}]})^2}{n} > 0 \tag{7}
$$

We now to proceed to show that the numerator of this fraction is positive. First note that

$$\begin{aligned}
\overline{C}_B &= \sum_{b\in B} \overline{C}_b \frac{\Pr(b)}{\Pr(B)} \\
&= \frac{\Pr(B^+)}{\Pr(B)} \sum_{b\in B^+} \overline{C}_b \frac{\Pr(b)}{\Pr(B^+)} + \frac{\Pr(B^-)}{\Pr(B)} \sum_{b\in B^-} \overline{C}_b \frac{\Pr(b)}{\Pr(B^-)} \\
\Pr(B)\overline{C}_B &= \Pr(B^+)\overline{C}_{B^+} + \Pr(B^-)\overline{C}_{B^-} \qquad (8)
\end{aligned}$$

Now using equation (2) we find

$$\begin{aligned}
\Pr(B)^2 V[C_B] &= \Pr(B) \left[\sum_{b\in B} [\sigma_b^2 + (\overline{C}_b - \overline{C}_B)^2] \Pr(b) \right] \\
&= \Pr(B) \sum_{b\in B^+} [\sigma_b^2 + (\overline{C}_b - \overline{C}_{B^+} + \overline{C}_{B^+} - \overline{C}_B)^2] \Pr(b) \\
&+ \Pr(B) \sum_{b\in B^-} [\sigma_b^2 + (\overline{C}_b - \overline{C}_{B^-} + \overline{C}_{B^-} - \overline{C}_B)^2] \Pr(b)
\end{aligned}$$

Simplifying a subexpression we find

$$\begin{aligned}
&\sum_{b\in B^+} (\overline{C}_b - \overline{C}_{B^+} + \overline{C}_{B^+} - \overline{C}_B)^2 \Pr(b) = \\
&\qquad \Pr(B^+) \sum_{b\in B^+} [(\overline{C}_b - \overline{C}_{B^+})^2 + 2(\overline{C}_b - \overline{C}_{B^+})(\overline{C}_{B^+} - \overline{C}_B) + (\overline{C}_{B^+} - \overline{C}_B)^2] \frac{\Pr(b)}{\Pr(B^+)} \\
&\qquad = \sum_{b\in B^+} (\overline{C}_b - \overline{C}_{B^+})^2 \Pr(b) + \Pr(B^+)(\overline{C}_{B^+} - \overline{C}_B)^2
\end{aligned}$$

And so

$$\begin{aligned}
\Pr(B)^2 V[C_B] &= \Pr(B) \left[\sum_{b\in B^+} [\sigma_b^2 + (\overline{C}_b - \overline{C}_{B^+})^2] \Pr(b) + \Pr(B^+)(\overline{C}_{B^+} - \overline{C}_B)^2 \right] \\
&+ \Pr(B) \left[\sum_{b\in B^-} [\sigma_b^2 + (\overline{C}_b - \overline{C}_{B^-})^2] \Pr(b) + \Pr(B^-)(\overline{C}_{B^-} - \overline{C}_B)^2 \right] \\
&= \Pr(B) \left[\Pr(B^+) V[C_{B^+}] + \Pr(B^+)(\overline{C}_{B^+} - \overline{C}_B)^2 \right] \\
&+ \Pr(B) \left[\Pr(B^-) V[C_{B^-}] + \Pr(B^-)(\overline{C}_{B^-} - \overline{C}_B)^2 \right]
\end{aligned}$$

The numerator of the LHS of equation (7) becomes

$$
\begin{aligned}
&\Pr(B^+)^2[V[C_{B^+}] + (\overline{C}_{B^+} - \overline{C}_B)^2] + \Pr(B^-)\Pr(B^+)[V[C_{B^+}] + (\overline{C}_{B^+} - \overline{C}_B)^2] \\
&\quad + \Pr(B^-)^2[V[C_{B^-}] + (\overline{C}_{B^-} - \overline{C}_B)^2] + \Pr(B^-)\Pr(B^+)[V[C_{B^-}] + (\overline{C}_{B^-} - \overline{C}_B)^2] \\
&\quad - \Pr(B^+)^2 V[C_{B^+}] - 2\Pr(B^+)\Pr(B^-)\sqrt{V[C_{B^+}]}\sqrt{V[C_{B^-}]} - \Pr(B^-)^2 V[C_{B^-}] \\
&= \Pr(B^-)\Pr(B^+)(\sqrt{V[C_{B^+}]} - \sqrt{V[C_{B^-}]})^2 + \Pr(B)[\Pr(B^+)(\overline{C}_{B^+} - \overline{C}_B)^2 + \Pr(B^-)(\overline{C}_{B^-} - \overline{C}_B)^2]
\end{aligned}
$$

which is strictly positive if either $V[C_{B^+}] \neq V[C_{B^-}]$ or $\overline{C}_{B^+} \neq \overline{C}_{B^-}$.

All this suggests the following strategy for choosing a partition: given any partition, keep splitting until every B has only one element. Of course this would lead to a partition with an infinite number of elements. Essentially we would be trying to estimate the following infinite sum by simulation

$$E[C] = \sum_{i=0}^{\infty} \overline{C}_i \Pr(i)$$

A practical way to deal with this is to approximate the infinite sum with the following partial sum $\sum_{i=0}^{I} \overline{C}_i \Pr(i)$, but we need a way to estimate the truncation error. We can do this by adapting the proof of the ratio test for convergent series. First, we argue that for sufficiently large i we have $\overline{C}_i = ai + b$ where a is a positive constant. This is because when $i = D^L$ is very large, then the only significant costs are the back order cost and the expected total back order cost is

$$
\begin{aligned}
&(1 - p_N)(i - s_N)b_N + (1 - p_{N-1})(p_N(i - s_N) - s_{N-1})b_{N-1} \\
&\qquad + (1 - p_{N-2})(p_{N-1}(p_N(i - s_N) - S_{N-1}) - s_{N-2})b_{N-2} + \cdots
\end{aligned}
$$

which clearly has the linear form postulated. Define now $a_i = (ai + b)\Pr(i) = (ai + b)e^{-\mu}\mu^i/i!$

and let $b_i = a_{i+1}/a_i$. Note that

$$E[C] = \sum_{i=0}^{I} a_i + \sum_{i=I+1}^{\infty} a_i$$

and so we'd like to get an error bound on the last term. As in the proof of the ratio test, we first compute the following limit:

$$\begin{aligned}
\lim_{i\to\infty} b_i &= \lim_{i\to\infty} a_{i+1}/a_i \\
&= \lim_{i\to\infty} \frac{[a(i+1)+b]e^{-\mu}\mu^{i+1}/(i+1)!}{[ai+b]e^{-\mu}\mu^i/i!} \\
&= \lim_{i\to\infty} \frac{a(i+1)+b}{ai+b}\frac{\mu}{i+1} \\
&= 0
\end{aligned}$$

Now choose a value for r, $0 < r < 1$ and then choose I so large that $b_i < r$ for $i > I$. Then we can write

$$\begin{aligned}
\sum_{i=I+1}^{\infty} a_i &= a_{I+1} + a_{I+2} + \cdots \\
&= a_{I+1}\left(1 + \frac{a_{I+2}}{a_{I+1}} + \frac{a_{I+2}}{a_{I+1}}\frac{a_{I+3}}{a_{I+2}} + \cdots\right) \\
&= a_{I+1}(1 + b_{I+1} + b_{I+1}b_{I+2} + \cdots) \\
&< a_{I+1}(1 + r + r^2 + \cdots) \\
&= a_{I+1}/(1-r)
\end{aligned}$$

If we take $r = 1/2$, then we get

$$\begin{aligned}
\sum_{i=I+1}^{\infty} a_i &< 2a_{I+1} \\
\sum_{i=I+2}^{\infty} a_i &< a_{I+1}
\end{aligned}$$

This suggests a way to determine I. Suppose we need to estimate $E[C]$ to $\pm\epsilon$. Then sequentially compute a_i to find the smallest integer k such that $a_k < a_{k-1}/2$ and $a_{k-1} < \epsilon$. If we take $I = k - 2$ and our estimate of the cost as $\sum_{i=0}^{k} a_i$ then we have

$$\begin{aligned} E[C] - \sum_{i=0}^{k} a_i &= \sum_{i=k+1}^{\infty} a_i \\ &< \sum_{i=k}^{\infty} a_i \\ &< a_{k-1} \\ &< \epsilon \end{aligned}$$

Of course this assumes there is no estimation error in computing the a_i, so we deal with that aspect of the problem next. Let's suppose that in determining k above, we initially sample n_0 times to estimate each a_i. Based on these n_0 samples let S_i denote the sample standard deviation of C_i. We then can compute an estimate of the variance of our estimator:

$$\sum_{i=s_N+2}^{k} \frac{S_i^2 \Pr(i)^2}{n_i}$$

where $n_i > n_0$ is the total number of times we eventually sample C_i. As before, to minimize this variance we should choose

$$n_i = \frac{S_i \Pr(i)}{\sum_{j=s_N+2}^{k} S_j \Pr(j)} n$$

where n is the total number of samples taken. Thus an estimate of the variance of the estimator when the sample sizes are optimally chosen is

$$\frac{\left(\sum_{j=s_N+2}^{k} S_j \Pr(j)\right)^2}{n}$$

Somewhat heuristically, if we want to estimate $E[C]$ to within $\pm d$ with confidence $100P^*\%$ we can try setting

$$n = z^2_{(1+P^*)/2} \left(\sum_{j=s_N+2}^{k} S_j \Pr(j) \right)^2 / d^2$$

where $z_{(1+P^*)/2}$ is the $(1 + P^*)/2$ fractional point of the unit normal distribution. If we let

$$f_i = \frac{S_i \Pr(i)}{\sum_{j=s_N+2}^{k} S_j \Pr(j)}$$

then the simplest and conservative thing to do is to set

$$n_i = \max\{n_0 + 1, \lceil f_i n \rceil\}$$

so that $n_i - n_0$ more observations will need to be taken at $D^L = i$.

Alternatively, a somewhat more rigorous, but complex, procedure could be followed. Let

$$\begin{aligned} i_{\min} &= \arg\min_i \{\Pr(i) S_i\} \\ n_{i_{\min}} &= 1 \\ n_i &= \lceil \Pr(i) S_i / \Pr(i_{\min}) S_{i_{\min}} \rceil, i \neq i_{\min} \end{aligned}$$

take as one replication of the simulation an observation of the following random variable

$$X = hR \Pr(D^L \leq s_N + 1) - h \sum_{i=1}^{s_N+1} i \Pr(i) + \sum_{i=s_N+2}^{k} \Pr(i) \sum_{j=1}^{n_i} C_i^j / n_i$$

Then the standard two stage procedure (Karian & Dudewicz) can be employed: to estimate $E[C]$ to with in $\pm d$ with confidence $100P^*\%$

1. Set n_0 as a positive integer ≥ 2. Set $w = t^{-1}_{n_0-1}((1 + P^*)/2)/d$.

2. Observe $X_1, \cdots, X_{n_0}$.

3. Calculate $\overline{X}(n_0) = \frac{X_1 + \cdots + X_{n_0}}{n_0}$, $S^2 = \sum_{i=1}^{n_0} (X_i - \overline{X}(n_0))^2/(n_0 - 1)$.

4. Set $n = \max\{n_0 + 1, \lceil w^2 S^2 \rceil\}$
5. Observe $X_{n_0+1}, \cdots, X_n$.
6. Calculate $\overline{\overline{X}} = \frac{X_1, \cdots, X_n}{n}$.
7. Claim with $100P^*\%$ confidence that $\overline{\overline{X}} - d \leq E[C] \leq \overline{\overline{X}} + d$.

2: HOSPITAL INVENTORY MODEL WITH INACCURATE RECORDS

Instead of focusing on grouped backorder costs to separate demand distributions as in Article 1, here we consider the case where accuracy shapes customer demand. Also, we shift from a rationing approach to a first-come, first-served demand environment.

We consider a "stockless" hospital supply chain with inaccurate inventory records. Recording demand with perfect, 100% accuracy was shown to be infeasible in practice, so we develop a model that is conditional on the level of accuracy in a particular hospital department, or point-of-use (POU).

Similar to previous literature on inventory inaccuracy, we consider both *actual* net inventory and *recorded* net inventory in developing the system performance measures. The resultant model is a periodic-review, cost minimization inventory model with full backordering that is centered at the POU.

Similar to the previous article, we assume a base stock ordering policy. However, in addition to choosing the optimal order-up-to level, we seek the optimal frequency of inventory counts to reconcile inaccurate records. We also present a version of the model that addresses availability, or the service level, of medical supplies.

2.1 INTRODUCTION

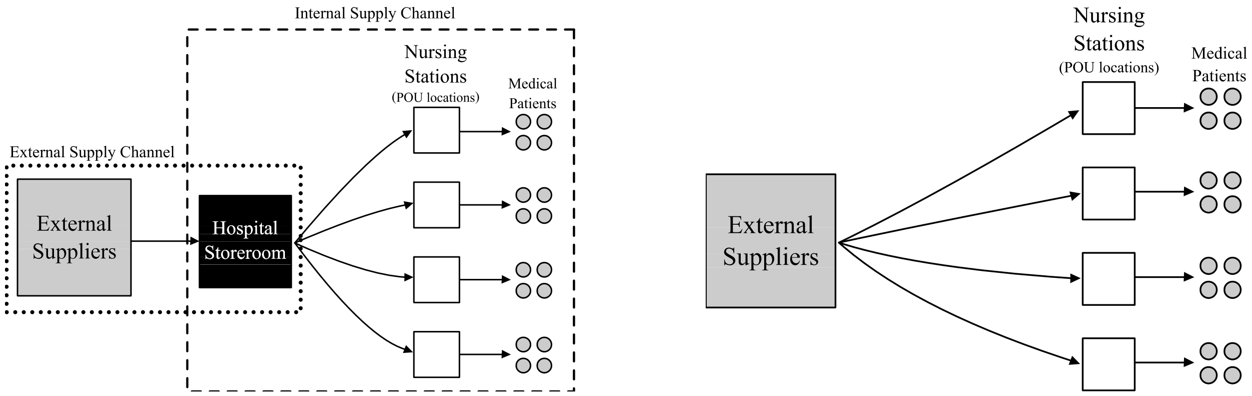
In industry, inaccurate inventory records can lead to increased costs, unexpected delays in demand fulfillment, and lost revenue. Inventory record inaccuracy is also one of the many sources of unecessary healthcare costs across the US. In a 583-bed Alabama hospital, inventory record inaccuracy was found to significantly affect the supply chain, leading to both overstock and daily shortages throughout the facility. The state of their internal supply chain led the Materials Manager to adopt a daily inventory counting policy, requiring additional employees. Inaccurate records failed to account for nearly half a million US dollars per year in consumed medical supplies, which represents a patient reimbursement value of about two million dollars annually.

This hospital's situation was the main motivation behind our research, however, this situation among hospitals is not unique. A representative from an automated inventory systems provider recently confirmed the existence of several southern US hospitals which struggle from similar inaccuracy issues. Other industry professionals in Illinois, Florida, Colorado, and Pennsylvania have reported similar stories.

2.1.1 The Hospital Supply Chain

In the past, most hospitals managed two simultaneous supply chains, where one was internal and the other external (Rivard-Royer et al. (2002)). The internal supply channel represents inventory and logistics from the hospital's on-site warehouse to the several point-of-use (POU) inventory locations within the hospital. The external supply channel represents the stream of inventory from the supplier to the hospital's on-site warehouse as shown in Figure 2.1a. In such cases, inventory control policies are required both for the warehouse and for each of the POU locations.

Rivard-Royer et al. (2002) discuss that, during the 1990s, hospitals in the US and Canada moved away from the classic hospital supply chain approach, and trended toward a stockless management system, as portrayed in Figure 2.1b. The stockless system has been compared to



(a) "Classic" Structure (b) "Stockless" Structure

Figure 2.1: Hospital Supply Chains

a just-in-time inventory system, and neither system was shown to be significantly better than the other (Rossetti (2008)).

The stockless system transforms the external supply channel into a direct replenishment stream from the supplier to the POU inventory locations, essentially removing the need to keep stock in an on-site warehouse. Based on direct demand information from each POU, the supplier ships materials in unit quantities directly to the POU. Stockless systems have been shown to drastically reduce inventories by as much as 70% to 80%, as well as reduce the material handler full time equivalents (FTEs) by as much as 30% to 45% (Rivard-Royer et al. (2002)). This accounts for several hundreds of thousands of dollars in annual savings for the hospitals involved.

However, the stockless system is not without its drawbacks. It requires up-to-date and accurate transmission of supply needs, a short lead time to ensure minimal inventory at each POU (or sufficient inventory capacity at each POU), and an ability to adapt to variation in product mixes and seasonality of demand. In many instances, hospitals are required to make costly systems upgrades or expand POU inventories to qualify for the stockless system. In other cases, hospitals that experience relatively longer lead times from the supplier are required to keep large amounts of inventory on-hand between available shipments. Rural hospitals and not-for-profit hospitals fall into the group of hospitals that may have difficulty in implementing or affording the changeover to a stockless system. (Rivard-Royer et al. (2002))

2.1.2 Inaccurate Inventory

In this paper, we assume the hospital operates under a stockless supply chain structure, as given by Figure 2.1b. As is the case in the Alabama hospital, nurses and staff at the POU inventory locations are responsible for recording (i.e. scanning) supply usage through the use of an electronic inventory tracking system. If the user fails to properly record the consumption of a particular supply item, the *recorded* inventory level within the electronic system remains unchanged while *actual* available inventory decreases. In the literature, this type of error has been called an overstatement error, where the recorded system inventory level is consistently greater than or equal to the actual inventory level.

We define p to represent the probability that a demand event is successfully recorded, and we call this probability the *accuracy level* or the level of recording accuracy. For example, if the accuracy level is 65% (i.e. $p = .65$), then there is a 65% probability that demand is recorded. This means that the recorded net inventory will be greater than actual inventory, since actual inventory will decrease at a faster rate than recorded inventory.

In an attempt to create a stockless hospital supply chain, the Alabama hospital that motivates this research relies on the electronic inventory system mentioned above to automatically place regular orders for replenishment. Thus, when the system is inaccurate, orders are delayed and the quantities ordered may be insufficient, causing backorders. In addition, the patients cannot be charged for the use of that supply item, as there is no record generated on their patient bill. This is how inaccurate inventory records led to failed reimbursement, to ineffective ordering, to stockouts, and sometimes to delay in patient care.

In this paper, we present two versions of an inventory model for a hospital that is attempting to operate according to a *stockless* supply chain, but must deal with inaccurate inventory records (given an accuracy level, p). We assume a periodic-review base stock inventory replenishment policy with full backordering, and we seek the optimal base stock level to minimize costs in the first model. In the second model, we seek the smallest base stock level required to meet a predetermined customer service level.

A review of related literature will be provided in Section 2.2. In Section 2.3 we present the cost-minimization model with its optimal solution. In Section 2.4 we discuss the service-level model with a solution approach. In Section 2.5 we discuss the solution approaches for both models, including heuristics for near-optimal solutions. In Section 2.6 we present the results of seeking optimal and near-optimal solutions for both models. We conclude in Section 2.7 with an Appendix in Section B.

2.2 LITERATURE

Literature relevant to this paper includes research focused on the health care supply chain, as well as inventory modeling where inaccuracy is addressed. So we will first discuss literature focused on the healthcare supply chain, and then discuss literature from the inventory management stream. More specifically, we include those articles focused on inventory models that address inaccurate inventory records.

2.2.1 Healthcare Supply Chain

Young (1989) and McKone-Sweet et al. (2005) address issues in the hospital supply chain from an empirical standpoint by surveying representatives from several hospitals. Young (1989) surveyed 22 Materials Managers regarding computer-based inventory control systems vs. manual systems, while McKone-Sweet et al. (2005) focused on the general obstacles to improving the hospital supply chain by surveying 26 individuals who were either hospital Materials Managers or representatives of supply distributors.

Both empirical studies cite a need for additional research regarding the hospital supply chain, as well as the complexity of such supply chain systems. However, advancement in computer-based control systems have outdated Young (1989), and McKone-Sweet et al. (2005) focus more on general hospital supply chain issues. Here we are concerned with developing a mathematical model that describes a specific hospital's supply chain, though other hospitals are known to share similar situations.

Other authors, such as Duclos (1993), Rivard-Royer et al. (2002), DeScioli (2005), and Jan (2006) provide specific case studies, where a hospital, or hospitals, are measured before and after an implementation of different supply chain modifications.

In Duclos (1993), they use a simulation to study the hospital in question, exploring operating conditions in a supply chain with a structure similar to the one considered here. However, Duclos (1993) is focused primarily on the effects of shock demand on such a system, and later authors (DeScioli (2005)) have argued that the advent of new technology has weakened the findings of

Duclos (1993).

Rivard-Royer et al. (2002) follow the implementation of a hybrid stockless hospital supply chain in a rural Canadian hospital. To implement the hybrid stockless policy, the Canadian hospital selected a portion of stock to be managed according to a stockless policy (replenished directly from the supplier), while the remaining portion was replenished through the warehouse. In a sense, this divided the medical supplies between the two different supply channels, and it required separate inventory policies for each channel.

Both DeScioli (2005) and Jan (2006) focus on the hospital supply chain, and especially on automated inventory control systems. DeScioli (2005) worked with a large medical supply distributor to analyze a new product and vendor-managed inventory (VMI) service offering. Using data from a specific hospital, the author compares the costs of several different types of replenishment policies, and applies Croston's method to forecast demand at the POU. The author argues that an (s, Q) policy provides the most stable replenishment method, and in their words, it is also the "most optimal" policy.

Jan (2006) considers a supply chain structure almost identical to the one considered here, but focuses more on a vendor-managed inventory approach, in conjunction with automated POU inventory systems. They look at a particular hospital's costs before and after installing POU inventory systems, and cite several problems in their systems such as products not being available and increased costs. However, instead of formulating a model and seeking optima as is done here, they compare explicit costs and recommend using an EOQ approach to select control parameters.

Most recently, Rossetti (2008) reported on several aspects of inventory management within the health care supply chain and how it has been addressed in the litereture. Nowhere in the report does the author discuss the effect of inaccuracy on the supply chain, or indicate that other authors have modeled or discused the issue.

In our model of the hospital supply chain, we will consider inventory decisions at a particular POU, where innacuracy plagues the automated replenishment system. Under inaccuracy, we consider a periodic-review, base stock replenishment policy and seek an optimal solution.

Initially, our objective is based on costs of holding and backordering, and then we present a modified version of the model that addresses inventory availability.

2.2.2 Inaccurate Inventory Records

Our research focuses on an inventory model when inventory records are inaccurate. We categorize related literature into four main categories: non-modeling, modeling for replenishment policy improvement only, modeling for inspection policy improvement only, and modeling for both inspection and replenishment improvement.

Non-modeling literature is either empirical in nature, or it does not present a specific model nor its optimization. Both Sheppard and Brown (1993) and Raman et al. (2001) are examples of empirical research focusing on inventory inaccuracy, while Fleisch and Tellkamp (2005) present a simulation study and share results and insights.

Sheppard and Brown (1993) present the first example of empirical research that focused on the causes of inventory record-keeping errors in a manufacturing setting. They found that, contrary to anecdotal speculation from previous literature, the number of transactions was not a significant indicator of the number of record errors. Rather, they found that items with high quantity on-hand or lower-value high-volume items are more prone to record inaccuracy. They also cite the underestimated impact of inventory record inaccuracy, and expect future literature to expand on the topic.

Raman et al. (2001) provide the results of an empirical study centered in the retail industry, and the impacts and prevalence of inventory inaccuracy. They found that profit lost due to inventory inaccuracy amounted to 10% of one company's profits over a certain period, which was roughly the profit from 100 of their stores over the same period.

Fleisch and Tellkamp (2005) conduct a simulation study on the effect of inventory inaccuracy on the performance of a three echelon supply chain. In their model, they consider several factors that may cause inventory inaccuracy, but do not include transaction errors. However, the simulation is not used to study improved management policies, but rather, they analyze the performance of the inaccurate supply chain when inspections occur, and when they do not occur. Then, the authors employ analysis of variance to find which cause of inaccurate inventory has the greatest impact on supply chain performance. This model is very different from our own, as they ignore transaction errors, and they do not study improved inspection policies, nor do they study improved replenishment policies.

