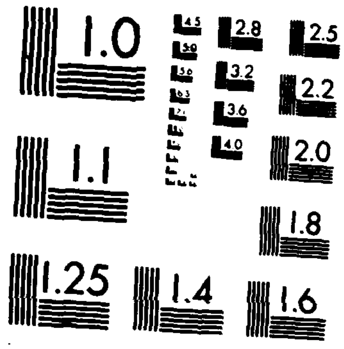
AD-A173 731 OPERATING POLICIES FOR NON- STATIONARY TWO-ECHELON INVENTORY SYSTEMS FOR REPARABLE ITEMS(U) ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY ABERDEEN PROVING GROU 1/2

UNCLASSIFIED M H KOTKIN MAY 86 AMSAA-SP-39 F/G 15/5 NL



!CROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(2)

AMSAA

ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY

AMSAA SPECIAL PUBLICATION NO. 39

OPERATING POLICIES FOR NON-STATIONARY TWO-ECHELON INVENTORY SYSTEMS FOR REPARABLE ITEMS

MEYER H. KOTKIN

MAY 1986

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

U S ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND, MARYLAND 21005

DISPOSITION

Destroy this report when no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so specified by other official documentation.

WARNING

Information and data contained in this document are based on the input available at the time of preparation. The results may be subject to change and should not be construed as representing the DARCOM position unless so specified.

TRADE NAMES

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. The report may not be cited for purposes of advertisement.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Special Publication No. 39	2. GOVT ACCESSION NO. AD-A173	3. RECIPIENT'S CATALOG NUMBER 731
4. TITLE (and Subtitle) OPERATING POLICIES FOR NON-STATIONARY TWO-ECHELON INVENTORY SYSTEMS FOR REPARABLE ITEMS		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) MEYER H. KOTKIN		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Inventory Research Office Room 800, Custom House, 2nd & Chestnut Streets Philadelphia, PA 19106-2976		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Materiel Command 5001 Eisenhower Avenue Alexandria, VA 22333-0001		12. REPORT DATE May 1986 13. NUMBER OF PAGES 177
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) U.S. Army Materiel Systems Analysis Activity Aberdeen Proving Ground, MD 21005-5071		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multi-Echelon
Non-Stationary
Transient Response
Inventory Control Policies
Non-Linear Integer Programming
Stock Policies
Customer Wait

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

There are many situations (for example, wartime and the introduction of a new product into the market) where the transient behavior of a multi-echelon inventory system is of paramount importance and in these situations, steady state models may be inadequate both in describing the relevant transient behavior of the system and in determining the least cost mix, quantity, and distribution of spares that meet specified performance objectives. In this report we develop a model and methodology for determining "cost effective"

20. ABSTRACT (continued)

inventory policies for a two-echelon non-stationary one-for-one inventory system for reparable items.

By studying the supply interactions between echelons, we analytically obtain the time dependent distributions of the number of units in the pipeline at each location in the system. We also obtain the time dependent distributions of customer wait at each echelon. These results allow us to define performance measures which adequately distinguish and rank different stock policies. Under the assumption that during the horizon management can not change the stock level of any item at any location, we formulate and solve the Fixed Asset Vector Problem (FAVP) for determining, over a catalog of items, the least cost stock levels that meet management specified catalog performance targets. We also develop a heuristic to use the FAVP to solve the general inventory control problem in which management is allowed to change the stock levels at one or more locations in the system.

Finally, we investigate approximations that reduce the computational burden of solving the FAVP. We demonstrate on data from three U.S. Army Weapon Systems that the pipeline distributions can be approximated by negative binomial distributions, thereby significantly reducing the computational effort. We also briefly discuss the sensitivity of the FAVP to changes in the input data. As a consequence of this discussion, we briefly address the issue of using the more convenient and facile steady state models to obtain stock levels for non-stationary two-echelon inventory systems.

CONTENTS

LIST OF FIGURES

Figure

Next page is blank.

LIST OF TABLES

Next page is blank.

GLOSSARY OF FREQUENTLY USED SYMBOLS

$a(y,t)$	1 if a unit that entered the depot pipeline at time y has left the pipeline by time t. $a(y,t) = 0$ otherwise.
aeb	Management target on the average expected total number of backorders at the bases over the horizon.
$a_i(t)$	Intensity of base_i resupply requests placed on the depot.
$A_i(t)$	Mean value function for the number of base_i resupply requests placed on the depot.
AVTP	Asset Vector Transformation Problem.
asset level	Maximum number of serviceable units that can be on-hand.
asset policy	Specification of an asset vector for every point in time during the horizon.
asset position	Asset level.
asset vector	Vector containing an asset level for every item at every location.
$b(t_1,t_2,t_3)$	Probability that a unit that entered the depot pipeline in $(t_1,t_2]$ is still in the pipeline at time t_3.
$B(n,p,k)$	Probability that a Binomial $[n,p]$ random variable is greater than or equal to k.
$B_j(t)$	Number of backorders outstanding at location_j at time t.
$bi(n,p,k)$	Probability that a Binomial $[n,p]$ random variable is equal to k.
C_i	Procurement/Holding cost for item_i.

$c_i(t)$	$a_i(t)/\lambda_0(t)$.
$D_j(t)$	Number of units in the location_j diagnostic facility at time t.
$DE_i(t)$	Number of units due-in to base_i from the depot at time t.
$e_i(s,y,t)$	Probability that a serviceable unit arrived at base_i by time t to replace a condemnation at time y of a unit that failed at base_i at time s.
$E_j(t)$	Number of units due-in to location_j from the external supplier at time t.
$ER_i(t)$	Number of units en route from the depot to base_i at time t.
FAVP	Fixed Asset Vector Problem.
$F_0(s,t)$	Probability that a unit that entered the depot pipeline at time s has left the pipeline by time t.
$G_i(s,t)$	Probability that a base_i diagnosis begun at time s has ended by time t.
$H(k,y,h,t)$	Number of units in the depot pipeline at time t given that k units entered in $(0,y]$, one unit entered at time y and h units entered in $(y,t]$.
$H_j(t)$	Number of units on-hand at location_j at time t.
HCP	Horizon Control Policy.
I	Number of items in the catalog.
$I_i(n)$	1 if the n^{th} demand on the depot was from base_i and $I_i(n) = 0$ otherwise.
$INT(x)$	Largest integer less than or equal to x.
$L(k,t_1,t_2,t_3)$	The number of units, out of the k units that entered the depot pipeline in $(t_1,t_2]$, that are still in the depot pipeline at time t_3.

$\lambda_j(t)$	Intensity of demand at location$_j$.
$m_0(t)$	MVF of demand at the depot.
meb	Management target on the maximum expected total number of backorders at the bases over the horizon.
MVF	Mean value function.
N	Number of bases.
$N_j(t)$	Number of demands at location$_j$ in (0,t].
$N_i^0(t)$	Number of base$_i$ resupply requests placed on the depot in (0,t].
NHPP	Non-Homogeneous Poisson Process.
NHCPP	Non-Homogeneous Compound Poisson Process.
1_N	A vector of ones with dimension N+1.
OST_i^*	Fixed order and ship time between the depot and base$_i$.
p[x;u]	Probability that a Poisson random variable with mean u is equal to x.
P[x;u]	Probability that a Poisson random variable with mean u is less than or equal to x.
$P_i(s,t)$	Probability that a unit that fails at time s and completes base$_i$ diagnosis at time t will be sent to the depot.
$P_R^i(s,t)$	Probability that a unit that fails at time s and completes base$_i$ diagnosis at time t will be sent to the base$_i$ repair facility.
$Q_i(t)$	Number of depot backorders outstanding at time t that are due-out to base$_i$.
$r_i(s,y,t)$	Probability that a unit that fails at base$_i$ at time s and is sent to the base$_i$ repair facility at time y will complete repair by time t.

$R_j(t)$	Number of units in the location_j repair facility at time t.
RTCP	Real-Time Control Policy.
$s_j(t)$	Asset level at location_j at time t.
$s_{++}(t)$	Asset vector at time t.
$s_{i+}(t)$	Vector containing the asset level of item_i at every location at time t.
T	Horizon length.
T(n)	Time of the n^{th} demand on the depot.
$V_0(t)$	Number of units en route to the depot from the bases at time t.
VAR	Variance.
VMR	Variance to mean ratio.
$W_j(t)$	Delay before satisfying a demand at location_j at time t.
$X_j(t)$	Pipeline quantity at location_j at time t.
$Y_0(t)$	Origination time of first base resupply request that will be satisfied by the depot after time t.

CHAPTER I

INTRODUCTION

Strictly speaking, real world inventory systems never reach steady state. A common characteristic of most inventory systems is that they are continually changing with time. The stochastic processes generating demands, order and ship times, and repair times may change with time as might the various costs that are of interest. In many instances, however, the changes may occur slowly enough or be subtle enough so that for considerably long periods of time the inventory system can be treated as if it were in a steady state mode of operation. For this reason, stationary steady state models have been developed and applied to many practical problems of inventory management and control in multi-echelon inventory systems. For example, the United States Army Materiel Support Commands use a stationary multi-echelon model (U.S. Army [1983]) to provision billions of dollars of reparable spare parts. For many inventory systems, including the Army system during peacetime, steady state models and the assumptions of stationarity embodied in them have been invaluable. These models have proven to be convenient in terms of input data collection and computational burden, and to be adequate for determining cost effective stock levels. However, there are many situations where the short-term behavior of the inventory system is of paramount importance and in these situations, stationary models may be inadequate both in describing the relevant transient behavior of the system and in determining the least cost mix, quantity and distribution of spares that meet specified performance objectives. For situations where steady state is not attained, for example, because of a short time horizon and/or because the underlying stochastic processes governing the behavior of the inventory system are not stationary, it may be necessary to use a non-stationary multi-echelon model that explicitly models the interactions among the echelons and properly describes the transient behavior of the system.

At the onset of a war we have a vivid example of the dynamic behavior that can be exhibited by the processes that generate demands, shipping times, repair times, etc., in a military inventory system. As hostilities begin, the demand for spare parts for weapon systems may show a significant increase over peacetime values. The demand for parts may then decrease as weapon systems are lost through attrition and combat damage and this reduction in demand may continue until replacement weapon systems can reach the combat area, at which time the demand for spare parts may again increase. Meanwhile, ship times between combat units, intermediate maintenance and supply echelons, and a centralized parts depot may fluctuate depending on enemy activity and the availability of different modes of transportation. Furthermore, the repair rate at a maintenance location may initially be small as the location awaits the arrival of specialized technicians and test equipment to be used in repairing damaged weapon systems. When the technicians and test equipment arrive, the repair rate may increase until, possibly, a maximum wartime repair rate is achieved. The repair rate may vary over the duration of hostilities if the repair facilities are subjected to enemy attacks and are rebuilt and remanned after these attacks. Therefore, even for wars of long duration, the dynamic behavior of the inventory system may be such as to preclude use of a stationary steady state model to determine optimal stock levels and resupply policy.

Furthermore, the initial state of the system (e.g., number of units in repair, in transit, on-hand, backordered, etc.), may have a pronounced effect on the achieved performance and thereby, the required stock. If all of the inventory system's spares are on-hand and available for a short combat contingency, we would expect a non-stationary model to yield different results than if all of the systems spares are in the repair facilities. Stationary steady state models ignore the initial state of the system and thereby may grossly misrepresent performance during the contingency.

Another example of the dynamic behavior that can be exhibited by inventory systems occurs when a new product is introduced into the market. (For military systems, this corresponds to the development and fielding of a new weapon system). As more and more units of the product are introduced into the market, the demand for spare parts increases. Repair times may be long until repairmen gain experience with the new items. Design changes may affect the demand and repair rates for the reparable parts of the product. Furthermore, as time goes by and demand and repair time data are collected, new estimates of reliability and maintainability factors such as mean time to failure and mean time to repair may be made. The accurate modelling of this type of dynamic behavior may be crucial to finding a cost effective inventory policy.

Multi-Echelon Inventory Systems

In a general N echelon inventory system, the highest echelon (echelon-N) consists of a lone installation referred to as the depot. Primary customer demands for reparable items usually occur only at the lowest echelon (echelon-1) locations. Echelon-1 is often referred to as the user echelon. Depending on the nature of the required repair, the failed unit is either repaired at the installation to which it has been brought, condemned as irreparable and removed from the system, or sent to a higher echelon for repair. (Locations on the same echelon generally have the same repair capability). At the higher echelon location a decision is again made whether to repair, condemn, or send the unit to a higher echelon. All failed units received at the depot are either condemned or repaired there. Condemned units may be replaced by procurement from an external supplier.

At each location in the system, there are continuous time stochastic processes which govern the behavior of the parameters of operation (PO) which are the probability distributions for the

a. order and ship time to each location that this particular location resupplies;

b. times to fault diagnose and isolate and to repair failed units at that location;

c. decision to condemn, repair, or send a failed unit to a particular higher echelon;

d. shipping time to each location from which this location seeks resupply;

e. costs to buy, hold, and scrap.

Furthermore, there are stochastic counting processes generating primary customer demands on the system which, in combination with the processes generating the PO, yield stochastic counting processes describing the demands for serviceable units at each location. There are many inventory systems for which the processes generating the PO are known and stationary and for which the demand process is suitably well-behaved so that passage to steady state is theoretically assured. Stationary models describe the steady state behavior of these stationary systems. As we have already seen there are also many important non-stationary inventory systems for which one or more of the PO change with time and/or for which the demand process, planning horizon and/or initial conditions are such that the transient behavior of the system must be analyzed. Models that describe these systems will be referred to as non-stationary models.

Inventory Control Policies

The status of an item at a location at time t contains the following information:

a. the number of units on-hand;

b. the number of units backordered;

c. the number of units in-repair;

d. a probability distribution for the remaining repair time of each unit in-repair;

e. the number of units on-order from the depot and the external supplier;

f. a probability distribution for the delivery time of each unit on-order.

The system condition at time t is defined as the status of every item at every location at t. Note that the initial system condition at 0 is simply the initial status of every item at every location.

Ideally, management would at all times like to have complete knowledge of the system condition. Inventory control decisions regarding resupply, procurements, disposals, and redistributions of spares could then be made in continuous time on an as-needed basis. These decisions would be based on the current system condition (or subset of the system condition) and the stochastic description of the behavior of the PO and the demand process over the rest of the horizon. This real time control policy (RTCP) offers management great flexibility in positioning spares in the system to improve inventory performance. Management can react quickly to updated estimates of the PO and demand parameters and can also quickly compensate for random phenomena such as unexpectedly small or large demand at one or more locations. Under a RTCP, management also has the ability to adapt future operating rules and policies based on the current system condition and knowledge of the PO and the demand process.

However, there are two major problems that make implementing a RTCP for a multi-item multi-echelon inventory system difficult, if not impossible. First, there is the need to continuously store and monitor the information on the system condition, PO and demand process upon which management decisions are based. For many systems this may be impossible because of the number of items and locations in the system and the cost of monitoring and storing this much data. Secondly, even if the data are continuously monitored and stored, the effort and cost involved in continuously determining and updating operating policies is prohibitive. In a very strict sense, there is an Uncertainty Principle involved in the calculation of operating policies under a RTCP. Since calculations are not instantaneous, by the time a new operating policy is determined the data may have changed so that this new policy is no longer optimal or desirable.

As an alternative to a RTCP, management can, at time 0, decide at which times during the horizon the system will be reviewed and data on the system condition, PO and demand process collected. Between system review times, an operating doctrine is followed at each location. The operating doctrine may be expressed in terms of any subset of system condition and/or item status at a location that is continuously available to the system and/or location. Only at system review times can the parameters of the operating doctrine at each location or the operating doctrine itself be changed. All of the classical operating doctrines of inventory theory such as two bin ordering policies are examples of this horizon control policy (HCP).

The U. S. Army has four maintenance and supply echelons. There are too many items and locations for a centralized facility even to monitor continuously just the number of units on-hand at every location. Rather, the Army uses an HCP. Each location monitors only its own inventory position and uses a (Q,r) policy between quarterly management directed system reviews. The reorder point and reorder quantity for each location are determined each quarter based on the information on the PO, demand process and system condition collected at the time of review. Redistributions of spares through the system may also be ordered at these times. While the HCP does not give the Army the full power of a RTCP, it represents the best viable alternative.

Literature Review

Stationary Multi-Echelon Models and Analysis

There has been extensive modelling and analysis of stationary multi-echelon inventory systems. Examples of some of these efforts can be found in Ignall and Veinott (1969), Silver (1972), Graves and Schwarz (1977), and Kim (1978). In 1968, Sherbrooke formulated the well known METRIC (Multi-Echelon Technique for Recoverable Item Control) model for stationary multi-echelon inventory systems for reparable items. METRIC's initial use was for military inventory systems but it has now been applied in private industry as well. In METRIC, Sherbrooke attempted to model explicitly the interactions among the various echelons in the inventory system. He assumed that all locations followed a continuous review (S-1,S) or one-for-one resupply (ordering) policy. Each time a location sent a failed unit to a higher echelon location for repair, the higher echelon location would resupply the lower echelon location with a serviceable unit as soon as possible. Therefore, an optimal policy required the determination of only the single critical number S for each location which was the constant asset position (number on-hand + on-order + in-repair - backorders) for that location. The (S-1,S) resupply assumption, along with the assumption of a homogeneous Poisson Process generating primary customer demands, greatly simplified the mathematical analysis and considerably reduced the computational burden involved in determining an optimal policy. Muckstadt (1973) extended Sherbrooke's work to include multi-indentured items. An (S-1,S) policy was followed at every location for every level of indenture (modules, components, sub-assemblies, etc.). Both METRIC and Muckstadt's MODMETRIC have been extensively analyzed in the literature (see, for example, Simon [1971], Shanker [1977] and Kotkin [1978]) and there have been many variants of these models (Mason [1975], Clark [1978] and Vincent [1980]). Simon (1971) corrected METRIC's misuse of Palm's Theorem for $M/G/\infty$ queues (Ross [1970]) by developing exact expressions for the number of units due-in to user locations in a two echelon system when the depot repair time was fixed. Kruse (1979) simplified Simon's expressions and extended Simon's work to more than two echelons. Approximations to Simon's computationally cumbersome model were developed by Slay (1980) and Kaplan (1980). Graves (1983) rediscovered Slay's negative binomial approximation and again demonstrated its effectiveness. Excellent reviews of stationary multi-echelon models may be found in Clarke (1972) and Nahmias (1981).

Non-Stationary Multi-Echelon Models and Analysis

Clark (1960) reported on a simulation for a periodic review non-stationary multi-echelon inventory system for reparable items. Unit purchase and holding costs were allowed to change by period and condemnations and transshipments between locations during a period were allowed. Stock could be redistributed among the locations at the beginning of each period. All customer backorders were passed up to the highest echelon location. No description of the simulation or of the heuristic used to determine stock levels for each period was given.

Bessler and Veinott (1966) studied a general arborescent multi-echelon periodic review system for consumable items. They assumed no delivery lags at any location; no fixed ordering (or setup) cost; and that any demands that could not be met at a location were passed on to the location's direct supplier at a higher echelon so that backorders existed only at the highest echelon location. By establishing a correspondence between the multi-product single facility problem studied by Veinott (1965) and their multi-facility single product problem, they determined conditions under which the optimal policy for an N-period problem could be expressed as the order-up to level solution of N single period problems.

Ignall and Veinott (1969) extended the work of Bessler and Veinott to include delivery lags. They also allowed a more general supply structure in which any facility could satisfy shortages at any other facility providing that the transferred stock was replaced from an exogenous source at the beginning of the next period. The authors gave sufficient conditions under which myopic single period order-up to policies were optimal for the N-period problem.