Morey (1985) conducted an analysis of a system with inaccuracies similar to Iglehart and Morey (1972) (discussed later), focusing on the situation where recorded inventory is greater than actual inventory (overstatement error). They describe how overstatement errors can cause a system to postpone replenishment, leading to lower service levels. The author develops an approach to estimate service levels, and then analyzes the effects of different managerial solutions (i.e. counting more frequently, increasing buffer stock, etc.). They argue that managers need to have the ability to estimate the impact of record inaccuracy, because using only a classic approach to manage inventory may significantly overestimate service levels, leading to costly stockouts.

Model-oriented research where only the replenishment policy is adjusted in response to inaccurate records includes the paper by Rekik et al. (2008). They develop a newsvendor model for a retail store where they consider misplacement errors as the only source of inventory inaccuracy. The main premise is to compare an optimized replenishment policy where inaccurate inventory is considered, with an optimized replenishment policy where the firm is assumed to have perfect inventory information by using RFID technology. They do not attempt to find an improved inspection policy, but only consider either inaccurate inventory data or perfect inventory data as they solve for the best order quantity.

Model-oriented research with a focus on only improving the inspection policy is given by Sandoh and Shimamoto (2001). The authors consider costs of both counting inventory and investigating the causes of inventory discrepancies. They include a probability that inventory counting leads to the additional costs of investigating significant discrepancies, depending on how frequently inventory is counted. The model seeks to find the periodical counting frequency that minimizes the total costs. Their work focuses on a supermarket setting where the SKUs are divided into brands, or types of product, and then the groups are assigned investigation costs based on group-specific inspection obstacles. They model the probabilities of inventory inaccuracy using a renewal process, where periodic cycle counts completely remove discrepancies and renew the process. They differentiate the objective to find the minimum cost, giving the optimal cycle count frequency between annual accounting audits.

The last category includes articles that present inventory models where both the replenishment and the counting policies are adjusted to address inaccurate records, and it represents research that is most similar to our own.

The first quantitative research regarding inaccurate inventory records was provided by Iglehart and Morey (1972), where they develop an optimization model where the replenishment policy is given. They modify the given replenishment policy by setting a buffer stock level sufficient to cover all probable inventory record errors until the next inventory count. They also set the frequency of inventory counts, as well as the depth, or quality, of counts. Costs are assigned to the different quality levels of counting and to carrying inventory, and they attempt to minimize total cost per unit time. They constrain their model to ensure that stock does not run out between inventory counts.

In their model, Iglehart and Morey (1972) do not allow for backordering, they assume inventory is replenished through a single channel, and they do not optimize the replenishment policy parameters. Also, Iglehart and Morey (1972) assume the magnitude of record errors is not dependent on the magnitude of demand. In our paper, errors are handled using a probability that particular demand event is recorded, though when physical stock is not available all demand is backordered without error. Also, we assume full backordering, and we seek an optimal base stock level rather than a buffer stock quantity under a given replenishment policy.

Kök and Shang (2007) develop a model for inaccurate inventory where they make joint inspection and replenishment decisions to minimize system costs. The inspection decision is made using a threshold level, so that when the inventory record falls to that level it triggers an inventory inspection that perfectly aligns the inventory record with on-hand inventory. A base-stock replenishment policy is shown to be optimal, and both the threshold level and the base-stock level are chosen based on the level of inaccuracy in the system. They first develop a single-period solution which they call an inspection-adjusted base-stock policy (IABS), and then use a dynamic programming formulation for the finite-period solution, termed the cycle-count adjusted base-stock policy (CCABS). They use a numerical example to compare the performance of several different situations, including the loss associated with not doing anything to circumvent the inaccuracy which was around a 12% loss in their example.

In our paper we also assume a base stock replenishment policy, but we choose a static counting frequency before the planning period begins, which is different than in Kök and Shang (2007) who utilize a dynamic inspection rule. Furthermore, our research includes an additional model where we seek to optimize based on the service-level requirements rather than on total cost.

DeHoratius et al. (2008) develop a model based in a retail setting, where they propose a modified inventory record using a Bayesian approach. Their research is different from our own, as they assume inventory inaccuracy with lost sales (instead of backordering) as they attempt to optimize the auditing frequency and replenishment policy. They provide methods by which retailers can estimate the model parameters and more easily implement the solution approach outlined in the paper. However, they do not optimize model parameters, but instead propose heuristic methods for improving the system and use a simulation study to analyze the quality of the heuristic solutions.

After reviewing the literature, we have found our research to be novel. We are the first to propose an inventory model for a hospital supply chain that addresses the inaccuracy that is prevalent in the healthcare industry. Among the inaccuracy-related inventory literature, we are the first to seek improvements to both replenishment and inspection policies where 1) the random demand is assumed to be discrete, 2) transaction errors are assumed to generate only positive (overstatement) error in recorded inventory, 3) both a cost-minimization model and a service-level model are presented, and 4) the situation is based on the healthcare industry.

2.3 COST MODEL

In this section, we introduce the inventory model with a cost objective which we will later seek to minimize. First, we discuss the physics and structure of the supply chain while introducing some of the notation and assumptions. Then we present the formulation with the cost objective we later minimize.

2.3.1 Model Structure

As stated previously, the following model is based on the inventory operations of the Alabama hospital that motivated this work. We develop the periodic-review inventory model for a single item at a single POU location, where replenishment comes directly from an external supplier. Thus, the replenishment of POU inventory bypasses the hospital's main storeroom.

To model the described inventory system, we assume that time is divided into daily periods, where each period, t, represents a 24-hour time interval. Orders are assumed to arrive 24 hours after being placed, so lead time, L, is one period, $L = 1$. At the beginning of each period, the POU receives the incoming order from the supplier. Immediately following order receipt, an order is placed with the external supplier through an automated system. Therefore, both receiving and ordering occur at the beginning of every period and in that sequence, respectively. This is basically what occurs in practice, but the events do not occur instantaneously, as is assumed here.

The automated replenishment system is responsible for electronically-generating orders each period with the external supplier according to a base stock policy with order-up-to level S. (Hospital staff and administrators call this order-up-to level the *par level.*) The actual order quantity is based on the difference between the current *recorded* inventory level, I', and S. For explanatory purposes, we define $\bar{I}'_i$ to be the recorded inventory at the end of period i. We also define $\underline{I}'_i$ to be the recorded inventory at the beginning of period i, *after* the outstanding order arrives. That means that when the system places an order at the beginning of period i, it will place an order of size $S - \underline{I}'_i$. This order will arrive at the beginning of the following period.

During a particular period(day), nurses and nursing staff personally remove inventory from the POU to use for patient care. Upon removing the inventory, the person taking the supplies is supposed to manually record which item is taken, and which patient will be billed for the item. There are several different automated POU inventory systems on the market, each with its own method for recording usage - some easier than others. At this hospital, the recording process takes a matter of seconds, but it requires the use of a "data wand" that is used to touch preprogrammed e-tags on the patient charge board and on the inventory supply bins. We have already defined p to be the probability that this usage-recording process occurs for a particular demand instance.

So, during each period, i, there is a portion of demand that is successfully recorded, which we define as the *recorded* demand in period i, RD_i. It is crucial to note, that the recorded inventory, $I^{'}$, is depleted only by recorded demand. Note the following figure:

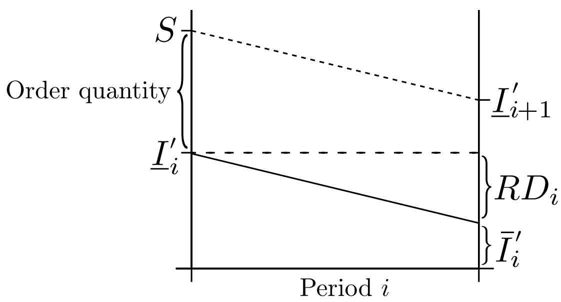


Figure 2.2: Recorded inventory over period i

Figure 2.2 demonstrates the following relationships, which are given for explanatory purposes:

$$\bar{I}'_i = \underline{I}'_i - RD_i \tag{2.1}$$

$$\underline{\mathrm{I}}'_{i+1} = S - RD_i \tag{2.2}$$

$$\underline{I}'_{i+1} = \bar{I}'_i + (S - \underline{I}'_i) = \bar{I}'_i + RD_{i-1} \tag{2.3}$$

Equation 2.1 shows how recorded inventory at the end of period i is affected by recorded demand. It is simply the recorded inventory at the beginning of period i, less the demand recorded during period i.

Equation 2.2 gives the relationship of the period's beginning recorded inventory with the prior period's recorded demand. The recorded inventory position is brought to S at the beginning of the period (after order is received), then recorded demand in i brings the recorded inventory level to $S - RD_i$ by the end of the period.

Equation 2.3 shows how the recorded inventory at the beginning of a period is affected by recorded demand two periods ago. This relationship follows from Equation 2.2 and Figure 2.2.

However, when inventory is removed, but its use is not recorded, the automated replenishment system has no way of knowing where it went, or if it was even removed. Due to this transaction-error inaccuracy, the recorded inventory level will differ from what is physically available or officially on backorder. This is why we define the *actual* net inventory, I. Here, we define actual net inventory at the end of a period, i, given by, I_i. Actual net inventory at the end of period i reflects the depletion of stock by both recorded demand (RD_i) and unrecorded demand (UD_i). This total demand in a period is given by:

$$D_i = RD_i + UD_i \tag{2.4}$$

We assume that total demand, D_i, occurs according to a Poisson distribution, with rate λ, and we call this day i *actual* demand, denoted D_i ~Poisson(λ). Since demand is recorded with probability p, we can say that RD_i ~Poisson($p\lambda$) and $UD_i \sim$ Poisson($(1-p)\lambda$) if demand may be recorded when no stock is on hand.

However, the above distributions are approximations of the actual distributions of RD_i and UD_i since demand is not recorded when there is no physical inventory on hand. In fact, the automated inventory system is not used to manage backorders in practice, and so the recorded inventory will never reflect backorders. However, backorders are real, so we must track actual net inventory. We provide a justification for our approximation of the distributions of RD_i and UD_i in the appendix.

Now that we have defined how demand occurs, we briefly mention the handling of incoming orders. We assume that the only inflow of stock to the POU is the incoming replenishment order at the beginning of the period. We assume that there are no errors in the receipt of incoming orders, and so both recorded inventory and actual inventory reflect identically-sized incoming order quantities each period. Therefore, transaction error (failure to record usage) is the only source of disparity between recorded inventory at the end of period i, $\bar{I}_i^{'}$, and actual net inventory at the end of period i, I_i. The following equations help to define the general relationship between actual net inventory at the end of period i, and recorded inventory.

$$I_i = I_{i-1} + (S - \underline{I}_{i-1}^{'}) - D_i = I_{i-1} + RD_{i-2} - D_i \tag{2.5}$$

$$\bar{I}_i^{'} = \bar{I}_{i-1}^{'} + (S - \underline{I}_{i-1}^{'}) - RD_i = \bar{I}_{i-1}^{'} + RD_{i-2} - RD_i \tag{2.6}$$

In Equation 2.5, we are showing actual net inventory at the end of some period i. Assuming that we know the actual net inventory at the end of period $i-1$, we can compute actual net inventory at the end of period i. Remembering the timing of events (as shown in Figure 2.3 and as discussed above), we know that the automated system will place an order of size $(S - \underline{I}_{i-1}^{'}) = RD_{i-2}$ at the beginning of period $i-1$, and it will arrive at the beginning of period i. So, the actual net inventory will be increased by this amount exactly, since we have assumed that orders are received without error. And finally, by the end of period i, total demand will have depleted actual net inventory by D_i.

Equation 2.6 is simply a rewrite of Equation 2.1, using the recorded inventory at the end of period $i-1$ as the reference point. Again, the arriving order is RD_{i-2}. However, in the case of recorded inventory, it is reduced by the recorded demand, $RD_i \leq D_i$. So if we assume that both recorded inventory and actual inventory are equal at the end of period $i-1$, then clearly, $I_i \leq \bar{I}_i^{'}$. This difference in values is the source and meaning of inaccurate inventory records, and it leads to greater shortages and an inability to maintain a "stockless" inventory system.

To reconcile the differences between recorded inventory and actual inventory at the POU, a physical inventory count must be performed. We assume that physically counting inventory is an error-free process that does not create additional inaccuracy. We define N to be the number of days between physical inventory counts.

Note in Figure 2.3 the timing of events over a planning cycle of N periods (days).

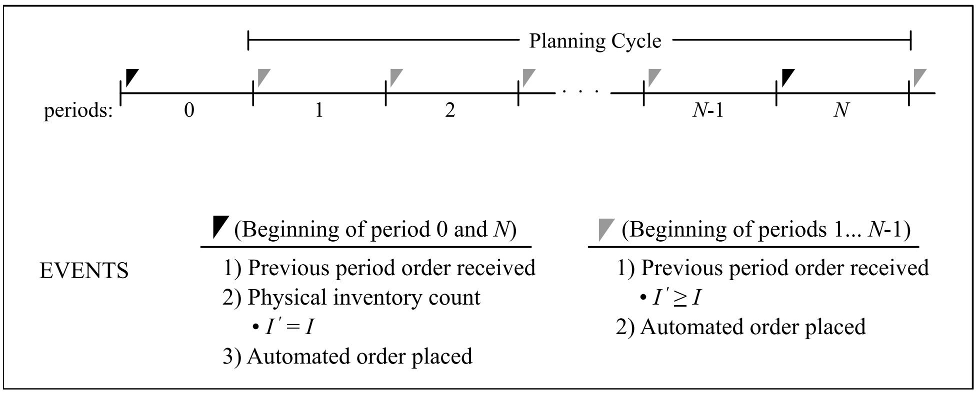


Figure 2.3: Event Timeline

Figure 2.3 shows that counting occurs at the beginning of period 0 after the incoming order is received, and before an order is placed. The next time counting occurs is at the beginning of period N. We will be modeling the daily expected cost over periods 1 through N.

After a count, recorded inventory is reconciled so that $I' = I$ at that instant. Thus, when the order is placed at the beginning of period 0, it is based on accurate inventory records. In other words, an order of size $S - \underline{I}_0'$ is placed when $I' = I$ at that instant. If no demand were to occur during period 0, then the incoming order would bring inventory (both actual and recorded) back up to the order-up-to level, S. Because demand does occur during period 0, *recorded* inventory at the beginning of period 1 will be $S - RD_0$, but *actual* inventory in the beginning of period 1 will be $S - D_0$. We can say precisely what the actual net inventory will be at the end of the first period after a count - simply the beginning period 1 inventory less period 1 total demand:

$$I_1 = S - D_0 - D_1 \tag{2.7}$$

Subsequent inventory values for actual net inventory at the end of periods $2 \ldots N$ can be derived recursively using Equation 2.5.

To apply cost consistently throughout the N periods of the planning cycle, we must define backorders and on-hand inventory at the end of each period. Backorders in a period are based on actual net inventory at the end of the period, and are given by the negative part, I_i^-. Similarly, on-hand inventory is based on actual net inventory at the end of the period, and is simply the positive part, I_i^+. We use the *actual* net inventory for an obvious reason - it represents the actual physical state of the inventory. System inventory costs are applied to the actual inventory, not to what the automated system reflects.

Having presented the physics of the inventory system, as well as the planning cycle, we now present the cost model formulation.

2.3.2 Model Formulation

In developing the cost minimization model, we now introduce the costs associated with this inventory system. We assume a fixed cost, k, for conducting physical inventory counts. This fixed cost is applied once over the planning cycle of N periods.

As we are assuming full backordering, we define c_b to be the backorder cost per unit per day. Similarly, holding cost per unit, per day, is given by c_h. Both the holding and backordering costs are applied at the end of each period, after demand occurs. Actual net inventory is the source of holding and backordering costs, as recorded inventory may not reflect the physically available inventory or the backlog. We combine Equations 2.7 and 2.5, applying the relationship given by Equation 2.4, to give the following definition for actual net inventory at the end of period i, given some count frequency N.

$$I_i = \begin{cases} S - D_0 - D_1 & \text{for } i = 1 \\ \\ S - \sum_{j=i-1}^{i} D_j - \sum_{j=0}^{i-2} UD_j & \text{for } i = 2 \dots N \end{cases} \tag{2.8}$$

You will note that for $i = 2 \dots N$, actual net inventory is not represented by a recursion as in Equation 2.5. The validity of the transformation can be easily shown and we include the derivation in the Appendix.

Then, using the definition for backorders and on-hand inventory, we can give the total cost of the system over N periods:

$$N \cdot C(S,N) = k + \sum_{i=1}^{N} c_h \left[E\left[I_i^+\right] + c_b E\left[I_i^-\right]\right] \tag{2.9}$$

Equation 2.9 includes the cost, k, to count once over the cycle N, and the expected holding costs and expected backorder costs that are incurred at the end of each period, i. Because we are assuming that demand occurs according to a Poisson distribution, we can simplify the expectations in the cost equation. Also, we will be minimizing the expected *daily* cost, $C(S,N)$, which will require both sides of Equation 2.9 to be divided by N. Simplifying the expectations and averaging over the N periods gives the following daily expected cost objective:

$$C(S,N) = \frac{k + (c_h + c_b)\sum_{i=1}^{N} G^1(S;\mu_i)}{N} + c_h\left(S - \mu - \frac{(N-1)(1-p)\lambda}{2}\right) \tag{2.10}$$

Recall that total demand $D_i \sim$ Poisson (λ). We define $\mu = 2\lambda$ and $\mu_i = \mu + (i-1)(1-p)\lambda$. So, $G^1(S;\mu_i)$ is the resultant first order loss function for a Poisson distribution with parameter μ_i. The derivation of the Equation 2.10 and the development of the new terms can be found in the Appendix.

We now want to find the optimal order-up-to level, S, for a given N. To do this, we take the first difference, over S, of Equation 2.10. That first difference is denoted $\Delta_S C(S,N)$ and is given below:

$$\Delta_S C(S,N) = -\frac{(c_h + c_b)\sum_{i=1}^{N} G^0(S;\mu_i)}{N} + c_h \tag{2.11}$$

Here, $G^0(S;\mu_i)$ is the complimentary cdf for the Poisson distribution, and results from taking the first difference of the Poisson first-order loss function. However, we would achieve the same result for any discrete demand distribution, as we have applied the properties of first order loss functions for discrete ransom variables. Now, we want to find the smallest S that makes $\Delta_S C(S,N) \geq 0$. Therefore, we define S_N^* as the smallest integer such that the following inequality holds:

$$\frac{\sum_{i=1}^{N} G^0(S;\mu_i)}{N} \leq \frac{c_h}{c_h + c_b} \tag{2.12}$$

Note that the LHS of the above equation is a non-increasing function of S. So for larger values of S than S_N^*, the inequality will continue to hold. Since the inequality holding is equivalent to Equation 2.11 remaining nonnegative, this shows that S_N^* is the global minimum, and not just a local minimum for a particular N.

To find the optimal N^*, we need to explore the optimal values, $C(S_N^*, N)$, for a range of values for N. It may be that when $C(S_{N+1}^*, N+1) \geq C(S_N^*, N)$ we have found the optimal solution, with order-up-to level, S_N^*, and counting frequency, $N^* = N$. We later explore the structure of the model computationally to see whether this is the case.

As backorder costs are notoriously difficult to estimate, we present a service level model under the given structure. We are still assuming an inventory system for a single item at a single POU location with probability, p, that demand is recorded. N represents the number of days between inventory counts, and S is the order-up-to level. All the parameters are identical, but we do not assume a backorder cost.

In this case, we define the actual net inventory at the *beginning* of each period, i, denoted NI_i. Since the actual net inventory at the end of some period i is given by Equation 2.8, we simply remove the term for period i demand and we get the actual net inventory at the beginning of period i, after the incoming order is received. Said in another way, since $NI_i - D_i = I_i$, Equation 2.8 implies the following:

$$NI_i = \begin{cases} S - D_0 & \text{for } i = 1 \\ S - D_{i-1} - \sum_{j=0}^{i-2} UD_j & \text{for } i = 2 \ldots N \end{cases} \tag{2.13}$$

To derive the fill rate for a particular period, we need to compute the number of units short each period, denoted US_i. If we begin a period with outstanding backorders ($NI_i < 0$), then clearly the number of units short will be the demand in that period, D_i. If there is on-hand inventory available, than the number of units short will be the positive part of the difference between demand and available inventory, $(D_i - NI_i)^+$. This idea is formalized in the following equation:

$$US_i = \begin{cases} D_i & \text{if } NI_i < 0 \\ (D_i - NI_i)^+ & \text{if } NI_i \geq 0 \end{cases} = \begin{cases} D_i & \text{if } NI_i < 0 \\ \left(D_i - S + D_{i-1} + \sum\limits_{j=0}^{i-2} UD_j \right)^+ & \text{if } NI_i \geq 0 \end{cases} \tag{2.14}$$

Note the special case when $i = 1$. Following the convention that the summation is zero when the upper limit falls below the lower limit, the relationships still hold true in the model.

From Equation 2.14 we can compute the expected number of units short in period i, which turns out to be:

$$
\begin{aligned}
E[US_i] &= \sum_{k=-\infty}^{S} E\left[US_i|NI_i = k\right] Pr(NI_i = k) \\
&= \sum_{k=-\infty}^{-1} \lambda g(S-k;\lambda_i+\lambda) + \sum_{k=0}^{S} E\left[(D_i-k)^+\right] g(S-k;\lambda_i+\lambda) \\
&= \lambda G^0\left(S;\lambda(i(1-p)+p)\right) + \sum_{k=0}^{S} G^1(k;\lambda) g(S-k;\lambda(i(1-p)+p))
\end{aligned}
\tag{2.15}
$$

In the above equation, we have $\lambda_i = (i-1)(1-p)\lambda$, similar to what was used previously. Also, $g(x;y)$, $G^1(x;y)$, and $G^0(x;y)$ represent a pmf, first-order loss function, and ccdf, respectively. The y parameter in the previous statement represents the mean of the respective Poisson distributions for each case, and the x parameter represents the independent variable in each case.

Now that we have the expected number of units short in period i, we can compute the fill rate for period i, denoted $FR(S,i)$. It is simply the compliment of the expected number of units short divided by the expected daily demand, as shown below:

$$
FR(S,i) = 1 - \frac{\lambda G^0\left(S;\lambda(i(1-p)+p)\right) + \sum_{k=0}^{S} G^1(k;\lambda) g(S-k;\lambda(i(1-p)+p))}{\lambda} \tag{2.16}
$$

With the fill rate defined for each period, i, we can now address the service level problem. So, instead of using an explicit backorder cost, we define a fill rate constraint. With N days in a cycle, we want the fill rate in the last day of the cycle, $FR(S,N)$ to be at least as much as some specified value, FR_{min}, for example. That constraint is given below:

$$
FR(S,N) \geq FR_{min} \tag{2.17}
$$

Notice that the holding costs are increasing in S, therefore we will be minimizing holding costs subject to a minimum service level constraint. So, we choose the smallest integer value of S, denoted as $S_N^{'}$, that satisfies the service level constraint. Then we simply choose the N to minimize the modified cost objective based on Equation 2.10, where we substitute $S_N^{'}$ for

S and remove the backorder costs, c_b. This "service level" cost is denoted $C_{SL}(S_N^{'}, N)$, and is given below:

$$C_{SL}(S_N^{'}, N) = \frac{k + c_h \sum_{i=1}^{N} G^1(S_N^{'}; \mu_i)}{N} + c_h \left(S_N^{'} - \mu - \frac{(N-1)(1-p)\lambda}{2} \right) \tag{2.18}$$

To minimize, we would like to choose the smallest N where $C_{SL}(S_{N+1}^{'}, N+1) \geq C_{SL}(S_N^{'}, N)$, if it is a unimodal function of N. However, the effectiveness of this approach needs to be explored computationally, as will be seen in the next section.

2.5 SOLUTION PROCEDURE

In this section, we describe the problem instances to be analyzed. We then explore the structure of the cost objective in each model, in relation to the length of the counting cycle, N. Based on the structural findings, we provide solution procedures for each model. In the next section, we explore the structure of the minimizing solutions.

2.5.1 Problem Definitions

For each model we considered an equal number of problem instances. The holding cost, c_h, counting cost, k, demand rate, λ, and accuracy level p, vary identically between both models. The difference between the models is the minimum fill rate constraint, FR_{min}, in the service level model, and the backorder cost, c_b, in the cost-only model. The parameters vary as follows:

$$\begin{aligned} c_h &= 0.05, 0.30, 0.60 \\ k &= 20, 40, 100 \\ \lambda &= 8, 15, 20 \\ p &= 0.45, 0.50, 0.55, \ldots, 0.90, 0.95 \\ FR_{min} &= 0.90, 0.95, 0.99 \\ c_b &= 3, 6, 12 \end{aligned}$$

Notice that the parameter for the accuracy level, p, ranges from 0.45 up to 0.95 in increments of 0.05, leading to 11 variations in p. Combined with the three levels of parameter settings for the remaining four parameters, gives a problem set with $11 \cdot 3^4$ (or 891) instances for each model. Before we can seek solutions for these sample problems, we must explore how the cost behaves in N.