Burns and Sivazlian (1978) used control theory to investigate the dynamic response of a non-stationary multi-echelon inventory system to demands placed upon it. They studied a cost free multi-echelon periodic review inventory system for consumable items. Locations on a particular echelon were resupplied only by a location on the next higher echelon with the highest echelon location receiving resupply from an exogenous source. The amount ordered at each location at the beginning of each period consisted of a replacement quantity for actual demands in the previous period and an inventory "adjustment" or hedging quantity which allowed a location to adjust its on-hand stock at the end of a period to a desired level of inventory ownership (safety level). This level of ownership was expressed as a certain number of periods worth of expected demand at that location: the expectation being a first order exponentially smoothed average of past demand. (This is similar to the ordering policy considered by Bessler and Zehna (1968) for single echelon systems and by Burns (1970) for multi-echelon systems). Burns and Sivazlian noted that under this ordering policy higher echelons would over react to lower echelon inventory adjustments. This was called a "false-order" effect. Minor variations in demand at the user echelon were amplified by the inventory system into major disturbances at the higher echelons. Using simulation they demonstrated the superiority of an ordering rule they developed which tried to eliminate these false-order effects.

Kotkin and Rhoads (1977) used a simulation to test a heuristic for using a stationary model to determine stock levels in a non-stationary three echelon multi-indentured inventory system for low demand items. All PO were assumed deterministic, known and constant over the horizon and demands were assumed to form a non-homogeneous Poisson Process whose intensity factor was monotone increasing over the horizon. The horizon was divided into convenient periods and in each period a stationary model (MODMETRIC, Muckstadt [1973]) was used to recalculate the asset position for the modules and components at each location. Redistributions and exogenous additions of module stock, when necessary, were instantaneously made at the beginning of each period. Over the horizon, module stock levels were monotone increasing but components levels fluctuated. Since no disposal of stock was allowed, component stock levels were fixed at their steady state levels based on the values of the PO and the demand rate at the end of the horizon. The authors found that judicious use of the stationary model yielded close to the optimal module stock levels. The success of the heuristic was credited to the fact that the product had very low demand so that redistributions and additions of module stock were not frequent over the horizon.

Muckstadt (1980), in studying a two-echelon inventory system for reparable items, assumed that all PO were deterministic, known and constant and that demands at the user level formed a non-homogeneous Poisson Process with known intensity. In order to calculate the non-stationary distribution of the number of units in resupply to a user location, he used an argument similar to Sherbrooke's (1968) argument for stationary two echelon systems. Given the depot stock level at time t, the time dependent version of Palm's Theorem for $M(t)/G(t)/\infty$ queues (Ross [1972], Hillestad and Carillo [1980]) was used to calculate the number of depot backorders outstanding at t. The delay at the depot before a serviceable unit could be shipped to a user location that requested resupply at t was calculated as the expected number of depot backorders outstanding at t divided by the average depot demand rate in $(t-R,t]$. R was the deterministic depot repair cycle time. Here, Muckstadt used the steady state queueing law $L = \lambda W$ as an approximation to the transient behavior of the system. Furthermore, no account was taken of the fact that depot stock might change after time t thereby affecting the delays experienced by user locations. (In Chapter III we derive the exact expressions for the depot delay). As in Sherbrooke's METRIC, the delay term was used to find the expected number of depot backorders belonging to a particular user location and this was added to the mean number of units in repair at and en route to the user location. Muckstadt then posited a Poisson distribution for the total number of units due-in to the user location. Muckstadt did not formulate an optimization problem for determining the optimal stock levels over time for the locations in the inventory system.

Muckstadt, possibly without realizing it, used a heuristic that apportioned the depot backorders outstanding at t to the various user locations according to the proportion of the depot demand in (t-R,t] that came from each user location. This was essentially what Simon (1971) did for stationary two echelon systems. As we shall see in Chapter IV, in non-stationary systems this is correct in only one very special case. Hillestad (1982) used a similar heuristic in his two-echelon Dyna-METRIC model by apportioning the depot backorders at t to the user locations by the proportion of the total depot demand from each location over some "empirically" determined though unspecified time interval.

Dyna-METRIC made provisions for indenture levels and various degrees of controlled substitution. (The Department of Defense differentiates between cannibalization and controlled substitution according to whether the weapon system/end item on which the unit is located will eventually be repaired). Hillestad proposed an optimization problem that considered the inventory system performance at times of interest specified by management. At each of these times, the cost of procuring additional stock beyond current system assets was minimized subject to a constraint on performance at this particular time only. Performance between the times specified was not considered.

Gross and Miller (1982) studied the transient behavior of a two-echelon Markovian system using the uniformization technique (Grassman [1977]) to obtain numerical solutions to the Chapman-Kolmogorov equations.

They allowed for finite repair capacity and state dependent failure rates. No disposals or external procurements were allowed. They used a one-for-one resupply rule and FCFS queue disciplines except when the depot and some user locations were backordered. In that case, when a unit completed depot repair it was sent to the user location with the highest number of outstanding backorders rather than the location at the head of the depot resupply queue.

Their state space grows as the square of the product of the number of locations in the system, the number of items in the system, and the stock of each item at each location. This quickly becomes unmanageable even after some state reduction techniques and therefore makes it prohibitive for inclusion in an optimization scheme.

Real-Time Multi-Echelon Models and Methods

In this section we review two approaches toward real-time control and management of inventory systems of the METRIC type. The problem considered was how best to utilize a given number of reparable spares in a two echelon system that consisted of a depot and user locations called bases. Instead of rigidly following an (S-1,S) resupply policy, these real time models made decisions about resupply on an as-needed basis. Therefore, even though a base sent a failed unit to the depot for repair, the depot was not obligated to ship a serviceable unit to the base to replace the failed unit. This allowed greater flexibility than the static resupply policy of stationary models like METRIC in redistributing stock to the various locations. The rationale for this was to provide the supply system the capability to respond to poor estimates of the PO and/or to transient effects such as, for example, an unexpectedly large number of demands at a particular base.

Miller (1968) developed a heuristic for real-time management called Real-Time METRIC (RTM). The initial stock level at each location was determined using METRIC. Stock was redistributed through the system via the depot. The decision of whether to ship a unit from the depot to a base depended on the depot "reluctance" to send available spares and the bases' "need" for serviceable stock. The depot's reluctance at time t was expressed as a heuristic function of on-hand stock at the depot at t. A base's need was defined as the number of backorders expected to be outstanding at the base a deterministic depot to base ship time into the future. A comparison of depot reluctance and base need was made whenever an event (demand at a base or a failed unit completing repair at the depot) occurred that caused either base need to increase or depot reluctance to decrease. A unit was shipped from the depot to the base with the largest need that exceeded depot reluctance. Miller reported a significant reduction in expected backorder days accumulated over a year for high demand items by using RTM instead of METRIC. For low demand items no significant difference was observed.

Each time that a comparison of depot reluctance and base need was made, RTM looked a transportation time ahead. Miller (1974) later showed that this "Transportation Time Look Ahead Policy" would be optimal if the depot repair cycle time were zero. While RTM could, in principle, handle additions to or depletions from system stock, no method was given for determining when to add or delete stock from the system. Clearly, in a non-stationary environment we would not only want to know when and how to redistribute stock but also when to change the overall system stock level.

Galliher and Wilson (1975) improved upon RTM by eliminating the need to use the RTM depot reluctance function and by use of different redistribution rules. As in RTM, stock was redistributed only through the depot, and redistribution decisions were made when a base requested resupply or when a unit finished depot repair. Instead of looking a deterministic transportation time ahead, Galliher and Wilson defined the length of the decision horizon at time t to be R(t), equal to a transportation time plus the expected time between demands at the depot. On-hand units were shipped from the depot spares pool until either no base was expected to have any backorders outstanding at t + R(t) or the depot ran out of stock. Remaining depot stock was shipped to the bases that were below their target levels (set by a stationary model) so that the probability of incurring a backorder in the system during (t,t+R(t)] was minimized.

Both of the above approaches heuristically determined decision horizons. The behavior of demand and the PO after these lengths of time was not considered. For example, if the demand rate at a base decreased considerably after the decision horizon, it may not have been best to ship a unit from the depot even though the base expected to be backordered at the end of the decision horizon. Similarly, no inventories were built up in anticipation of an increase in demand after the decision horizon. More forward looking rules might not have made the same redistribution decisions and might have improved performance.

Scope of Dissertation

The objective of this dissertation is to develop a model for determining "cost" effective stock policies for non-stationary two echelon inventory systems. The model consists of two basic components. The first component is the analytical description of the important stochastic processes, such as on-hand inventory, that determine inventory effectiveness. The second component is an optimization scheme that selects a least "cost" stock policy subject to constraints on inventory performance. We shall also examine approximations, where necessary, that reduce the computational burden and thereby aid in real world implementation.

Throughout this dissertation we deal with the time dependent stochastic nature of the PO and the demand process. However, we do not concern ourselves with uncertainty in the basic parameters of the underlying stochastic processes that govern the behavior of the PO and the demand process. This problem has not been addressed explicitly even in stationary multi-echelon models because it severely complicates the analysis. A non-stationary model and heuristics dealing with uncertainty in the basic parameters would be a logical and worthwhile extension of the work presented here.

Data Availability

The increased data burden in using a non-stationary instead of a stationary model can range from the effort involved in determining the length of the scenario to a mammoth data collection effort if all parameters are continuously changing with time. In this section we cite two situations where data were obtained from U. S. Army sources for use in the non-stationary model developed in this dissertation. These examples do not establish the validity of the work herein. Rather, they demonstrate that data for effective use of non-stationary models in the Army are available.

Example 1.1: Contingency Fly-Away Kits

In a Grenada type contingency, U. S. combat forces are deployed outside the continental U. S. for a short period of time. No external resupply to these forces is possible. The problem is to determine fly-away kits of spares that combat units should carry in order to achieve some specified weapon system performance targets.

The U. S. Army Aviation Systems Command wanted a method for determining the fly-away kits for a division (2 echelons) of Blackhawk helicopters. Contingencies were expected to arise with little or no warning. Therefore, there would not be enough time to run a detailed combat simulation to obtain daily part demand rates or daily information on the PO. The non-stationary data that are available are the length of the scenario and the initial system condition which usually will reflect that all spares will be on-hand at time 0. These data, along with the stationary part data obtained from the standard Army data base (U. S. Army [1983]), form the input to the non-stationary model.

Example 1.2: Wartime Requirements Determination

The U. S. Army Training and Doctrine Command (TRADOC) and the U. S. Army Materiel Systems Analysis Activity have developed detailed combat simulations. From these simulations we obtained, for a specific scenario, daily demand rates and changes in the repair time distributions based on the attack-defense posture of each location in a two echelon system. These data, along with scenario length and initial conditions, were input to the non-stationary model to obtain recommended stock quantities and the expected delays in satisfying resupply requests. These delays were input to another TRADOC simulation of the wartime logistics support structure to evaluate, from a logistics point of view, various combat and budgetary strategies.

Organization of Dissertation

Chapter II catalogs and discusses the basic assumptions made in this dissertation. The chapter also contains background material on non-homogeneous Poisson Processes and one-for-one inventory systems, as well as new results on the recursive calculation of the moments of the distributions of backorders and on-hand inventory. Chapters III and IV derive the time dependent probability distributions of the pipeline and other important stochastic processes at the depot and bases, respectively. Chapter V discusses evaluation of inventory performance in non-stationary systems and also proposes two performance measures to be considered in an optimization problem. The optimization problem, the Fixed Asset Vector Problem (FAVP), assumes that the asset position at each location is fixed at the beginning of the horizon and that a one-for-one resupply policy is followed thereafter. In Chapter VI we examine approximations to the bases' computationally cumbersome pipeline distributions and propose approximating the actual base pipeline distribution with a negative binomial distribution. In Chapter VII we discuss computational experience with the FAVP on weapon systems managed by the U. S. Army

Materiel Command. We also compare the stock lists and costs obtained from the FAVP with those obtained from a stationary steady-state model. Chapter VIII introduces the Asset Vector Transformation Problem which, together with the FAVP, can be used in a heuristic to try to determine the optimal asset levels over time in a non-stationary inventory system. Chapter IX contains a summary of the major results of this dissertation and brief discussions on various extensions to the work presented herein.

CHAPTER II

PRELIMINARIES

This chapter presents some basic definitions and assumptions used throughout this study as well as background material on non-homogeneous Poisson Processes and one-for-one [(S-1,S)] inventory systems. The final section contains new results on the recursive calculation of moments of the distributions of backorders and on-hand inventory in one-for-one inventory systems.

Non-Homogeneous Poisson Process

The stochastic counting process $\{N(t),\ t \geq 0\}$ is said to be a non-homogeneous Poisson Process (NHPP) with intensity function $\lambda(t)$ if

(i) $N(0) = 0$;

(ii) $\{N(t),\ t \geq 0\}$ has independent increments;

(iii) $\Pr(2 \text{ or more events in } [t,t+h]) = O(h)$;

(iv) $\Pr(\text{exactly one event in } [t,t+h]) = \lambda(t)h + O(h)$.

These four conditions are often referred to as the axioms defining a NHPP. Parzen (1962), among others, shows that these conditions ensure that for any half-open interval $(s,t]$, $0 \leq s \leq t$ for which $\lambda(\cdot)$ is not identically zero, and $n = 0, 1, 2, \ldots,$

$$\Pr(N(t) - N(s) = n) = e^{-[m(t) - m(s)]} \frac{[m(t) - m(s)]^n}{n!}$$

where $m(t) = \int_0^t \lambda(x)dx$ is the mean value function (MVF) so referred to since $E[N(t) - N(s)] = m(t) - m(s)$ and in particular, $E[N(t)] = m(t)$. If $\lambda(x) = 0$ for all x in $(s,t]$ then $N(s) = N(t)$. Throughout this study we assume that $m(t)$ is differentiable and the derivative of $m(t)$ is thus simply the intensity of the NHPP at t. We also assume that $\lambda(t)$ is finite for all t. Note that $m(t)$ is non-decreasing and always right continuous because of the definition of a NHPP but in this study because of the assumption of the differentiability of $m(t)$, $m(t)$ is necessarily continuous. (See Cinlar [1975] for a brief discussion of non-continuous MVF and reference to the more complete work of Khinchine [1960] on non-continuous MVF).

Appendix A catalogs some useful properties of NHPP. Most notable are P1, ensuring that the superposition of independent NHPP is itself a NHPP, and P10, the Splitting Property, concerning the decomposition of a NHPP into independent constituent NHPP by means of an independent, though possibly time varying, splitting mechanism. These, and many of the other properties listed in Appendix A are analogs of results for homogeneous Poisson Processes. These results might suggest a relationship between non-homogeneous and homogeneous Poisson Processes. In fact, such a relationship does exist.

For a NHPP $\{N(t), t \geq 0\}$ with MVF $m(t)$ let $m^{-1}(u) = \inf \{s: m(s) \geq u\}$ be the inverse of the MVF. If $\lambda(t) > 0$ for all t so that the MVF is strictly increasing, then $m^{-1}(u)$ is the classical inverse. Otherwise, $m^{-1}(u)$ is the inverse of the function $m(\cdot)$ restricted to a domain consisting of points where the MVF strictly increases. In either case, $m^{-1}(u)$ is strictly increasing and it is continuous because of the continuity of the MVF (Kitchen [1968]). If $N^*(u) = N[m^{-1}(u)]$ then it is easy to see that $\{N^*(u), u \geq 0\}$ is a homogeneous Poisson Process with MVF $E[N^*(u)] = E[N(m^{-1}(u))] = m[m^{-1}(u)] = u$ and intensity 1. The counting process $\{N^*(u), u \geq 0\}$ records the number of events occurring on a transformed time scale that measures u when the real time, in the sense of the NHPP $\{N(t), t \geq 0\}$, is $m^{-1}(u)$. Cinlar (1975) gives the following theorem restating the above result in terms of the times of arrivals.

<u>Theorem 2.1</u>: $t_1, t_2, \ldots$ are the arrival times in a NHPP with MVF $m(t)$ if and only if $m(t_1), m(t_2), \ldots$ are the arrival times in a homogeneous Poisson Process with intensity 1. //

Therefore, every NHPP with continuous MVF can be converted to a homogeneous Poisson Process. Theorem 2.1 can be used to prove many of the properties listed in Appendix A by appealing to the corresponding result for homogeneous Poisson Processes. Theorem 2.1 is extremely useful in Monte Carlo simulations involving NHPP (Cinlar [1975]).

In a NHPP, the magnitude of an event is always one. Consider the case where the event magnitude at time t is independent of the magnitude of other events and has distribution function $F(x,t)$ $x \geq 0$, $t \geq 0$, with Laplace-Stieltjes transform $F^*(w,t)$. If we replace axioms (iii) and (iv) with

$$\text{(iiic)} \quad \Pr(N(t+h) - N(t) = x) = \lambda(t)d_x F(x,t)h + O(h)$$

then we have a set of axioms defining a non-homogeneous compound Poisson Process (NHCPP). By writing the forward or backward Kolmogorov equations it is straightforward to show that

$$E[e^{-wN(t)}] = e^{-m(t)[1-C(w,t)]} \qquad (2.1)$$

where

$$m(t) = \int_0^t \lambda(s)ds$$

$$C(w,t) = \frac{1}{m(t)} \int_0^t F^*(w,s)\ (s)ds.$$

Note that $C(w,t)$ is the "intensity-weighted" average of the Laplace-Stieltjes transforms over $(0,t]$ (Galliher [1975]). If $d_x F(1,t) = 1$ for all t, then setting $e^{-w} = z$ reduces (2.1) to the generating function of a NHPP.

Analogously to homogeneous compound Poisson Processes, a NHCPP can be viewed as the stochastic process that records the sum of the independent jump magnitudes of a process whose jump times occur in accordance with a NHPP. Thus, if $0 < s_1 \leq s_2 \ldots \leq s_{N(t)} \leq t$ are the epochs of the $N(t)$ events that have occurred by time t in a NHPP $\{N(s),\ s \geq 0\}$ and $y(s_j)$ is the magnitude of the event at time s_j with distribution $F(x,s_j)$, $j = 1, 2, \ldots, N(t)$, then as long as the jump magnitudes are mutually statistically independent random variables, $\{N_c(t) = \sum_{j=1}^{N(t)} y(s_j),\ t \geq 0\}$ is a NHCPP and the transform (2.1) is readily obtained by direct arguments.

Major Assumptions

For expository purposes, throughout most of this study we focus attention on a two echelon inventory system consisting of N user echelon locations called bases and a central second echelon resupply and maintenance depot. The flow of units of an item through a two base system is depicted in Figure 1. In this section we list and discuss the major assumptions made regarding the operation of the system.

Assumption 1: Primary customer demands at $base_i$, $i = 1,2,\ldots N$, form

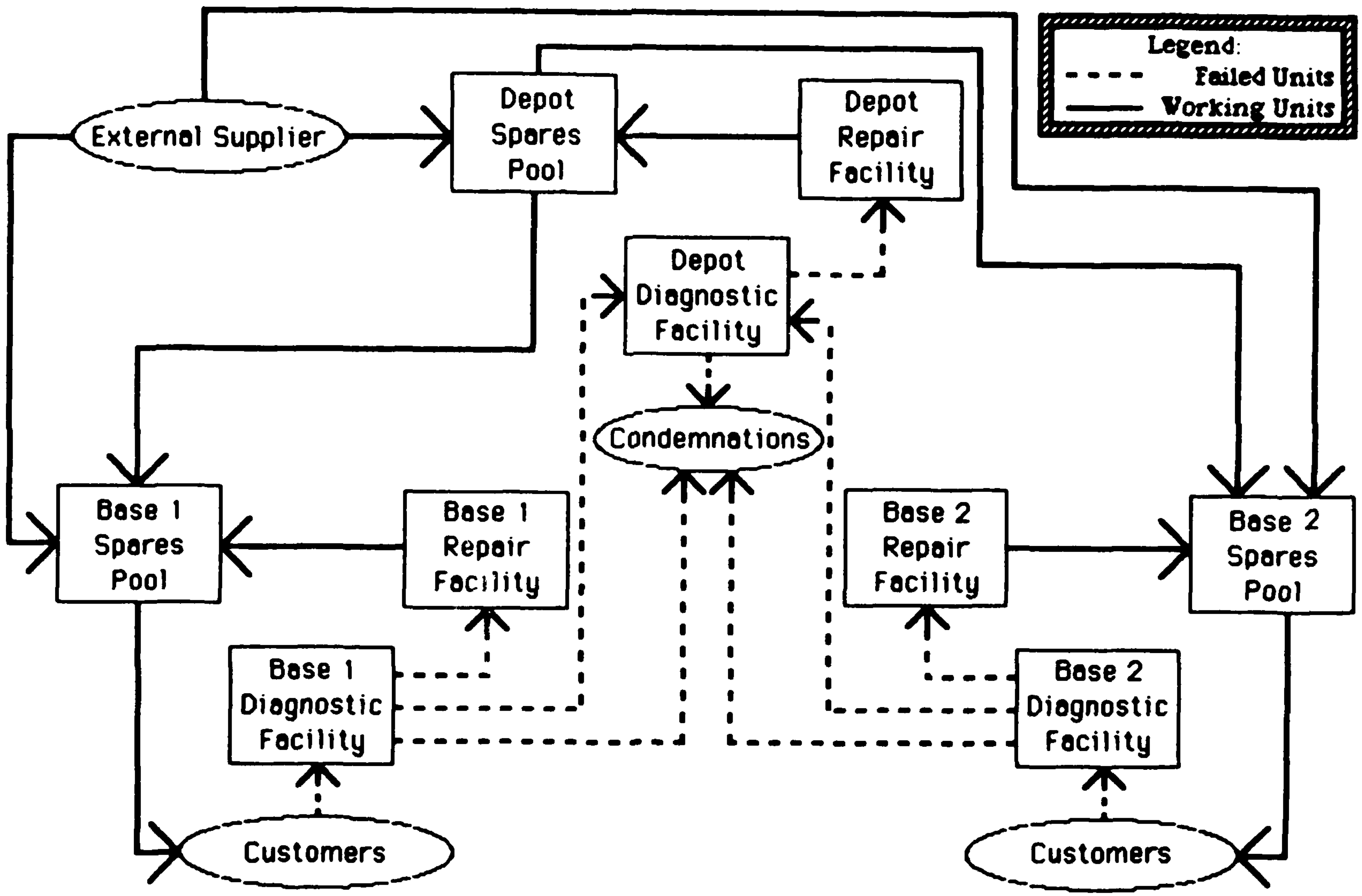


Figure 1: Flow of Units in a Two-Echelon Two Base System

a NHPP with known intensity $0 \leq \lambda_i(t) < \infty$ and differentiable MVF $m_i(t) = \int_0^t \lambda_i(x)\,dx$. The demand processes at the bases are mutually statistically independent.

To the author's knowledge, no studies of the demand processes of any real world non-stationary inventory systems have been performed. The Armed Forces often conduct training exercises and war games but as of yet no reports on any data collected have appeared. In combat situations, however, item failures not due to combat damage are often assumed to follow a homogeneous or non-homogeneous Poisson Process (Coggins [1983]). Studies of stationary systems by Galliher and Wilson (1975) [aircraft engines] and Mitchell et al (1980) [10,000 aircraft parts] accepted the hypothesis that failures of an item at the user level followed a homogeneous Poisson Process. Other studies by Johnson and McCoy (1978), Metzner (1981), and Proschan (1983) on aircraft parts rejected the same hypothesis although in each study an explanation is given as to why the hypothesis was rejected assuming it was likely to be true.

A priori, there are three possible objections to Assumption 1. The first is that while customers might arrive according to a NHPP, order sizes might be larger than one. If this were true, a NHCPP model of demand is more appropriate. Most of the results of this study can be extended to systems in which primary customer demand follows a NHCPP. However, customers usually demand only single units of an expensive item and it is the expensive items that usually account for the largest inventory investment (Orr [1977], Peterson and Silver [1979], p73-80). The second objection is that the variance to mean ratio (VMR) of the demand in any interval might be greater than one.

If this were due only to non-unit order sizes, a NHCPP model of demand would be appropriate. If the increased variability is due to other factors, such as uncertainty of the intensity function or a truly more variable underlying demand process, using a NHCPP to model demand is only an approximation whose effectiveness would need to be investigated. (This issue will be discussed in more detail in Chapter IX). The third objection arises from the fact that if there were a finite calling population of customers at the bases, the arrival rate of failed units at time t may depend upon the number of operating units at time t. The U. S. Army assumes for most weapon systems that a certain operating tempo (hours flown, miles driven, etc.) is maintained each month regardless of the number of operational weapon systems (Kaplan [1980]). Each operational weapon system may be used more than originally planned to compensate for the downed systems and therefore, the number of failures of an item will tend to be consistent with original projections. For systems with large calling populations and/or high performance targets in which the percentage of total customers that are down is expected to be small, the infinite population assumption should cause no harm. For very small customer populations, the assumptions of a state independent NHPP generating demand may be inappropriate. However, Zmurkewycz (1984) observed that even for small populations (in one case 10 customers with a medium performance target and in another case 2 customers with a high performance target) the expected number of customer backorders obtained from an infinite population two-echelon stationary model was within 1% of the backorders from a finite population Monte Carlo simulation.

Assumption 2. When a failed unit is brought to base$_i$ the decision, made by the $base_i$ diagnostic facility, either to repair the failed unit at the $base_i$ repair facility, to send it to the depot for repair, or to condemn it as irreparable is made independently of the decisions made on other units at other times and depends only upon the complexity of the required repair.

The $base_i$ diagnostic facility examines a failed unit and determines the repair action required to restore the unit to a serviceable condition. When the required repair can be accomplished at the $base_i$ repair facility, the failed unit is sent there. If the failed unit requires depot action, either because the repair is beyond $base_i$ capabilities or because the diagnostic facility cannot determine the extent of the required repair, the diagnostic facility sends the unit to the depot. Otherwise, the diagnostic facility condemns the unit as irreparable and removes it from the system. Condemnations result in the loss of a system asset that can be replaced by a procurement from the external supplier. The decision of the diagnostic facility depends only upon the estimated ability of the $base_i$ and depot repair facilities to effect the necessary repairs and is independent of all else including the decisions made on other units at other times, the number of serviceable spares on-hand at $base_i$ and the depot, and the number of units already in the $base_i$ and depot repair facilities.

<u>Assumption 3</u>: When a failed unit is sent from a base to the depot the decision, made by the depot diagnostic facility, either to repair the unit or to condemn it and remove it from the system, depends only upon the complexity of the required repair. This decision is made independently of the decisions made on other units and without regard to the number of units on-hand or in repair at the depot.

Assumption 3 imposes an independence condition on failed units sent to the depot similar to the one imposed by Assumption 2 on failed units brought to the bases. Both Assumption 2 and Assumption 3 allow the probabilities of a particular diagnosis to vary with time as long as each diagnosis is independent of other diagnoses and the number of units on-hand and in repair. This is particularly useful in wartime scenarios where the various combat situations and missions may, over different time intervals, result in different types of damage to the item and/or affect the ability of the various repair facilities to perform the necessary repairs.

Assumption 4: The times spent in the diagnostic facilities are mutually statistically independent. Repair times are mutually statistically independent.

There is no restriction on the distribution of repair times at a repair facility as long as the independence of individual repair times is maintained. In fact, both the diagnostic time and repair time distributions are allowed to change with the time of failure, and the repair time distribution may also depend upon the time at which repair was initiated after fault diagnosis and isolation. For example, a wartime scenario that calls for $base_i$ to be in an intense combat zone in $(s,t]$ may alter the repair time distribution according to whether the unit failed before or after s and according to the time at which repair was initiated after fault isolation was completed by the appropriate diagnostic facility. In summary, the diagnostic and repair times of a particular failed unit need not be independent but the times of different units are mutually independent.

Assumption 5: There is no batching of units before fault diagnosis begins. There is no batching of units before repair begins.

It is usually uneconomical to adopt a batch diagnostic and/or repair policy for expensive reparable items (Peterson et al [1959]) so for most items of interest this assumption is not restrictive.

Assumption 6: There are an infinite number of servers at the diagnostic and repair facilities.

Assumptions 5 and 6 ensure that diagnosis begins immediately upon receipt of a failed unit at a diagnostic facility and that likewise, repair begins immediately upon receipt of a failed unit at a repair facility. Together with Assumption 4 they preserve the statistical independence of the times different units spend in the diagnostic and repair facilities. If units were allowed to queue and wait for diagnosis and/or repair there would be a correlation between the times successive units spend in the diagnostic and repair facilities.

The effect of an ample service assumption on finite server systems has been studied directly by Gross (1982), who compared the steady state behavior of an $M/M/\infty$ queue and an $M/M/c$ queue with the same mean time in the repair facility (waiting time plus repair time), and indirectly by Slay (1980) and Kaplan (1980), both of whom compared the $M/G/\infty$ queue arising from Sherbrooke's (1968) METRIC model and the infinite server queue with Poisson arrivals and correlated service times arising from Simon's (1971) two-echelon model. They all found that for low utilization factors (Kleinrock [1975], p18) the ample service assumption induced little error in determining the equilibrium number of customers waiting. As expected, as the utilization factor increased, the error increased. Gross (1982) reports significant errors only as the utilization factor approached 1. We conclude from these studies that the ample service assumption should not be made lightly but that there are many situations for which this assumption is adequate.

In particular, based on intuition and the stationary studies above, we expect the ample service assumption to be satisfactory for items for which the number of units in the repair (diagnostic) facility at any time during the horizon is not large compared to the maximum number of available servers. Such items might have low failure rates, small repair (diagnostic) times and/or be part of multi-item systems where the servers and test equipment are not dedicated to any particular item. In the latter case, since servers can work on any of a number of different items, the chances of a backlog in repairing (diagnosing) a particular item are considerably reduced. For these reasons, the ample service assumption is routinely used in the stationary multi-echelon models employed by the Armed Forces (Sherbrooke [1968], Clark [1978], U. S. Army [1983]). We will discuss this issue again in Chapter IX.

<u>Assumption 7</u>: There is no lateral resupply among the bases.

By Assumption 2, the $base_i$ diagnostic facility never sends a failed unit to another base repair facility. Complementing this, Assumption 7 prohibits $base_i$ from ever seeking resupply from another base for serviceable units $base_i$ has issued to customers. $Base_i$ will seek resupply only from its own repair facility, the depot or, in the case of condemnations, the external supplier. Sherbrooke (1968) reports very little lateral transfer for Air Force Systems and the Army (1983) and Navy (Clark [1978]) do not consider it a major factor in determining stock levels. In "optimally" controlling a non-stationary multi-echelon inventory system, it might of course be desirable to transfer assets from a base with a lot of on-hand stock to another base that has little or no on-hand stock. The data required for making this decision "optimally" are similar to the data required for optimal real time control, which, as was pointed out in Chapter I, presents an almost insurmountable data management problem. In fact, Miller's Real Time METRIC (1968) does not allow for lateral resupply. Optimal policies or heuristics for handling lateral resupply are beyond the scope of this study.

Assumption 8: All locations follow a one-for-one ($[S(t)-1,S(t)]$) resupply policy at all times.

Every location seeks resupply by exchanging a failed unit for a serviceable unit from a resupplier on a one-for-one basis. If the $base_i$ diagnostic facility sends the unit for $base_i$ repair, resupply comes from the $base_i$ repair facility in the form of repairing the failed unit and returning it to the $base_i$ serviceable spare stock pool. If the unit is condemned, resupply comes from the external supplier. Otherwise, resupply is from the depot. In a sense we are assuming that at all times the economic order quantity for $base_i$ from each of its resuppliers is 1. Similarly, the depot seeks resupply on a one-for-one basis from its repair facility or the external supplier according to whether the unit is repaired at the depot or condemned. For most expensive or low demand items, a one-for-one resupply policy is usually optimal for stationary continuous review inventory systems (Hadley and Whitin [1963], p204) and during many dynamic scenarios of interest it is reasonable to assume that an $[S(t)-1,S(t)]$ policy is still the best ordering policy for these types of items. Furthermore, the various models used by the Armed Forces for peacetime spares determination assume a one-for-one ordering policy and there is no reason to believe this policy would change during a surge or wartime situation (Kruse and Cohen [1983], Hillestad [1982]).

Ultimately, we would like to make real-time resupply decisions rather than always following a one-for-one resupply policy. Even under an HCP, we might want to change the resupply policy at management intervention times to reflect what we predict will happen during the remainder of the horizon. For example, if we expect $base_i$ to be destroyed at time t, there seems to be very little reason for the depot to honor a resupply request from $base_i$ at t^-, thereby giving the depot another asset to use in resupplying the surviving bases. As we shall see, a change in the resupply policy corresponds to a change in the asset levels (Definition 2.2 in the next section) at one or more locations. Hence, we can view the system as continually using a one-for-one resupply policy except for management directed changes in S(t). This will be discussed in detail in Chapter VIII where we develop a model for determining the management intervention times and the decisions to be made at these times.

Assumption 9: Unfilled demand at a location is backordered and eventually satisfied on a first come-first served (FCFS) basis.

If there is no on-hand stock at the base at the time of a customer's demand, the demand is backordered and eventually satisfied on a FCFS basis. At the depot, resupply requests from the bases are handled similarly.

Assumption 10: The external supplier has an infinite supply of serviceable spares.

In summary then, the two echelon inventory system depicted in Figure 1 behaves as follows: primary customer demands occur at the bases. When a failed unit is brought to a base, the base issues the customer a serviceable spare if one is available or else backorders the demand (FCFS). The base diagnostic facility either condemns the failed unit or sends it to either the base repair facility or the depot. If the unit is condemned, an order is placed with the external supplier for a new unit to be added, upon arrival, to the base serviceable spare stock pool. If the unit is sent to the depot, the depot is obliged to send a serviceable spare to the base as soon as one is available with all the unfilled base resupply requests being backordered and satisfied on a FCFS basis. The decision by the base diagnostic facility depends only upon the complexity of the repair required and is independent of the decisions on other units at other times. Likewise, when the depot diagnostic facility receives a failed unit, it decides, independently of the decisions on other units at other times and depending only upon the complexity of the required repair, whether to condemn or repair the unit. A condemnation results in an order being placed on the external supplier while if the unit is repaired at the depot, upon completion the serviceable unit is put into the depot spare stock pool and is available to satisfy base resupply requests.

Pipelines and Asset Levels

Throughout this study, we number the bases from 1 through N and 0 will refer to the depot. At location$_i$, $i = 0,1,2,\ldots N$, at time $t \geq 0$ let

$H_i(t)$ = number of units on-hand;

$B_i(t)$ = number of backorders outstanding;

$D_i(t)$ = number of unserviceable units in the location$_i$ diagnostic facility;

$R_i(t)$ = number of unserviceable units in the location$_i$ repair facility;

$E_i(t)$ = number of serviceable units still en route at t to location$_i$ from the external supplier to replace condemnations at location$_i$ in $(-\infty,t]$.

For base$_i$, $i = 1,\ldots, N$, also define $DE_i(t)$ as the number of serviceable units en route from the depot to base$_i$ at time t plus the number of resupply requests placed by base$_i$ in $(-\infty,t]$ that are backordered at the depot at t (because no serviceable units are available at the depot spares pool for shipment). $DE_i(t)$ is the number of resupply requests placed on the depot by base$_i$ in $(-\infty,t]$ for which base$_i$ has not received a serviceable spare by t: either because the depot has not yet shipped one and has thus backordered the request or because the serviceable unit the depot shipped is still in transit to base$_i$ at t.

For base$_i$, let $X_i(t) = D_i(t) + R_i(t) + E_i(t) + DE_i(t)$. Because of the one-for-one resupply policy, $X_i(t)$, called the pipeline quantity or pipeline for short, represents the number of customer demands at base$_i$ in $(-\infty,t]$ for which the base$_i$ spares pool has not received a serviceable unit by t as resupply for the demanded serviceable unit that it either has issued or will issue to satisfy the customer request. Note that $X_i(\cdot)$ does not change when diagnosis on a unit is completed since $D_i(\cdot)$ decreases by one but because of the one-for-one resupply policy, one and only one of $E_i(\cdot)$, $R_i(\cdot)$ or $DE_i(\cdot)$ increases by 1. In fact, $X_i(\cdot)$ increases by 1 if and only if $D_i(\cdot)$ increases by 1. $X_i(\cdot)$ decreases by 1 if and only if one and only one of $R_i(\cdot)$, $E_i(\cdot)$ or $DE_i(\cdot)$ decreases by 1 representing the receipt of a serviceable unit at the base_i spares pool from either the base_i repair facility, the external supplier or the depot, respectively. $X_i(t)$ reflects the initial system condition since it includes units that were in the pipeline at 0 and are still in the pipeline at t.

At the depot, let $V_0(t)$ be the number of unserviceable units en route to the depot diagnostic facility from the bases. Similarly to the bases, define the pipeline at the depot as $X_0(t) = D_0(t) + R_0(t) + E_0(t) + V_0(t)$. The depot pipeline has a similar interpretation as the number of serviceable units due-in to the depot from the depot's resupply process. Note that $X_0(\cdot)$ changes only when $V_0(\cdot)$ increases or when $R_0(\cdot)$ or $E_0(\cdot)$ decreases.

When a location receives a demand, it immediately issues a serviceable unit if there is one on-hand or it backorders the request to be filled on a FCFS basis. The failed unit immediately enters the resupply process by being sent to the location's diagnostic facility. This implies that for all $t \geq 0$, and all $i = 0,1,\ldots,N$, $H_i(t)B_i(t) = 0$. Since resupply is on a one-for-one basis it also implies that if $X_i(t) = 0$ all serviceable units that location_i is authorized to have would be on-hand at t. This leads to Definition 2.2.

<u>Definition 2.2</u>: The asset level at location_i, $i = 0,1,\ldots N$, at time $t \geq 0$, $s_i(t) \equiv H_i(t) - B_i(t) + X_i(t)$, is the maximum number of serviceable units that can be on-hand at location_i at time t.

After base_i receives a demand at t, it either issues a serviceable unit from its spares pool or backorders the demand, thereby decreasing $H_i(t) - B_i(t)$ by precisely 1. The failed unit immediately enters the diagnostic facility, increasing $X_i(t)$ by 1 and thereby resulting in no change to $s_i(t)$. Furthermore, if $base_i$ receives a serviceable unit from the resupply process at t, $X_i(t)$ decreases by 1 but the unit just received either satisfies an existing backorder or is put into the $base_i$ spares pool resulting in an increase in $H_i(t) - B_i(t)$ of 1 and thereby again causing no change in $s_i(t)$. Similarly, a resupply request placed on the depot at t decreases $H_0(t) - B_0(t)$ by 1. Since concurrent with the request a failed unit is sent from a base, $V_0(t)$ and therefore $X_0(t)$ increases by 1 resulting in no change to $s_0(t)$. Furthermore, if $X_0(t)$ decreases by 1, $H_0(t) - B_0(t)$ increases by 1. Thus, at every location the asset level varies with time only through management directive and not because of demands at a location or the interactions of the resupply system. This is a direct consequence of the one-for-one resupply policy and is characteristic of all (S-1,S) inventory systems.

Of course, management decisions to change the asset level at a location could be based on an optimization model that reflects the demand history, anticipated future demand and the current system condition. (See Chapter VIII). If management does not interfere with the normal workings of the inventory system, then $s_i(t)$ is constant through time and in particular, $s_i(t) = s_i(0)$ for all t and $i = 0,1,2,..N$. If management has sent additional spares to $location_i$, either through new external procurement or redistribution of system assets, $s_i(t)$ reflects only those units that have arrived at $location_i$ by t and were therefore available to satisfy demands that occurred before t at $location_i$. Likewise, $s_i(t)$ reflects only those management cutbacks that have been implemented by t. Therefore, $s_i(t)$ differs from $s_i(0)$ only through the total of the assumed known management directives to either increase or decrease assets at location_i that have been implemented by time t.

From Definition 2.2 we have that

$$B_i(t) = [X_i(t) - s_i(t)]^+ \tag{2.2}$$

$$H_i(t) = [s_i(t) - X_i(t)]^+. \tag{2.3}$$

Since $s_i(t)$ is a known function of time, (2.2) and (2.3) show that the pipeline quantity $X_i(t)$ provides all the necessary information to determine the on-hand stock and backorder position at location_i at any time t. Chapters III and IV are devoted to describing the stochastic processes, $\{\{X_i(t), t \geq 0\}\ i = 0,1,\ldots N\}$.