2.5.2 Structural Findings

For both models described in this paper, we have defined optimality conditions for finding S_N^* for a given N and $S_N^{'}$ for a given N and a given fill-rate constraint. However, we need to explore how the related cost objectives for both models behave over changing values of N. If the objectives $C(S_N^*, N)$ and $C_{SL}(S_N^{'}, N)$ are unimodal over values of N, then we can claim that our solutions are optimal using the previously described approaches. Therefore, we will explore each model computationally to test whether they might be unimodal over N.

After we know the structure of each model, we can proceed with a solution procedure that will minimize the cost objective. For the service level problem we will minimize cost, subject to the fill-rate constraint for period N.

Shortage Cost Model

To begin our look at the structure of the cost objective, we apply the basic algorithm as described previously. That is, we incremented N until $C(S_{N+1}^*, N+1) > C(S_N^*, N)$, for each of the 891 problem instances. Then, once we had the potential N^*, we continued to increment N beyond N^* until it was clear that the cost would not decrease below the minimum value. In this way, we explored unimodality while seeking optimal solutions.

The results of this effort showed that by incrementing N until $C(S_{N+1}^*, N+1) > C(S_N^*, N)$ we were able to find the optimal solution in all but 15 problem instances. On average, the cost generated by the heuristic approach was 0.03% above the cost of the optimal solution for those 15 non-optimal cases. Furthermore, while we found that the cost function was not unimodal in N for all problem cases, it *was* non-decreasing in our search range beyond N^* for all but 26 problem instances. We cannot say, then, that the function is unimodal in N for all problem instances.

We found that the above method is easy to implement and it also performs quite well, as it found the optimal solution in 98.3% of the test cases and was only off by an average of 0.03% the other 1.7% of the time. Therefore, in the next section we will analyze the results given by using the above method to find optimal or near-optimal solutions to the shortage cost model.

Service Level Model

For the service level model, we will be looking at how the cost, $C_{SL}(S_N^{'}, N)$, behaves over values of N. Initially, suppose that we set N^* to be the smallest integer such that $C_{SL}(S_{N+1}^{'}, N+1) > C_{SL}(S_N^{'}, N)$. We want to see how close we can get to the optimal with this simple approach, as with the previous model.

The Service Level model was not found to be unimodal over N, and just over half of the problem instances were found to be non-decreasing after finding N^*. Those that were non-decreasing in our search area found the optimal solution using the simple incrementing approach. For the remaining problem instances, we incremented N beyond N^* enough to see the function turn upwards in a seemingly non-decreasing path. More specifically, we increased N beyond N^* until there were ten consecutive values for N with increased costs. Over the complete range we took the minimum cost solution to be the optimal.

With this approach, we found that about 74% of the problem instances were able to find optimal solution at N^*. The remaining 26% of the problems had an N^* that was, on average, about 2.4% away from the optimal solution.

As such, the incremental approach performed modestly, but additional work should be done to improve the minimization technique. We leave this additional work for later research, and proceed with our analysis of the results. We hope to gain both structural and managerial insights through our analysis that are useful to both practitioners and academicians.

2.6 NUMERICAL RESULTS

In this section we explore how the parameters affect the minimizing solutions to both the service level problem and the shortage cost problem. As discussed previously, the parameters were largely the same for the separate models, but the shortage costs and fill rate constraints were not present in both models.

2.6.1 Shortage Cost Model Analysis

As the first model is based on minimizing shortage costs as well as holding and counting costs, we will first look at how these parameters influence cost. Obviously, the base stock level and counting frequency will also have an effect on cost, and so it will be explored later. We are also interested in the fill rate associated with the different shortage costs, and so we will discuss that later on.

Accuracy level, p

When p was greater, the optimal cost was lower. In fact, as p increases linearly from 45% towards 95% accuracy, the cost decreases at a somewhat parabolic descent. If we look at the average cost over all other parameters, and test the sensitivity to p, as given by

$$\frac{\Delta C/C}{\Delta p/p}$$

we find that as p increases, the sensitivity to changes in p also increases. Specifically we found that for small p, a 1% increase in the accuracy level equates to only a 0.31% reduction in cost, but when p is large, a 1% increase can mean up to a 3.6% decrease in cost.

Managerially, one would see only minor cost benefit when moving from an accuracy level of, say 45% up to 60%. Once accuracy has been improved sufficiently, the cost is more easily influenced by further improvements in accuracy.

The accuracy level also strongly influences the counting cycle length, N^*. The cycle length seems to increase exponentially as accuracy increases beyond 80% or 85%, as it is highly sensitive to the value of p. Using the same type of measure as above, we find that for large p, a 1%

change in accuracy can lead to an 8% change in N, whereas 1% changes to lesser values of p may only imply 0.45% increases in N. This is an intuitive result, since perfect accuracy would, in theory, require no counting. So, as p approaches 100%, N^* should approach ∞, which would rapidly reduce the average daily cost of counting stock.

The opposite is true for the base stock inventory level. As p increases, we are required to hold less stock. However, since there is a lead time to replenish stock, we are still required to hold stock to meet demand between daily order arrivals, even when accuracy is very high. The data suggest that p influences the base stock level very similarly to the way cost is influenced.

Shortage cost, c_b

The backorder cost, c_b, was found to have less influence on the minimizing solutions than intuition suggests. That is, as the shortage cost increases almost 400%, there is only a slight increase (16%) in the average cost of the minimizing solution. This finding demonstrates that the decision variable solutions are successfully reducing shortages, which was the goal of the model.

The shortage cost had opposing effects on the cycle length, N, and the base stock level, S. The base stock level increases slightly with the shortage cost parameter's increase, while the cycle length decreases slightly. With a higher cost for backorders, it is wise to increase inventory and count more frequently to guard against shortages due to inaccurate records and low on-hand inventory quantities.

While computing cost for the shortage cost model, we also computed the resultant fill rate for each minimizing solution. We found the shortage cost to have a significant influence on the resulting service level, in period N, of the cost-minimizing solution. As the shortage cost increases, so does the fill rate. There doesn't appear to be a specific shortage cost value in the cost-minimizing model that corresponds to a specific fill rate constraint in the service level problem. Instead, it seems as though a combination of counting frequency, holding cost, and shortage cost values would correspond to specific fill rate constraints.

It is also interesting to note, that the shortage cost value has no influence on how other parameters affect the resulting cost. We find this same type of relationship later when discussing the service level model.

Holding cost, c_h

We found that holding cost does not greatly influence the results of the minimization. As holding cost is increased by a factor of 12, the cost also increases, by only a factor of 2.5. If we look at cost's sensitivity to c_h, we find that for every 1% increase in holding cost there is only a 0.35% increase in total cost.

An interesting aspect about the holding cost, is that as c_h is increased, both the cycle length, N, and the base stock level, S, are decreased. This means that for higher holding costs, total cost is minimized by counting more frequently and holding less in stock. Basically, this result shows the tradeoff between holding costs and counting costs. If we count more frequently, then we pay the fixed counting cost more often. Therefore, we would reduce holding costs by holding less inventory, which would be safe since counting more frequently ensures that the records maintain enough accuracy to keep shortage costs low.

If we look at the combined influence of holding cost and accuracy level, we find that lower values of holding cost, c_h, lead to a lower cost sensitivity to the accuracy level. That is, for higher c_h, improving the accuracy has a greater affect on cost. This is probably due to the relationship between accuracy and the optimal base stock level, S.

Fixed counting and reconciliation cost, k

As just mentioned for holding cost, an increase in counting costs affects both cycle length and base stock level in a similar fashion. For k, however, the influence is in the opposite direction. So, when k is increased, both S and N are *increased.* Again, this shows the tradeoff between holding cost and counting cost. If it costs more to count, then one would count less and hold more inventory as a result of the decreased accuracy of the ordering system stemming from less frequent reconciliation.

The fixed counting cost also affects total cost in a direct fashion. As k increases, so does the minimized total cost. Although, the overall cost is about as sensitive to k as it is to c_h. This means that a 1% change in k equates to about a 0.3% change in total cost.

We can also see a second-order influence between counting cost and accuracy level. Similar to the relationship between c_h and p, when k is higher, then changes in p have a stronger impact than when k is lower.

Daily demand rate, λ

The Poisson rate for random daily demand was an interesting parameter to explore. It actually had opposite effects on the minimizing solutions of S and N. That is, as the demand rate increased, S was increased to meet increased demand, while the cycle length decreased to ensure reconciliation happened more frequently. This can be explained by the fact that the rate of unrecorded demand increases with an increase in the rate of overall demand.

We also find that as demand rate increases, overall cost is also increased. However, it is interesting that as demand increases, the period N fill rate (for the minimized solution) is actually increased. So as demand increases, the minimizing solution is more likely to alleviate shortages. This result is almost counter-intuitive.

General insights

These results indicate that holding costs, shortage costs, and counting costs can be successfully balanced using the model structure employed here. Managerially, it is wise to have a dynamic cycle length, N, and base stock level, S, based on the accuracy level, p.

Some of the minimizing base stock levels were found to be quite high, with some cases requiring over 250 units (recall that maximum lead time demand is just 20 units). This amount of stock seems excessive where there is a one period lead time, and where we assume demand has a highest average demand rate of 20 units per day. As a result, a lot of space may be necessary to store certain items in stock. In the hospital setting, where space is valuable and scarce, this may not be feasible. Another type of model, or an adjustment of this model, may be useful to explore for those items which require a lot of space to store the minimizing quantity.

Furthermore, the backorder costs and holding costs chosen in this analysis were arbitrary. A more accurate estimate of actual costs would provide additional insight into this inventory model, and may prove to be more useful to practitioners. We now move on to discuss the service level model.

2.6.2 Service Level Model Analysis

In analyzing the results of minimizing cost subject to a fill rate constraint, recall that our solution approach found the optimal solutions approximately 74% of the time. As a result, about 26% of our results are near-optimal solutions that were about 2% higher than the optimal solution on average.

We now proceed with our analysis, we use the same pattern as was used in the previous model. We consider the parameters of holding cost, c_h, counting cost, k, demand rate, λ, and the fill rate constraint, FR_{min}, as applied to period N.

Fill rate analysis

It was validated, in every case, that our solution approach did find solutions that met fill rate constraints in every 891 problem instance considered. Remember that the fill rate constraint was applied to the last period of the cycle, period N, but we are ensured that the resultant fill rates in previous periods (i.e. periods $1, \ldots, N-1$) were higher than the constraint.

In analyzing just the fill rate, only the constraint, FR_{min}, had noticeable influence on the resultant fill rate of the minimizing solution. In fact, there is almost a one-to-one correspondence to between the constraint and the resultant fill rate. This means that the fill rate was a binding constraint in the model. While this seems intuitively obvious, never was there an example where meeting a higher fill rate was less costly than meeting a lower fill rate.

As FR_{min} appeared to be a binding constraint, it is not unexpected that we found that inventory inaccuracy, counting cost and holding cost were all shown to have little, if any, influence on the resultant fill rate in period N. However, each of the parameters did affect the minimized cost value and the value of the decision variables, S and N, in the minimizing solution. We explore these influences next.

Other parameters: p, c_h, k, λ

Compared to the shortage cost model, the other parameters in the service level model behave nearly identically. For instance, not only does cost decrease as accuracy increases, but as accuracy increases, so does the cost sensitivity to the accuracy parameter, as was seen in the previous model. Similarly, the holding cost and counting cost have opposite effects on S and N, as in the previous model. Even the second order effects, which were discussed previously, are found to generate basically the same results in the minimizing solutions.

Similar to the previous model, the resulting base stock levels were quite high, in relation to how a hospital stores and uses inventory. In fact, as in the previous model, there were several cases of the minimizing solution to require a base stock level higher than 200 units (recall that the lead time demand was no greater than 20 units in our problem set). This may not be feasible in a real industry setting, and so there may be some require modifications to the model to address this issue.

2.7 CONCLUSION

We develop two healthcare inventory models that allow for full backordering and are based on the use of an automated inventory replenishment system that is prone to inaccuracy. While the premise of the model is similar to other inaccuracy-related literature, we are the first to introduce both a cost-minimization model and a fill-rate constrained model where demand is random and discrete.

The solution approach we develop was shown to find the optimal base-stock policy for a given counting cycle length. In most test cases, our solution approach also finds the optimal counting cycle length, thus minimizing cost under either shortage costs or fill rate constraints. More specifically, our approach found the optimal solution in 98% of the test cases for the shortage cost model and about 74% of the test cases for the service level model. On average, the near-optimal solutions were off by about 2%.

The models developed here are unique to healthcare management literature, as the unique challenges of the healthcare industry supply chain are underrepresented in the literature. The hospital supply chain considered here operates according to a stockless inventory system, similar to the type defined in Rivard-Royer et al. (2002).

To analyze the effectiveness of the inventory system and of our solution approach, we considered a total of 1782 problem instances, evenly split between the service level model and the shortage cost model. We found that where accuracy, p, is low, there is a crippling affect on the inventory system performance, that is - service level is decreased and cost is increased. We also saw that initial efforts in addressing accuracy (i.e. improving from $p = 0.45$ to $p = 0.70$) will not see as significant results as when improvements are made at higher values (i.e. from $p = 0.80$ to $p = 0.95$). This should be motivation for adopters of a stockless inventory system to strive for only high accuracy values, as the cost savings have been shown to be significant at higher levels of improvement.

An application of such findings could likely provide a basis for putting monetary value on inventory accuracy in similar settings in industry. There is currently a thriving industry of inventory system providers that purport to reduce inaccuracy in the inventory system by recording inventory in innovative ways. Having a way to validate these cost savings may strengthen the position of the hospital in making system purchase decisions. Similarly, understanding the model presented here may lead systems providers to assess the effectiveness and value of their own offerings.

The results of the model also show that the base stock inventory system, which is generally used in healthcare, may require a significantly higher base stock level than space allows. In the healthcare industry this would not be feasible, as space is often costly or limited. In a later article currently in progress, we hope to present a model with a modified structure to support lower inventory levels while still keeping costs low.

2.7.1 Future Work

It will also be interesting to study the pervasiveness of inaccuracy in the healthcare industry, as it is already planned as future research. The results of such a study could show how inaccuracy effects both profitability (as in the shortage cost model) and patient care (as in the service level model). As medical supplies are in high demand and are of high value, it is likely that the healthcare industry can see enormous savings in addressing the accuracy issue, as shown in our results and throughout our interaction with industry.

Other future work that applies to what has been done here, may include a modified heuristic approach to increase the possibility of finding optimal solutions. In the research presented here, we employed Maple 12 mathematical programming language to perform our analysis and computation. It may be that other methods of implementing our model my prove to be more efficient.

REFERENCES FOR ARTICLE 2

DeHoratius, N., A. J. Mersereau, L. Schrage. 2008. Retail inventory management when records are inaccurate. *Manufacturing and Service Operations Management* **10**(2) 257–277.

DeScioli, D. 2005. Differentiating the hospital supply chain for enhanced performance. Ph.D. thesis, Massachusetts Institute of Technology.

Duclos, L. 1993. Hospital inventory management for emergency demand. *International Journal of Purchasing and Materials Management* 30–37.

Fleisch, E., C. Tellkamp. 2005. Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain. *International Journal of Production Economics* **95** 373–385.

Friedman, J. L. 1994. Justification of a point-of-use inventory management system. *Hospital Materiel Management Quarterly* **15**(3) 57–62.

Iglehart, D. L., R. C. Morey. 1972. Inventory systems with imperfect asset information. *Management Science* **18**(8) 388–394.

Jan, P. J.-H. 2006. Tailored hospital supply chain for greater return on investment. Ph.D. thesis, Massachusets Institute of Technology.

Kök, A. G., K. H. Shang. 2007. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing and Service Operations Management* **9**(2) 185–205.

McKone-Sweet, K. E., P. Hamilton, S. B. Willis. 2005. The ailing healthcare supply chain: A prescription for change. *The Journal of Supply Chain Management* (Winter) 4–17.

Morey, R. C. 1985. Estimating service level impacts fom changes in cycle count, buffer stock, or corrective action. *Journal of Operations Management* **5**(4) 411–418.

Perrin, R. A. 1994. Exchange cart and par level supply distribution systems: Form follows function. *Hospital Materiel Management Quarterly* **15**(3) 63–76.

Raman, A., N. DeHoratius, Z. Ton. 2001. Execution: The missing link in retail operations. *California Management Review* **43**(3) 136–152.

Rekik, Y., E. Sahin, Y. Dallery. 2008. Analysis of the impact of the rfid technology on reducing product misplacement errors at retail stores. *International Journal of Production Economics* **112** 264–278.

Rivard-Royer, H., S. Landry, M. Beaulieu. 2002. Hybrid stockless: A case study: Lessons for health-care supply chain integration. *International Journal of Operations and Production Management* **22**(4) 412–424.

Rossetti, M. D. 2008. Inventory management issues in health care supply chains. Tech. rep., University of Arkansas.

Sandoh, H., H. Shimamoto. 2001. A theoretical study on optimal inventory-taking frequency for retailing. *Journal of Retailing and Consumer Services* **8** 47–52.

Sheppard, G. M., K. A. Brown. 1993. Predicting inventory record-keeping errors with discriminant analysis: A field experiment. *International Journal of Production Economics* **32** 39–51.

Young, S. T. 1989. Hospital material management: Systems and performance. *Journal of Purchasing and Materials Management* (Fall) 31–35.

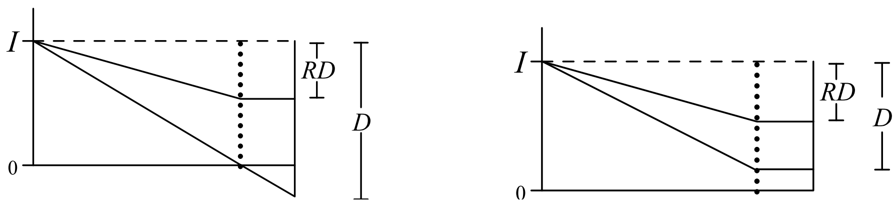
Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.

B APPENDIX

B.1 Justification of the approximate distributions used for UD_i and RD_i

The Alabama hospital is using an automated replenishment system that does not record backorders, while the actual net inventory can incur backorders. Because of this, recorded demand, RD_i is dependent on the amount of actual inventory, I, on hand in the POU. Once actual inventory is depleted, demand can no longer be recorded in the system. Similarly, the recorded demand cannot exceed total demand, D_i. In other words, $0 \leq RD_i \leq \min(I, D_i)$. The Figures below demonstrate this idea when $D_i > I$ and $D_i < I$.

When a demand event occurs, and stock is available in the POU inventory location, then the probability that demand is recorded is p. However, if there is no stock on hand when a demand occurs at the POU, the demand is not recorded in the automated ordering system. In practice, the demand is backordered until the next incoming replenishment arrives, leading to a delay in patient care. In the end, the automated system sees neither the backorder, nor its fulfillment. As a result, the recorded demand is dependent on the level of on-hand stock and the amount of total demand seen. The figures below demonstrate this concept, where the recorded demand, RD is constrained by either the total demand, D, or the available on-hand inventory, I.



(a) Constrained by inventory (b) Constrained by demand

Figure B.1: Recorded Demand dependency

Basically, the recorded demand can be no larger than the smaller of on-hand inventory (Figure B.1a) versus total demand (Figure B.1b). This can be written as the following relation:

$$0 \leq RD \leq \text{Min}(D, I) \tag{B.1}$$

From a demand standpoint, we want to derive the distribution for recorded demand in a particular shift i and day j. Due to the complexity of the actual distribution of recorded demand, we use an approximation of this distribution. We condition on total shift i demand, D_{ij}, which follows a Poisson distribution. Doing this will allow us to derive the approximate distribution of recorded demand, RD_{ij}. The basic distribution equation follows, though we ignore the ij index without loss of generality:

$$Pr[RD = j] = \sum_{k=0}^{\infty} Pr[RD = j|D = k] * Pr[D = k] \tag{B.2}$$

Since we need to also know the value of available inventory, I, we will be using its value as a reference point as we expand Equation B.2, as follows. Note that we also substitute the Poisson probability as total demand follows a Poisson distribution with parameter λ.

$$Pr[RD = j] = \sum_{k=0}^{I} Pr[RD = j|D = k]\frac{e^{-\lambda}\lambda^k}{k!} + \sum_{k=I+1}^{\infty} Pr[RD = j|D = k]\frac{e^{-\lambda}\lambda^k}{k!} \tag{B.3}$$

Since each demand event has probability, p, of being recorded, we can say that the conditional distribution of recorded demand, RD is simply binomial with probability p. If $D \leq I$ then it will be

$$Pr[RD = j|D] = \begin{pmatrix} D \\ j \end{pmatrix} p^j(1-p)^{D-j}$$

and if $D > I$, then it will be

$$Pr[RD = j|I] = \begin{pmatrix} I \\ j \end{pmatrix} p^j(1-p)^{I-j}$$

This relationship affects Equation B.3 as follows:

$$Pr[RD = j] = \sum_{k=j}^{I} \begin{pmatrix} k \\ j \end{pmatrix} p^j(1-p)^{k-j}\frac{e^{-\lambda}\lambda^k}{k!} + \begin{pmatrix} I \\ j \end{pmatrix} p^j(1-p)^{I-j} \sum_{k=I+1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \tag{B.4}$$

Note that the first term in Equation B.4 is in effect when $D = k \leq I$, and the second term when $D = k > I$. The first term can be simplified as follows:

$$\sum_{k=j}^{I} \begin{pmatrix} I \\ j \end{pmatrix} p^j (1-p)^{k-j} \frac{e^{-\lambda}\lambda^k}{k!} = \frac{p^j e^{-\lambda}}{j!} \sum_{k=j}^{I} \frac{[(1-p)\lambda]^{k-j}\,\lambda^j}{(k-j)!} = \frac{(p\lambda)^j e^{-\lambda}}{j!} \sum_{k=0}^{I-j} \frac{[(1-p)\lambda]^k}{k!} \quad \text{(B.5)}$$

Equation B.5 can now be written as

$$Pr[RD = j] = \frac{(p\lambda)^j e^{-\lambda}}{j!} \sum_{k=0}^{I-j} \frac{[(1-p)\lambda]^k}{k!} + \begin{pmatrix} I \\ j \end{pmatrix} p^j (1-p)^{I-j} \left(1 - \sum_{k=0}^{I} \frac{e^{-\lambda}\lambda^k}{k!}\right) \quad \text{(B.6)}$$

Because we cannot condition on I since we do not know its distribution, we consider how Equation B.6 behaves as $I \to \infty$. This will provide us with an asymptotic approximation. The second term will clearly become zero as I approaches infinity. The summation element of the first term has the limit: $e^{\lambda - p\lambda}$ as $I \to \infty$. So, we finally get the following approximate distribution for recorded demand:

$$Pr[RD = j] = \frac{(p\lambda)^j e^{-p\lambda}}{j!} \quad \text{(B.7)}$$

Therefore, we approximate the recorded demand distribution using the Poisson distribution with parameter $p\lambda$. Remembering the indexing used, we define the recorded demand in shift i of day j to be $RD_{ij} \sim$ Poisson $(p\lambda_i)$. To be consistent with $D_{ij} \sim$ Poisson(λ_i), we see that for unrecorded demand, $UD_{ij} \sim$ Poisson$((1-p)\lambda)$.

In our simulation and our approximate cost versions of the model, we use the above approximations for the distributions of recorded demand and unrecorded demand.

B.2 Justification of the actual net inventory equation

The equation below is a reprint of Equation 2.8 from the text and is derived using the recursive relationship given in Equation 2.5 from the text.

$$I_i = \begin{cases} S - D_0 - D_1 & \text{for } i = 1 \\ S - \sum_{j=i-1}^{i} D_j - \sum_{j=0}^{i-2} UD_j & \text{for } i = 2 \dots N \end{cases}$$

We want to show that the two forms are identical. First, we look at the case for I_2 and I_3. Then we can prove by induction by assuming true for I_i, $i = 1, \dots, n$, and then by showing true for I_{n+1}.