Backorder and On-hand Moments

In this section we present results on the calculation of the moments of the distributions of on-hand inventory and outstanding backorders for one-for-one inventory systems with integer pipelines. The results are true for all locations and all times so we temporarily suppress the notation indicating location and time and define

B[s] = backorders given an asset level of s;

H[s] = on-hand inventory given an asset level of s;

X = pipeline quantity.

From (2.2) we have that

$$E(B[s]) = \sum_{j>s} (j-s)\Pr(X=j)$$

$$= \sum_{j>s} \Pr(X \geq j) \tag{2.4}$$

and

$$E(B^2[s]) = \sum_{j>s} (j-s)^2 \Pr(X=j)$$

$$= \sum_{j>s} j^2 \Pr(X=j) + s^2 \Pr(X \geq s+1) - 2s[E(X) - \sum_{j=1}^{s} j\Pr(X=j)].$$

For all distributions on the non-negative integers,

$$\sum_{j=1}^{k} \Pr(X \geq j) = \sum_{j=1}^{k} j \Pr(X=j) + k \Pr(X \geq k+1)$$

so that using this and (2.4) yields

$$E(B^2[s]) = E(X^2) - \sum_{j=1}^{s} j^2 \Pr(X=j) - s^2 \Pr(X \geq s+1) - 2sE(B[s]). \qquad (2.5)$$

Note that $E(B[0]) = E(X)$ and $E(B^2[0]) = E(X^2)$ as we expected from (2.2) since for $s = 0$, $B = X$.

Equations (2.4) and (2.5) allow the calculation of the mean and variance of the backorder distribution for a given asset level based on knowledge of the pipeline distribution. As we have already mentioned and shall see in Chapter V, the asset level is often the decision variable in an optimization problem. Therefore, many trial values may be examined in an algorithm (Kotkin [1978]) and a recursive calculation of the mean and variance of the backorder distribution is desirable.

Theorem 2.3: (a) $E(B[s]) = E(B[s-1]) - \Pr(X>s-1)$

(b) $\mathrm{Var}(B[s]) = \mathrm{Var}(B[s-1]) - \Pr(X \leq s-1)\{E(B[s]) + E(B[s-1])\}$

Proof: Let $I(z) = 1$ if $z > 0$ and let $I(z) = 0$ otherwise. Then,

$$B[s] = B[s-1] - I(B[s-1]) \qquad (2.6)$$

since increasing the asset level from $s-1$ to s decreases the number of outstanding backorders by precisely 1 whenever $B[s-1] \geq 1$. From (2.2) for $n = 1,2,\ldots$

$$E\{I^n(B[s-1])\} = \Pr(B[s-1] \geq 1) = \Pr(X>s-1) \qquad (2.7)$$

so that (a) follows immediately upon taking the expectation of both sides of (2.6) [as well as from (2.4)]. Furthermore, from (2.6)

$$\mathrm{Var}(B[s]) = \mathrm{Var}(B[s-1]) + \mathrm{Var}\{I(B[s-1])\} - 2\,\mathrm{Cov}\{B[s-1], I(B[s-1])\}$$

But

$$\begin{aligned}\mathrm{Var}\{I(B[s-1])\} &= E\{I^2(B[s-1])\} - E^2\{I(B[s-1])\} \\ &= \Pr(X>s-1)\Pr(X\le s-1)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}\{B[s-1], I(B[s-1])\} &= E\{B[s-1]I(B[s-1])\} - E(B[s-1])E\{I(B[s-1])\} \\ &= E(B[s-1])\Pr(X\le s-1)\end{aligned}$$

from (2.7) and the fact that $B[s-1]I(B[s-1]) = B[s-1] \ge 0$. Therefore,

$$\mathrm{Var}(B[s]) = \mathrm{Var}(B[s-1]) + \Pr(X\le s-1)\{\Pr(X>s-1) - 2E(B[s-1])\}$$

From (a) we have that $\Pr(X>s-1) = E(B[s-1]) - E(B[s])$ and substituting this above we immediately obtain (b). //

Note that as expected both the mean and variance of the backorder distribution are decreasing functions of the asset level and both go to zero as s goes to infinity.

Using (2.6) we can obtain a recursive formula for any moment of the backorder distribution.

Theorem 2.4: For $n = 1,2,\ldots$

$$E(B^n[s]) = E(B^n[s-1]) + (-1)^n\Pr(X>s-1) + \sum_{j=1}^{n-1}(-1)^{n-j}\binom{n}{j}E(B^j[s-1]).$$

Proof: From (2.6) we have

$$\begin{aligned}E(B^n[s]) &= E\{[B[s-1] - I(B[s-1])]^n\} \\ &= E\{\sum_0^n \binom{n}{j} B^j[s-1]I^{n-j}(B[s-1])(-1)^{n-j}\} \\ &= E(B^n[s-1]) + (-1)^n E\{I^n(B[s-1])\} \\ &\quad + \sum_{j=1}^{n-1}(-1)^{n-j}\binom{n}{j}E\{B^j[s-1]I^{n-j}(B[s-1])\}.\end{aligned}$$

For $n \geq j > 0$, $B^j[s-1]\ I^{n-j}(B[s-1]) = B^j[s-1] \geq 0$. Using this and (2.7) in the above expression establishes Theorem 2.4. //

Realizing that

$$H[s] = H[s-1] + I(s-X) \tag{2.8}$$

and using arguments similar to the ones used to prove Theorems 2.3 and 2.4 we can prove Theorem 2.5 which gives results on the moments of the distribution of on-hand inventory.

<u>Theorem 2.5</u>: (a) $E(H[s]) = E(H[s-1]) + Pr(X \leq s-1)$

(b) $Var(H[s]) = Var(H[s-1]) + Pr(X>s-1)\{E(H[s]) + E(H[s-1])\}$

(c) for $n = 1,2,\ldots$

$$E(H^n[s]) = E(H^n[s-1]) + Pr(X \leq s-1) + \sum_{j=1}^{n-1} \binom{n}{j} E(H^j[s-1]). \quad //$$

Note that the mean and variance of the distribution of on-hand inventory are increasing functions of s and they grow without limit as s goes to infinity.

The above results can be extended to arbitrary distributions for X so that the results can be used in many of the classical inventory models where X will represent demand during a lead time and analogs of (2.2) and (2.3) are valid (Kotkin [1983]).

The results in this section are used extensively in Chapter V where we formulate an optimization problem that requires the determination, for various depot and base asset levels, of the expected number of base backorders outstanding at each point in time during the horizon. We present these results here as background material so that the reader can keep them in mind as we derive the distributions of the depot and base pipelines in the next two chapters.

Next page is blank.

CHAPTER III

DEPOT CHARACTERISTICS

In this chapter we derive the time dependent probability distribution of the depot pipeline and we also study other important stochastic processes that arise at the depot. The results of this chapter are used extensively in Chapter IV to derive the distributions of the bases' pipelines. However, the material in this chapter is interesting and useful in its own right: it pertains to the analysis of non-stationary single echelon inventory systems.

Demand at the Depot

Recall that a demand at the depot for a serviceable unit occurs when some base, upon completion of its diagnostic procedures, seeks resupply for the unserviceable unit it has just sent to the depot. Define $P_i(s,t)$ as the probability that a unit that fails at time $s > 0$ and completes $base_i$ diagnosis at $t \geq s$ will be sent to the depot for further diagnosis and/or repair. Furthermore, let $G_i(s,t)$ be the probability that a unit that fails at $s > 0$ will complete $base_i$ diagnosis by $t \geq s$. For fixed $t > 0$, we classify a failure at time $0 < s \leq t$ as one of three mutually exclusive and exhaustive types:

Type I: Diagnosis on the unit was completed by t and a decision was made to send the unit to the depot. This occurred with probability

$$P_I(s) = \int_s^t P_i(s,y)\; dG_i(s,y).$$

Type II: Diagnosis on the unit was completed by t and a decision was made either to condemn the unit or to repair the unit at the base. This occurred with probability

$$P_{II}(s) = \int_s^t (1-P_i(s,y))dG_i(s,y).$$

Type III: Diagnosis on the unit was not completed by t. This occurred with probability $P_{III}(s) = 1 - G_i(s,t)$.

Note that for all $0 < s \leq t$, $P_I(s) + P_{II}(s) + P_{III}(s) = 1$. For a particular unit the decision made by the diagnostic facility may depend upon the time of failure and upon the time spent in the diagnostic facility. However, by assumption, the decisions made on different units, as well as the diagnostic times of different units, are mutually statistically independent. Therefore, the Splitting Property for NHPP, P10, guarantees that the number of resupply requests placed on the depot by base_i forms a NHPP with MVF

$$A_i(t) = \int_0^t \lambda_i(s)\; P_I(s)\; ds. \tag{3.1}$$

Since demands at each base form independent NHPP and the diagnostic times and decision making mechanisms at the bases are also independent, we have, as a consequence of the Superposition Property P1, that the demands for serviceable units at the depot form a NHPP with MVF $m_0(t) = \sum_{i=1}^{N} A_i(t)$.

Depot Asset Position

Because of the one-for-one resupply assumptions, the depot asset position changes over time only through management directive. At time $t \geq 0$, management may change the asset position by directing a change either in the depot's net inventory, $H_0(t) - B_0(t)$, or in the depot pipeline quantity, $X_0(t)$. (See Definition 2.2). Management may change the net inventory by delivering or removing serviceable units and "creating" or "cancelling" outstanding backorders. A change in $X_0(t)$ can be effected either by discarding a unit already in the pipeline or by adding a unit to the depot pipeline without a corresponding resupply request from a base.

Management may elect to increase the depot's net inventory by delivering serviceable units to the depot spares pool. These deliveries, as well as all other management directives that change the depot asset position, are assumed to be scheduled at the beginning of the horizon and are also assumed not to depend on the system condition at any time in $(0,t]$ or on the depot demand process in $(0,t]$.

Essentially, management decreases the depot's net inventory by "demanding" a serviceable unit from the depot spares pool. If there is stock on-hand, a serviceable unit is immediately issued to management. Otherwise, we assume the creation of a backorder due-out to management which, along with the other demands on the depot, will be satisfied in accordance with the FCFS discipline. Since the times of the management directed changes are determined at the beginning of the horizon, these "management demands" result in a non-continuous MVF for the NHPP describing depot demand. For ease of exposition, the subsequent analyses in

Chapter III assume a continuous MVF arising from base resupply requests. The methods of Cinlar (1975) for non-continuous MVF can be applied to easily extend the analyses.

Management may increase $X_0(t)$ by placing an unserviceable unit in the depot pipeline without a corresponding resupply request from a base. (Presumably, these unserviceable units are available as the result of previous management decisions not to diagnose/repair units that failed somewhere in the inventory system). This type of management directive does not affect the depot demand process. However, there is no longer a one-to-one correspondence between units in the pipeline and base resupply requests. The NHPP describing the number of units that enter the depot pipeline now has a non-continuous MVF to represent both the base resupply requests and the entries scheduled by management. Again, for ease of exposition, in the subsequent analyses we deal only with a continuous MVF arising from base resupply requests. The methods of Cinlar (1975) can again be applied to extend the analyses to include this type of management directive.

The management directives to cancel an existing backorder or to discard a unit already in the pipeline are more cumbersome to deal with. First, there is no assurance that these directives can be implemented at their scheduled times. Therefore, we would need rules to cover the possibility that there may be a delay in implementing these changes. Depending on the rules chosen, $s_0(t)$ may become a random variable. Secondly, the very fact of whether these directives can or cannot be implemented at t provides information about the pipeline and the demand process in (0,t]. For example, cancelling a backorder due-out to a particular base implies, among other things, that $X_0(t) > s_0(t)$ and also that there was at least one resupply request from that particular base backordered at t. The subsequent analyses in this chapter do not allow for these two types of asset position changes. The analyses can be extended to deal with these types of management directives by carefully and tediously conditioning on the relevant information obtained at their originally scheduled times for implementation and the effects on the asset position of a delay in their implementation. In a similar manner, the analyses could be extended to include management directives that depend on the system condition at times in (0,t] and/or on the depot demand process in (0,t].

Depot Pipeline

The depot pipeline at time $t \geq 0$, $X_0(t)$, consists of: units that were in the depot pipeline at time 0 and are still in the pipeline at t; units that were in some base diagnostic facility at time 0, were sent to the depot in (0,t] and are still in the depot pipeline at t; and units that failed at the bases in (0,t], were sent to the depot in (0,t] and are still in the depot pipeline at t. Because of our assumptions, these three components of $X_0(t)$ are independent so that $X_0(t)$ can be obtained from the convolution of three random variables. The first two random variables can be easily, albeit tediously, computed from knowledge of: the initial system condition; the distributions of the diagnostic, shipping and external procurement lead times; and the diagnostic decision making mechanisms at the bases and the depot. Since the stochastic description of the first two components is not a goal of this dissertation, we will, for ease of exposition, assume that $X_0(0) = 0$ and also that all base pipelines are empty at 0. Without loss of generality, we concentrate here on the stochastic description of the component of $X_0(t)$ arising from failures at the bases in $(0,t]$.

Fix some time $t > 0$ and consider a demand placed on the depot at $0 < s \leq t$. The demand is classified as one of the following mutually exclusive and exhaustive types:

Type 1: The failed unit sent to the depot at s is still en route to the depot at t.

Type 2: The failed unit arrived at the depot by t but has not completed diagnosis by t.

Type 3: The failed unit completed depot diagnosis by t. A decision was made to repair the unit at the depot but the repair was not completed by t.

Type 4: The failed unit completed depot diagnosis by t. A decision was made to condemn the unit thereby generating an order on the external supplier. A serviceable replacement from the external supplier has not arrived at the depot by t.

Type 5: The failed unit completed depot diagnosis by t. If the decision made was to repair, the unit completed depot repair by t. If the decision made was to condemn, a serviceable replacement from the external supplier arrived at the depot by t.

Since each demand is classified independently of other demands, P10 guarantees that the number of demands classified into each type form mutually statistically independent NHPP. In particular, define:

$v(s,t) = \Pr(\text{a shipping time begun at } s \text{ has ended by } t)$;

$c(s,t_1,t_2)$ = Pr(a unit that was shipped to the depot at s and which started diagnosis at t_1 has left the diagnostic facility by t_2);

$r(s,t_1,t_2,t_3)$ = Pr(a unit that was shipped to the depot at s, started diagnosis at t_1 and entered the repair facility at t_2 has left the repair facility by t_3);

$e(s,t_1,t_2,t_3)$ = Pr(a serviceable unit arrived by t_3 in response to an order placed on the external supplier at t_2 to replace a unit that was shipped to the depot at s and which started diagnosis at t_1);

$P_0(s,t_1,t_2)$ = Pr(a unit that was shipped to the depot at s, arrived at t_1 and completed diagnosis at t_2 was sent to the depot repair facility).

$V_0(t)$ is the number of Type 1 demands in $(0,t]$. As a consequence of P10, $V_0(t)$ has a Poisson distribution with mean

$$\int_0^t \lambda_0(s)\ [1-v(s,t)]\ ds \tag{3.2}$$

where $\lambda_0(s)$ is the intensity of the NHPP describing demands at the depot. $D_0(t)$ [Type 2 demands] has a Poisson distribution with mean

$$\int_0^t \lambda_0(s) \int_s^t d[v(s,y)][1-c(s,y,t)]\ ds. \tag{3.3}$$

$R_0(t)$ [Type 3 demands] has a Poisson distribution with mean

$$\int_0^t \lambda_0(s) \int_s^t d[v(s,y)] \int_y^t d[c(s,y,z)]P_0(s,y,z)[1-r(s,y,z,t)]\ ds. \tag{3.4}$$

Finally, $E_0(t)$ [Type 4 demands] has a Poisson distribution with mean

$$\int_0^t \lambda_0(s) \int_s^t d[v(s,y)] \int_y^t d[c(s,y,z)][1-P_0(s,y,z)][1-e(s,y,z,t)]\ ds. \quad (3.5)$$

Since for all $t > 0$, $V_0(t)$, $D_0(t)$, $R_0(t)$ and $E_0(t)$ are mutually statistically independent, $X_0(t)$ has a Poisson distribution with a mean obtained by summing (3.2) to (3.5).

Define $F_0(s,t)$ as the probability that a failed unit that was shipped to the depot at time $s > 0$ (and thereby entered the depot pipeline at s) is not in the depot pipeline at $t \geq s$. Then, for all $t > 0$, $X_0(t)$ has a Poisson distribution with mean

$$\int_0^t \lambda_0(s)\ [1-F_0(s,t)]\ ds \qquad (3.6)$$

where, using (3.2) to (3.5), we have that

$$\begin{aligned} F_0(s,t) = \int_s^t \int_{y_1}^t & d[v(s,y_1)]d[c(s,y_1,y_2)] \\ & \cdot \{P_0(s,y_1,y_2)\ r(s,y_1,y_2,t) \qquad (3.7) \\ & + [1-P_0(s,y_1,y_2)]\ e(s,y_1,y_2,t)\}. \end{aligned}$$

For ease of exposition and notational convenience, we have assumed that the depot shipping, diagnostic, repair and external procurement lead times, as well as the depot repair/condemn decision making mechanism, do not depend upon the original time of failure of the unit or upon the base at which it failed. As long as the independence among different failed units is maintained, it is a simple, straightforward matter to extend the above analysis to the case where the time and location of failure affect the relevant depot processes. We leave it to the reader to verify that (3.7) can be modified to handle these new dependencies so that $X_0(t)$ remains a Poisson random variable with mean given by (3.6).

Delay at The Depot

Let $W_0(t)$ be the delay before the depot sends a serviceable unit to the base that requested resupply from the depot at $t^+ > 0$. To calculate the distribution of $W_0(t)$ we extend the method used by Kruse (1980) to determine the distribution of customer wait in a stationary, single location (S-1,S) inventory system with arbitrary, independent resupply times and demands forming a homogeneous Poisson Process.

Clearly, $W_0(t) = 0$ if and only if $H_0(t) > 0$ since one of the $H_0(t)$ serviceable units on-hand at the depot spares pool will be immediately sent to the base that requested resupply. If $X_0(t) \geq s_0(t)$ then, from (2.3), $H_0(t) = 0$ and, from (2.2), there are $B_0(t) = X_0(t) - s_0(t) \geq 0$ backorders outstanding at the depot at t. Because of the FCFS policy, these $B_0(t)$ backordered base resupply requests must be satisfied before the resupply request at t^+ can be satisfied. Therefore, if $X_0(t) \geq s_0(t)$, the base that requested resupply will receive the $(X_0(t) - s_0(t) + 1)^{th}$ serviceable unit that becomes available for issue at the depot after t.

There are two ways that serviceable units become available for issue by the depot spares pool. First, a serviceable unit will enter the depot spares pool when it leaves the depot pipeline because the unit either just arrived from the external supplier or just completed repair at the depot. Secondly, management may send additional serviceable units to the depot in order to increase the depot asset level. Therefore, $W_0(t) > w \geq 0$ if and only if

$$X_0(t) - s_0(t) \geq M^+(t,t+w) + a(t,t+w) + f(0,t) + f(t,t+w) \qquad (3.8)$$

where

$M^+(t,t+w)$ = number of serviceable units sent by management that arrived at the depot in $(t,t+w]$.

$a(t,t+w)$ = 1 if the unit that entered the depot pipeline because of the demand at t^+ has left the pipeline by $t + w$. Otherwise, $a(t,t+w) = 0$.

$f(t_1,t_2)$ = number of units that entered the depot pipeline in $(t_1,t_2]$ and left the pipeline in $(t,t+w]$.

Since units that left the pipeline in $(t,t+w]$ must have originally entered the pipeline either in $(0,t]$ or in $(t,t+w]$, the right hand side of (3.8) is precisely the non-negative number of serviceable units that becomes available for issue at the depot spares pool in $(t,t+w]$ when we know a unit entered the pipeline at t^+. Assuming $F_0(t,t) = 0$, (3.8) assures us that $W(t) > 0$ if and only if $X_0(t) \geq s_0(t)$.

Rearranging (3.8) we have that $W_0(t) > w \geq 0$ if and only if

$$X_0(t) - f(0,t) \geq s_0(t) + M^+(t,t+w) + a(t,t+w) + f(t,t+w). \quad (3.9)$$

$X_0(t) - f(0,t)$ is the number of units that entered the depot pipeline in $(0,t]$ and are still in the depot pipeline at $t+w$. These units are precisely the Type 1, 2, 3 and 4 demands at $t+w$ that occurred in $(0,t]$. Using (3.6), we have that $X_0(t) - f(0,t)$ has a Poisson distribution with mean

$$\int_0^t \lambda_0(s) \ [1-F_0(s,t+w)] \ ds.$$

Furthermore, since $f(0,t)$ is the number of Type 5 demands at $t+w$ that occurred in $(0,t]$, $X_0(t) - f(0,t)$ is independent of $f(0,t)$. Using Assumptions 2 through 6 and the fact that a NHPP has independent increments, it is easy to show that $X_0(t) - f(0,t)$ is independent of $a(t,t+w)$ and $f(t,t+w)$.

$a(t,t+w)$ is a Bernoulli random variable with mean $F_0(t,t+w)$

and by Assumptions 2 through 6 is independent of $f(t,t+w)$. $f(t,t+w)$ is precisely the number of Type 5 demands at $t+w$ that occurred in $(t,t+w]$. Therefore, from (3.6), $f(t,t+w)$ has a Poisson distribution with mean

$$\int_t^{t+w} \lambda_0(s)\, F_0(s,t+w)\, ds$$

and $f(t,t+w)$ is independent of $X_0(t)$ and $X_0(t+w)$.

Since $s_0(t)$ and $M^+(t,t+w)$ are management parameters, all of the random variables in (3.9) have been identified and stochastically described. It is now a straightforward but tedious task to use (3.9) to calculate the probability distribution and expected value of $W_0(t)$. Figure 2 illustrates the behavior of the distribution of $W_0(t)$ for various values of $s_0(t)$. In constructing Figure 2 we assumed that:

$$X_0(0) = 0$$

$$M^+(t,t+w) = 0 \text{ for all } w \geq 0$$

$$t = 30$$

$$\lambda_0(s) = 4 \sin^2 \pi s \text{ for } s \geq 0$$

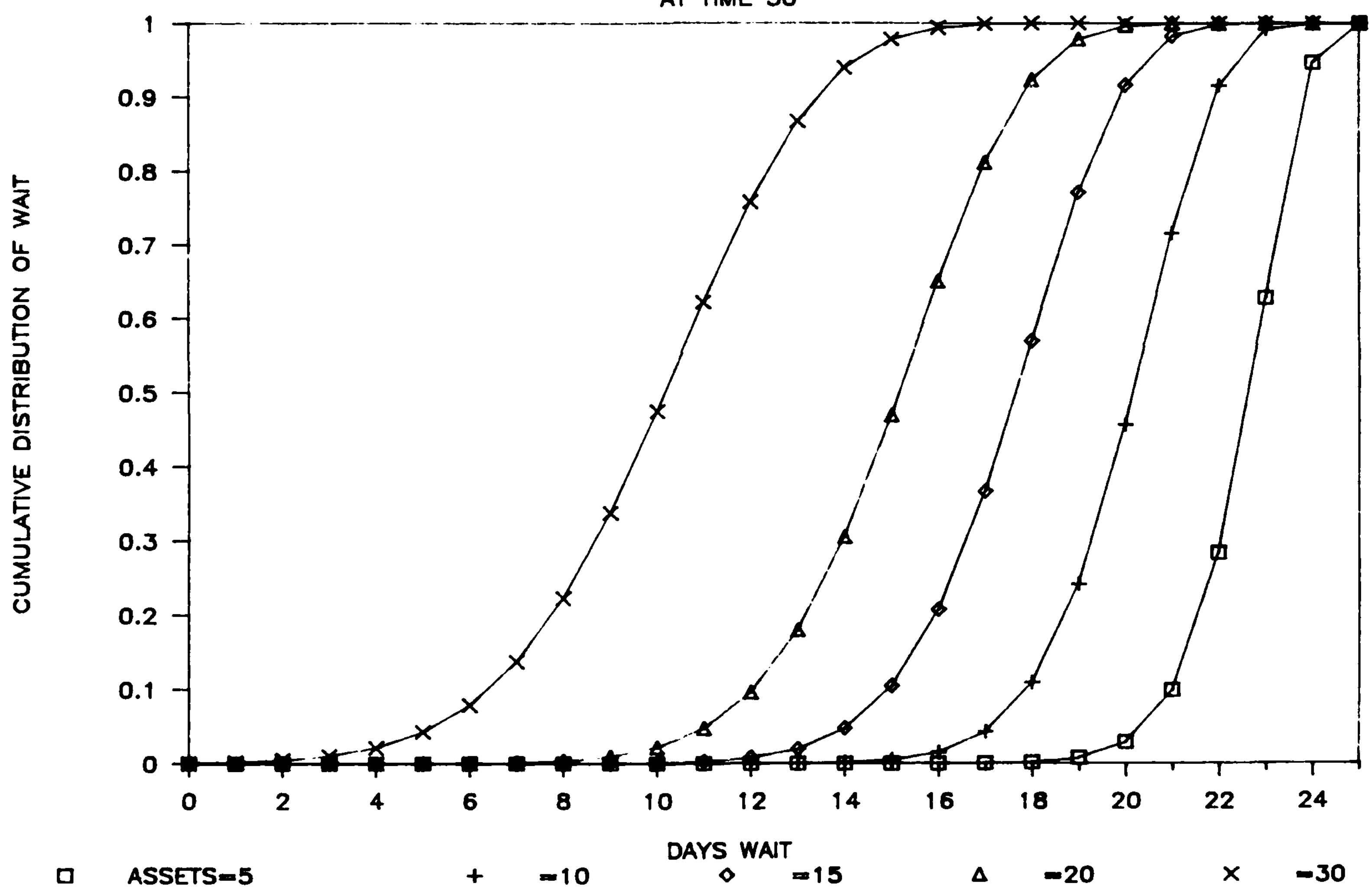
$$F_0(s,y) = 1 \text{ if } y \geq s + 25 \text{ and } 0 \text{ otherwise.}$$

Origination Time of Oldest Backorder

The origination time of a base's resupply request on the depot is defined as the time at which the base officially requested depot resupply by sending a failed unit to the depot. Let $Y_0(t)$ be the origination time of the first base resupply request that will be satisfied after time $t > 0$. $Y_0(t) > t$ if and only if $B_0(t) = 0$. If $B_0(t) = n > 0$ and $0 < t_1, t_2, \ldots, t_n \leq t$ are the origination times of the n resupply requests backordered at t, then $Y_0(t) = \min \{t_1, t_2, \ldots, t_n\}$. $t-Y_0(t)$

FIG. 2: CUMULATIVE DISTRIBUTION OF WAIT

AT TIME 30



is the length of time during which the bases requested but were unable to receive resupply from the depot by t. In this sense, $t-Y_0(t)$ is a measure of the ability of the depot to perform its supply mission. Intuitively, we expect that as $s_0(t)$ increases (decreases), $t-Y_0(t)$ decreases (increases) and more (less) base resupply requests can be satisfied by the depot by t. Therefore, $t-Y_0(t)$ is a measure of inventory effectiveness at the depot that is of interest to managers and strategic planners especially when t is set to the length of the scenario (HZ). In fact, it is quite plausible for management to consider alternate resupply sources for the bases whenever $HZ-E[Y_0(HZ)]$ (or possibly HZ - some percentile of $Y_0(HZ)$) is alarmingly high. In this section we derive the distribution of $Y_0(t)$ for any $t > 0$.

For the remainder of this dissertation, we shall for convenience use $\Pr(Y = y)$ to represent both the probability that a continuous random variable Y is in the interval $[y,y+dy]$ and the probability density function of Y. It will be clear from the context which meaning to assign.

By the Law of Total Probability we have for $y \le t$ that

$$\Pr(Y_0(t)=y) = \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} \Pr(Y_0(t)=y,\ B_0(t)=h,\ N_0(t)=k)$$

where $N_0(t)$ is the number of base resupply requests placed on the depot in $(0,t]$. Let $T(n)$ be the time of the n^{th} demand on the depot, $n = 1,2,\ldots$. The event $(Y_0(t)=y, B_0(t)=h, N_0(t)=k)$ occurs if and only if the event $(T(k-h+1)=y, B_0(t)=h, N_0(t)=k)$ occurs (see Figure 3).

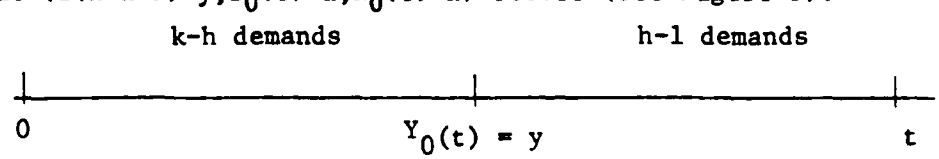


Figure 3: The event $(Y_0(t)=y,\ B_0(t)=h,\ N_0(t)=k)$

Therefore,

$$Pr(Y_0(t)=y) = \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} Pr(T(k-h+1) = y,\ B_0(t)=h,\ N_0(t)=k)$$

$$= \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} \{Pr(B_0(t)=h \mid T(k-h+1) = y,\ N_0(t)=k)$$

$$\cdot\ Pr(T(k-h+1)=y \mid N_0(t)=k)\ Pr(N_0(t)=k)\}. \quad (3.10)$$

From the analysis leading to (3.1) we know that $N_0(t)$ has a Poisson distribution with mean $m_0(t)$. From P5 we have that

$$Pr(T(k-h+1)=y \mid N_0(t)=k) =$$

$$\frac{k!}{(k-h)!(h-1)!} \left(\frac{m_0(y)}{m_0(t)}\right)^{k-h} \left(1 - \frac{m_0(y)}{m_0(t)}\right)^{h-1} \left(\frac{\lambda_0(y)}{m_0(t)}\right). \quad (3.11)$$

The event $(B_0(t)=h \mid T(k-h+1)=y, N_0(t)=k)$ occurs if and only if there are $s_0(t) + h$ units remaining in the depot pipeline at t when h-1 units entered the pipeline in (y,t], one unit entered at y, and k-h units entered in (0,y]. Let $L(n,t_1,t_2,t_3) \leq n$ be the number of units, out of the n units that entered the pipeline in $(t_1,t_2]$, that remain in the pipeline at t_3. The number of units that enter the pipeline forms a NHPP and the times different units spend in the pipeline are mutually statistically independent. Hence, it follows from P3 that $L(n,t_1,t_2,t_3)$ is a Binomial random variable with parameters n and $b(t_1,t_2,t_3)$ = probability that a unit that entered the pipeline in $(t_1,t_2]$ is still in the pipeline at t_3. Using P2 and (A.2) we have that

$$b(t_1,t_2,t_3) = \int_{t_1}^{t_2} [1-F_0(s,t_3)] \frac{\lambda_0(s)}{[m_0(t_2)-m_0(t_1)]} ds. \quad (3.12)$$

Recalling the definition of a(y,t) from the previous section, we then have that

$$\Pr(B_0(t) = h \mid T(k-h+1) = y,\ N_0(t) = k)$$

$$= \Pr[L(k-h,0,y,t) + [1-a(y,t)] + L(h-1,y,t,t) = s_0(t)+h]$$

$$\equiv \Pr(H(k-h,y,h-1,t) = s_0(t) + h). \qquad (3.13)$$

In (3.13) we have defined $H(k-h,y,h-1,t)$ as the number of units that are still in the depot pipeline at t given that k-h units entered in (0,y], one unit entered at y and h-1 units entered in (y,t]. Combining the above into (3.10) we have that

$$\Pr(Y_0(t)=y) = \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} \{e^{-m_0(t)} \frac{m_0(y)^{k-h}}{(k-h)!} \frac{[m_0(t)-m_0(y)]^{h-1}}{(h-1)!} \lambda_0(y)$$

$$\cdot \Pr[H(k-h,y,h-1,t) = s_0(t)+h]\}. \qquad (3.14)$$

In order to show that (3.14) defines a true probability distribution we need to use Theorem 3.1 below. Let $p[x;u] = e^{-u} u^x/x!$ be the probability a Poisson random variable with mean u is equal to x.

Theorem 3.1: For $h \geq 1$, $k \geq s_0(t) + h$ and $-(h-1) \leq n \leq k - h$,

$$\int_0^t \{p[k-h-n;m_0(y)]\ p[h-1+n;m_0(t)-m_0(y)]\lambda_0(y)$$

$$\Pr[H(k-h-n,y,h-1+n,t) = s_0(t) + h]\}\ dy$$

$$= \Pr(B_0(t)=h, N_0(t)=k).$$

Proof: For convenience, let $Q(s_0(t)+h)$ be the integral on the left side of the equation in Theorem 3.1. Since $H(k-h-n,y,h-1+n,t) \leq k$ we have that the generating function of $Q(j)$ is

$$\sum_{j=0}^{k} Q(j)z^j = \int_0^t p[k-h-n;m_0(t)]\ p[h-1+n;m_0(t)-m_0(y)]\lambda_0(y)H^*(z)\ dy$$

where $H^*(z)$ is the generating function of $H(k-h-n,y,h-1+n,t)$ which is the sum of three independent random variables: a Binomial with parameters k-h-n and b(0,y,t); a Bernoulli with mean $1-F_0(y,t)$; and a Binomial with parameters h-1+n and b(y,t,t). After some rearrangement, we get that

$$\sum_{j=0}^{k} Q(j)z^j = \int_0^t \{e^{-m_0(t)} [1-(1-z)(1-F_0(y,t))] \lambda_0(y)dy$$

$$\cdot \frac{[m_0(y)(1-(1-z)b(0,y,t))]^{k-h-n}}{(k-h-n)!} \frac{[(m_0(t)-m_0(y))(1-(1-z)b(y,t,t))]^{h-1+n}}{(h-1+n)!} \}.$$

Let I(k-h-n, h-1+n) be the integral on the right side of the equation above. Since

$$\frac{d}{dy} \{m_0(y)[1-b(0,y,t)(1-z)] = \lambda_0(y) \{1-(1-z)[1-F_0(y,t)]\}$$

$$\frac{d}{dy} \{[m_0(t)-m_0(y)][1-b(y,t,t)(1-z)]\} = - \frac{d}{dy} \{m_0(y)[1-b(0,y,t)(1-z)]\}$$

we can integrate I(k-h-n,h-1+n) by parts to get

$$I(k-h-n,h-1+n) = I(k-n+1,h-2+n)$$

$$+\{e^{-m_0(t)} \frac{(m_0(y)[1-b(0,y,t)(1-z)])^{k-h+1-n}}{(k-h+1-n)!}$$

$$\cdot \frac{([m_0(t)-m_0(y)][1-b(y,t,t)(1-z)])^{h-1+n}}{(h-1+n)!} \}.$$

Evaluating at y = 0 and y = t yields that

$$I(k-h-n,h-1+n) = I(k-h-n+1,h-2+n).$$

Continuing, we find that I(k-h-n,h-1+n) = I(k-1,0) which is equal to

$$\int_0^t e^{-m_0(t)} \lambda_0(y)[1-(1-z)(1-F_0(y,t))] \frac{[m_0(t)(1-b(0,y,t)(1-z))]^{k-1}}{(k-1)!} dy$$

$$= p[k;m_0(t)][1-b(0,t,t)(1-z)]^k. \qquad (3.15)$$

Thus, the generating function of Q(j) is equal to a constant times the generating function of a Binomial random variable with parameters k and b(0,t,t). Equating coefficients of z^j we have that

$$Q(s_0(t)+h) = p[k;m_0(t)] \binom{k}{s_0(t)+h} b(0,t,t)^{s_0(t)+h} [1-b(0,t,t)]^{k-s_0(t)-h}$$

$$= \Pr(N_0(t)=k) \Pr(L(k,0,t,t) = s_0(t)+h). \quad (3.16)$$

But $L(k,0,t,t) = s_0(t) + h$ if and only if $(X_0(t)=s_0(t)+h \mid N_0(t)=k)$ which occurs if and only if $(B_0(t)=h \mid N_0(t)=k)$ [See (2.2)]. Using this in (3.16) establishes Theorem 3.1. //

Theorem 3.1 effectively conditions the event $(B_0(t)=h, N_0(t)=k)$ on T(k-h+1-n). When n = k - h we are conditioning on T(1) so there must have been k-1 demands in (T(1),t] and $H(0,T(1),k-1,t) = s_0(t) + h$. When n = -(h-1) we are conditioning on T(k) so that $H(k-1,T(k),0,t) = s_0(t) + h$. Therefore, Theorem 3.1 verifies that $\Pr(B_0(t)=h, N_0(t)=k)$ can be calculated by conditioning on the time of any of the k demands that occurred in (0,t].

From (3.14), Theorem 3.1 (setting n = 0) and Fubini's Theorem (since the summand in (3.14) is the product of probability terms, all required interchanges are justified) we have that

$$\int_0^t \Pr(Y_0(t)=y) = \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} \Pr(B_0(t) = h,\ N_0(t) = k)$$

$$= 1 - \Pr(B_0(t) = 0). \quad (3.17)$$

By definition, $\Pr(Y_0(t) > t) = \Pr(B_0(t) = 0)$ and therefore the density of $Y_0(t)$ integrates to 1.

<u>Example 3.2</u>: Consider the case of a fixed known depot pipeline residence time R_0, so that for all s > 0, $F_0(s,t) = 1$ if $s + R_0 \geq t$ and $F_0(s,t) = 0$ otherwise.

For $y \leq t - R_0$ and all $n \geq 0$ we have that $L(n,0,y,t) = 0$ and that $1-a(y,t) = 0$. Therefore, for all $s_0(t)$, $h \geq 1$ and $k \geq s_0(t) + h$ we have that

$$Pr(H(k-h,y,h-1,t) = s_0(t)+h) = Pr(L(h-1,y,t,t) = s_0(t)+h) = 0$$

and from (3.14) we then have that $Pr(Y_0(t) \leq t - R_0) = 0$.

For $t - R_0 < y \leq t$ we have that $1 - a(y,t) = 1$ and $L(h-1,y,t,t) = h - 1$. Therefore, for $k - h > 0$

$$Pr(H(k-h,y,h-1,t) = s_0(t)+h) = Pr(L(k-h,0,y,t) = s_0(t))$$

$$= Pr(N_0(y)-N_0(t-R_0)=s_0(t)\,|\,N_0(y)=k-h)$$

$$= \binom{k-h}{s_0(t)} \left(\frac{m_0(y) - m_0(t-R_0)}{m_0(y)}\right)^{s_0(t)} \left(\frac{m_0(t-R_0)}{m_0(y)}\right)^{k-h-s_0(t)}$$

from P4 or (3.12). Inserting this in (3.14) yields, after some rearrangement and cancellation,

$$Pr(Y_0(t)=y) = \{e^{-m_0(t)} \lambda_0(y) \frac{[m_0(y) - m_0(t-R_0)]^{s_0(t)}}{s_0(t)!}$$

$$\cdot \sum_{h=1} \frac{[m_0(t) - m_0(y)]^{h-1}}{(h-1)!} \sum_{k=s_0(t)+h} \frac{m_0(t-R_0)^{k-h-s_0(t)}}{[k-h-s_0(t)]!} \}$$

$$= p[s_0(t);\ m_0(y) - m_0(t-R_0)]\lambda_0(y). \tag{3.18}$$

(3.18) is precisely the probability density that $T(N_0(t-R_0)+s_0(t)+1)$ occurred at y. (See P14). A little thought should reveal that if $B_0(t) > 0$ then the origination time of the first demand satisfied after t is precisely the time of the $s_0(t)$ + 1st demand after $t - R_0$.

For $t > 0$, let $Z_0(t) = T(N_0[Y_0(t)]-1)$ be the origination time of the last base resupply request that the depot satisfies by t. ($Z_0(t) = 0$ implies that the depot has not satisfied any base resupply requests in (0,t]). $t - Z_0(t)$ is an alternate measure of performance [related to $t-Y_0(t)$] that management can use to gauge the ability of the depot to perform its supply mission. Using arguments similar to those used to obtain (3.14), we can obtain the density of $Z_0(t)$, $t \geq 0$.