We already have that the actual net inventory at the end of the first period is given by $I_1 = S - D_0 - D_1$. At the beginning of period 2 we know that an order will arrive based on the recorded inventory use from period 0, RD_0. Briefly explained, this order of size RD_0 is placed at the beginning of period 1 with a one period lead time. Then, during period 2, actual demand reduces the actual net inventory by D_2. This gives the following recursive equation for the actual net inventory at the end of period 2:

$$I_2 = I_1 + RD_0 - D_2 \tag{B.8}$$

Substituting the value of actual net inventory from the previous period we have:

$$I_2 = S - D_0 - D_1 + RD_0 - D_2 = S - \sum_{i=1}^{2} D_i - UD_0 \tag{B.9}$$

which we can say by applying $D_i = UD_i + RD_i$.

Now, for I_3 we have, by the same reasoning:

$$I_3 = I_2 + RD_1 - D_3 \tag{B.10}$$

$$= S - D_0 - D_1 + RD_0 - D_2 + RD_1 - D_3 r \tag{B.11}$$

$$= S - \sum_{i=2}^{3} D_i - \sum_{i=0}^{1} UD_i \tag{B.12}$$

Now we can see the form given in Equation 2.8. We now assume it is true for all periods i from $i = 1, \ldots, n$. This means, that at the end of period $n+1$, we have the following recursive equation:

$$I_{n+1} = I_n + RD_{n-1} - D_{n+1} \tag{B.13}$$

The above equation shows the incoming order at the beginning of period $n+1$ based on period $n-1$ recorded demand. It also shows total demand in period $n+$, D_{n+1}, reducing the actual net inventory. Substituting the value of period n ending actual net inventory, we have:

$$I_{n+1} = S - \sum_{i=n-1}^{n} D_i - \sum_{i=0}^{n-2} UD_i + RD_{n-1} - D_{n+1} \tag{B.14}$$

$$= S - \sum_{i=n}^{n+1} D_i - \sum_{i=0}^{n-2} UD_i + RD_{n-1} - D_{n-1} \tag{B.15}$$

$$= S - \sum_{i=n}^{n+1} D_i - \sum_{i=0}^{n-1} UD_i \tag{B.16}$$

where we get Equation B.15 by pulling a demand element from the second term of the RHS, and include the demand element from period $n+1$ in the second term. Then, we apply $D_i = UD_i + RD_i$ to get Equation B.16. And we see that the relationship holds for period $n+1$, and we have shown that the recursive relationship for end-of-period actual net inventory is identical to Equation 2.8. ■

B.3 Justification of the expected daily cost, $C(S, N)$

The notation is the same here as in the main text, though we introduce some new terms. The idea in this section of the Appendix, is to begin by deriving cost for cycle lengths of 1 and 2, as given by $C(S, 1)$ and $C(S, 2)$. From there we use induction to derive cost for any cycle length, N, given by $C(S, N)$, and throughout we develop the optimality conditions for the model solution.

Before we begin, we reiterate some of the assumptions of the model. We are assuming that inventory is counted and reconciled at the beginning of the day prior to the start of the cycle. That means that inventory would have reached S when the order arrived at the beginning of the day 1, but during day 0 there was a total demand seen of D_0. This means that net inventory at the *beginning* of day 1 will always be given by:

$$NI_1 = S - D_0 \tag{B.17}$$

With the above starting point, we can now discuss the development of the expected daily cost objective we seek to minimize.

B.3.1 Counting inventory daily

The cost to count inventory every day includes the fixed cost to count, k, added to the holding and backorder costs, both applied to the end of period actual net inventory, I_1. Remember that:

$$I_1 = \underline{I}_1 - D_1 = S - D_0 - D_1 \tag{B.18}$$

Therefore, when we compute the cost, it will be based on actual net inventory at the end of the period. Note that actual net inventory is a random variable, which is why we seek to optimize the expected daily cost, $C(S, 1)$, as given below:

$$C(S, 1) = k + E\left[c_h I_1^+ + c_b I_1^-\right] \tag{B.19}$$

The above equation gives the expected daily cost, since our cycle is of length $N = 1$ days. To generate the change in daily expected cost over changes in S, we have to first expand the positive part of net inventory into a computable form. But first, we define the following term:

$$\mu = E\left[D_{i-1} + D_i\right] = 2\lambda \tag{B.20}$$

where D_{i-1} and D_i are independent and identically distributed Poisson random variables as previously defined for all periods i.

We now have the derivation of expected daily cost in a computable form:

$$\begin{aligned} C(S,1) &= k + E\left[c_h I_1^+ + c_b I_1^-\right] \\ &= k + E\left[c_h\left(I_1 + I_1^-\right) + c_b I_1^-\right] \\ &= k + c_h(S-\mu) + (c_h + c_b)G^1(S;\mu) \end{aligned} \tag{B.21}$$

where the first-order loss function is defined as $G^1(S;\mu) = E\left[(D-S)^+\right]$, and $D \sim$Poisson(μ).

We can define the change in daily expected cost over decision variable, S, when counting inventory daily. The following equation applies the first difference for a first-order loss function to give the first difference of expected daily cost:

$$\Delta_S C(S,1) = c_h - (c_h + c_b)G^0(S;\mu) \tag{B.22}$$

where $G^0(S;\mu)$ is the ccdf of the Poisson distribution with rate μ. To optimize the cost function, we find the smallest integer solution of $\Delta_S C(S,1) \geq 0$. This last inequality is equivalent to:

$$G^0(S;\mu) \leq \frac{c_h}{c_h + c_b} \tag{B.23}$$

Let S_1^* denote the smallest integer solution to Equation B.23. We will now take the same approach as we look at counting every other day.

B.3.2 Counting inventory every two days

Now we look at the case when $N = 2$. We still have that inventory is counted and reconciled at the beginning of day 0, so that our starting inventory on day 1 is the same as when we counted daily. Similarly, we still have the same value for day 1 ending inventory, I_1, but now we need actual net inventory at the end of day 2. This is given below:

$$I_2 = I_1 + RD_0 - D_2 = S - D_1 - D_2 - UD_0 \tag{B.24}$$

Since the order placed at the beginning of day 1 is based on recorded inventory, that is the size of the order that arrives at the beginning of day 2. Then, throughout day 2, regular demand occurs which reduces net inventory by D_2. When we expand and apply the relationship $D_0 = RD_0 + UD_0$, we end up with what is given for I_2. It is important to note that the three random variables in Equation B.24 are in different periods, so they are independent by assumptions. Also, since they are Poisson, their sum is also Poisson.

Having a way to describe actual net inventory at the end of day 2, we can now compute the expected cost over a two-day cycle. Recall that we have defined $C(S, j)$ to be the expected daily cost over a cycle of length j. Therefore, we can write the following equation:

$$2 \cdot C(S, 2) = k + E\left[\sum_{i=1}^{2} \left(c_h I_i^+ + c_b I_i^-\right)\right] \tag{B.25}$$

From the above equation, we want to now derive the expected daily cost over a cycle of length $N = 2$. The following shows the intermediate steps taken to derive $C(S, 2)$.

$$2 \cdot C(S, 2) = k + E\left[\sum_{i=1}^{2} c_h I_i\right] + E\left[\sum_{i=1}^{2} (c_h + c_b) I_i^-\right] \tag{B.26}$$

$$
\begin{aligned}
2 \cdot C(S,2) &= k + E\left[c_h I_1\right] + E\left[c_h I_2\right] + E\left[(c_h + c_b) I_1^-\right] + E\left[(c_h + c_b) I_2^-\right] && \text{(B.27)} \\
&= k + c_h(S - \mu) + c_h(S - (\mu + (1-p)\lambda)) + (c_h + c_b) G^1(S;\mu) \\
&+ (c_h + c_b) G^1(S;\mu + (1-p)\lambda) && \text{(B.28)} \\
&= k + c_h(2(S - \mu) - (1-p)\lambda) + (c_h + c_b) \sum_{i=1}^{2} G^1(S;\mu_i) && \text{(B.29)} \\
C(S,2) &= k/2 + c_h((S - \mu) - 1/2(1-p)\lambda) + 1/2(c_h + c_b) \sum_{i=1}^{2} G^1(S;\mu_i) && \text{(B.30)}
\end{aligned}
$$

To get Equation B.26 we simply apply the definition of $I_i = I_i^+ - I_i^-$, as was done when counting daily, and then combine like terms.

Equation B.27 expands the summations for the next step, given by Equation B.28, where we substitute $E\left[I_2\right] = E\left[S - (D_1 + D_2 + UD_0)\right] = S - \mu - (1-p)\lambda$ as defined by Equation B.24. Note that the values relating to period 1 are the same as when counting daily. Similarly, while the first order loss function is the same for period 1, we introduce $G^1(S, \mu + (1-p)\lambda)$ as the first order loss function for period 2.

To explain Equation B.29, we introduce the following definition which will be used hereafter:

$$
\mu_i = \mu + (i-1)(1-p)\lambda \tag{B.31}
$$

Equation B.31 defines the Poisson rate at which actual net inventory, I_i, is depleted over the counting cycle.

We substitute B.31 into B.28 to get Equation B.29. After doing so, we can then solve for the expected daily cost for a cycle of length $N = 2$ days. This is given in Equation B.30. From this point, we can employ the same approach to compute the difference, and we have that:

$$
\Delta_S C(S,2) = c_h - 1/2(c_h + c_b) \sum_{i=1}^{2} G^0(S;\mu_i) \tag{B.32}
$$

Similar to the previous case, we define S_2^* to be the smallest integer value for S that satisfies the following inequality:

$$\frac{\sum_{i=1}^{2} G^0(S;\mu_i)}{2} \leq \frac{c_h}{c_h + c_b} \tag{B.33}$$

B.3.3 Counting inventory every N days

We now conjecture that the following equation represents the expected daily cost of the inventory system with a counting frequency of N. It is given below:

$$C(S,N) = \frac{k}{N} + \frac{(c_h + c_b)\sum_{i=1}^{N} G^1(S;\mu_i)}{N} + c_h(S - \mu - 1/2(N-1)(1-p)\lambda) \tag{B.34}$$

We prove the above conjecture by induction. Since we have already shown the above for $N = 1$ and $N = 2$, we will assume that it is true for $N = 1, \ldots, n$, and then show it is true for $N = n + 1$.

We first consider the actual net inventory at the end of day $n + 1$, as given by:

$$I_{n+1} = S - \sum_{i=0}^{n} D_i + \sum_{i=0}^{n-1} RD_i - D_{n+1} \tag{B.35}$$

which can be explained using the same arguments for period 2, and we have shown this relationship in the previous section in this Appendix. We have seen actual demand, D_i for every period through $n + 1$, and we have seen replenishment arrive based on recorded demand, RD_i, from all periods up to $n - 1$. If we apply the relationship in Equation 2.4 we can reduce the above equation to:

$$I_{n+1} = S - \sum_{i=0}^{n-1} UD_i - D_n - D_{n+1} \tag{B.36}$$

We can then compute the total cost of the individual day $n + 1$ by using the following two formulas as in the previous cases:

$$c_h E\left[I_{n+1}\right] = c_h\left(S - \mu - n(1-p)\lambda\right) \tag{B.37}$$

$$(c_h + c_b)E\left[I_{n+1}^-\right] = (c_h + c_b)G^1(S; \mu_{n+1}) \tag{B.38}$$

Now, we will add the costs for day $n+1$ to the total cost over n days, $nC(S,n)$. This sum we will denote $(n+1)C(S,n+1)$, and we will simplify and combine terms to complete the proof.

$$\begin{aligned}(n+1)C(S,n+1) &= nC(S,n) + c_h\left[S - \mu - n(1-p)\lambda\right] + (c_h+c_b)G^1(S;\mu_{n+1}) \\ &= k + (c_h+c_b)\sum_{i=1}^{n} G^1(S;\mu_i) + c_h\left(nS - n\mu - \frac{n}{2}(n-1)(1-p)\lambda\right) \qquad \text{(B.39)} \\ &+ c_h\left(S - \mu - n(1-p)\lambda\right) + (c_h+c_b)G^1(S;\mu_{n+1})\end{aligned}$$

$$(n+1)C(S,n+1) = k + (c_h+c_b)\sum_{i=1}^{n+1} G^1(S;\mu_i) + c_h\left((n+1)S - (n+1)\mu - (n)(n+1)\frac{1}{2}(1-p)\lambda\right) \tag{B.40}$$

Finally, we divide both sides of Equation B.40 by $n+1$ and we have the form we need, as given below:

$$C(S,n+1) = \frac{k}{n+1} + \frac{(c_h+c_b)\sum_{i=1}^{n+1} G^1(S;\mu_i)}{n+1} + c_h\left(S - \mu - (n)\frac{1}{2}(1-p)\lambda\right) \tag{B.41}$$

which proves the conjecture from above. ■

Knowing the expected daily cost for any counting cycle length, N, allows us to derive the optimality conditions for S_N^*.

3: INVENTORY MODEL WITH EMERGENCY ORDERING WHEN INVENTORY RECORDS ARE INACCURATE

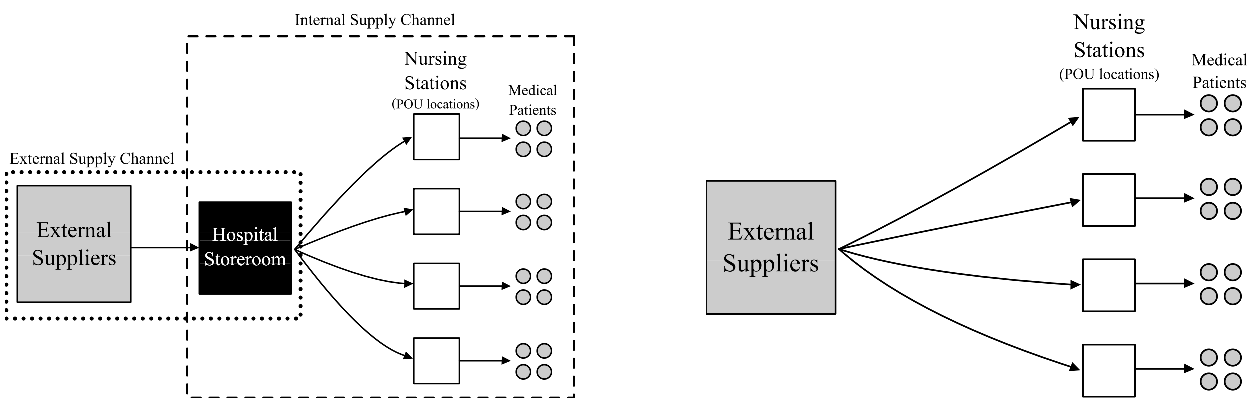
In our final article, we extend the work of the previous article as we consider a hybrid hospital supply chain with both regular and emergency ordering when inventory records are inaccurate. This model mimics what several hospitals in the United States have experienced first hand, where inaccuracy or other factors have forced them to operate two simultaneous supply chains. The unique structure presented here has not been modeled in the literature, and our solution approach is adapted from both inventory inaccuracy literature and from emergency replenishment literature. As such, this final paper serves as a bridge between two previously segregated streams of inventory research.

This final set of models follow the common theme of Poisson demand and base-stock replenishment, but the addition of an emergency ordering option requires that we use simulation both to compute cost and to optimize. Similar to the first article we employ a type of perturbation analysis, but by another name, marginal analysis, as the process to estimate cost differences in an approximate model. Through this approach, combined with simulation, we seek an optimal solution to both the approximate cost model and the simulated model.

3.1 INTRODUCTION

The hospital supply chain has received limited attention in the literature, though interest has been increasing due to recent political and social influences. In the past, hospital supply chains required external supply replenishment to an on-site warehouse, and internal replenishment to inventory stores in each patient care unit. Periodically, bulk orders were placed with the external supplier to replenish the warehouse, and then during each shift throughout every 24-hour day, material handlers moved supplies from the warehouse to point of use (POU) inventories within each care unit.

With the advent of new technology, and because of increasing costs for medical supplies, the 1990s saw hospitals in the US push to transform the structure of their supply chains (Rivard-Royer et al. (2002)). They moved from the classic structure to what is now termed a *stockless* supply chain structure. The stockless hospital supply chain relies on up-to-date inventory data from each POU inventory, with automated systems placing orders with the external suppliers. The stockless system also requires short replenishment lead times from the external suppliers. The figures below demonstrate both the classic and stockless hospital supply chain structures.



(a) "Classic" Structure (b) "Stockless" Structure

Figure 3.1: Hospital Supply Chains

Neve and Schmidt (2011) present an inventory model for a regional medical center that had recently transformed their supply chain structure. By relying on an automated ordering system to place external orders for each POU, the hospital was able to operate according to a *stockless* supply chain. Each POU followed a daily-review base stock inventory replenishment policy, and this required the external supplier to deliver orders within 24 hours.

However, inaccurate inventory records within the automated system required hospital staff to periodically count the physical inventory at each POU to reconcile the discrepancies at each location. Without addressing the inaccuracy, the system would fail to maintain reasonable inventories, leading to shortages and delayed medical care. The authors present both a service-level problem, and a cost minimization problem assuming backorder costs, with optimal solutions given for both cases.

In practice, however, the regional medical center is unable to successfully manage all their supplies according to a strictly stockless supply structure. Inventory space constraints at the POU, coupled with inventory record inaccuracy leads to continued shortages throughout the day without additional inventory support. Similar to other hospitals facing the same issues, the regional medical center uses a central storeroom for inventory to fill emergency orders when stock at the POU is dangerously low. There is still an automated replenishment system that places regular orders with the external suppliers on a daily basis, but now emergency orders can be placed manually each *shift* and arrive to the POU almost instantly. The updated supply chain structure can be termed a *hybrid* supply chain, with a similar structure denoted *hybrid stockless* supply chain by Rivard-Royer et al. (2002). Figure 3.2 shows an example of the hybrid supply chain structure.

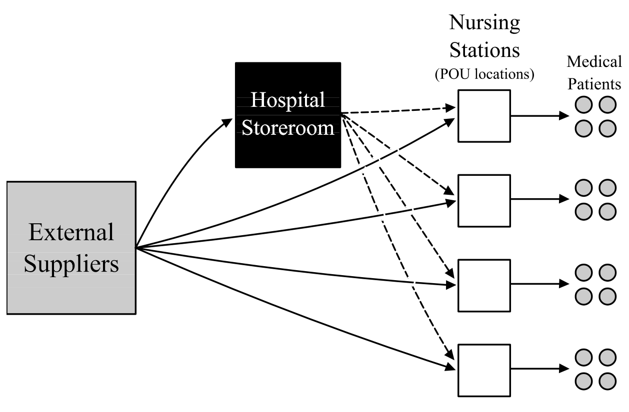


Figure 3.2: "Hybrid" Hospital Supply Chains

In the figure, the dashed-line arrows represent the emergency supply flow between the hospital storeroom and each POU. Here we assume that emergency orders are based on emergency order-up-to levels. The solid-line arrows from the external suppliers to each POU depict regular replenishment flow, and these daily regular orders are based on a regular order-up-to level, different from the emergency policy. In this paper, we present the inventory model used at the POU under the hybrid supply chain structure. We address the issue of inventory inaccuracy by allowing for physical inventory counts, while allowing for emergency orders to be placed at shift-intervals during the day.

The hybrid supply chain, as given in Figure 3.2, is the result of the regional medical center in Alabama that attempted to move from a *classic* hospital supply chain structure (Figure 3.1a) to a *stockless* structure (Figure 3.1b). However, due to inaccurate inventory records and lack of storage space at each POU, the system never fully transitioned to the stockless system. This paper is an extension of Neve and Schmidt (2011), where we allow for emergency orders.

A more detailed description of how the hybrid supply structure operates will be given in Section 3, as well as a general formulation of the model. But first, Section 3.2 will provide a review of the related literature. In Section 3.4 we provide the formulation and solution of an approximate cost version of the model, and in Section 3.5 we discuss the simulation approach for the main model. The results and analysis are given in Section 3.6 with a conclusion in Section 3.7. Included at the end of the paper, in Section C, is the Appendix which will be referred to throughout the course of the paper.

3.2 LITERATURE

The work presented here is unique to three streams of research, which can be generally characterized as the following areas: health care supply chain, inventory models with regular and emergency ordering, and inventory models with inaccurate records.

3.2.1 Healthcare Supply Chain

Papers like Young (1989) and McCone-Sweet et al (2005) address issues in the hospital supply chain from an empirical standpoint by surveying representatives from several hospitals. Young (1989) surveyed 22 Materials Managers regarding computer-based inventory control systems vs. manual systems, while McCone-Sweet et al (2005) focused on the general obstacles to improving the hospital supply chain by surveying 26 individuals who were either hospital Materials Managers or representatives of supply distributors. Both empirical studies cite a need for additional research regarding the hospital supply chain, as well as the complexity of such supply chain systems.

Other authors, such as Duclos (1993), Rivard-Royer et al (2002), DeScioli (2005), and Jan (2006) provide specific case studies where hospital performance is measured before and after an implementation of some supply chain modification.

In Duclos (1993), they use a simulation to study the effects of shock demand on the hospital in question, exploring operating conditions in a classic hospital supply chain. Rivard-Royer et al (2002) follow the implementation of a hybrid stockless hospital supply chain in a rural Canadian hospital.

Both DeScioli (2005) and Jan (2006) focus on the hospital supply chain, and some of the different automated inventory control systems in use. Both papers also deal with a single hospital case and do not formulate a generalized model, and neither paper addresses the issue of inaccurate records. Even these later papers again cite the lack of current research on the topics of supply chain and inventory as applied to health care.

Our contribution to the health care supply chain literature is the consideration of a POU inventory system where inaccurate records impact the availability of stock. Also, we allow for two supply modes at the POU level, which has also not been addressed from a modeling perspective in the hospital supply chain literature.

3.2.2 Inaccurate Inventory Records

Inventory literature dealing with inaccurate records was not found to have any models that allow for emergency ordering, nor to address issues in the health care industry. Our research is unique, then, if taking the perspective of inventory research dealing with inaccuracy.

Sheppard and Brown (1993) and Raman et al. (2001) are examples of empirical research focusing on inventory inaccuracy, while Fleisch and Tellkamp (2005) present a simulation study and share results and insights. While we do include a simulation approach, the model we introduce here is much different from the previous work, as previously discussed.

Rekik et al. (2008) present a model where the only the replenishment policy is modified to address inaccuracy, and the model in Sandoh and Shimamoto (2001) focuses on only improving the inspection policy. Iglehart and Morey (1972), Kök and Shang (2007), and DeHoratius et al. (2008) present models that are more like our research, where both the replenishment and the counting policies are adjusted to address inaccuracy in the records.

In Iglehart and Morey (1972), they modify the given replenishment policy by setting a buffer stock level to cover probable errors. In addition, they set the frequency and quality of the inventory counts but do not allow backorders and do not minimize cost.

Kök and Shang (2007) develop a model for inaccurate inventory where they make joint inspection and replenishment decisions to minimize system costs. The inspection decision is made using a threshold level, so that when the inventory record falls to that level it dynamically triggers an inventory inspection. A base-stock replenishment policy is shown to be optimal, and both the threshold level and the base-stock level are chosen based on the level of inaccuracy in the system.

DeHoratius et al. (2008) develop an interesting model based in a retail setting, where they propose a modified inventory record using a Bayesian approach. Their research is different from our own, as they assume inventory inaccuracy with lost sales (instead of backordering) as they attempt to optimize the auditing frequency and replenishment policy.

In our previous work, Neve and Schmidt (2011), we addressed inaccuracy by choosing both a counting frequency and a replenishment policy. It was the first example of an inaccuracy-related model considering both a cost minimization approach and an approach using a service level constraint. This paper is an extension of our previous work, where we allow for emergency ordering between regular replenishments.

3.2.3 Emergency Ordering

Within the inventory literature, there is an area of research dealing with multiple supply modes. More specifically, the addition of an emergency supply mode to supplement the availability of stock in an inventory system. Most of the articles considered here assume a base stock replenishment policy for both regular orders and emergency orders. The main difference between these periodic-review base stock inventory systems, is the timing of the emergency orders.

The models presented in Chiang and Gutierrez (1996), Lawson and Porteus (2000), and Scheller-Wolf et al. (2007) only allow for emergency ordering at the same time as regular ordering. The decision maker chooses between the supply modes, weighing both the shortened lead time of the emergency order and the lower cost of the regular order. Both Lawson and Porteus (2000) and Scheller-Wolf et al. (2007) allow for the placement of both types of orders (emergency and regular) each review period, but Chiang and Gutierrez (1996) requires the decision maker choose only one supply mode each period.