Summary

In this chapter we obtained the time dependent distributions of $X_0(t)$, $W_0(t)$ and $Y_0(t)$. All of these distributions are useful tools for evaluating inventory performance at the depot when viewing the depot as a single location inventory system. However, for the purposes of this dissertation, our interest in the depot as an inventory system unto itself is limited. Rather, we are primarily concerned with the impact of inventory decisions at the depot on inventory performance and customer satisfaction at the bases. In Chapter IV we shall use the arguments and results developed in this chapter to explicitly examine and define the supply interactions between the depot and bases.

Next page is blank.

CHAPTER IV

BASE CHARACTERISTICS

From (2.2) we see that increasing $s_0(t)$ reduces the number of base resupply requests backordered at the depot at t. As $B_0(t)$ decreases, we intuitively expect that inventory performance and customer satisfaction at the bases increase. Therefore, stock at the depot has a definite impact on inventory performance at the bases. In this chapter we precisely define this impact by deriving the time dependent probability distributions of the bases′ pipelines as functions of the depot′s asset level. We shall also study other important stochastic processes arising at the bases.

Demand at the Bases

Upon completion of the diagnostic procedures at $base_i$, a failed unit is either condemned as irreparable, sent to the $base_i$ repair facility, or sent to the depot for further diagnosis and action. The decision on each unit is made independently of the decisions on other units. In the first section of Chapter III we used these facts, along with the Splitting Property, P10, to show that the number of resupply requests placed on the depot by $base_i$ forms a NHPP with MVF $A_i(t)$ [given by (3.1)] and intensity $a_i(t)$, $i = 1,\ldots,N$. Similarly, it is straightforward to show that the number of units that enter the

base_i repair facility forms a NHPP with MVF

$$\int_0^t \lambda_i(s) \int_s^t dG_i(s,y) P_R^i(s,y)ds$$

where $P_R^i(s,y)$ is the probability that a unit that failed at time $s > 0$ and completed diagnosis at $y \geq s$ was be sent to the base_i repair facility. Furthermore, the number of base_i condemnations forms a NHPP with MVF

$$\int_0^t \lambda_i(s) \int_s^t dG_i(s,y)[1-P_i(s,y)-P_R^i(s,y)]ds.$$

As a further consequence of P10, these three NHPP are mutually statistically independent.

Base Asset Position

Because of the one-for-one resupply policy, the base_i asset position, analogously to the depot asset position, changes over time only through management directive. Many of the comments in Chapter III regarding the depot asset position apply to the base_i asset position as well. In particular, for ease of exposition, we assume that management decisions to change $s_i(t)$, the base_i asset position at $t \geq 0$, are made at the beginning of the horizon. We also assume these decisions do not depend upon the system condition at any times in $(0,t]$ or upon the base_i demand process in $(0,t]$. Therefore, the delivery of serviceable units to the base_i spares pool in order to increase base_i's net inventory yields no information on the system condition at any time in $(0,t]$ or on the base_i demand process. Without this assumption, our analyses would have to contain tedious arguments conditioning on any relevant information obtained at scheduled management intervention times.

Decisions to increase $X_i(t)$ by placing an unserviceable asset in the $base_i$ pipeline without a corresponding customer demand for a serviceable unit result in a non-coutinuous MVF for the NHPP describing the number of units that enter the components of the $base_i$ pipeline. Decisions to decrease $X_i(t)$ by a "management demand" result in a non-continuous MVF for the NHPP describing the demand at $base_i$. For ease of exposition, the subsequent analyses assume continuous MVF. The methods of Cinlar (1975) for non-continuous MVF can be applied to extend the analyses. Furthermore, for reasons similar to those given in the second section of Chapter III, the analyses in this chapter do not consider management directives to discard units already in the $base_i$ pipeline or to cancel a depot backorder due-out ("belonging") to $base_i$.

Base Pipeline

$X_i(t)$ consists of units that are still in the $base_i$ pipeline at t and either were in the $base_i$ pipeline at 0 or entered the $base_i$ pipeline in (0,t]. Because of our assumptions these two components of $X_i(t)$ are statistically independent so $X_i(t)$ can be found by the convolution of two random variables. The first random variable can be calculated directly from knowledge of the initial system condition and of the behavior of the PO. For ease of exposition we shall, as in Chapter III, assume that $X_j(0) = 0$, $j = 0,1,\ldots,N$. We concentrate here on the stochastic description of the component of $X_i(t)$, $i = 1,\ldots,N$, arising from failures at $base_i$ in (0,t].

Recall from Chapter II that

$$X_i(t) = D_i(t) + R_i(t) + E_i(t) + DE_i(t). \tag{4.1}$$

From the arguments in the first section of Chapter III, we know that $D_i(t)$ has a Poisson distribution with mean

$$\int_0^t \lambda_i(s)\ [1-G_i(s,t)]ds.$$

We also know from Chapter III that $D_i(t)$ is statistically independent of the number of units that completed diagnosis in $(0,t]$. Hence, $D_i(t)$ is independent of $R_i(t)$, $E_i(t)$ and $DE_i(t)$. In the first section of this chapter we showed that the output of the diagnostic facility in $(0,t]$ is split into three independent NHPP. $R_i(t)$ is a function of the number of units entering the repair facility. $E_i(t)$ is a function of the number of condemnations while $DE_i(t)$ is a function of the number of units sent to the depot. Therefore, $R_i(t)$, $E_i(t)$ and $DE_i(t)$ are mutually independent and hence, all the random variables on the right in (4.1) are mutually statistically independent.

Let $r_i(s,t_1,t_2)$ be the probability that a unit which failed at $s > 0$ and was sent to the base_i repair facility at $t_1 \geq s$ has left the base_i repair facility by $t_2 \geq t_1$. By Assumptions 2 through 6 the base_i repair facility acts as an $M(t)/G(t)/\infty$ queue with service time distribution $r_i(s,t_1,t_2)$. Therefore, $R_i(t)$, the number of busy "servers", has a Poisson distribution with mean

$$\int_0^t \lambda_i(s) \int_s^t dG_i(s,y)\ P_R^i(s,y)[1-r_i(s,y,t)]ds.$$

Let $e_i(s,t_1,t_2)$ be the probability that a serviceable unit has arrived from the external supplier by t_2 in response to a condemnation at $t_1 \leq t_2$ of a unit that failed at $0 < s \leq t_1$. Since the external supplier has infinite stock and order and ship times are independent, the external supplier functions as an $M(t)/G(t)/\infty$ queue with service time distribution

$e_i(s,t_1,t_2)$. Therefore, $E_i(t)$ has a Poisson distribution with mean

$$\int_0^t \lambda_i(s)\int_s^t dG_i(s,y)[1-P_i(s,y)-P_R^i(s,y)][1-e_i(s,y,t)]ds.$$

Until now, the analysis of the components of $X_i(t)$ has been similar to the analysis for the corresponding components of $X_0(t)$. $DE_i(t)$ is the unique component of the base_i pipeline through which the supply interactions of the depot and base_i manifest themselves.

Unless either the depot asset level is infinite or $F_0(t,t) = 1$ for all $t > 0$, we see from (3.8) that there is a positive probability that there will be a delay before the depot sends a serviceable unit in response to a base_i resupply request. Since the depot satisfies base resupply requests in a FCFS manner, $W_0(t)$ and $W_0(t+y)$, $y \geq 0$, are generally not statistically independent. In fact, $W_0(t+y) \geq W_0(t) - y$ so that for smaller y there will tend to be more correlation between $W_0(t)$ and $W_0(t+y)$ than for larger values of y. Therefore, if base_i submits two resupply requests to the depot, there may be some correlation in the times spent in the base_i pipeline by the failed units that accompanied these resupply requests. An analysis based upon P10 or, equivalently, upon treating the depot resupply process as an $M(t)/G(t)/\infty$ queue is therefore not appropriate.

$DE_i(t)$ has two components: $ER_i(t)$, the number of serviceable units en route at t from the depot to base_i; and $Q_i(t)$, the number of depot backorders outstanding at t that belong to base_i. Define

$T(n)$ = origination time of the n^{th} demand on the depot, $n = 1,2,\ldots;$

$I_i(n)$ = 1 if the n^{th} demand on the depot was from base_i and 0 otherwise, $n = 1,2,\ldots;$

$IS_i(n,t) = 1$ if $T(n) \leq t$, $I_i(n) = 1$ and the depot has sent a serviceable unit by t which is still en route to base_i at t. $IS_i(n,t) = 0$ otherwise.

Then,

$$ER_i(t) = \sum_{n=1}^{N_0(t)-B_0(t)} IS_i(n,t) \tag{4.2}$$

$$Q_i(t) = \sum_{n=N_0(t)-B_0(t)+1}^{N_0(t)} I_i(n) \tag{4.3}$$

and from P7,

$$E[I_i(n)\,|\,T(n)] = a_i[T(n)]/\lambda_0[T(n)] \equiv c_i[T(n)]. \tag{4.4}$$

From (4.3) we see that $Q_i(t)$ is a function of $B_0(t)$ which, from (2.2), is a function of $s_0(t)$. Theorem 4.2 guarantees that increasing $s_0(t)$ will stochastically reduce $Q_i(t)$.

Definition 4.1: (Lehmann [1959]). A random variable Q (or, equivalently, the distribution of Q) is stochastically decreasing (increasing) with respect to a parameter p if for $p_1 \leq p_2$ and all $q > 0$,

$$Pr(Q[p_1] \geq q) \geq (\leq)\ Pr(Q[p_2] \geq q).$$

Theorem 4.2: For all $t > 0$, $Q_i(t)$ is stochastically decreasing with respect to $s_0(t)$.

Proof: $0 \leq Q_i[t;s_0(t)] \leq Q_i[t;s_0(t)-1]$ since the extra depot asset may reduce the number of depot backorders belonging to base_i. The theorem now follows straightforwardly. //

From (4.2) we see that $ER_i(t)$ is also a function of $s_0(t)$. However, $ER_i(t)$ also depends upon the (time-dependent) distribution of the depot to base$_i$ order and ship times. One can construct cases where at t_1, $ER_i(t_1)$ is stochastically increasing with respect to $s_0(t_1)$; at $t_2 \neq t_1$, $ER_i(t_2)$ is stochastically decreasing with respect to $s_0(t_2)$; and at t_3, $t_1 \neq t_3 \neq t_2$, $ER_i(t_3)$ is neither stochastically increasing or decreasing with respect to $s_0(t_3)$. Corollary 4.3 is the only general statement we can make.

Corollary 4.3: Let $\{N_i^0(t), t \geq 0\}$ be the NHPP describing base$_i$ resupply requests placed on the depot. Then, for all $t > 0$, $[N_i^0(t) - Q_i(t)]$, the number of base$_i$ resupply requests the depot has satisfied by time t, is stochastically increasing with respect to $s_0(t)$.

Proof: Since $N_i^0(t)$ does not depend on $s_0(t)$, the corollary follows directly from Theorem 4.2. //

In general, $I_i(n)$ may provide information on $T(n)$ and since the $T(n)$, $n=1,2,\ldots$, are not independent, we see from (4.4) that the $I_i(n)$, $n = 1,2,\ldots$, are usually not independent. Similarly, for $t > 0$, the $IS_i(n,t)$, $n = 1,2,\ldots$, are generally not independent. Furthermore, $B_0(t)$ is generally not independent of the $I_i(n)$ or $IS_i(n,t)$. As we shall see, the above facts lead to complex expressions for the distributions of $Q_i(t)$ and $ER_i(t)$. Before examining the general case we shall therefore study two special cases which arise frequently in practice and for which tractable expressions can be obtained.

Proportionate Bases' Ownership of Depot Backorders

Base$_i$ is said to be a proportionate base with parameter c_i if there exists a constant $0 \leq c_i \leq 1$ such that for all $t \geq 0$, $c_i = c_i(t)$.

The "proportionate base case" arises frequently when scenarios call for the deployment of "identical" bases or for the demand intensity at each base to vary over time by the same multiplicative factor (U.S. Army [1983]). From (4.4), for n = 1,2,... we now have that

$$\Pr(I_i(n) = 1) = c_i \tag{4.5}$$

regardless of T(n). Since in the proportionate case base $I_i(n)$ does not provide information on T(n), $B_0(t)$ is independent of the $I_i(n)$, n=1,2,.... Furthermore, the $I_i(n)$, n=1,2,..., are themselves independent, identically distributed Bernoulli random variables with mean c_i. From (4.3) we have that $Q_i(t)$ is the sum of $B_0(t)$ of these i.i.d. random variables. Using (2.2), (3.6) and (4.5) we therefore have that

$$\begin{aligned}\Pr(Q_i(t)=q) &= \sum_{n=q}^{\infty} \Pr(Q_i(t)=q \mid B_0(t)=n)\Pr(B_0(t)=n) \\ &= \Pr(X_0(t) < s_0(t))\delta_0(q) \qquad (4.6) \\ &+ \sum_{n=q}^{\infty} \binom{n}{q}(c_i)^q(1-c_i)^{n-q}p[s_0(t)+n;\ E[X_0(t)]]\end{aligned}$$

where $\delta_0(q)$ is 1 or 0 according to whether q = 0 or q > 0. Clearly,

$$\sum_{q=0}^{\infty} \Pr(Q_i(t)=q) = \Pr(X_0(t)<s_0(t)) + \sum_{n=0}^{\infty} p[s_0(t)+n;\ E[X_0(t)]] = 1;$$

$$\begin{aligned}&E[Q_i(t)] = c_i E[B_0(t)]; \\ &VAR[Q_i(t)] = (1-c_i)E[Q_i(t)] + c_i^2\ VAR[B_0(t)]; \qquad (4.7) \\ &VMR[Q_i(t)] = (1-c_i) + c_i VMR[B_0(t)] .\end{aligned}$$

Let $B(n,c_i,q)$ be the probability that a Binomial (n,c_i) random variable is greater than or equal to q. For q > 0,

$$\Pr(Q_i(t) \geq q | s_0(t)=s_0+1) - \Pr(Q_i(t) \geq q | s_0(t)=s_0)$$

$$= \sum_{n=q}^{\infty} B(n,c_i,q)\{p[s_0+1+n;\ E[X_0(t)]] - p[s_0+n;\ E[X_0(t)]]\}$$

$$= \sum_{n=q}^{a_0} B(n,c_i,q)\ \{p[s_0+1+n;\ E[X_0(t)]] - p[s_0+n;\ E[X_0(t)]]\}$$

$$+ \sum_{n=a_0+1} B(n,c_i,q)\{p[s_0+1+n;\ E[X_0(t)]] - p[s_0+n;\ E[X_0(t)]]\}$$

where $a_0 = \max[q,\ INT(E[X_0(t)]) - s_0 - 1]$ and $INT(X)$ = largest integer less than or equal to X. Since $B(n,c_i,q)$ is increasing in n and the probability mass function of a Poisson random variable with mean u is unimodal with a peak at INT(u) (and at u-1 if u is an integer), the above is

$$\leq B(a_0,c_i,q)\{\Pr(X_0(t) \geq s_0+1+q) - \Pr(X_0(t) \geq s_0+q)\} < 0.$$

Hence, in the proportionate base case, the distribution of $Q_i(t)$ given by (4.6) satisfies Theorem 4.2.

Ownership of Depot Backorders for Fixed Depot Pipeline Times

We now replace the assumption that $base_i$ is a proportionate base with the assumption of a fixed, known depot pipeline residence time, R_0. Therefore, for all $s > 0$, $F_0(s,t) = 1$ if $s + R_0 \geq t$ and $F_0(s,t) = 0$ otherwise. This is a common assumption found in many of the inventory models used by the Army (U.S. Army [1983]).

Consider first the case where $s_0(t) = 0$. An unserviceable unit that enters the depot pipeline at $y > 0$ will leave the pipeline at precisely $y + R_0$. Since there is no depot stock and units leave the pipeline in the same order that they entered, the n^{th} demand on the depot will be satisfied at precisely $T(n) + R_0$, $n=1,2,\ldots$. Furthermore, the n^{th} demand will be backordered at the depot for the entire interval $[T(n),T(n)+R_0)$, $n=1,2,\ldots$. Therefore, a base_i resupply request will be backordered at the depot at t if and only if the failed unit was sent to the depot in $(t-R_0,t]$. Hence,

$$\Pr(Q_i(t) = q) = p[q;\ A_i(t) - A_i(t-R_0)] \tag{4.8}$$

that is, $Q_i(t)$ has a Poisson distribution with mean $A_i(t) - A_i(t-R_0)$. In this case, $\mathrm{VMR}(Q_i(t)) = 1$.

Now assume $s_0(t) > 0$. $X_0(t)$ consists of precisely the units sent to the depot in $(t-R_0,t]$. Hence, $T^*(s_0(t)) = T(N_0(t-R_0) + s_0(t)) > t$ if and only if $X_0(t) < s_0(t)$ which, from (2.2), implies that $B_0(t) = 0$ and hence that $Q_i(t) = 0$. All $B_0(t) > 0$ backorders outstanding at the depot at t must have resulted from resupply requests in $(T^*(s_0(t)),t]$. Therefore, for $q \geq 0$,

$$\Pr(Q_i(t)=q) = \Pr(T^*(s_0(t))>t)\ \delta_0(q)$$

$$+ \int_{t-R_0}^{t} \Pr(Q_i(t)=q|T^*(s_0(t))=y)\ \Pr(T^*(s_0(t))=y)dy. \tag{4.9}$$

Clearly, $\Pr[T^*(s_0(t)) > t] = \Pr[X_0(t) < s_0(t)]$ and from <u>P14</u> we have for $y \geq t - R_0$ that

$$\Pr(T^*(s_0(t)) = y) = p[s_0(t) - 1;\ m_0(y) - m_0(t-R_0)]\lambda_0(y). \tag{4.10}$$

Since $T^*(s_0(t))$ is a Markov time with respect to the depot demand process we have, for $t - R_0 \leq y \leq t$,

$$\Pr(Q_i(t) = q \geq 0|T^*(s_0(t))=y)$$

$$= \Pr(q\ \text{base}_i\ \text{resupply requests in}\ (y,t])$$

$$= p[q;\ A_i(t) - A_i(y)]. \tag{4.11}$$

Using (4.10) and (4.11) in (4.9) we obtain $Pr(Q_i(t) = q \geq 0)$.

Note that

$$\sum_{q=0}^{\infty} Pr(Q_i(t)=q) = Pr(T^*(s_0(t))>t)$$

$$+ \int_{t-R_0}^{t} \{p[s_0(t)-1;\ m_0(y)-m_0(t-R_0)]\lambda_0(y)$$

$$\cdot \sum_{q=0}^{\infty} p[q;\ A_i(t) - A_i(y)]\}dy$$

where the interchange of integral and sum is justified since the integrand in (4.9) is positive. Using (4.10) the above becomes

$$= Pr(T^*(s_0(t))>t) + \int_{t-R_0}^{t} Pr(T^*(s_0(t)) = y)dy$$

$$= Pr(T^*(s_0(t)) \geq t-R_0) = 1.$$

From (4.9), (4.10) and (4.11)

$$E[Q_i(t)] = \int_{t-R_0}^{t} \{p[s_0(t)-1;\ m_0(y) - m_0(t-R_0)]\lambda_0(y)$$

$$\cdot \sum_{q=1}^{\infty} qp[q;\ A_i(t) - A_i(y)]\}dy \tag{4.12}$$

$$= \int_{t-R_0}^{t} [A_i(t) - A_i(y)]\ p[s_0(t)-1;\ m_0(y)-m_0(t-R_0)]\lambda_0(y)dy$$

$$E[Q_i^2(t)] = \int_{t-R_0}^{t} \{([A_i(t)-A_i(y)]^2 + [A_i(t)-A_i(y)])\lambda_0(y)$$

$$p[s_0(t)-1;\ m_0(y)-m_0(t-R_0)]\}dy. \tag{4.13}$$

By the Cauchy-Schwarz inequality, $VAR[Q_i(t)] \geq E[Q_i(t)]$ and hence, $VMR[Q_i(t)] \geq 1$.