More like our own model, are the models presented by Chiang and Gutierrez (1998), Vlachos and Tagaras (2001), Teunter and Vlachos (2001), and Tagaras and Vlachos (2001). In each of the periodic-review base stock inventory systems, emergency ordering occurs during sub-periods between regular ordering. However, both Tagaras and Vlachos (2001) and Vlachos and Tagaras (2001) only allow for one emergency order between regular orders. Whereas, Chiang and Gutierrez (1998) and Tagaras and Vlachos (2001) allow for several sub-periods of emergency review and ordering, with Chiang and Gutierrez (1998) also allowing for emergency ordering at the same time as a regular review.

The research presented here is most like Tagaras and Vlachos (2001) in that we allow for emergency reviews in sub-periods between regular reviews. Our approximate cost model mimics much of what they presented, however we have the added issue of inventory inaccuracy, which has not been addressed in the multiple supply mode literature. Not only do we see that our approach is novel in the literature, but it is also timely and applicable. We have witnessed this type of inventory system in actual operation within the health care industry, including both inventory inaccuracy and emergency ordering. We continue now with our formulation of a general model.

3.3 GENERAL MODEL FRAMEWORK

As this paper is an extension of Neve and Schmidt (2011), our model is also motivated by the case of a regional medical center in Alabama, though we do not assume a strictly stockless structure as in our previous work. In this paper, we assume the hospital operates under a hybrid supply chain structure, as given by Figure 3.2. Time is divided into periods of 24-hour length, with three shift-length sub-periods defined for each 24-hour period. Notationally, we use a double subscript to designate each period by both the day and the shift.

For example, we define D_{ij} to be the total demand that occurs during shift i on day j. Similar to Neve and Schmidt (2011), we are also concerned with both unrecorded demand UD_{ij} and recorded demand RD_{ij}. Total demand, D_{ij}, is the sum of these demands as given below:

$$D_{ij} = RD_{ij} + UD_{ij} \tag{3.1}$$

We assume that total demand during a shift follows a Poisson distribution with rate λ_i for each shift $i = 1, \ldots, 3$. We can write that $D_{ij} \sim \text{Poisson}(\lambda_i)$.

The probability of a particular demand event being recorded is p, which should allow us to define the distributions of UD_{ij} and RD_{ij}. In actual practice, however, if there is no inventory available to fill a demand then it is backordered. When demand is backordered it is not, or cannot be, recorded in the automated system.

More specifically, the current automated system installed at the Alabama Hospital, as well as at other hospitals throughout the United States, is not designed to track shortages. As such, when inventory is unavailable but is demanded by regular patient care requirements, the backorders are handled manually by hospital staff. Our model attempts to mimic the actual operations of the inventory system using a similar automated replenishment system. Therefore, backorders are tracked using the actual net inventory, I, and are ignored by the recorded inventory level, $I^{'}$, both of which are introduced further below.

This implies that the actual distributions of UD_{ij} and RD_{ij} are dependent on the amount of inventory available when a demand occurs. Due to the complexity of modeling this system behavior, we develop approximate distributions for RD_{ij} and UD_{ij} that do not rely on the amount of inventory available. The detailed derivation of the approximation is found in the Appendix of the previous article, Neve and Schmidt (2011). From this approximation, we assume the following Poisson distributions for unrecorded demand and recorded demand:

$$RD_{ij} \sim \text{Poisson}(p\lambda_i) \tag{3.2}$$

$$UD_{ij} \sim \text{Poisson}\left((1-p)\lambda_i\right) \tag{3.3}$$

Using these approximations still supports the actual distribution of total demand, D_{ij}, given the relationship defined in Equation 3.1. We now turn our attention to how the different types of demand affect the system.

The automated replenishment system uses an electronic inventory record to track supply flow and place orders with an external supplier on a daily basis. This *recorded* inventory, at some arbitrary instant in time, is denoted $I^{'}$, and we also define it specifically at the end of some shift i of day j by $I^{'}_{ij}$.

The recorded inventory level, $I^{'}$, is depleted when demand is successfully recorded, or by recorded demand, RD_{ij}. For example, if the recorded inventory is $I^{'}$ at the beginning of some shift i of day j, then at the end of the shift (after demand occurs) the recorded inventory would be $I^{'}_{ij} = I^{'} - RD_{ij}$. Which means that $I^{'}_{ij}$ gives the recorded inventory level after demand occurs during shift i.

Due to the way demand is recorded and received by the automated system, *recorded* inventory is a non-negative performance measure. As such, there is not a recorded *net* inventory defined for this inventory system, nor are there recorded backorders. Instead, we use the *actual* net inventory to track the actual state of available inventory. Actual net inventory for some arbitrary instant of time is given by I. Similar to how we defined recorded inventory, we define *actual* net inventory specifically at the end of each shift i of each day j by I_{ij}.

Actual net inventory is depleted each shift i of day j by both recorded demand, RD_{ij}, and unrecorded demand, UD_{ij}. Here, I_{ij} gives the value of actual net inventory in shift i of day j, *after* demand occurs. Since total demand is the sum of unrecorded and recorded demand, we use total demand in a shift, D_{ij}, to impact the actual net inventory, I.

Backorders at the end of a shift are given simply by the negative part of actual net inventory, $[I_{ij}]^-$, and the end-of-shift available inventory is given by the positive part, $[I_{ij}]^+$. We apply the per-unit cost of holding inventory, c_h, and the per-unit cost of backordering, c_b, to their respective values at the end of each shift.

Replenishment occurs along two channels - regular and emergency, and each replenishment channel uses a separate performance index. Regular replenishment is managed by the automated replenishment system, and it uses the recorded inventory level, $I^{'}$, for its performance index. Emergency replenishment is a manual process, relying on the actual net inventory, I, as the performance index.

Regular orders are placed, after incoming orders are received, at the beginning of the first shift of every day. As regular replenishment follows an order-up-to policy with base stock level S, we can define the order quantity to be $RO_j = \left[S - I^{'}\right]^+$. Here, we are assuming that $I^{'}$ is the recorded inventory level at the beginning of the first shift of day j, after any outstanding orders are received. Note that the lead time for regular orders is assumed to be one day, or three shifts. Thus, an order of size RO_j will arrive at the beginning of the first shift of day $j + 1$.

Emergency orders are placed only at the end of the first and second shifts each day. They are based on emergency order-up-to levels, with E_i being the order-up-to level applied to the end of shift $i = 1, 2$. We define the emergency order quantity to be $EO_{ij} = [E_i - I_{ij}]^+$, and it is placed at the end of shift i. The lead time is practically zero for emergency orders, so they are given to arrive at the beginning of the shift immediately following the order. So, an order of size EO_{ij} will arrive at the beginning of shift $i + 1$.

The cost of emergency orders is addressed in a similar fashion as Teunter and Vlachos (2001) and Tagaras and Vlachos (2001). Basically, we assume a fixed per-unit emergency ordering cost, c_e, that is applied to each order. Technically, there is a similar cost for regular orders, but here we just define c_e to be the per-unit difference in cost between regular and emergency orders.

To simplify, we refer to c_e as the emergency ordering cost, per unit ordered.

The automated system, since it does not see backorders, may see an incoming order that is less than what was expected. To illustrate, consider the system state when there are outstanding backorders, that is $I < 0$. Now, suppose that the an order of size RO_{j-1} was placed the previous day. When that order arrives, it will fill some or all of the backorders before it is received into on-hand stock, depending on whether $RO_{j-1} > [I]^-$. The automated system will only "see" incoming stock if it is more than sufficient to fill the outstanding backorders. In other words, the automated system will receive the quantity $[RO_{j-1} - [I]^-]^+$ for incoming regular orders and $[EO_{i-1,j} - [I]^-]^+$ for incoming emergency orders.

Lastly, we need a way to address the inaccuracy that is embedded in this model. As this is an extension of Neve and Schmidt (2011), we assume the same type of physical inventory count to reconcile the discrepancies in the automated system. We will therefore define N to be the number of days between physical inventory counts, where we attempt to minimize the average daily cost over a cycle of N periods.

Our planning cycle begins one day following a reconciliatory inventory count and concludes at the end of the N^{th} day of the cycle. It is important to note that the these physical inventory counts occur at the beginning of the first shift of the day, *after* the incoming order arrives, and *before* the regular order is placed. At that instant, $I^{'} = I$ if $I \geq 0$ and $I^{'} = 0$ if $I < 0$, so that recorded inventory is perfectly aligned with actual net inventory if there are no backorders.

Having defined both regular and emergency ordering, an important assumption is that an optimal policy requires $S \geq E_1 \geq E_2 \geq 0$. This monotonicity property of the optimal solution was also assumed by Teunter and Vlachos (2001) and shown by Chiang and Gutierrez (1998). Here, we will be seeking near optimal policies as did Teunter and Vlachos (2001), and will focus our attention on solutions that follow the same structure.

3.3.1 Simple Illustrative Example

As stated previously, our planning cycle consists of N days of operation, where N is the number of days between reconciliatory inventory counts. We have already explained the timing of inventory counts, as well as the timing of regular and emergency orders and their respective arrivals. The figure below demonstrates the normal timing of events for an example case when $N = 3$ days.

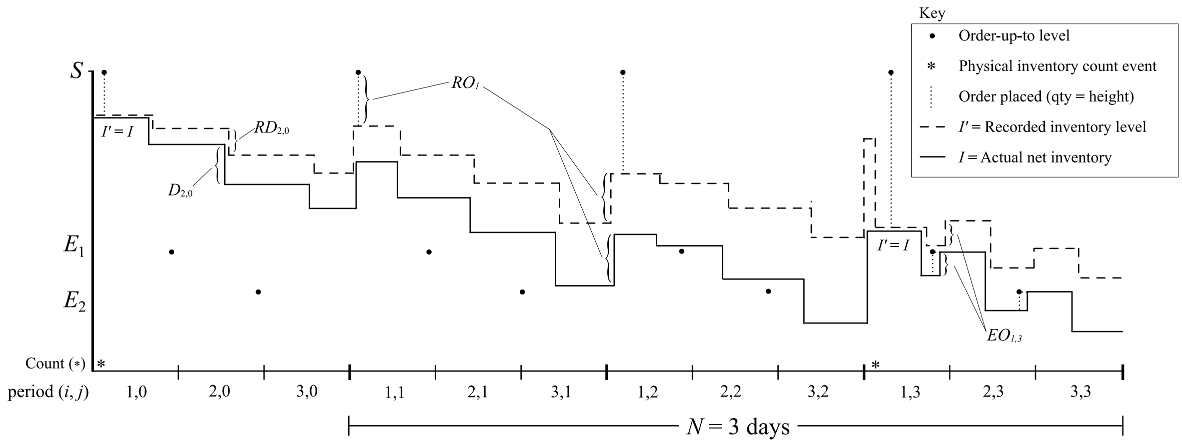


Figure 3.3: Inventory Event Time Line

Periods are labeled according to shift i and day j, denoted i, j. Note in Figure 3.3 how counting occurs at the beginning of shift 1 of day zero, and again at the beginning of shift 1 of day $N = 3$. More generally, our planning cycle begins at the start of day 1 and ends at the conclusion of day N, for any N. Also note how recorded inventory, I', drifts further away from actual net inventory, I, until counting occurs at the beginning of shift 1 on day 3. The graph is labeled $I' = I$ where counting occurs as the inventory records are reconciled to match actual net inventory.

For illustrative purposes, the total demand, $D_{2,0}$, and recorded demand, $RD_{2,0}$, are labeled in shift 2 of day 0. This indicates how demand influences the inventory performance measures during the period of time covered on the graph.

The emergency and regular order-up-to levels, $S \geq E_1 \geq E_2 \geq 0$, are labeled on the y-axis and are indicated on the graph as solid dots. Using those marks, notice that the regular order (up to S) is placed at the beginning of day 1, labeled RO_1, and it also indicates the arrival of the order at the beginning of the following day. Finally, an emergency order (up to E_1) is also indicated on the graph as $EO_{1,3}$, occurring at the end of shift 1 of day 3 and arriving at the beginning of shift 2 of day 3.

The original emergency replenishment model, as presented by Teunter and Vlachos (2001), was stated to be practically infeasible to solve analytically. The added complexity of dealing with inaccuracy only increases the difficulty in addressing the problem analytically. We therefore follow the same course as the previous work and introduce an approximate cost version of the model.

3.4 APPROXIMATE COST MODEL

In our approximate cost model, we make some simplifying assumptions that allow us to derive optimality conditions. We later compare the resultant approximate solutions to the solutions from a simulation model of the same system.

Our main assumption is that emergency ordering does not occur until the final day of the cycle. This also means that we assume that no emergency ordering occurs during the day before our planning cycle. Similar assumptions are made by Teunter and Vlachos (2001) and Tagaras and Vlachos (2001), though our planning cycle is constructed differently. Our lead time is equal to the length of the regular review (a special case of the model from Teunter and Vlachos (2001) when $k = 2$), and our planning cycle may contain several regular review periods (an extension of the previous work).

We will explore whether this is a satisfactory approximation for the complete model, as found in the previous work. We follow the same approach as Teunter and Vlachos (2001) in constructing the optimality conditions for the approximate model. We will condition on the different states of the system and recursively solve for E_2^*, E_1^*, and S^*, in that order.

We begin by assuming that inventory is counted every N days, so our planning cycle is also N days. We know that, at the beginning of the day preceding our cycle, a physical inventory count reconciled the recorded inventory level so that $I' = I$ at that time. We also know that no emergency orders were placed during the day preceding our cycle, due to our approximation assumption.

When the order arrives at the beginning of shift 1 of day 1 of the cycle, we know the initial distribution of inventory parameters. That is, the starting recorded inventory in shift 1 of day 1 is given by: $I' = S - (RD_{1,0} + RD_{2,0} + RD_{3,0})$. Similarly, actual net inventory at the beginning of shift 1 of day 1 is given by $I = S - (D_{1,0} + D_{2,0} + D_{3,0})$. From these starting points, we can begin our recursive derivation of the expected incremental costs.

It should be clearly remembered, that we have assumed that emergency orders do not occur until the final day of our planning cycle, day N (see Figure 3.3). We go to the end of the planning cycle to begin our recursive process of finding E_2^*, E_1^*, and S^*, in that order. We further assume that $S \geq E_1 \geq E_2 \geq 0$, similar to Teunter and Vlachos (2001).

We define the general Poisson cmf and pmf for ease of notation throughout this discussion:

$$G(x;\omega) = \sum_{k=0}^{x} \frac{e^{-\omega}\omega^k}{k!} \tag{3.4}$$

$$g(x;\omega) = \frac{e^{-\omega}\omega^x}{x!} \tag{3.5}$$

3.4.1 Finding E_2^*

Using marginal analysis, we will compute the expected incremental cost used for the optimality condition to find E_2^*. We do not need to know E_1^* or S^* since we condition on the event of an emergency order being placed at the end of the second shift up to E_2. Note that if an emergency order is not placed, then changing the value of E_2 will have no impact on the marginal costs.

Because we condition on there being an emergency order at the end of the second shift, we know that actual net inventory at the beginning of the third shift is $I = E_2$. At the end of the third shift, after demand occurs, there may or may not be backorders at that time. We consider both possibilities as discrete cases, and then derive the expected incremental cost of increasing E_2. The cases are denoted as follows:

Case 0α: There are NOT backorders at the end of shift three

Case 0β: There ARE backorders at the end of shift three

For each case we can derive probabilities, denoted $P_{0\alpha}$ and $P_{0\beta}$, and then develop the incremental costs for each case. The probabilities are given for each case below, and their derivation can be found in the Appendix.

$$P_{0\alpha} = Pr[D_{3,N} \leq E_2] = G(E_2;\lambda_3) \tag{3.6}$$

$$P_{0\beta} = Pr[D_{3,N} > E_2] = 1 - G(E_2;\lambda_3) \tag{3.7}$$

The incremental cost must also be defined for each case. For case 0α, if E_2 were increased by one unit, then there would be an additional unit ordered using the emergency channel at a cost of c_e, and there will be an additional unit on-hand at the end of the third shift at a cost of c_h. Thus, the incremental cost for this case is $(c_e + c_h)$. The same line of reasoning is used to develop the incremental cost for case 0β, which is given by $(c_e - c_p)$, and a more detailed description can be found in the Appendix.

The expected incremental cost is simply the probabilities for each case multiplied by the incremental costs for each case, summed over all cases. We denote this expected incremental cost for E_2 as C_{E_2}, and its reduced form is given below.

$$\begin{aligned} C_{E_2} &= (c_e + c_h)P_{0\alpha} + (c_e - c_p)P_{0\beta} \\ &= c_e - c_p + (c_h + c_p)G(E_2; \lambda_3) \end{aligned} \tag{3.8}$$

Because it is the expected incremental cost, Equation 3.8 gives the optimality condition for E_2^*. The optimal value is found by setting E_2^* to the smallest value for E_2 where the inequality holds: $C_{E_2} \geq 0$. This can also be written in a more familiar format, where we set E_2^* to the smallest value for E_2 such that the following holds:

$$G(E_2; \lambda_3) \geq \frac{c_p - c_e}{c_h + c_p} \tag{3.9}$$

3.4.2 Finding E_1^*

Now that we can find E_2^*, we use the outcome to find the emergency order-up-to level for the first shift. We will condition on the event that an emergency order is placed up to E_1 at the end of the first shift. Then we define specific cases, or system states, conditioning to find an expected incremental cost. We will use the expected incremental cost for the optimality condition and find E_1^*.

At the end of the first shift of the last day of the cycle, we assume that we are placing an order up to the level E_1. However, we do not know if we will need to place an emergency order at the end of the second shift (as discussed previously). As a result, we now have four cases to consider in our development of the expected incremental cost for E_1, and they are denoted:

Case 0α: NO emergency order in shift two, NO backorders in shift three

Case 0β: NO emergency order in shift two, there ARE backorders in shift three

Case 1α: Emergency order in shift two, NO backorders in shift two

Case 1β: Emergency order in shift two, there ARE backorders in shift two

In the Appendix, we develop the probabilities and incremental costs for each of the four cases in detail. Again, combining the probabilities and incremental costs gives the expected incremental cost over E_1, and it is given below, denoted C_{E_1}.

$$C_{E_1} = (c_h + c_p) \sum_{k=0}^{E_1 - E_2^*} G(E_1 - k; \lambda_3) g(k; \lambda_2) + (c_e - c_p) G(E_1 - E_2^*; \lambda_2) + (c_h + c_p) G(E_1; \lambda_2) - c_p \tag{3.10}$$

Again, using the optimality condition provided by the expected incremental cost, we set E_1^* to the smallest value of E_1 where $C_{E_1} \geq 0$. In this case, the expected incremental cost does not simplify as cleanly as the previous case.

Although we do not attempt to prove it in this paper, we have strong numerical evidence that suggests Equation 3.10 is non-decreasing in E_1, as is Equation 3.9. In fact, the problem instances considered in our analysis, as well as additional cases that were explored, all support the idea of a non-decreasing structure for the incremental cost equations over the respective variables, E_2 and E_1.

3.4.3 Finding S^*

We now use E_2^* and E_1^* in our derivations to find S^*, with the derivation steps shown in detail in the Appendix. The derivations differ from the previous approach for emergency orders, since the automated system affects inventory every day of the cycle. Here, we define inventory at the beginning of the final day of the cycle, and then introduce the cases associated with finding S^*.

During the time leading up to the final day of the planning cycle, net inventory will increase with the arrival of regular orders based on recorded demand, $RD_{i,j}$, while total demand, $D_{i,j}$, will decrease net inventory. We can say that actual net inventory at the beginning of period N, denoted in this instance by I_N, is given below.

$$\begin{aligned} I_N &= S - \sum_{j=0}^{N-1} \sum_{i=1}^{3} D_{i,j} + \sum_{j=0}^{N-2} \sum_{i=1}^{3} RD_{i,j} \\ &= S - \left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} \right) \end{aligned} \tag{3.11}$$

Since the total demand, $D_{i,j}$, and unrecorded demand, $UD_{i,j}$, are random variables from independent Poisson distributions, we can define the following distribution from the above equation:

$$\left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \right) \sim \text{ Poisson} \left((N-1)(1-p) \sum_{i=1}^{3} \lambda_i + \sum_{i=1}^{3} \lambda_i + \lambda_1 \right) \tag{3.12}$$

For ease of reference, we define $\mu = (N-1)(1-p)\sum_{i=1}^{3} \lambda_i + \sum_{i=1}^{3} \lambda_i + \lambda_1$. With this defined, we now proceed with the cases used to find the expected incremental cost for S. With I_N as the starting point, we look at the final day of the cycle. The cases considered are as follows, where emergency orders occur during this last day of the cycle:

Case 0α: NO emergency order in shifts one and two, NO backorders in shift three

Case 0β: NO emergency order in shifts one and two, there ARE backorders in shift three

Case 1α: First emergency order in shift two, NO backorders in shift two

Case 1β: First emergency order in shift two, there ARE backorders in shift two

Case 2α: First emergency order in shift one, NO backorders in shift one

Case 2β: First emergency order in shift one, there ARE backorders in shift one

The derivations for the probabilities and incremental costs are in the Appendix, however we give the expected incremental cost below:

$$
\begin{aligned}
C_S = {} & (3N-3)c_h - c_e - c_p + (c_h + c_p)G(S;\mu) - (c_p)G(S - E_1^*;\mu) \\
& + (c_h + c_p)\sum_{x=0}^{S-E_1^*} G(S - x;\lambda_2) \cdot g(x;\mu) \\
& + (c_e - c_p)\sum_{x=0}^{S-E_1^*} G(S - x - E_2^*;\lambda_2) \cdot g(x;\mu) \\
& + (c_h + c_p)\sum_{x=0}^{S-E_1^*} \sum_{y=0}^{S-x-E_2^*} G(S - x - y;\lambda_3) \cdot g(y;\lambda_2) \cdot g(x;\mu)
\end{aligned}
\tag{3.13}
$$

We set S^* to be the smallest integer S where $C_S \geq 0$.

Again, we explored for each of the problem instances considered here whether the incremental cost equation, C_S was non-decreasing in S for given values of N. We found in these cases that the C_S was indeed non-decreasing for all values of N that were considered. The proof of this model structure will not be attempted here.

So, we have an approach to find the optimal solution to the approximate cost version of the model, for a given number of days between physical inventory counts, N. However, it should be noted, as was noted in Tagaras and Vlachos (2001) where a similar structure was discussed, that for extreme values for the cost parameters the structure of the model may be adversely affected. We discuss the sensitivity to the shortage cost in the Appendix.

3.4.4 Solving the Approximate Model

To solve the approximate cost model, we employ C++ to compute the incremental costs for given parameters and variables. We begin by incrementing E_2 from zero until the above optimality conditions hold for C_{E_2}. Based on that solution, E_2^*, we will increment from $E_1 = E_2^*$ until E_1^* is found to satisfy the optimality condition for C_{E_1}.

We would like to take the same approach for C_S to find S^* for a given N. However, the incremental costs defined for the approximate cost model do not take into consideration the fixed cost to count inventory, k. So, we need a way to compute overall costs to know when we have found a cost-minimizing cycle length, N. In the next section, we discuss a simulation approach that allows us to estimate the total cost of a given solution.

Once we have a way to compute a cost estimate through simulation, we can increment S from $S = E_1^*$ until the optimality condition for C_S is met for a given N. Then, based on the cost estimate, we will increment N from $N = 1$, computing E_1^*, E_2^*, and S^* at each step, until we see a minimum cost. We will address this issue in the next section, but it is important to note that we are assuming the cost function is basically unimodal in N when we find E_1^*, E_2^*, and S^* for each value of N. All of the cases considered here numerically support the assumption regarding the unimodality (in N) of the approximate cost model.

Later, we use the solution from the approximate cost model as the starting point for the simulation-optimization approach presented in the next section.

3.5 SIMULATION APPROACH

The approximate cost version of the model by itself has no way to minimize overall cost in a way that includes the fixed counting cost. Also, the assumptions of the approximate cost model that 1) emergency orders can only be placed in the last day of the cycle, 2) that no emergency orders are placed the day before the cycle, and 3) that recorded demand is approximate, are unreasonable assumptions in practice.

Therefore, in this section we develop a simulation approach to provide a way to compute cost and to model a more realistic inventory system with relaxed, more realistic assumptions. To minimize the relaxed model, we will utilize simulation-optimization to generate minimizing solutions.

In simulating our model, we want to closely represent the actual operations of the inventory system in use. We relax the assumptions made in the approximate cost model, so that emergency ordering can occur in any day of the cycle. Naturally, this also means that we allow for emergency ordering during the day before the beginning of the planning cycle of N days. Furthermore, we will be generating the recorded demand, $RD_{i,j}$, based on available inventory and the probability, p, rather than using an approximate distribution as in the previous model.