Let

$$a_0(s_0) = \min[t,\ \inf\{y \geq t - R_0;\ m_0(y) - m_0(t-R_0) \geq s_0\}].$$

For $q > 0$,

$$\Pr(Q_i(t)\geq q|s_0(t)=s_0+1) - \Pr(Q_i(t)\geq q|s_0(t)=s_0)$$

$$= \int_{t-R_0}^{a_0(s_0)} \{(p[s_0;m_0(y)-m_0(t-R_0)] - p[s_0-1;\ m_0(y)-m_0(t-R_0)])$$

$$\cdot \sum_{j=q}^{\infty} p[j;\ A_i(t) - A_i(y)]\lambda_0(y)\}dy$$

$$+ \int_{a_0(s_0)}^{t} \{(p[s_0;m_0(y)-m_0(t-R_0)] - p[s_0-1;\ m_0(y)-m_0(t-R_0)])$$

$$\cdot \sum_{j=q}^{\infty} p[j;\ A_i(t)-A_i(y)]\lambda_0(y)\}dy.$$

$p[s_0;\ m_0(y) - m_0(t-R_0)] - p[s_0-1;\ m_0(y) - m_0(t-R_0)]$ is ≤ 0 or ≥ 0 according to whether $y \leq a_0(s_0)$ or $y > a_0(s_0)$. Since a Poisson random variable is stochastically increasing with respect to its mean, the above is

$$\leq \sum_{j=q}^{\infty} p[j;\ A_i(t)-A_i(a_0(s_0))]$$

$$\cdot \int_{t-R_0}^{t} p[s_0;\ m_0(y)-m_0(t-R_0)] - p[s_0-1;\ m_0(y)-m_0(t-R_0)]\lambda_0(y)\ dy$$

$$= \sum_{j=q}^{\infty} p[j;\ A_i(t)-A_i(a_0(s_0))]\{\Pr(T^*(s_0+1)\leq t) - \Pr(T^*(s_0)\leq t) \leq 0$$

where the last inequality follows from the definition of $T^*(s_0)$. In a similar manner it is straightforward to use (4.8) and (4.9) to show that $\Pr(Q_i(t) \geq q > 0)$ decreases when $s_0(t)$ increases from 0 to 1. Therefore, the given distribution of $Q_i(t)$ satisfies Theorem 4.2.

<u>Example 4.4</u>: Let $a_i(t) = a_i$ for all $t \geq 0$ so that the number of base_i resupply requests placed on the depot forms a homogeneous Poisson Process. Furthermore, let $\lambda_0(t) = \lambda_0$ for all $t \geq 0$. Then, $c_i(t) = c_i$ for all $t \geq 0$ and $X_0(t)$ has a Poisson distribution with mean $\lambda_0 R_0$. For $s_0(t) \geq 1$, we have from (4.9) that

$$\Pr(Q_i(t)=0) = \Pr(X_0(t)<s_0(t)) + \int_{t-R_0}^{t} p[0;(t-y)a_i]\, p[s_0(t)-1;\ (y-(t-R_0))\lambda_0]\lambda_0 dy$$

$$= \Pr(X_0(t)<s_0(t)) + \sum_{k=s_0(t)}^{\infty} p[k;\lambda_0 R_0](1-c_i)^{k-s_0(t)} \tag{4.14}$$

after some rearrangement and after using Property 16 (for the integral of Poisson probabilities) in Appendix 3 of Hadley and Whitin (1963). Note the correspondence to (4.6) since the stationary case is a special case of the proportionate base case. From (4.9),

$$\Pr(Q_i(t)=q\ 0)$$

$$= \int_{t-R_0}^{t} p[q;(t-y)a_i]\, p[s_0(t)-1;(y-(t-R_0))\lambda_0]\lambda_0 dy. \tag{4.15}$$

Letting $z = t - y$, rearranging and using the power series expansion for the exponential terms, (4.15) becomes

$$= \frac{(\lambda_0)^{s_0(t)}}{q!} \left(\frac{c_i}{1-c_i}\right)^q (e^{-\lambda_0 R_0}) \int_0^{R_0} \sum_{k=0}^{\infty} \frac{[(\lambda_0-a_i)z]^{k+q}}{k!} \frac{(R_0-z)^{s_0(t)-1}}{(s_0(t)-1)!} dz.$$

For $0 \le z \le R_0$ the summand above is non-negative and by Fubini's Theorem we can interchange the order of integration and summation. After integrating by parts we have that

$$\Pr(Q_i(t)=q) = \sum_{k=q}^{\infty} p[s_0(t)+k;\lambda_0 R_0]\binom{k}{q}(c_i)^q(1-c_i)^{k-q}. \tag{4.16}$$

Again, note the correspondence with (4.6) for the proportionate base case.

(4.14) and (4.16) agree with the results of Simon (1971) for the stationary case. These formulae involve infinite sums so in actual computations there must be some truncation. Using (4.15) we can, however, develop an equivalent of (4.16) that contains only finite sums. Note that after some rearrangement, (4.14) can be expressed as a finite sum as

$$\Pr(X_0(t)<s_0(t)) + e^{-a_i R_0} \left(\frac{1}{1-c_i}\right)^{s_0(t)} [1-P[s_0(t)-1;(\lambda_0-a_i)R_0]] \qquad (4.17)$$

where $P[x;u]$ is the cumulative distribution function of a Poisson random variable with mean u. Letting $z = y - (t-R_0)$ in (4.15) and rearranging yields

$$\Pr(Q_i(t)=q) = e^{-a_i R_0} \frac{(a_i)^q}{q!} \left(\frac{1}{1-c_i}\right)^{s_0(t)-1} (\lambda_0)$$

$$\cdot \int_0^{R_0} (R_0-z)^q \, p[s_0(t)-1;\ (\lambda_0-a_i)z]dz.$$

Using Property 20 (expressing the integral above as a finite sum) in Appendix 3 of Hadley and Whitin (1963) the above becomes

$$= e^{-a_i R_0} \left(\frac{1}{1-c_i}\right)^{s_0(t)} \frac{(a_i)^q}{q!}$$

$$\cdot \sum_{k=0}^{q} (-1)^k \binom{q}{k} \frac{(R_0)^{q-k}(s_0(t)-1+k)!}{(s_0(t)-1)!\ (\lambda_0-a_i)^k} [1-P[k+s_0(t)-1;(\lambda_0-a_i)R_0]]. \qquad (4.18)$$

(4.18) looks more formidable than (4.16) yet (4.18) can be computed more quickly than (4.16) [Kotkin (1982)]. However, one must exercise caution in using (4.18) because of problems involving numerical stability and accuracy that arise from operating on numbers that differ considerably in magnitude. Kotkin (1982) examined the computational issues involved in using either (4.17) and (4.18) or (4.14) and (4.16) to determine the limiting probability distribution of $Q_i(t)$ for stationary systems. He also used the Vandermonde Convolution (Riordan [1971]) to directly show the equivalence of the two sets of formulae.

<u>Example 4.5</u>: Let base_i be a proportionate base with parameter c_i. Then, for all $z \geq 0$, $A_i(z) = c_i m_0(z)$. After some rearrangement in

(4.9) we have that

$$\Pr(Q_i(t)=q) = \Pr(X_0(t)<s_0(t))\ \delta_0(q) + K_1(s_0(t),q)K_2(s_0(t)-1,q) \quad (4.20)$$

where for j=0,1,2,... and q=0,1,2...,

$$K_1(j,q) = \exp[-c_i(m_0(t)-m_0(t-R_0))]\ c_i^q/(1-c_i)^{j+q}$$

$$K_2(j,q) = \int_{t-R_0}^{t} \{p[j;(1-c_i)(m_0(y)-m_0(t-R_0))](1-c_i)\lambda_0(y)$$

$$\cdot\ [(1-c_i)(m_0(t)-m_0(y))]^q/q!\ dy.$$

Note from P14 that for all j, $K_2(j,0)$ is simply the probability that in a NHPP with MVF $(1-c_i)m_0(z)$, $z \geq 0$, the (j+1)st event after $t-R_0$ occurred no later than t. Therefore, for j = 0,1,2,..

$$K_2(j,0) = 1 - P[j;(1-c_i)(m_0(t)-m_0(t-R_0))]. \quad (4.21)$$

It is now straightforward to show that for q = 0, (4.20) reduces to (4.6).

By induction on q we will show that for all j and q

$$K_2(j,q) = \sum_{n=j+1+q}^{\infty} p[n;(1-c_i)(m_0(t) - m_0(t-R_0))]\binom{n-j-1}{q}. \quad (4.22)$$

We have already seen that (4.22) holds for all j when q = 0. Assume (4.22) is true for all j when q = m - 1. From P14 we note that

$$p[j;(1-c_i)(m_0(y)-m_0(t-R_0))]\ (1-c_i)\ \lambda_0(y) \quad (4.23)$$

is the probability density that in a NHPP with MVF $(1-c_i)m_0(z)$, $z \geq 0$, the (j+1)st event after $t-R_0$ occurred at y. The antiderivative of (4.23) is thus simply the probability that the (j+1)st event after $t-R_0$ occurred no later than y. Therefore, integrating $K_2(j,m)$ by parts yields

$$K_2(j,m) = \sum_{w=j+1}^{\infty} K_2(w,m-1)$$

since the interchange of integral and sum is clearly justified. Using the induction hypothesis and (4.22) yields

$$K_2(j,m) = \sum_{w=j+1}^{\infty} \sum_{n=w+m}^{\infty} p[n;(1-c_i)(m_0(t)-m_0(t-R_0))]\binom{n-w-1}{m-1}$$

and the result now follows by interchanging the order of summation and using a basic combinatorial identity.

Using (4.22) in (4.20) establishes that for $q > 0$, (4.20) reduces to (4.6) in the proportionate base case.

Ownership of Depot Backorders in the General Case

In this section we remove the assumptions of the previous two sections and calculate the distribution of $Q_i(t)$, $t > 0$, for the general case. The analysis herein is complex and tedious so to help understand what we must do, it may be beneficial to first understand what we cannot do in order to find the distribution of $Q_i(t)$, $t > 0$.

One approach we might try would be, as in the proportionate base case, to apportion to the bases the $B_0(t)$ backorders outstanding at the depot at t based on the ratios $A_i(t)/m_0(t)$, $i=1,2,\ldots,N$. However, consider the two base case at $t = 2$ where:

$m_0(1) = 1$; $m_0(2) = 2$; $F_0(s,2) = 0$, $s > 0$;

$A_1(1) = 1$; $A_2(1) = 0$; $A_1(2) = 1$; $A_2(2) = 1$.

If $s_0(2) = 10$ and $B_0(2) = 1$, there is a better than even chance that the backorder belongs to base_2 even though $A_i(2)/m_0(2) = .5$, $i = 1,2$.

A second approach might be to claim, as we did in the previous section, that $Q_i(t) = N_i^0(t) - N_i^0(Z_0(t))$ where $Z_0(t) = T(N_0(Y_0(t)) - 1)$ is the origination time of the last base resupply request satisfied at the depot by t. In the case of a fixed depot pipeline residence time, R_0, $Z_0(t) = \min[T(N_0(t)),T^*(s_0(t))]$ is a Markov time with respect to the depot demand process. In general, however, this is not true as $Z_0(t)$ depends upon demands at the depot in $(Z_0(t),t]$ (See Chapter III. The distribution of $Z_0(t)$ can be obtained by arguments similar to those used to obtain the distribution of $Y_0(t)$. Using those arguments one clearly sees that $Y_0(t)$ and $Z_0(t)$ are not, in general, Markov times). Specifically, $Z_0(t)$ generally provides information on whether the units sent to the depot in $(Z_0(t),t]$ are still in the pipeline at t and thereby also provides information on the time and the number of demands in $(Z_0(t),t]$.

Certainly, the $B_0(t)$ depot backorders outstanding at t were from demands in $(Z_0(t),t]$. If we knew $Z_0(t)$ and if for each of the $B_0(t)$ demands in $(Z_0(t),t]$ we knew whether the failed unit that accompanied the demand was or was not in the depot pipeline at t, we could use P8 to find the probability that the demand was from base_i. This is the approach we will take in this section. Before proceeding with the details we need some additional definitions and results.

Fix some time $t > 0$ and let $NP_i(y_1,y_2)$ be the number of units sent to the depot by base_i in $(y_1,y_2]$ that are still in the depot pipeline at t. From arguments similar to the ones used in Chapter III we can show that $\{NP_i(0,y),\ y \geq 0\}$ is a NHPP with MVF

$$mp_i(0,y) = \int_0^y a_i(s)[1-F_0(s,t)]ds$$

and that $\{NP_0(0,y) = \sum NP_i(0,y),\ y \geq 0\}$ is a NHPP with MVF

$$mp_0(0,y) = \int_0^y \lambda_0(s)[1-F_0(s,t)]ds.$$

Note that $X_0(t) = NP_0(0,t)$. Similarly, $\{NE_i(0,y)=N_i^0(y)-NP_i(0,y),\ y \geq 0\}$, the stochastic process counting the number of units sent to the depot by base_i that have left the depot pipeline by t, is a NHPP with MVF

$$me_i(0,y) = \int_0^y a_i(s)F_0(s,t)ds.$$

Finally, $\{NE_0(0,y) = \sum NE_i(0,y),\ y \geq 0\}$ is a NHPP with MVF

$$me_0(0,y) = \int_0^y \lambda_0(s)F_0(s,t)ds.$$

By the Law of Total Probability, when $s_0(t) > 0$,

$$\Pr(Q_i(t)=q\geq 0) = \Pr[X_0(t)\leq s_0(t)]\ \delta_0(q)$$

$$+ \sum_{h=\max(1,q)} \ \sum_{k=s_0(t)+h} \ \sum_{h_p=\max[0,h-(k-[s_0(t)+h])]}^{h} \int_0^t \{1$$

$$\cdot \Pr(Q_i(t)=q,T(k-h)=y,B_0(t)=h,N_0(t)=k,NP_0(y,t)=h_p)\}dy. \quad (4.24)$$

When $s_0(t) = 0$ we must also account for the atom arising from the fact that all demands at the depot in $(0,t]$ may be backordered at t. Therefore, when $s_0(t) = 0$,

$$\Pr(Q_i(t)=q\geq 0) = \Pr[X_0(t)=0]\ \delta_0(q)$$

$$+ \sum_{h=\max(1,q)}^{\infty} \ \sum_{h_p=0}^{h} \Pr(Q_i(t)=q,T(0)=0,B_0(t)=h,N_0(t)=h,NP_0(0,t)=h_p)$$

$$+ \sum_{h=\max(1,q)}^{\infty} \ \sum_{k=h+1}^{\infty} \ \sum_{h_p=\max[0,h-(k-h)]}^{h} \int_0^t \{1$$

$$\cdot \Pr(Q_i(t)=q,T(k-h)=y,B_0(t)=h,N_0(t)=k,NP_0(y,t)=h_p)\}dy. \quad (4.25)$$

The event (EV_1)

$$\{T(k-h) = y,\ B_0(t) = h,\ N_0(t) = k,\ NP_0(y,t) = h_p\}$$

occurs if and only if the event (EV_2)

$$\{T(k-h) = y,\ NP_0(0,y) = s_0(t) + h - h_p,\ NP_0(y,t) = h_p,$$
$$NE_0(0,y) = k - (s_0(t)+h) - (h-h_p),\ NE_0(y,t) = h - h_p\}$$

occurs. Given EV_2, the demands backordered at t are precisely the h demands that were placed on the depot in $(y,t]$. Therefore, $Q_i(t) = q$

if and only if q of these h demands were from base$_i$. Hence,

$$\Pr(Q_i(t)=q|EV_2) = \Pr(NP_i(y,t) + NE_i(y,t)=q|EV_2)$$

$$= \Pr(NP_i(y,t) + NE_i(y,t)=q|NP_0(y,t) = h_p,\ NE_0(y,t)=h-h_p) \quad (4.26)$$

where the last equality follows from the fact that NHPP have independent increments.

For any $y \geq 0$, P10 guarantees that $NP_i(y,t)$ is independent of $NE_i(y,t)$ and $NE_0(y,t)$. Furthermore, from P8, $[NP_i(y,t)|NP_0(y,t)]$ has a Binomial distribution with parameters $NP_0(y,t)$ and

$$cp_i(t) = mp_i(y,t)/mp_0(y,t).$$

$NE_i(y,t)$ is independent of $NP_0(y,t)$ and, from P8, $[NE_i(y,t)|NE_0(y,t)]$ has a Binomial distribution with parameters $NE_0(y,t)$ and

$$ce_i(t) = me_i(y,t)/me_0(y,t).$$

(4.26) can now be obtained from the convolution of two independent Binomial random variables.

All that remains to be done in order to evaluate (4.24) and (4.25) is to find $\Pr[EV_1]$($=\Pr[EV_2]$). Recalling the notation introduced in the last two sections of Chapter III we have that:

(1) For $k > h$ (i.e. $s_0(t)>0$):

$$\Pr(B_0(t) = h|T(k-h)=y, N_0(t)=k, NP_0(y,t)=h_p)$$

= Pr{out of the k-h-1 demands in (0,y) and the demand at y, $s_0(t)+h-h_p$ are still in the depot pipeline at t}

$$= \Pr(1-a(y,t)+L(k-h-1,0,y,t)=s_0(t)+h-h_p). \quad (4.27a)$$

Note that for $0 \leq h_p < h - [k-(s_0(t)+h)]$, (4.27a) is zero.

For $k = h$ (i.e. $s_0(t)=0$): $T(0) = 0$ by definition. Also,

$$\Pr(B_0(t)=h|T(0)=0, N_0(t)=h, NP_0(0,t)=h_p)$$

equals 1 if $h = h_p$ and 0 otherwise. (4.27b)

(2) For $y > 0$, $Pr(NP_0(y,t)=h_p|T(k-h)=y,N_0(t)=k)$

$$= Pr(NP_0(y,t)=h_p|N_0(t)-N_0(y)=h)$$

$$= Pr(L(h,y,t,t)=h_p) \tag{4.28}$$

$$= \binom{h}{h_p}\left(\frac{mp_0(y,t)}{m_0(t)-m_0(y)}\right)^{h_p}\left(\frac{me_0(y,t)}{m_0(t)-m_0(y)}\right)^{h-h_p}.$$

(3) For $k > h$, using P8,

$$Pr(T(k-h)=y,N_0(t)=k)$$

$$= \lambda_0(y)p[k-h-1;m_0(y)]\ p[h;m_0(t)-m_0(y)]. \tag{4.29a}$$

For $k = h$, $Pr(T(0)=0,N_0(t)=h) = p[h;m_0(t)]$. (4.29b)

For $k > h$, $Pr(EV_1)$ is given by the product of (4.27a), (4.28) and (4.29a). For $k = h$, $Pr(EV_1)$ is given by the product of (4.27b), (4.28) and (4.29b).