The simulation is used to estimate the expected daily cost over the N days of the planning cycle. In order to get a "good" estimate for the expected daily cost for a specific set of parameters, we are required to use several simulation runs of terminating N-day planning cycles. This means that beginning inventory on day 1 will have great influence on the outcome of each simulation replication. To address this starting inventory level, we follow the approach suggested by Law and Kelton (2000). We utilize a warm up period so that the beginning inventory level in each cycle achieves a "typical" range of values.

In addition to a warm-up period, we must simulate enough cycles (in steady state) of length N over which to find the average expected daily cost with sufficiently small error. These strategies are necessary to reduce the error of the estimate for expected daily cost for a given set of parameters. We continue with our method of carrying out these strategies.

3.5.1 Basic Simulation

We define l to be the number days in the simulation warm-up period. Using Welch's method as suggested by Law and Kelton (2000), pp. 518-25, we find after $l = 60$ days that the simulation has reached a steady state. So, when simulating expected daily cost over some cycle length of N days, we require $\lceil l/N \rceil = \lceil 60/N \rceil$ cycles for the warm-up period. Basically, we are transforming the number of *days* of warm up needed into the number of *cycles* of warm up, given the decision variable value for cycle length, N. We use ceiling notation to show that we round up the needed value for warm up length. Once the simulation has generated $\lceil l/N \rceil$ cycles, we begin recording data from each simulation replication to estimate cost. However, we still require enough replications to achieve a certain estimate error.

We define n_0 to be the initial number of cycles simulated (after the warm-up period) for each particular set of parameters and decision variables. We use the cost results from the first n_0 observations to estimate the appropriate sample size, n, required to achieve an error of $\epsilon \leq 0.1$ at a 95% level of confidence.

The basic simulation thus requires the following parameters and decision variables:

Parameters

λ_i : The Poisson rate of total demand during each shift, $i = \{1, 2, 3\}$

p : Probability that a demand event is successfully recorded

k : Fixed cost of physically counting inventory to reconcile inaccuracy

c_e : Emergency ordering cost per unit ordered through emergency channel

c_h : Holding cost per unit, per shift

c_p : Backorder cost per unit backordered, per shift

l : The number of days required for the Warm-up phase

Decision Variables

S : Regular order-up-to level applied at the beginning of each day

E_i : Emergency order-up-to level applied to the end of shift $i = \{1, 2\}$

N : Number of days in each cycle

After inputing the parameters and variables necessary to run, the simulation will conduct three phases, or loops, of discrete-event simulation. They are as follows:

Initialization Phase: Set $I' = I = S$, as though inventory had been counted, and then simulate one day, tracking both recorded inventory and actual net inventory.

Warm-up phase: Using the results of the Initialization phase, simulate the next $\lceil l/N \rceil$ cycles, again tracking inventory measures while ignoring costs.

Replication phase: Starting from where the Warm-up phase completed, the Replication phase simulates the next $n_0 = 50$ cycles, computing and recording expected daily cost over each of the n_0 cycles. At the end of the first n_0 cycles, we compute the average daily expected cost and the variance, given by $\bar{C}(n_0)$ and $S^2(n_0)$, respectively. We then find the appropriate n so that after sampling the next $n - n_0$ cycles, we can claim with 95% confidence that the expected cost over a cycle of length N is within an error of $\epsilon = \pm 0.01$.

A more detailed simulation algorithm is found in the appendix which shows the step-by-step progression of each phase of the basic simulation. In the next section we discuss the solution methodology, where a solution from the approximate model is used as the starting point for the simulation-optimization approach.

3.5.2 Simulation Optimization

Before we can begin optimization, we need a good solution for a starting point. So, to generate a starting solution for the simulation, we optimize the Approximate model, discussed in the previous section, over values of N, computing the cost, $\bar{C}^*(N)$, via the simulation approach described above. For all test cases considered here, the cost generated from the Approximate Model solution, $\bar{C}^*(N)$, is unimodal in N. The minimum cost generated from this approach is given by $\bar{C}^*(N^*)$, and we use the solution as the starting point, given by S^*, E_1^*, E_2^*, and N^*.

After generating the starting point, we employ simulation-optimization to seek an optimal or near-optimal solution to the model. We want to minimize the average of the expected daily cost over the n replications required for each set of parameters and variables.

Basically, we use a neighborhood search procedure, generating n simulation replications to compute the expected daily cost for each point in the immediate neighborhood of the previous best solution. This approach is basically a direction-of-steepest-descent heuristic. We continue by generating costs for all neighbors of a point (i.e. unit-valued integer perturbations over the variables N, S, E_1, E_2), and then choose the lowest cost point to be the center of the next neighborhood search.

We assume for each decision variable, x, that we have only $\Delta x = -1, 0$, or 1 - where x represents any of the four decision variables. This requires that we generate at most $3^4 - 1$ costs from the neighborhood surrounding each point, as the simulation will have already generated the center point. Previous solutions for a given set of decision variable values will be stored, so that the simulation will not have to re-run to generate those values again. To further reduce computation time, we apply our previously defined assumption that for the optimal solution, $S' \geq E_1' \geq E_2' \geq 0$, which also decreases the number of required neighborhood points to consider.

This approach continues until there is no adjacent neighbor having a lower cost, at which point the simulation ends, returning the minimized total expected cost, $\bar{C}'$ and its respective decision variable values, S', E_1', E_2', and N'. (the simulated optimal solution).

In the next section we will present the analysis of the results of the approximate cost model and the simulation-optimization model.

3.6 NUMERICAL RESULTS

The hospital provided data regarding the demand rates, inventory accuracy levels, and other operational processes to aid us in our choosing realistic demand and accuracy parameters for the computational analysis. For the costs however, each POU inventory location may have more than a hundred items in stock with widely varying respective costs. Therefore, we arbitrarily selected holding, ordering, and backorder costs that would provide sufficient insight regarding the model's structure and performance. For the fixed counting cost, we tried to use values that represent the hospital's actual counting process. The specific problem instances are given next, followed by an analysis of the results.

3.6.1 Problem Instances

We will therefore be concerned with accuracy levels of $p = .55, .70, .85, .97$. These rates are a good representation of actual values seen in the operation of a real-world hospital.

Similarly, we define demand rates for each shift of each day, as it is highly possible that demand rates change throughout the day. We consider five demand patterns in our analysis, demand increases throughout the day, demand decreases through out the day, two cases of uneven demand with spikes of demand in either the second or third shifts, and demand that is equally distributed throughout the day. Notationally, this means:

Case 1	Case 2	Case 3	Case 4	Case 5
$\lambda_1 = 3$	$\lambda_1 = 8$	$\lambda_1 = 5$	$\lambda_1 = 5$	$\lambda_1 = 5$
$\lambda_2 = 5$	$\lambda_2 = 5$	$\lambda_2 = 8$	$\lambda_2 = 3$	$\lambda_2 = 5$
$\lambda_3 = 8$	$\lambda_3 = 3$	$\lambda_3 = 3$	$\lambda_3 = 8$	$\lambda_3 = 5$

Choosing realistic cost estimates for both holding and emergency ordering costs can also be based on the data provided, but shortage costs are difficult to estimate. This is especially true in a hospital setting, where lack of a particular supply can have adverse affects on a particular patient's health. If a patient receives poor treatment as a result of a missing supply, the cost to the hospital can increase substantially under such circumstances - especially if a patient needs additional care or seeks monetary settlement through legal channels.

Thus, as our model applies to the hospital inventory system, we can justify a shortage cost that is higher than either the emergency ordering cost or the holding cost. Generally, we can also argue that the emergency ordering cost is higher than the holding cost, $c_e > c_h$. We will therefore assume that $c_p > c_e > c_h$ in most of our numerical studies. We do however, consider one case where $c_e > c_p$.

There is also the fixed cost, k, to count inventory and reconcile the *recorded* inventory level with the *actual* inventory level. This reconciliation occurs once per cycle of length N days. At the hospital, counting inventory for reconciliation requires either one or two full-time employees. We therefore consider two values for k that roughly represent either one or two entry level employee wages for about 70% of a working day.

In total, we considered 100 problem instances in our numerical study, where each of the five demand patterns were paired with 20 combinations of the following cost values:

$$c_h = 0.1,\ 0.3$$

$$c_p = 1,\ 3,\ 6$$

$$c_e = 1,\ 3$$

$$k = 30,\ 60$$

A complete list of all 100 problem instances is shown in the appendix in Table **??**.

3.6.2 Analysis of Results

In our analysis we would like to first compare the approximate cost solutions with the simulation-optimization cost solutions. Due to the assumptions made for the approximate cost model, we expect to see significantly higher costs than the simulation-optimized solution.

We will also look at how the cost, demand, and accuracy parameters affect the resultant solutions from both the approximate model and the simulated model. In considering all these analyses, we frame the discussion with management insights as applied to inventory systems with inaccurate records and emergency ordering.

Table 3.1: Processing Times

	Processing Time (Minutes)				
	Total	Min	Average	Max	Instances
Approximate Cost Model	12.53	0.05	0.13	0.35	100
Simulation Optimization Model	894.27	1.17	8.94	24.28	100

Computations were performed on a 64-Bit Windows Vista Home Premium OS PC, with 4 GB of RAM, an AMD Turion X2 Dual-Core Mobile RM-70 2.0 GHz processor. The time spent finding the solutions for both approaches is given below:

In the following analysis, we will first consider the influence of all parameters on the magnitude of the differences between the starting solution (provided by the approximate cost model) and the simulated optimal solution. Then, we look at the overall influence of demand, cost, and accuracy parameters on both the minimum costs and the solution values of the decision variables.

3.6.3 Approximate Model vs. Simulation Model

In every case, the simulation-optimization approach improved from the approximate cost model starting solution by at least 3.6%, up to an improvement of at most 62.3%. The median cost reduction was 22.3%, and three-fourths of the problem instances saw a reduction of cost by at least 10% over the approximation-generated solution.

Some larger improvements made by the simulation-optimization stem from the nature of the optimality conditions for the approximate cost model, which are sensitive to the counting frequency, N, and the shortage cost, c_p. As mentioned in Section 3.4, the optimization of the approximate cost model may be hindered in some cases of extreme values for c_p or N. A detailed analysis regarding the relationship between the optimality conditions and the shortage cost, c_p, can be found in the Appendix.

Some parameters have a greater influence on the ability of the simulation-optimization to find improvements. In just considering the main effects, we find that the demand pattern (i.e.

Case 1, Case 2, etc. as defined above), the shortage cost, c_p, the counting cost, k, and the accuracy level, p, are the most influential.

First, we found that a higher level of accuracy, p, provided the most influence on opportunity for improvement over the approximate cost solution. When $p = .55$, there is an average improvement of only 7.5%. As p increases from $p = .55$ to .70, .85, and .97, average improvement increases to 15.1%, 33.4%, and 55.2% respectively. This shows that the approximate cost model is a better approximation of actual system performance when the accuracy level is low.

This result is counterintuitive, as the higher accuracy would make it less likely to utilize emergency ordering during the course of the cycle. In that case, it would seem that both the approximate model and the simulation-optimization would use emergency ordering in a similar fashion. However, only the simulation model shows a sensitivity to accuracy in N.

In fact, the approximate cost model saw very little difference in cycle lengths over parameter p, which might explain the parity between models for low levels of accuracy. Since N is much lower (i.e. counting more frequently) for low accuracy levels in both models, the systems seem to be behave similarly. However, since higher p does not indicate significantly higher N for the approximate model, the simulation model is able to show greater cost improvements as it allows for much larger values of N.

The other factors, c_p, k, and the demand pattern, had more influence than the rest of the factors, but were still nominally influential in comparison to the influence of the accuracy level, p.

Looking at the demand patterns, we see that when there is a spike in demand in the third shift (i.e. Case 1 and Case 4), the amount of improvement between models decreases by about 2%. This can likely be attributed to the fact that emergency ordering is not allowed at the end of the third shift in either the approximate model or the full simulated model. In other words, not allowing for third-shift emergency orders affects both models in a similar fashion, leading to less likelihood of improvement opportunities.

This means that both inventory systems are required to place additional emergency orders at the end of the second shift, or similarly, the regular orders that are placed at the beginning of the following shift must be large enough to carry inventory until the spike occurs in the third shift. In either event, that is, increased emergency ordering or increased inventory, both models will likely see similar issues due to structural constraints. Later, we discuss the effects of demand patterns on emergency order-up-to levels, providing evidence to support this idea.

Lastly, the costs c_p and k both influence the ability of the simulation optimization to improve upon the solution to the approximate model. For larger c_p the improvement is not as successful as for smaller values. More specifically, when c_p is doubled from $c_p = 3$ to 6, the average improvement percentage is reduced by nearly eight percentage points (to about 22$ average cost improvement). As we have assumed that in most cases $c_p \geq c_e$, we know that the minimizing solution in both cases should favor emergency orders over shortages. However, we do not have enough numerical evidence to support any hypotheses regarding this phenomena.

For the fixed counting cost, k, we see a differences in improvement from the approximate cost solution. This ties to the inability of the approximate cost model to fully utilize a longer counting cycle. By not allowing for emergency ordering, the approximate cost model tends to favor shorter cycle times to reduce shortage costs. So, for larger k, the simulation model is better able to reduce total costs and increase cycle length, as it allows for emergency ordering any day during the cycle.

Next, we consider the overall effects of the cost, demand, and accuracy parameters on both decision variables and costs.

3.6.4 Influence of Cost Parameters

The influence of cost parameters relating to minimum costs are, in a large part, presented above in the discussion about the simulation model's improvement performance. What follows is a look at how the cost parameters influence the decision variables of the minimum-cost solutions for each model.

First, we consider the holding cost, c_h, and how it impacts the different decision variables. When holding costs are higher, the system should try to hold less inventory, meaning a lower S, which could possibly affect cycle length (a shorter cycle length to maintain a lower inventory level). It is counter-intuitive, then, that the simulation model does not see significant changes in the counting frequency, N, for differing values for the holding cost, c_h. The approximate cost model, however, does show reductions in N for larger values of c_h. Later, we see stronger influence on N by shortage and emergency costs.

The holding cost seems to have similar influences on S in both models. For higher holding costs, S is significantly lower in both models. A similar effect is seen on the emergency order-up-to levels, where higher c_h implies slightly lower E_1 and E_2 in both models.

The greater influence on emergency order-up-to levels is the emergency ordering cost, c_e, and the shortage cost, c_p. In both models, higher c_e leads to significantly lower order-up-to levels, while higher shortage cost implies significantly higher emergency order-up-to levels. This is an expected trade-off, where increased emergency ordering is used to reduce shortages when emergency ordering is less expensive than shortages. The converse is true when emergency ordering is more costly than shortages. In such a case, shortages are allowed more frequently because it is too expensive to order through emergency channels. The numerical data from both models support this idea.

In the approximate cost model, we find that c_e and c_p have little influence on the regular order-up-to level S, and only a slightly greater impact in the simulation model. In fact, there is actually a slight decrease in the regular order-up-to level S when emergency order costs, c_e, are greater. It is important to recall, that the emergency order cost, c_e, is only larger for the case when $c_e > c_p$. In such a case, more shortages are allowed at the reduced cost level (i.e. S is smaller), which decreases the need to place emergency orders which are at an increased cost level.

The last effect from the shortage and emergency ordering costs is on the cycle length, N. As expected, for higher shortage and emergency ordering costs, N is reduced to ensure that counting and reconciliation occurs more frequently. Doing so allows the system to guard against increased shortages and increased emergency ordering that stem from inaccurate records that normally grow worse over time.

The cost of counting, k, should have a significant impact on the length of the cycle, N.

The approximate cost model was unable to successfully utilize longer cycle lengths, due to the sensitivity of the optimality conditions to extreme values of N. As a result, N was not shown to be influenced significantly by counting cost k in the approximate cost model. The simulation model, however, saw the most substantial increase in cycle length, N, over values of counting cost, as one would expect. Counting cost, k, had very little effect on the other decision variables.

3.6.5 Influence of Demand Parameters

We have grouped demand parameters into five distinct Cases representing daily demand patterns, and we frame our analysis accordingly. This makes more sense than doing a shift-by-shift analysis of the influence of demand rate on each model. Note that Case 5 represents even demand throughout the day. Therefore, we are interested in how demand patterns that are different than Case 5 influence the model solutions.

Minimum costs do show influence by the differing demand patterns in both models, and the average costs between the demand patterns differ by as much as 11%. If demand is heaviest at the beginning of the day, as in Case 2, both models achieve lower minimum costs. Case 5 performance is similar, though slightly higher than the costs for Case 2 demand. The most expensive cases are those with demand patterns with heavy demand in shift three, as in Cases 1 and 4. Case 3, where demand is heaviest in the second shift, performs better than Cases 1 and 4, but the minimum costs from Case 3 are higher than Cases 2 or 5.

These differences are likely caused by the need for emergency ordering in later shifts of the day. As demand increases throughout the day, inventory must be available or shortages are more likely to occur, and emergency ordering is more likely to be required. Basically, higher demand means balancing higher inventory, ordering, and shortage costs.

In the approximate cost model, the demand pattern does affect the cycle length somewhat. It is more interesting how demand patterns affect cycle length in the simulation model. For demand case 1, we see cycle length, N, that is almost 25% higher on average than the other cases. It is likely that for increasing demand throughout the day, the model fairs better as a result of the inventory system design.

With increasing demand, we would rely more heavily on emergency ordering later in the day. This means that since emergency ordering more immediately addresses the increasing demand at the end of the day, we can seek to minimize counting costs in counter-balance with emergency ordering. Since the approximate cost model does not fully utilize emergency ordering, this relationship between demand pattern and cycle length is not as obvious.

When we look at how the demand pattern influences the emergency ordering level, E_2, it is obvious that Cases 1 and 4 requires an increased E_2 for both models. In the approximate model, E_2 is an average of 5-6 units higher (2-3 times higher) than other cases. In the simulation-optimized solutions, we find that E_2 is about 4 units higher (about 2 times higher), on average, for demand cases 1 and 4. This supports the idea that with higher demand later in the day, both models are required to hold more inventory or incur additional emergency ordering costs, as discussed previously.

Since demand in the later shifts affects emergency ordering, we might assume that heavier demand at the beginning of the day would increase regular ordering. To an extent, we see that happening in both models, with demand Case 2 leading to higher values of S than for Case 1 or 5. We also see higher S for Cases 3 and 4 in the simulation model. This may be that since the lowest demand rate occurs sometime after the first shift, a similar tradeoff is being made. Higher inventory at the beginning of the day from larger S is used quickly, reducing holding costs. However, for Case 4 it is a little strange to see this occurring since there is still a spike in demand at the end of the day. What we have seen, is that Demand patterns have a significant influence on the minimum cost solutions in many ways.

3.6.6 Influence of Accuracy Parameters

By far, the most influential parameter is the assumed level of accuracy, p. In every case in the simulation model we see significant influence on both cost, cycle length, N, and the regular order-up-to level, S. As p increases, S decreases, N increases, and cost decreases. In a way, this shows the value of an accurate inventory system. Surprisingly, the accuracy level has very little influence on the emergency order levels, though, as p increases there is a slight decrease in E_1 and a slight increase in E_2. This lack of direct influence on emergency order-up-to levels might be explained by an influence on S that is sufficient to push emergency ordering to occur in a similar fashion over all cases of p.

This influence of the accuracy of the inventory system could be quantified in a way to assign monetary value to levels of accuracy. In industry applications, it is often difficult to assign precise costs to shortages and to other elements of inventory management. The fact that accuracy was a similarly significant influence for all of the cost parameters considered here, shows that our model may be very useful in quantifying the value of different inventory systems that promise certain levels of accuracy, no matter the cost parameters.

3.7 CONCLUSION

In our paper, we have introduced an inventory system that is currently found in the health care industry, and has been called a hybrid-stockless inventory system. The inventory system is unique, in that inventory is replenished by both regular ordering and emergency ordering, and is prone to inventory record inaccuracy. Counting is required each cycle to reconcile recorded inventory levels with actual inventory levels, thus removing inaccuracy in the inventory records. Under these operating conditions, we developed two inventory models in an attempt to minimize the average daily holding, ordering, shortage, and counting costs.

The first model is an approximate model for which we are able to find optimal solutions using derived optimality conditions. The optimality conditions were found to be sensitive to shortage costs and counting cycle length. The assumptions of the approximate cost model limited the effectiveness of the solutions, and we found that it performed best for lower levels of accuracy.

The second model relaxed the constrictive assumptions of the approximate cost model, and we used simulation to minimize the associated costs of counting, holding, shortages, and emergency ordering. To start the solution procedure, we used the minimizing solution from the approximate cost model, and then applied a simple neighborhood search procedure that found the direction of steepest descent. We found the simulation model to significantly improve the approximate model solution, though for lower levels of accuracy the improvement was less.

Overall, the health-care-based inventory system was unique to the literature, as was the utilization of emergency ordering in a system with inaccuracy. We found that accuracy levels had the largest impact on system performance, and that counting cycle length and the inventory replenishment policies all needed modification to minimize the system costs. Eventually, such a model could be used to estimate the value of an inventory management system that could guarantee high levels of accuracy.

3.7.1 Future Work

While the models presented here were based on a single industry case, we have seen other instances of similar issues in health care. The prevalence and cost-of-care impact of such issues is a focus of future research, as is modeling other types of inventory replenishment systems prone to error. It would be interesting to use the simulation study to incorporate the service level impacts of the optimal solutions, especially where the shortage costs in a hospital setting can be impossible to estimate.

Chiang, C., G. J. Gutierrez. 1996. A periodic review inventory system with two supply modes. *European Journal of Operational Research* **94** 527–547.

Chiang, C., G. J. Gutierrez. 1998. Optimal control policies for a periodic review inventory system with emergency orders. *Naval Research Logistics* **45** 187–204.

DeHoratius, N., A. J. Mersereau, L. Schrage. 2008. Retail inventory management when records are inaccurate. *Manufacturing and Service Operations Management* **10**(2) 257–277.

Fleisch, E., C. Tellkamp. 2005. Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain. *International Journal of Production Economics* **95** 373–385.

Iglehart, D. L., R. C. Morey. 1972. Inventory systems with imperfect asset information. *Management Science* **18**(8) 388–394.

Kök, A. G., K. H. Shang. 2007. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing and Service Operations Management* **9**(2) 185–205.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. McGraw-Hill.

Lawson, D. G., E. L. Porteus. 2000. Multistage inventory management with expediting. *Operations Research* **46**(6) 878–893.

Neve, B., C. P. Schmidt. 2011. A hospital inventory model with inaccurate inventory records. *Working Paper* .

Raman, A., N. DeHoratius, Z. Ton. 2001. Execution: The missing link in retail operations. *California Management Review* **43**(3) 136–152.

Rekik, Y., E. Sahin, Y. Dallery. 2008. Analysis of the impact of the rfid technology on reducing product misplacement errors at retail stores. *International Journal of Production Economics* **112** 264–278.

Rivard-Royer, H., S. Landry, M. Beaulieu. 2002. Hybrid stockless: A case study: Lessons for health-care supply chain integration. *International Journal of Operations and Production Management* **22**(4) 412–424.

Sandoh, H., H. Shimamoto. 2001. A theoretical study on optimal inventory-taking frequency for retailing. *Journal of Retailing and Consumer Services* **8** 47–52.

Scheller-Wolf, A., S. K. Veeraraghavan, G.-J. v. Houtum. 2007. Effective dual sourcing with a single index policy.

Sheppard, G. M., K. A. Brown. 1993. Predicting inventory record-keeping errors with discriminant analysis: A field experiment. *International Journal of Production Economics* **32** 39–51.

Tagaras, G., D. Vlachos. 2001. A periodic review inventory system with emergency replenishments. *Management Science* **47**(3) 415–429.

Teunter, R., D. Vlachos. 2001. An inventory system with periodic regular review and flexible emergency review. *IEE Transactions* **33** 625–635.

Vlachos, D., G. Tagaras. 2001. An inventory system with two supply modes and capacity constraints. *International Journal of Production Economics* **72** 41–58.

C APPENDIX

C.1 Deriving the expected incremental costs

If we assume that inventory is counted every N days, then our planning cycle is also N days. We know that, at the beginning of the day preceding our cycle, a physical inventory count reconciled the recorded inventory level so that $I' = I$ at that time. We also assume that no emergency orders were placed during day preceding our cycle.

When the order arrives at the beginning of shift 1 of day 1 of the cycle, we know the initial inventory parameters. That is, the starting recorded inventory in shift 1 of day 1 is given by: $I' = S - (RD_{1,0} + RD_{2,0} + RD_{3,0})$. Similarly, actual net inventory at the beginning of shift 1 of day 1 is given by $I = S - (D_{1,0} + D_{2,0} + D_{3,0})$. From these starting points, we can begin our recursive derivation of the expected incremental costs.

It should be clearly stated, that we have assumed that emergency orders do not occur until the final day of our planning cycle, day N. We jump to the end of the planning cycle to begin our recursive process of finding E_2^*, E_1^*, and S^*, in that order.

We also define the general Poisson cmf and pmf for ease of notation throughout this discussion:

$$G(x;\mu) = \sum_{k=0}^{x} \frac{e^{-\mu}\mu^k}{k!} \tag{C.1}$$

$$g(x;\mu) = \frac{e^{-\mu}\mu^x}{x!} \tag{C.2}$$

C.1.1 Derivations for E_2^*

Similar to Teunter and Vlachos (2001), we begin the recursive process focusing on when we place our final emergency order (up to E_2), in this case, at the end of the second shift of day N. After the order is placed, it will arrive almost instantaneously at the beginning of the third shift, and we know that actual net inventory will be $I = E_2$ at that time. At the end of the third shift, after demand occurs, there may or may not be backorders at that time. We condition on both possibilities, and then derive the expected incremental cost of increasing E_2.

The cases are denoted as follows, assuming that we will place an emergency order at the end of the second shift:

Case 0α: There are NOT backorders at the end of shift three

Case 0β: There ARE backorders at the end of shift three

Having defined the cases above, we will systematically derive the expected incremental costs for each case, starting with the first case:

Case 0α - NO backorders in the third shift

In this first case, the probability of having no backorders is simply $Pr[E_2 - D_{3,N} \geq 0] = Pr[D_{3,N} \leq E_2]$. Since $D_{3,N} \sim$ Poisson(λ_3), the probability that this case occurs is given by the Poisson distribution. Using Equation C.1, we can write that the probability that case 0α occurs is:

$$P_{0\alpha} = Pr[D_{3,N} \leq E_2] = G(E_2; \lambda_3) \tag{C.3}$$

The incremental cost must also be defined for each case. If E_2 were increased by one unit, then there would be an additional unit ordered using the emergency channel at a cost of c_e, and there will be an additional unit on-hand at the end of the third shift at a cost of c_h. Thus, the incremental cost for this case is $c_e + c_h$.

Case 0β - There ARE backorders in the third shift

Remember that we are conditioning on the fact that we place an emergency order at the end of the second shift. In this case, we assume that there are backorders that remain unfilled at the end of the third shift. Basically, we are assuming that $I_{3,N} < 0$, which is equivalent to saying $E_2 - D_{3,N} < 0$. The probability of this occurring, $Pr[E_2 - D_{3,N} < 0] = Pr[D_{3,N} > E_2]$, can be written using the definition in Equation C.1 as follows:

$$P_{0\beta} = Pr[D_{3,N} > E_2] = 1 - G(E_2; \lambda_3) \tag{C.4}$$

The incremental cost of increasing E_2 in this case includes an additional unit cost for the emergency order. However, since we assume that the system is in a backordered state at the end of the shift, the increase in E_2 leads to a reduction in the number of backorders at the end of the shift. This reasoning shows that the incremental cost for this case is $c_e - c_p$.

Expected incremental cost for E_2

To give the expected incremental cost, we simply multiply the probabilities of each case, 0α and 0β, by their respective individual incremental costs. We give the expected incremental cost for E_2 below:

$$\begin{aligned} C_{E_2} &= (c_e + c_h)P_{0\alpha} + (c_e - c_p)P_{0\beta} = (c_e + c_h)G(E_2; \lambda_3) + (c_e - c_p)(1 - G(E_2; \lambda_3)) \\ &= c_e - c_p + (c_h + c_p)G(E_2; \lambda_3) \end{aligned} \tag{C.5}$$

Optimality condition As we are using marginal analysis, we can use the expected incremental cost as an optimality condition to find E_2^*. We use the smallest value for E_2^* such that the following inequality holds:

$$C_{E_2} = (c_e + c_h)G(E_2; \lambda_3) + (c_e - c_p)(1 - G(E_2; \lambda_3)) \geq 0 \tag{C.6}$$

We now move to the next emergency order-up-to level, E_1, and use the same approach.

C.1.2 Derivations for E_1^*

Now that we can find E_2^*, we use the outcome to find the emergency order-up-to level for the first shift. Again, we will define specific cases, or system states, and then condition to find an expected incremental cost. We will then use the expected incremental cost for the optimality condition and find E_1^*.

At the end of the first shift of the last day of the cycle, we assume that we are placing an order up to the level E_1. However, we do not know if we will need to place an emergency order at the end of the second shift (as discussed previously). As a result, we now have four cases to consider in our development of the expected incremental cost for E_1, and they are denoted:

Case 0α: NO emergency order in shift two, NO backorders in shift three

Case 0β: NO emergency order in shift two, there ARE backorders in shift three

Case 1α: Emergency order in shift two, NO backorders in shift two

Case 1β: Emergency order in shift two, there ARE backorders in shift two

As done previously, we will systematically derive the expected incremental costs for each case, starting with the first case:

Case 0α - NO emergency order in the second shift, NO backorders in the third shift

If we do not place an emergency order in the second shift, it means that the selected value for E_1 will influence inventory until the end of the third shift. Generally, you might consider $E_i < 0$, which would mean there could be backorders at the end of the second shift, if $E_2 < 0$, with no emergency order generated. In that case, we would need to account for the cost of the backorders that would be carried through two shifts. However, in practice, the hospital never allows backorders to go more than a shift without being filled, and where no emergency order-up-to level is defined, the default value is zero. (This was implicitly assumed in Teunter and Vlachos (2001), though a system might be considered in the future where backorders may carry over through more than one emergency review period.)

In this case (as well as in case 0β), it is assumed that there are no backorders at the end of the second shift since $E_2 \geq 0$ and we assume no emergency order in the second shift. As we are also assuming there are no backorders at the end of the third shift, the first event is: $E_1 - D_{2,N} \geq E_2^*$, which means that no emergency order is placed at the end of the second shift, and the next event is $E_1 - (D_{2,N} + D_{3,N}) \geq 0$, so that there are no backorders at the end of the third shift.

To compute the probability, we will condition on the value of shift 2 total demand, $D_{2,N} \sim \text{Poisson}(\lambda_2)$. This gives the following:

$$
\begin{aligned}
P_{0\alpha} &= \sum_{k=0}^{\infty} Pr[E_1 - E_2^* \geq D_{2,N} \text{ and } D_{3,N} \leq E_1 - D_{2,N} | D_{2,N} = k] \cdot Pr[D_{2,N} = k] \\
&= \sum_{k=0}^{\infty} Pr[E_1 - E_2^* \geq k \text{ and } D_{3,N} \leq E_1 - k] \cdot Pr[D_{2,N} = k] \\
&= \sum_{k=0}^{E_1 - E_2^*} Pr[D_{3,N} \leq E_1 - k] \cdot Pr[D_{2,N} = k] \\
&= \sum_{k=0}^{E_1 - E_2^*} G(E_1 - k; \lambda_3) \cdot g(k; \lambda_2)
\end{aligned}
\tag{C.7}
$$

Now, we need only define the incremental cost associated with case 0α. In this case, if E_1 were to increase, it would increase inventory by one unit for two shifts, incrementing cost by $2c_h$. Also, there would be an additional unit ordered via the emergency channel at a cost of c_e. Thus, the expected incremental cost in this case is $(c_e + 2c_h) \sum_{k=0}^{E_1 - E_2^*} G(E_1 - k; \lambda_3) \cdot g(k; \lambda_2)$. This will be added to the other expected incremental costs from the other cases.

Case 0β - NO emergency order in second shift, there ARE backorders in the third shift

Similar to the previous case, we assume no emergency order is placed at the end of the second shift. But now, we assume that there are backorders at the end of the third shift. For this to occur, the following must be true: $D_{2,1} \leq E_1 - E_2^*$ and $D_{2,1} + D_{3,1} > E_1$. The probability of this occurring is given by: $Pr[E_1 - E_2^* \geq D_{2,1} \text{ and } D_{3,1} > E_1 - D_{2,1}]$. Taking the same approach as in the previous case, we find the following case probability:

$$
\begin{aligned}
P_{0\beta} &= \sum_{k=0}^{\infty} Pr[E_1 - E_2^* \geq D_{2,N} \text{ and } D_{3,N} > E_1 - D_{2,N} | D_{2,N} = k] \\
&= \sum_{k=0}^{E_1 - E_2^*} (1 - G(E_1 - k; \lambda_3))\, g(k, \lambda_2)
\end{aligned}
\tag{C.8}
$$

The incremental cost for case 0β must include the reduced cost of backorders ($-c_p$), the increased cost of an additional unit ordered through the emergency channel (c_e), and the cost for the additional unit of on-hand inventory available at the end of the second shift (c_h). So, the expected incremental cost for case 0β is $(c_e + c_h - c_p) \sum_{k=0}^{E_1 - E_2^*} (1 - G(E_1 - k; \lambda_3))\, g(k, \lambda_2)$.

Case 1α - Emergency order in second shift, NO backorders in the *second* shift

In this case, the demand in the second shift must reduce net inventory to a level that falls between zero and E_2^*. In other words, $0 \le I_{2,N} < E_2^*$. Since net inventory at the beginning of shift 2 is given by $I = E_1$, the probability of this case occurring is given by:

$$\begin{aligned} P_{1\alpha} &= Pr[0 \le E_1 - D_{2,N} < E_2^*] \\ &= Pr[E_1 - E_2^* < D_{2,N} \le E_1] \\ &= \sum_{k=E_1 - E_2^* + 1}^{E_1} Pr[D_{2,N} = k] \\ &= \sum_{k=E_1 - E_2^* + 1}^{E_1} g(k, \lambda_2) \end{aligned} \tag{C.9}$$

Increasing E_1 in this case will increase the available inventory at the end of shift 2 at a cost of c_h. This increased availability will reduce the amount ordered through the emergency channel at the end of the second shift, reducing cost by c_e. However, the reduced emergency ordering cost is offset by the identical increase in ordering an additional unit at the end of shift 1. This gives an expected incremental cost of $(c_h) \sum_{k=E_1 - E_2^* + 1}^{E_1} g(k, \lambda_2)$ for the case 1α.

Case 1β - Emergency order in second shift, there ARE backorder in the *second* shift

As stated previously, $E_2^* \ge 0$, and for the elements of the case to be satisfied, the inventory at the end of shift 2 $(E_1 - D_{2,N})$ must be negative. So the probability for this case is given by:

$$\begin{aligned} P_{1\beta} &= Pr[E_1 - D_{2,N} < 0] \\ &= (1 - G(E_1; \lambda_2)) \end{aligned} \tag{C.10}$$

Here, the cost to increase E_1 is a reduction in both backorder costs (c_p) and in emergency ordering costs (c_e) at the end of shift 2. Again, the reduced emergency ordering cost is negated due to the increased emergency ordering cost at the end of shift 1. We can write the expected incremental cost for this case as $(-c_p)\,(1 - G(E_1; \lambda_2))$

Expected incremental cost for E_1

We sum the expected incremental costs for each of the cases, 0α, 0β, 1α, and 1β, to get the expected incremental cost for E_1.

$$C_{E_1} = (c_e + 2c_h) \sum_{k=0}^{E_1 - E_2^*} G(E_1 - k; \lambda_3) \cdot g(k; \lambda_2) + (c_e + c_h - c_p) \sum_{k=0}^{E_1 - E_2^*} (1 - G(E_1 - k; \lambda_3))\, g(k, \lambda_2)$$
$$+ c_h \sum_{k=E_1 - E_2^* + 1}^{E_1} g(k, \lambda_2) - c_p\,(1 - G(E_1; \lambda_2)) \tag{C.11}$$

To reduce the above equation, we expand the second and fourth terms, and use the relationship between Equations C.1 and C.2 to remove summations and combine terms. The reduced equation is given below, which we will use in our optimality condition for finding E_1^*:

$$C_{E_1} = (c_h + c_p) \sum_{k=0}^{E_1 - E_2^*} G(E_1 - k; \lambda_3) g(k; \lambda_2) + (c_e - c_p) G(E_1 - E_2^*; \lambda_2) + (c_h + c_p) G(E_1; \lambda_2) - c_p \tag{C.12}$$

To find the best emergency order-up-to level for the first shift, we will use the expected incremental cost from above. Given E_2^*, we will set E_1^* to the smallest value of E_1 where $C_{E_1} \geq 0$. Our last step is to find S^*, given our previous solutions for E_2^* and E_1^*.

C.1.3 Derivations for S^*

We now use E_2^* and E_1^* in our derivations to find S^*. Similar to the previous derivations, we will organize the system states into specific cases, and compute the incremental costs for S. Our assumption that emergency ordering only occurs in the last day of the cycle is limiting. However, if we consider another approximate model where emergency ordering may occur in the last *two* days of the cycle, we must consider another 6 cases for each case considered in our current approach. The number of cases required will increase exponentially as we increase the number of days where emergency ordering is allowed. Therefore, we continue with the assumption that emergency ordering only occurs during the final day of the cycle.

The derivations here will differ from the previous approach for emergency orders, since the automated system manages regular ordering. Remember that the automated system uses the recorded inventory level, I', as the performance index, instead of net inventory, I, as used in the manual emergency ordering system. The incoming regular orders that increase net inventory will be based on recorded demand, $RD_{i,j}$, while total demand, $D_{i,j}$, will decrease net inventory. This is important, as operating costs are tied to the actual net inventory, so we continue using it as the performance measure while seeking S^*, even though the automated system does not know the actual net inventory.

Since we have assumed that no emergency ordering occurs during the first $N-1$ days of the cycle, we can say that actual net inventory at the beginning of period N, denoted in this instance by I_N, is given below.

$$\begin{aligned} I_N &= S - \sum_{j=0}^{N-1}\sum_{i=1}^{3} D_{i,j} + \sum_{j=0}^{N-2}\sum_{i=1}^{3} RD_{i,j} \\ &= S - \left(\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} \right) \end{aligned} \tag{C.13}$$

As discussed previously, at the beginning of the first day of the cycle, actual net inventory is given by $I = S-(D_{1,0}+D_{2,0}+D_{3,0})$. This is reduced each shift, i, during days $j = 1, \ldots, N-1$, by total demand $D_{i,j}$. Regular orders are placed at the beginning of days $j = 1, \ldots, N-1$, and are based on the recorded demand, $RD_{i,j}$, in each shift i of each day $j = 0, \ldots, N-2$, respectively (recall the one day lead time for regular orders). At the beginning of day N, all of these previous regular orders will have arrived. After defining net inventory, we apply the relationship given in Equation 3.1 and rewrite the beginning day N net inventory using unrecorded demand, $UD_{i,j}$ (recall Equation 3.3).

With I_N as the starting point, we now look at the final day of the cycle. The cases considered are as follows, where emergency orders occur during this last day of the cycle:

Case 0α: NO emergency order in shifts one and two, NO backorders in shift three

Case 0β: NO emergency order in shifts one and two, there ARE backorders in shift three

Case 1α: First emergency order in shift two, NO backorders in shift two

Case 1β: First emergency order in shift two, there ARE backorders in shift two

Case 2α: First emergency order in shift one, NO backorders in shift one

Case 2β: First emergency order in shift one, there ARE backorders in shift one

Again, we will systematically derive the expected incremental costs for each case, starting with the first case, as was done previously:

Case 0α - NO emergency order in the first and second shifts, NO backorders in the third shift

In this case, the demand throughout the last day of the cycle does not reduce net inventory enough to cause backorders by the end of shift 3, nor does demand in the first shift and second shifts trigger any emergency orders. The probability that this occurs is given by:

$$Pr[I_N - D_{1,N} \geq E_1^* \; ; \; I_N - D_{1,N} - D_{2,N} \geq E_2^* \; ; \; I_N - D_{1,N} - D_{2,N} - D_{3,N} \geq 0] \tag{C.14}$$

We use conditioning to compute the above probability, but first, we expand I_N to better see the distributions involved. Each event term from the above probability is expanded individually below:

$$I_N - D_{1,N} \geq E_1^* \Rightarrow S - \left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \right) \geq E_1^* \tag{C.15}$$

$$I_N - D_{1,N} - D_{2,N} \geq E_2^* \Rightarrow S - \left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \right) - D_{2,N} \geq E_2^* \tag{C.16}$$

$$I_N - D_{1,N} - D_{2,N} - D_{3,N} \geq 0] \Rightarrow S - \left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \right) - D_{2,N} - D_{3,N} \geq 0 \tag{C.17}$$

Since the total demand, $D_{i,j}$, and unrecorded demand, $UD_{i,j}$, are random variables from specific distributions, we will condition on their related distributions to compute the probability. We will first condition on the value of the more complex term found in each of the above equations, given below with its associated Poisson distribution:

$$\left(\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N}\right) \sim \text{ Poisson}\left((N-1)(1-p)\sum_{i=1}^{3}\lambda_i + \sum_{i=1}^{3}\lambda_i + \lambda_1\right) \tag{C.18}$$

To simplify the notation, we define λ without a subscript to be the sum of the λ_i for one day, giving $\lambda = \sum_{i=1}^{3}\lambda_i$. We can then express the above distribution as:

$$\left(\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N}\right) \sim \text{ Poisson}\left((N+p-Np)\lambda + \lambda_1\right) \tag{C.19}$$

For notational purposes, we define $\mu = (N+p-Np)\lambda + \lambda_1$.

Going back to Equations C.15, C.16, and C.17, we see that both C.16 and C.17 include the random variable for total demand in shift two, $D_{2,N}$. We will therefore also condition on the value of $D_{2,N} \sim$ Poisson(λ_2) to compute the probability for this first case, 0α.

So, conditioning on $\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} = x$ and $D_{2,N} = y$, we can give the probability of case 0α:

$$\begin{aligned}
P_{0\alpha} &= \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} Pr[S-x \geq E_1^* \ ; \ S-x-y \geq E_2^* \ ; \ S-x-y-D_{3,N} \geq 0]\cdot g(y;\lambda_2)\cdot g(x;\mu) \\
&= \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} Pr[x \leq S-E_1^* \ ; \ y \leq S-x-E_2^* \ ; \ D_{3,N} \leq S-x-y]\cdot g(y;\lambda_2)\cdot g(x;\mu) \\
&= \sum_{x=0}^{S-E_1^*}\sum_{y=0}^{S-x-E_2^*} G(S-x-y;\lambda_3)\cdot g(y;\lambda_2)\cdot g(x;\mu)
\end{aligned} \tag{C.20}$$

Now that we have the probability for this case, what is the incremental cost? Since we do not assume a unit cost for regular orders, the additional item that may be ordered each day along the regular channel does not affect incremental cost. Also, since this case assumes no backorders, we are only concerned with holding cost, c_h. There are N days in the cycle, with three shifts each day where c_h is applied to each unit on hand, so the incremental cost for increasing the order-up-to level S for case 0α is $(3N \cdot c_h)$.

Case 0β - NO emergency order in the first and second shifts, there ARE backorders in the *third* shift

This case is very similar to the previous case, but now we are assuming that there are backorders. We can use nearly the same events as before, but in the last shift, we assume that demand brings inventory into a backordered state. So instead of using Equation C.17 as the last term of the probability, we use the following event, where we swap the ≥ 0 with a < 0.

$$I_N - D_{1,N} - D_{2,N} - D_{3,N} < 0 \Rightarrow S - \left(\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \right) - D_{2,N} - D_{3,N} < 0 \tag{C.21}$$

The probability for case 0β is nearly identical to the previous case, but it uses the complimentary Poisson c.m.f instead of the c.m.f. This probability is given below:

$$P_{0\beta} = \sum_{x=0}^{S-E_1^*} \sum_{y=0}^{S-x-E_2^*} (1 - G(S-x-y;\lambda_3)) \cdot g(y;\lambda_2) \cdot g(x;\mu) \tag{C.22}$$

The incremental cost in this case, would require one shift less worth of increased holding cost than the previous case, $(3N-1)c_h$. Also, there will be one less backorder at the end of the third shift after increasing S, so the cost would be reduced by c_p. The total incremental cost for case 0β is $((3N-1)c_h - c_p)$.

Case 1α - First emergency order in the second shift, NO backorders in the *second* shift

Here, we are looking at the first time an emergency order is placed. This is the same approach taken by Teunter and Vlachos (2001) when developing the cases for the regular order-up-to level, S. With this case, we don't look beyond the end of the second shift because costs there are unaffected by changing S, given that we order to E_2^* at the end of shift 2. Therefore, the probability of this case is given as follows.

$$P_{1\alpha} = Pr[I_N - D_{1,N} \geq E_1^* \ ; 0 \leq I_N - D_{1,N} - D_{2,N} < E{*}_2] \tag{C.23}$$

The first event of the above joint probability ensures that no emergency order is placed at the end of the first shift. The second event ensures, that at the end of the second shift, an emergency order will be placed and there are no backorders at that time. Similar to the previous case, we will condition on $\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} = x$. After expanding I_N in the same fashion as before, we get the following joint probability:

$$
\begin{aligned}
P_{1\alpha} &= \sum_{x=0}^{\infty} Pr[S - x \geq E_1^* \; ; \; 0 \leq S - x - D_{2,N} < E_2^*] \cdot g(x;\mu) \\
&= \sum_{x=0}^{\infty} Pr[x \leq S - E_1^* \; ; \; S - x - E_2^* < D_{2,N} \leq S - x] \cdot g(x;\mu) \\
&= \sum_{x=0}^{S-E_1^*} Pr[S - x - E_2^* < D_{2,N} \leq S - x] \cdot g(x;\mu) \\
&= \sum_{x=0}^{S-E_1^*} \sum_{y=S-x-E_2^*+1}^{S-x} g(y;\lambda_2) \cdot g(x;\mu) \\
&= \sum_{x=0}^{S-E_1^*} (G(S-x;\lambda_2) - G(S-x-E_2^*;\lambda_2)) \cdot g(x;\mu)
\end{aligned}
\tag{C.24}
$$

The incremental cost for this case must include holding costs for all shifts during the cycle except shift three of day N, giving $(3N-1)c_h$. Since we have assumed an emergency order in the second shift, by increasing S we effectively reduce the emergency order by one unit, so that cost would decrease by c_e. The incremental cost for case 1α is $((3N-1)c_h - c_e)$.

Case 1β - First emergency order in the second shift, there ARE backorders in the *second* shift

This case is similar to the previous case since we assume the first emergency order occurs in the second shift. However, as we are assuming backorders in the second shift, the final event in the joint probability would simply be $I_N - D_{1,N} - D_{2,N} < 0$. This is simpler than the previous case, though we still condition on $\sum_{j=0}^{N-2}\sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} = x$. This gives the following joint probability:

$$
\begin{aligned}
P_{1\beta} &= \sum_{x=0}^{\infty} Pr[S - x \geq E_1^* \ ; \ S - x - D_{2,N} < 0] \cdot g(x, \mu) \\
&= \sum_{x=0}^{S-E_1^*} Pr[S - x < D_{2,N}] \cdot g(x; \mu) \\
&= \sum_{x=0}^{S-E_1^*} (1 - G(S - x; \lambda_2)) \cdot g(x; \mu)
\end{aligned}
\tag{C.25}
$$

The incremental cost will now include the holding costs for all shifts of the cycle, excluding the last *two* shifts, as there are backorders at the end of the second shift. So the holding cost term is $(3N-2)c_h$, and the backorder costs are decreased by c_p. Incrementing S also decreases the amount ordered along the emergency channel, decreasing cost by c_e. The total incremental cost for case 1β is $((3N-2)c_h - c_p - c_e)$. We move now to the final two cases, after which we will derive the expected incremental cost for the regular order-up-to level, S.

Case 2α - First emergency order in the first shift, NO backorders in the *first* shift

In this case, we must have that $0 \leq I_N - D_{1,N} < E_1^*$. The resultant probability will not require conditioning in this case, and it is given below:

$$
\begin{aligned}
P_{2\alpha} &= Pr[0 \leq I_N - D_{1,N} < E_1^*] \\
&= Pr[S - E_1^* < \sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} \leq S] \\
&= \sum_{x=S-E_1^*+1}^{S} g(x; \mu) \\
&= G(S; \mu) - G(S - E_1^*; \mu)
\end{aligned}
\tag{C.26}
$$

We can reason through the incremental cost of this case as follows: Because there are no backorders, an increase in S will cost $(3N-2)c_h$, since the increased inventory will be held through the end of shift 1. Also, there will be one less item ordered through the emergency channel, leading to a cost reduction of c_e. The total incremental cost in this case is $(3N - 2)c_h - c_e$.

Case 2β - First emergency order in the first shift, there ARE backorders in the *first* shift

This case is the simplest to consider, since the only requirement is that $I_N - D_{1,N} < 0$ due to the assumption that we do have backorders at the end of the first shift and it naturally requires an emergency order to be placed. The probability is given by:

$$\begin{aligned} P_{2\beta} &= Pr[I_N - D_{1,N} < 0] \\ &= Pr[\sum_{j=0}^{N-2} \sum_{i=1}^{3} UD_{i,j} + \sum_{i=1}^{3} D_{i,N-1} + D_{1,N} > S] \\ &= 1 - G(S;\mu) \end{aligned} \tag{C.27}$$

The incremental cost in this case parallels the cost in case 1β, where both emergency ordering and backordering costs decrease $(-c_e - c_p)$. However, inventory is not held at the end of shift 1 of day N, so the incremental inventory cost is $(3N-3)c_h$. This gives the total incremental cost for case 2β of $((3N-3)c_h - c_e - c_p)$.

Expected incremental cost for S^*

We can now write the expected incremental cost for S by multiplying the probabilities with the respective incremental cost of each case. The full equation is given below.

$$\begin{aligned} C_S = {} & (3N \cdot c_h) \sum_{x=0}^{S-E_1^*} \sum_{y=0}^{S-x-E_2^*} G(S-x-y;\lambda_3) \cdot g(y;\lambda_2) \cdot g(x;\mu) \\ & + ((3N-1)c_h - c_p) \sum_{x=0}^{S-E_1^*} \sum_{y=0}^{S-x-E_2^*} (1 - G(S-x-y;\lambda_3)) \cdot g(y;\lambda_2) \cdot g(x;\mu) \\ & + ((3N-1)c_h - c_e) \sum_{x=0}^{S-E_1^*} (G(S-x;\lambda_2) - G(S-x-E_2^*;\lambda_2)) \cdot g(x;\mu) \\ & + ((3N-2)c_h - c_p - c_e) \sum_{x=0}^{S-E_1^*} (1 - G(S-x;\lambda_2)) \cdot g(x;\mu) \\ & + ((3N-2)c_h - c_e) \left(G(S;\mu) - G(S-E_1^*;\mu)\right) \\ & + ((3N-3)c_h - c_e - c_p) \left(1 - G(S;\mu)\right) \end{aligned} \tag{C.28}$$

After combining terms and restructuring different elements we can reduce this expected incremental cost for S as follows:

$$
\begin{aligned}
C_S = {} & (3N-3)c_h - c_e - c_p + (c_h + c_p)G(S;\mu) - (c_p)G(S - E_1^*;\mu) \\
& + (c_h + c_p) \sum_{x=0}^{S-E_1^*} G(S - x;\lambda_2) \cdot g(x;\mu) \\
& + (c_e - c_p) \sum_{x=0}^{S-E_1^*} G(S - x - E_2^*;\lambda_2) \cdot g(x;\mu) \\
& + (c_h + c_p) \sum_{x=0}^{S-E_1^*} \sum_{y=0}^{S-x-E_2^*} G(S - x - y;\lambda_3) \cdot g(y;\lambda_2) \cdot g(x;\mu)
\end{aligned}
\tag{C.29}
$$

We set S^* to be the smallest integer S where $C_S \geq 0$. So we can now find the optimal solution to the approximate cost version of the model, for a given number of days between physical inventory counts, N.

C.2 Conditions on the shortage cost

During our numerical study of the incremental costs C_{E_2}, C_{E_1}, and C_S, we found that for certain values of N and c_p the optimality conditions for the approximate model were met instantly. That is, $S^* = E_1^*$, $E_1^* = E_2^*$, or $E_2^* = 0$.

Analytically, we can define the values for c_p (also based on values of N for C_S) where the optimality conditions will be met instantly. Recall that the optimal decision variables satisfy the constraint $S^* \geq E_1^* \geq E_2^* \geq 0$. That means, that when $S^* = E_1^*$, $E_1^* = E_2^*$, or $E_2^* = 0$, the model instance would be on the"boundary" for values of the parameter c_p for the respective incremental cost equation.

As derived below, we find that increasing c_p would cause the optimality conditions to remain satisfied for $S^* = E_1^*$, $E_1^* = E_2^*$, or $E_2^* = 0$. Decreasing c_p from that "boundary" point leads to $E_2^* > 0$, $E_1^* > E_2^*$, or $S^* > E_1^*$ - depending on the value of N.

Therefore, depending on the cost parameters it may be that $S^* = E_1^*$. If so, then it must be true that $C_S \geq 0$ for $S \geq E_1^*$. Similarly, it may be that $E_1^* = E_2^*$, where $C_{E_1} \geq 0$ for $E_1 \geq E_2^*$ and that $E_2^* = 0$, where $C_{E_2} \geq 0$ for $E_2 \geq 0$. This reasoning leads to conditions on c_p where this is true in each individual case.

Case: $E_2^* = 0$

In this first case, we use Equation C.6, and set $E_2 = 0$. We get the following incremental cost equation:

$$\begin{aligned} C_{E_2} &= c_e - c_p + (c_h + c_p)e^{-\lambda_3} \\ &= c_h e^{-\lambda_3} + c_e - (1 - e^{-\lambda_3})c_p \end{aligned} \tag{C.30}$$

Using it as an optimality condition, $E_2^* = 0$ requires $C_{E_2} \geq 0$, which implies that:

$$c_p \leq \frac{c_h e^{-\lambda_3} + c_e}{1 - e^{-\lambda_3}} \tag{C.31}$$

Case: $E_1^* = E_2^*$

Here, we use Equation C.12, setting $E_1 = E_2^*$.

$$\begin{aligned}
C_{E_1} &= (c_h + c_p)G(E_2^*, \lambda_3)e^{-\lambda_2} + (c_e - c_p)e^{\lambda_2} + (c_h + c_p)G(E_2^*, \lambda_2) - c_p \\
&= \left(G(E_2^*, \lambda_2) + G(E_2^*, \lambda_3)e^{-\lambda_2}\right) c_h + (e^{-\lambda_2})c_e \\
&- \left(1 + e^{-\lambda_2} - G(E_2^*, \lambda_2) - G(E_2^*, \lambda_3)e^{-\lambda_2}\right) c_p
\end{aligned} \tag{C.32}$$

Using the same reasoning as before, we see that $E_1^* = E_2^*$ requires that $C_{E_1} \geq 0$. This gives the condition for c_p:

$$c_p \leq \frac{\left(G(E_2^*, \lambda_2) + G(E_2^*, \lambda_3)e^{-\lambda_2}\right) c_h + (e^{-\lambda_2})c_e}{1 + e^{-\lambda_2} - G(E_2^*, \lambda_2) - G(E_2^*, \lambda_3)e^{-\lambda_2}} \tag{C.33}$$

Which is dependent on the value of E_2^*. So, for a given set of parameters, Equation C.33 gives the values of c_p that would require the optimal order up to level be $E_1^* = E_2^*$.

Case: $S^* = E_1^*$

We set $S = E_1^*$, and then Equation C.29 can be written as:

$$\begin{aligned}
C_S = {} & (3N-3)c_h - c_e - c_p + (c_h + c_p)G(E_1^*, \mu) - c_p e^{-\mu} + (c_h + c_p)G(E_1^*, \lambda_2)e^{-\mu} \\
& + (c_e - c_p)G(E_1^* - E_2^*, \lambda_2)e^{-\mu} + (c_h + c_p) \sum_{y=0}^{E_1^*-E_2^*} G(E_1^* - y, \lambda_3)g(y, \lambda_2)e^{-\mu} \\
& = \left[G(E_1^* - E_2^*, \lambda_2)e^{-\mu} - 1\right] c_e \\
& + \left[3N - 3 + G(E_1^*, \mu) + e^{-\mu}\left(G(E_1^*, \lambda_2) + \sum_{y=0}^{E_1^*-E_2^*} G(E_1^* - y, \lambda_3)g(y, \lambda_2)\right)\right] c_h \\
& - \left[1 + e^{-\mu}\left(1 + G(E_1^* - E_2^*, \lambda_2) - G(E_1^*, \lambda_2) - \sum_{y=0}^{E_1^*-E_2^*} G(E_1^* - y, \lambda_3)g(y, \lambda_2)\right)\right] c_p
\end{aligned} \tag{C.34}$$

Similar to the previous cases, if $C_S \geq 0$ for $S \geq E_1^*$, then the largest that c_p can be such that $S^* = E_1^*$ is given by:

$$
\begin{aligned}
c_p \leq & \frac{\left[G(E_1^*-E_2^*,\lambda_2)e^{-\mu}-1\right]c_e+\left(3N-3+G(E_1^*,\mu)\right)c_h}{1+e^{-\mu}\left(1+G(E_1^*-E_2^*,\lambda_2)-G(E_1^*,\lambda_2)-\sum\limits_{y=0}^{E_1^*-E_2^*}G(E_1^*-y,\lambda_3)g(y,\lambda_2)\right)} \\
& +\frac{e^{-\mu}\left(G(E_1^*,\lambda_2)+\sum\limits_{y=0}^{E_1^*-E_2^*}G(E_1^*-y,\lambda_3)g(y,\lambda_2)\right)c_h}{1+e^{-\mu}\left(1+G(E_1^*-E_2^*,\lambda_2)-G(E_1^*,\lambda_2)-\sum\limits_{y=0}^{E_1^*-E_2^*}G(E_1^*-y,\lambda_3)g(y,\lambda_2)\right)}
\end{aligned}
\tag{C.35}
$$

In the above condition, a larger N allows for higher c_p while the optimal order-up-to level remains $S^* = E_1^*$. In the numerical study and attempted optimization of the model, we use these conditions to shape our analysis, so that we can experiment over a greater diversity of solution space.

C.3 Simulation Algorithm

The following simulation algorithm was programmed in C++ and was used to simulate the model.

Initialization phase:

Step 0: Initialize

Set beginning day 0 inventory to $I^{'} = I = S$

Set $i = 1$ for the first shift of day 0

Go to Step 2

Step 1: Receive stock

Receive emergency order (physical inventory): Set $I = I_{i-1,0} + EO_{i-1,0}$

Receive emergency order (recorded inventory): Set $I^{'} = I^{'}_{i-1,0} + (EO_{i-1,0} - [I_{i-1,0}]^{-})$

Go to Step 2

Step 2: Simulate actual demand

Generate shift i total demand, $D_{i,0} \sim$ Poisson(λ_i)

Ending shift i actual inventory given by $I_{i,0} = I - D_{i,0}$

Go to Step 3

Step 3: Simulate recorded demand

Generate shift i recorded demand, $RD_{i,0} \sim$ Binomial(Min$[I, D_{i,0}], p$)

Compute ending shift i recorded inventory: $I^{'}_{i,0} = I^{'} - RD_{i,0}$

Go to Step 4

Step 4: Place emergency order, advance shift

If $(i < 3)$

Place emergency order of size $EO_{i,0} = [E_i - I_{i,0}]^{+}$

Advance to next shift: $i++$

Go to Step 1

Else

Go to Step 5

Step 5: Initialize warm-up

Set shift to $i = 1$, set day $j = 1$

Set number of cycles in warm-up phase, $w = \lceil l/N \rceil$

Set the cycle number,z, to 1: $z = 1$

Set $RO_N = 0$, $I_{3,N} = I_{3,0}$, and $I^{'}_{3,N} = I^{'}_3, 0$

Go to Warm-up Phase

Warm-up phase:

Step 6: Receive regular order in shift $i = 1$ of day j, count on day $j = N$

If $j = 1$ and $N > 1$

Receive regular order (physical inventory): Set $I = I_{3,N} + RO_N$

Receive regular order (recorded inventory): Set $I^{'} = I^{'}_{3,N} + [RO_N - [I_{3,N}]^-]^+$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 8

Else If $j = N = 1$

Receive regular order (physical inventory): Set $I = I_{3,N} + RO_N$

Reconcile inventory: Set $I^{'} = I$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 8

Else If $j = N > 1$

Receive regular order (physical inventory): Set $I = I_{3,j-1} + RO_{j-1}$

Reconcile inventory: Set $I^{'} = I$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 8

Else

Receive regular order (physical inventory): Set $I = I_{3,j-1} + RO_{j-1}$

Receive regular order (recorded inventory): Set $I^{'} = I^{'}_{3,j-1} + [RO_{j-1} - [I_{3,j-1}]^-]^+$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 8

Step 7: Receive emergency order in shift $i = \{2, 3\}$ *of day* j

Receive emergency order (physical inventory): Set $I = I_{i-1,j} + EO_{i-1,j}$

Receive emergency order (recorded inventory): Set $I' = I'_{i-1,j} + (EO_{i-1,j} - [I_{i-1,j}]^-)$

Go to Step 8

Step 8: Simulate actual demand in shift i *of day* j

Generate shift i total demand, $D_{i,j} \sim \text{Poisson}(\lambda_i)$

Ending shift i actual inventory given by $I_{i,j} = I - D_{i,j}$

Go to Step 9

Step 9: Simulate recorded demand in shift i *of day* j

Generate shift i recorded demand, $RD_{i,j} \sim \text{Binomial}(\text{Min}[I, D_{i,j}], p)$

Compute ending shift i recorded inventory: $I'_{i,j} = I' - RD_{i,j}$

Go to Step 10

Step 10: Place emergency order in shift $i = \{1, 2\}$ *of day* j*, advance shift*

If $(i < 3)$

Place emergency order of size $EO_{i,j} = [E_i - I_{i,j}]^+$

Advance to next shift: $i++$

Go to Step 7

Else

Go to Step 11

Step 11: Advance to next day

If $(j < N)$

Advance to next day: $j++$

Set $i = 1$

Go to Step 6

Else

Go to Step 12

Step 12: Advance to next cycle

If $(z < w)$

Advance to next cycle: $z++$

Set $j = 1$ and $i = 1$

Go to Step 6

Else

Set number of cycles in initial replication phase: $n_0 = 50$

are required to ensure the absolute estimate error is less than β

Reset cycle number: $z = 1$

Set $j = 1$ and $i = 1$

Go to Replication Phase

Replication phase:

Step 13: Receive regular order in shift $i = 1$ of day j, count on day $j = N$

If $j = 1$ and $N > 1$

Receive regular order (physical inventory): Set $I = I_{3,N} + RO_N$

Receive regular order (recorded inventory): Set $I' = I'_{3,N} + [RO_N - [I_{3,N}]^-]^+$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Reset cost: $C = 0$

Go to Step 15

Else If $j = N = 1$

Receive regular order (physical inventory): Set $I = I_{3,N} + RO_N$

Reconcile inventory: Set $I' = I$

Assign fixed counting cost for the current cycle: $C+ = k$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 8

Else If $j = N > 1$

Receive regular order (physical inventory): Set $I = I_{3,j-1} + RO_{j-1}$

Reconcile inventory: Set $I' = I$

Assign fixed counting cost for the current cycle: $C+ = k$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 15

Else

Receive regular order (physical inventory): Set $I = I_{3,j-1} + RO_{j-1}$

Receive regular order (recorded inventory): Set $I' = I'_{3,j-1} + [RO_{j-1} - [I_{3,j-1}]^-]^+$

Place regular order: $RO\_j = \left[S - I'\right]^+$

Go to Step 15

Step 14: Receive emergency order in shift $i = \{2, 3\}$ *of day* j

Receive emergency order (physical inventory): Set $I = I_{i-1,j} + EO_{i-1,j}$

Receive emergency order (recorded inventory): Set $I' = I'_{i-1,j} + (EO_{i-1,j} - [I_{i-1,j}]^-)$

Go to Step 15

Step 15: Simulate actual demand in shift i *of day* j

Generate shift i total demand, $D_{i,j} \sim \text{Poisson}(\lambda_i)$

Ending shift i actual inventory given by $I_{i,j} = I - D_{i,j}$

Assign backorder and holding costs for shift i: $C+ = c_h[I^+_{i,j}] + c_p[I^-_{i,j}]$

Go to Step 16

Step 16: Simulate recorded demand in shift i *of day* j

Generate shift i recorded demand, $RD_{i,j} \sim \text{Binomial}(\text{Min}[I, D_{i,j}], p)$

Compute ending shift i recorded inventory: $I'_{i,j} = I' - RD_{i,j}$

Go to Step 17

Step 17: Place emergency order in shift $i = \{1, 2\}$ *of day* j*, advance shift*

If $(i < 3)$

Place emergency order of size $EO_{i,j} = [E_i - I_{i,j}]^+$

Assign emergency ordering cost: $C+ = c_e(EO_{i,j})$

Advance to next shift: $i++$

Go to Step 14

Else

Go to Step 18

Step 18: Advance to next day

If $(j < N)$

Advance to next day: $j++$

Set $i = 1$

Go to Step 13

Else

Go to Step 19

Step 19: Advance to next cycle or end simulation

If $(z < n_0)$

Advance to next cycle: $z++$

Add cycle cost to total cost: $TC+ = C$

Set $j = 1$ and $i = 1$

Go to Step 13

Else If $(z = n_0)$

Compute average expected daily cost over n_0 cycles: $\bar{C}(n_0) = \frac{TC}{n_0}$

Compute Sample Variance over n_0 cycles: $S^2(n_0) = \sum_{k=1}^{n_0} \frac{(C_k - \bar{C}(n_0))^2}{n_0 - 1}$

Find $n = \arg\min_i \left\{ i : i > n_0, t_{i-1,1-\alpha/2} \cdot \sqrt{\frac{S^2(n_0)}{i}} \leq \epsilon \right\}$

Advance to next cycle: $z++$

Add cycle cost to total cost: $TC+ = C$

Set $j = 1$ and $i = 1$

Go to Step 13

Else If $(z < n)$

Advance to next cycle: $z++$

Add cycle cost to total cost: $TC += C$

Set $j = 1$ and $i = 1$

Go to Step 13

Else

Add cycle cost to total cost: $TC += C$

Compute average expected daily cost over n cycles: $C_{avg} = \frac{TC}{n}$

End Simulation

Table C.1: Problem Instances for Numerical Analysis

#	λ_1	λ_2	λ_3	c_h	c_e	c_p	k	p		#	λ_1	λ_2	λ_3	c_h	c_e	c_p	k	p
1	3	5	8	0.3	1	3	30	0.55		51	5	8	3	0.3	1	3	60	0.85
2	3	5	8	0.3	1	3	30	0.70		52	5	8	3	0.3	1	3	60	0.97
3	3	5	8	0.3	1	3	30	0.85		53	5	8	3	0.3	3	1	30	0.55
4	3	5	8	0.3	1	3	30	0.97		54	5	8	3	0.3	3	1	30	0.70
5	3	5	8	0.1	1	3	30	0.55		55	5	8	3	0.3	3	1	30	0.85
6	3	5	8	0.1	1	3	30	0.70		56	5	8	3	0.3	3	1	30	0.97
7	3	5	8	0.1	1	3	30	0.85		57	5	8	3	0.3	1	6	30	0.55
8	3	5	8	0.1	1	3	30	0.97		58	5	8	3	0.3	1	6	30	0.70
9	3	5	8	0.3	1	3	60	0.55		59	5	8	3	0.3	1	6	30	0.85
10	3	5	8	0.3	1	3	60	0.70		60	5	8	3	0.3	1	6	30	0.97
11	3	5	8	0.3	1	3	60	0.85		61	5	3	8	0.3	1	3	30	0.55
12	3	5	8	0.3	1	3	60	0.97		62	5	3	8	0.3	1	3	30	0.70
13	3	5	8	0.3	3	1	30	0.55		63	5	3	8	0.3	1	3	30	0.85
14	3	5	8	0.3	3	1	30	0.70		64	5	3	8	0.3	1	3	30	0.97
15	3	5	8	0.3	3	1	30	0.85		65	5	3	8	0.1	1	3	30	0.55
16	3	5	8	0.3	3	1	30	0.97		66	5	3	8	0.1	1	3	30	0.70
17	3	5	8	0.3	1	6	30	0.55		67	5	3	8	0.1	1	3	30	0.85
18	3	5	8	0.3	1	6	30	0.70		68	5	3	8	0.1	1	3	30	0.97
19	3	5	8	0.3	1	6	30	0.85		69	5	3	8	0.3	1	3	60	0.55
20	3	5	8	0.3	1	6	30	0.97		70	5	3	8	0.3	1	3	60	0.70
21	8	5	3	0.3	1	3	30	0.55		71	5	3	8	0.3	1	3	60	0.85
22	8	5	3	0.3	1	3	30	0.70		72	5	3	8	0.3	1	3	60	0.97
23	8	5	3	0.3	1	3	30	0.85		73	5	3	8	0.3	3	1	30	0.55
24	8	5	3	0.3	1	3	30	0.97		74	5	3	8	0.3	3	1	30	0.70
25	8	5	3	0.1	1	3	30	0.55		75	5	3	8	0.3	3	1	30	0.85
26	8	5	3	0.1	1	3	30	0.70		76	5	3	8	0.3	3	1	30	0.97
27	8	5	3	0.1	1	3	30	0.85		77	5	3	8	0.3	1	6	30	0.55
28	8	5	3	0.1	1	3	30	0.97		78	5	3	8	0.3	1	6	30	0.70
29	8	5	3	0.3	1	3	60	0.55		79	5	3	8	0.3	1	6	30	0.85
30	8	5	3	0.3	1	3	60	0.70		80	5	3	8	0.3	1	6	30	0.97
31	8	5	3	0.3	1	3	60	0.85		81	5	5	5	0.3	1	6	30	0.55
32	8	5	3	0.3	1	3	60	0.97		82	5	5	5	0.3	1	6	30	0.70
33	8	5	3	0.3	3	1	30	0.55		83	5	5	5	0.3	1	6	30	0.85
34	8	5	3	0.3	3	1	30	0.70		84	5	5	5	0.3	1	6	30	0.97
35	8	5	3	0.3	3	1	30	0.85		85	5	5	5	0.3	1	3	30	0.55
36	8	5	3	0.3	3	1	30	0.97		86	5	5	5	0.3	1	3	30	0.70
37	8	5	3	0.3	1	6	30	0.55		87	5	5	5	0.3	1	3	30	0.85
38	8	5	3	0.3	1	6	30	0.70		88	5	5	5	0.3	1	3	30	0.97
39	8	5	3	0.3	1	6	30	0.85		89	5	5	5	0.1	1	3	30	0.55
40	8	5	3	0.3	1	6	30	0.97		90	5	5	5	0.1	1	3	30	0.70
41	5	8	3	0.3	1	3	30	0.55		91	5	5	5	0.1	1	3	30	0.85
42	5	8	3	0.3	1	3	30	0.70		92	5	5	5	0.1	1	3	30	0.97
43	5	8	3	0.3	1	3	30	0.85		93	5	5	5	0.3	1	3	60	0.55
44	5	8	3	0.3	1	3	30	0.97		94	5	5	5	0.3	1	3	60	0.70
45	5	8	3	0.1	1	3	30	0.55		95	5	5	5	0.3	1	3	60	0.85
46	5	8	3	0.1	1	3	30	0.70		96	5	5	5	0.3	1	3	60	0.97
47	5	8	3	0.1	1	3	30	0.85		97	5	5	5	0.3	3	1	30	0.55
48	5	8	3	0.1	1	3	30	0.97		98	5	5	5	0.3	3	1	30	0.70
49	5	8	3	0.3	1	3	60	0.55		99	5	5	5	0.3	3	1	30	0.85
50	5	8	3	0.3	1	3	60	0.70		100	5	5	5	0.3	3	1	30	0.97

OVERALL CONCLUSIONS

This research studied inventory management within three non-traditional supply chain structures. To model the unique structures, we assumed base stock replenishment and Poisson demand in all three articles. We used computer simulation to estimate costs, and then minimized costs using perturbation analysis, marginal analysis, and heuristics.

Where backorder costs are significantly different across different groups of customers, we developed, in Article 1, an inventory rationing model with a cost objective. Using the structure of a serial stage inventory system, we formulated a new model that can be solved quickly by selecting an intelligent starting solution.

Within the hospital supply chain, we were able to demonstrate, in Articles 2 and 3, how inaccurate inventory records can lead to increased costs and decreased service levels. Where counting was not enough to completely alleviate this affect, we demonstrated, in Article 3, the usefulness of emergency ordering to support the patient care efforts within the supply chain.