Therefore, when $s_0(t) > 0$,

$$Pr(Q_i(t)=q) = Pr[X_0(t)\le s_0(t)]\ \delta_0(q)$$

$$+ \sum_{h=\max(1,q)}^{\infty}\ \sum_{k=s_0(t)+h}^{\infty}\ \sum_{h_p=\max[0,h-(k-[s_0(t)+h])]}^{h} \int_0^t \{1$$

$$\cdot\ Pr[NP_i(y,t)+NE_i(y,t) = q|NP_0(y,t)=h_p,NE_0(y,t)=h-h_p]$$

$$\cdot\ Pr[1-a(y,t)+L(k-h-1,0,y,t)=s_0(t)+h-h_p]$$

$$\cdot\ Pr[L(h,y,t,t)=h_p]$$

$$\cdot\ \lambda_0(y)\ p[k-h-1;m_0(y)]\ p[h;m_0(t)-m_0(y)]\}dy. \tag{4.30}$$

When $s_0(t) = 0$, we note that

$$\sum_{h=\max(1,q)}^{\infty} \{Pr[NP_i(0,t)=q|NP_0(0,t)=h]$$

$$\cdot\ Pr[L(h,0,t,t)=h]\ p[h;m_0(t)]\}$$

$$= e^{-m_0(t)}\ \frac{[mp_i(0,t)]^q}{q!}\ [e^{mp_0(0,t)-mp_i(0,t)} - \delta_0(q)]$$

$$= p[0;me_0(0,t)]\ p[q;mp_i(0,t)] - \delta_0(q)e^{-m_0(t)}. \tag{4.31}$$

since $m_0(t) = me_0(0,t) + mp_0(0,t)$. Therefore, when $s_0(t) = 0$,

$$\Pr(Q_i(t)=q) = \Pr[X_0(t)=0]\ \delta_0(q)$$
$$+ p[0;\ me_0(0,t)]\ p[q;\ mp_i(0,t)] - \delta_0(q)e^{-m_0(t)}$$
$$+ \sum_{h=\max(1,q)}^{\infty} \sum_{k=h+1}^{\infty} \sum_{h_p=\max[0,h-(k-h)]}^{h} \int_0^t \{1$$
$$\cdot \Pr[NP_i(y,t)+NE_i(y,t) = q \mid NP_0(y,t)=h_p, NE_0(y,t)=h-h_p]$$
$$\cdot \Pr[1-a(y,t)+L(k-h-1,0,y,t)=h-h_p]$$
$$\cdot \Pr[L(h,y,t,t)=h_p]$$
$$\cdot \lambda_0(y)\ p[k-h-1;m_0(y)]\ p[h;m_0(t)-m_0(y)]\}dy. \tag{4.32}$$

First we verify that (4.30) defines a proper probability mass function. Since the integrand in (4.30) is the product of probabilities, all required interchanges are justified by Fubini's Theorem. When summing (4.30) from $q = 0$ to $q = \infty$ we note that after interchanging the order of some summations and integration we have inside the integral

$$\sum_{q=0}^{h} \Pr[NP_i(y,t)+NE_i(y,t)=q \mid NP_0(y,t)=h_p, NE_0(y,t)=h-h_p]=1,$$

$$\sum_{h_p=\max[0,h-(k-[s_0(t)+h])]}^{h} \{\Pr[L(h,y,t,t)=h_p]$$
$$\cdot \Pr[1-a(y,t)+L(k-h-1,0,y,t)=s_0(t)+h-h_p]\}$$
$$= \Pr[H(k-h-1,y,h,t)=s_0(t)+h].$$

Therefore, when $s_0(t) > 0$,

$$\sum_{q=0}^{\infty} \Pr(Q_i(t)=q) = \Pr[X_0(t)\le s_0(t)]$$
$$+ \sum_{h=1} \sum_{k=s_0(t)+h} \int_0^t \{\lambda_0(y)\ p[k-h-1;m_0(y)]$$
$$p[h;m_0(t)-m_0(y)]\ \Pr[H(k-h-1,y,h,t)=s_0(t)+h]\}dy$$
$$= \Pr[X_0(t)\le s_0(t)] + \sum_{h=1}^{\infty} \sum_{k=s_0(t)+h}^{\infty} \Pr(B_0(t)=h, N_0(t)=k)=1$$

where the next to the last equality follows from Theorem 3.1. Since $(NP_i(0,t) | NP_0(0,t)=h)$ is a Binomial $[h, cp_i(0,t)]$ random variable and

$$\Pr[L(h,0,t,t)=h]\ p[h;m_0(t)] = \Pr[B_0(t)=h, N_0(t)=h]$$

we can, in a manner similar to the above, show that (4.32) also defines a proper probability mass function.

Let

$$SM[Z;j] = \sum_{h=1}^{\infty} \sum_{k=j+h}^{\infty} \sum_{h_p=\max[0,h-(k-[s_0(t)+h])]}^{h} \int_0^t Z\Pr(EV_1)dy$$

$$\begin{aligned} m_1(h,h_p) &= E[NP_i(y,t) + NE_i(y,t) | NP_0(y,t) = h_p,\ NE_0(y,t) = h - h_p] \\ &= h_p[cp_i(y,t)] + (h-h_p)\ ce_i(y,t) \\ m_2(h,h_p) &= E[\{NP_i(y,t) + NE_i(y,t)\}^2 | NP_0(y,t) = h_p,\ NE_0(y,t) = h - h_p] \\ &= h_p[cp_i(y,t)][1-cp_i(y,t)] + [m_1(h,h_p)]^2 \\ &\qquad + (h-h_p)ce_i(y,t)[1-ce_i(y,t)] \\ &= m_1(h,h_p) + [m_1(h,h_p)]^2 - h_p[cp_i(y,t)]^2 - (h-h_p)[ce_i(y,t)]^2. \end{aligned}$$

Then from (4.30) and (4.32) we have

$$\begin{aligned} E[Q_i(t);s_0(t)>0] &= SM[m_1(h,h_p);s_0(t)]; \\ E[Q_i^2(t);s_0(t)>0] &= SM[m_2(h,h_p);s_0(t)]; \\ E[Q_i(t);s_0(t)=0] &= p[0;me_0(0,t)]\ mp_i(0,t) + SM[m_1(h,h_p);1]; \\ E[Q_i^2(t);s_0(t)=0] &= p[0;me_0(0,t)]\ \{[mp_i(0,t)]^2 + mp_i(0,t)\} \\ &\qquad + SM[m_2(h,h_p);1]. \end{aligned} \tag{4.33}$$

$VMR(Q_i(t);s_0(t)>0) \geq 1$ if and only if

$$\begin{aligned} SM[\{m_1(h,h_p)\}^2;s_0(t)] &\geq \{SM[m_1(h,h_p);s_0(t)]\}^2 \\ &\quad + SM[h_p[cp_i(y,t)]^2 + (h-h_p)[ce_i(y,t)]^2;s_0(t)] \end{aligned} \tag{4.34}$$

and $VMR(Q_i(t);s_0(t)=0) \geq 1$ if and only if

$$\begin{aligned} &SM[\{m_1(h,h_p)\}^2;1] + p[0;me_0(0,t)]\{[mp_i(0,t)]^2 + mp_i(0,t)\} \\ &\geq \{SM[m_1(h,h_p);1]\}^2 + SM[h_p[cp_i(y,t)]^2 + (h-h_p)[ce_i(y,t)]^2;1] \\ &\quad + \{p[0;me_0(0,t)]mp_i(0,t)\}^2 + SM[2p[0;me_0(0,t)]mp_i(0,t)m_1(h,h_p);1]. \end{aligned} \tag{4.35}$$

We have found (4.30) and (4.32) to be analytically intractable. In fact, we have been unable to analytically verify either that (4.30) is stochastically decreasing with respect to $s_0(t)$ or that (4.34) and (4.35) hold. However, empirical evidence from calculations on the weapon systems described in Appendix B indicates not only that (4.30) and (4.32) are stochastically decreasing with respect to $s_0(t)$ but also that $VMR[Q_i(t)] \geq 1$. We shall return to this point in Chapter VI where we investigate approximations that reduce the computational burden involved in calculating $Q_i(t)$ and $X_i(t)$.

<u>Example 4.6</u>: Let $base_i$ be a proportionate base with parameter c_i. Then, $[NP_i(y,t) | NP_0(y,t)]$ has a Binomial $[NP_0(y,t),c_i]$ distribution and is independent of $[NE_i(y,t) | NE_0(y,t)]$ which has a Binomial $[NE_0(y,t),c_i]$ distribution. Therefore,

$$[NP_i(y,t) + NE_i(y,t) | NP_0(y,t) = h_p,\ NE_0(y,t) = h - h_p]$$

has a Binomial $[h,c_i]$ distribution regardless of y. Let $bi(q,h,c_i)$ be the probability that a Binomial $[h,c_i]$ random variable equals q. Removing this probability from inside the integral in (4.30) and recalling the definition of $H(k-h-1,y,h,t)$ we have that

$$\begin{aligned} Pr(Q_i(t)=q) &= Pr[X_0(t) \leq s_0(t)]\ \delta_0(q) \\ &+ \sum_{h=\max(1,q)}^{\infty} \ \sum_{k=s_0(t)+h}^{\infty} bi(q,h,c_i) \\ &\cdot \int_0^t \{\lambda_0(y)\ p[k-h-1;m_0(y)]\ p[h;m_0(t)-m_0(y)] \\ &\qquad \cdot Pr[H(k-h-1,y,h,t) = s_0(t)+h]\}dy \\ &= Pr[X_0(t) \leq s_0(t)]\ \delta_0(q) + \sum_{h=\max(1,q)}^{\infty} bi(q,h,c_i)\ Pr(B_0(t)=h) \end{aligned}$$

from Theorem 3.1. The above is easily seen to be equivalent to (4.6). Similarly, (4.32) reduces to (4.6) when $base_i$ is a proportionate base.

<u>Example 4.7</u>: Let $F_0(s,t) = 1$ if $t - s \geq R_0 \geq 0$ and $F_0(s,t) = 0$ otherwise. For all $y \leq t - R_0$ and $k > h$, $1 - a(y,t) + L(k-h-1,0,y,t) = 0$ so there is no contribution to $\Pr(Q_i(t)=q)$ in (4.30). For all $t \geq y > t - R_0$, $L(h,y,t,t) = h$ so there is no contribution to $\Pr(Q_i(t)=q)$ in (4.30) except when $h_p = h$. Furthermore,

$$\Pr(NP_i(y,t)+NE_i(y,t)=q|NP_0(y,t)=h,NE_0(y,t)=0)$$
$$= \Pr(NP_i(y,t)=q|NP_0(y,t)=h)$$
$$= bi(q,h,[a_i(t)-a_i(y)]/[m_0(t)-m_0(y)]).$$

Since $a(y,t) = 0$ for $t \geq y > t - R_0$ we have

$$\Pr[1-a(y,t)+L(k-h-1,0,y,t)=s_0(t)]=\Pr[L(k-h-1,0,y,t)=s_0(t)-1]$$
$$= bi[s_0(t)-1,k-h-1,[m_0(y)-m_0(t-R_0)]/m_0(y)]. \tag{4.36}$$

Multiplying (4.36) by $p[k-h-1;m_0(y)]$ and summing this product over the range of k yields $p[s_0(t)-1;m_0(y)-m_0(t-R_0)]$. Combining all of the above in (4.30) we have for $s_0(t) > 0$,

$$\Pr(Q_i(t)=q) = \Pr[X_0(t) \leq s_0(t)]\, \delta_0(q)$$
$$+ \sum_{h=\max(1,q)}^{\infty} \int_{t-R_0}^{t} \{\lambda_0(y)\; p[s_0(t)-1;m_0(y)-m_0(t-R_0)]$$
$$\cdot\, bi(q,h,[a_i(t)-a_i(y)]/[m_0(t)-m_0(y)])\; p[h;m_0(t)-m_0(y)]\}dy.$$

Summing over the range of h yields (4.9).

Note that

$$me_0(0,t) = m_0(t-R_0)$$
$$mp_i(0,t) = a_i(t) - a_i(t-R_0) = mp_i(t-R_0,t)$$

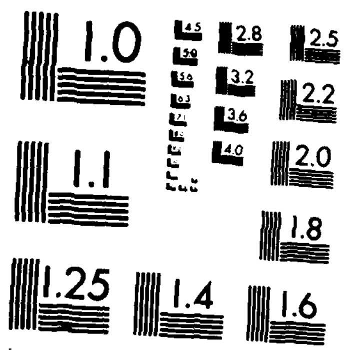
so that (4.31) becomes

$$p[q;a_i(t)-a_i(t-R_0)]\; p[0;m_0(t-R_0)] - \delta_0(q)e^{-m_0(t)}.$$

For $t \geq y > t - R_0$, $L(h,y,t,t) = h$ so we need only consider the case where $h = h_p$. Since $a(y,t) = 0$, $1 - a(y,t) + L(k-h-1,0,y,t) > 0 = h - h_p$ so there is no contribution to $\Pr(Q_i(t)=q)$ in (4.32) when $t \geq y > t - R_0$. For

AD-A173 731 OPERATING POLICIES FOR NON- STATIONARY TWO-ECHELON INVENTORY SYSTEMS FOR REPARABLE ITEMS(U) ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY ABERDEEN PROVING GROU 2/2

UNCLASSIFIED M H KOTKIN MAY 86 AMSAA-SP-39 F/G 15/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

$y \le t - R_0$, $1 - a(y,t) + L(k-h-1,0,y,t) = 0$ so we again need only consider the case where $h = h_p$. Noting that

$$\Pr(NP_i(y,t)=q \mid NP_0(y,t)=h)$$
$$= bi(q,h,[a_i(t)-a_i(t-R_0)]/[m_0(t)-m_0(t-R_0)])$$
$$\Pr(L(h,0,y,t)=h) = \{[m_0(t)-m_0(t-R_0)]/[m_0(t)-m_0(y)]\}^h$$

the last term in (4.32) becomes, after summing over the range of h,

$$\int_0^t \{\lambda_0(y)[mp_i(t-R_0,t)]^q e^{-[m_0(t)-m_0(y)]}$$
$$[e^{m_0(t)-m_0(t-R_0)-mp_i(t-R_0,t)} - \delta_0(q)]\}dy$$
$$= p[q;a_i(t)-a_i(t-R_0)]\ \{1-p[0;m_0(t-R_0)]\}$$
$$+ \delta_0(q)\ \{e^{-m_0(t)} - p[0;m_0(t)-m_0(t-R_0)]\}.$$

Since $p[0;m_0(t)-m_0(t-R_0)] = \Pr[X_0(t)=0]$, (4.32) reduces, as expected, to $p[q;a_i(t)-a_i(t-R_0)]$.

Number of Units Due-In From the Depot

Let $OST_i(y)$ be the order and ship time for a unit sent from the depot to $base_i$ at $y \ge 0$. For $t \ge 0$ and n=1,2,..., $IS_i(n,t) = 1$ if and only if:

a. $I_i(n) = 1$

b. $T(n) + W_0[T(n)] \le t$ (4.37)

c. $T(n) + W_0[T(n)] + OST_i\{T(n) + W_0[T(n)]\} > t$.

Since

$$E[IS_i(n,t) \mid T(n) = y \le t] =$$
$$\int_0^{t-y} c_i(y)\Pr[W_0(y)=z]\ \Pr[OST_i(y+z) > t-(y+z)]\ dz$$

we have that

$$E[IS_i(n,t)] = \int_0^t \int_0^{t-y} \{p[n-1;m_0(t)]\lambda_0(y)c_i(y)$$

$$\cdot\ Pr[W_0(y) = z]\ Pr[OST_i(y+z) > t-(y+z)]\}dzdy.$$

Of course, $Var[IS_i(n,t)] = E[IS_i(n,t)]\{1-E[IS_i(n,t)]\}$.

Using (4.2) we have that

$$E[ER_i(t)] = \int_0^t \int_0^{t-y} \{a_i(y)\ Pr[W_0(y) = z]$$

$$\cdot\ Pr[OST_i(y+z) > t-(y+z)]\}dzdy. \tag{4.38}$$

Consider a stationary system where, for all $y \geq 0$, $a_i(y) = a_i$, $i=1,2,...N$, $F_0(s,t)$ depends only on t-s and $OST_i(y) = OST_i$ does not depend on y. Using (3.8) we can show that as t goes to infinity, $W_0(t)$ has a limiting distribution. Then we can use (4.38) to show $E[ER_i(t)] = a_i E[OST_i]$ which can also be obtained, in a stationary system, from Little's Formula.

Unfortunately, other than (4.38) it is cumbersome to obtain any general results for $ER_i(t)$. To verify the conditions in (4.37) we must determine $T(n)$, $W_0[T(n)]$ and $OST_i\{T(n)+W_0[T(n)]\}$ for n=1,2,.... Even if the order and ship times are independent, the $T(n)$, n=1,2,..., are not independent and there may also be some correlation in the times different resupply requests wait for satisfaction at the depot. Therefore, the $IS_i(n,t)$, n=1,2,..., are usually not mutually independent. In order to determine the joint distribution of the $IS_i(n,t)$ [and thereby the distribution of $ER_i(t)$] we need to determine the joint distribution of the origination and waiting times of the $N_0(t)$ demands on the depot. To find even the variance of $ER_i(t)$ requires determining the joint waiting time distribution of any two of the $N_0(t)$ resupply requests. This is extremely cumbersome and impractical for any realistic implementation.

Since

$$DE_i(t) = Q_i(t) + ER_i(t) \tag{4.39}$$

the inability to obtain general results for $ER_i(t)$ prohibits finding general results for $DE_i(t)$ (other than using (4.33) and (4.38) to find $E[DE_i(t)]$). Furthermore, because of the correlation in the $T(n)$, $n=1,2,...$ and the correlation in the waiting times at the depot, $ER_i(t)$ and $Q_i(t)$ are not independent. This fact further complicates the task of obtaining general results for $DE_i(t)$ from (4.39).

There is, however, a useful and important special case, embodied in Assumption 11 below, for which we can get tractable expressions for the distribution of $DE_i(t)$.

<u>Assumption 11</u>: $OST_i(y)$ does not depend on y and is a fixed known value, OST_i^*, $i=1,2,...,N$.

The assumption of a constant, deterministic order and ship time between the depot and $base_i$ does not in itself provide a convenient aid in obtaining the distribution of $ER_i(t)$. However, Theorem 4.8 shows how Assumption 11 allows us to calculate directly the distribution and other properties of $DE_i(t)$, $t \geq 0$.

<u>Theorem 4.8</u>: For constant, deterministic order and ship times and $t \geq 0$,

$$DE_i(t) = [N_i^0(t) - N_i^0(t-OST_i^*)] + Q_i(t-OST_i^*)$$

where, by assumption, $N_i^0(y) = Q_i(y) = 0$ for $y \leq 0$.

<u>Proof</u>: $\{N_i^0(t) - Q_i(t),\ t \geq 0\}$ is the counting process describing the number of $base_i$ resupply requests satisfied by the depot. Since order and ship times are constant and deterministic, only $base_i$ resupply requests satisfied in $(t-OST_i^*, t]$ will be en route to $base_i$ at $t \geq 0$. Hence

$$ER_i(t) = [N_i^0(t) - Q_i(t)] - [N_i^0(t-OST_i^*) - Q_i(t-OST_i^*)].$$

Substituting for $ER_i(t)$ in (4.39) establishes the theorem. //

While Assumption 11 is somewhat restrictive, it seems that it is necessary to make this assumption in order to obtain any tractable analytic results. However, there is another reason why one might wish to adopt Assumption 11. In practice, it is almost always true that a base will receive units from the depot in the same order that the units were shipped. Hence, deliveries to a particular base do not cross. This makes it unlikely that order and ship times are independent random variables. Without developing a detailed model of the depot shipping process, it is difficult to describe the dependence among shipping times to a particular base. By Assumption 11, we prevent deliveries from crossing by removing any variation in the shipping times. Accounting for shipping time variation is a difficult and unsolved problem even for a single location (Q,r) inventory system (Hadley and Whitin [1963]).

For the remainder of this dissertation we shall use Assumption 11. Using Theorem 4.8 we can then write the base_i pipeline for $t \geq 0$ as

$$X_i(t) = D_i(t)+R_i(t)+E_i(t)+[N_i^0(t)-N_i^0(t-OST_i^*)]+Q_i(t-OST_i^*). \qquad (4.40)$$

All the random variables on the right in (4.40) are independent. Therefore, $X_i(t)$, $t \geq 0$, can be obtained from the convolution of the distributions that we have derived in the previous sections of this chapter.

Delay at a Base

An interesting and oft times useful measure of inventory performance at a base is the expected delay until a customer who has brought in a failed unit is resupplied from the base's spares pool. In a stationary system, the expected wait may be obtained by using Little's

Formula. In a non-stationary system, however, we need to determine the distribution of the base$_i$ waiting time, $W_i(t)$, in order to calculate the mean waiting time at $t \geq 0$. The limiting distribution of $W_i(t)$ has not previously been derived for stationary multi-echelon systems but can be obtained as a special case of the results of this section.

When a customer arrives at base$_i$ at $t \geq 0$, he will not wait if $X_i(t) < s_i(t)$. However, if $X_i(t) \geq s_i(t)$, the customer will receive the $(X_i(t) - s_i(t) +1)$th serviceable unit that becomes available for issue at base$_i$ after t. For $w \geq 0$, let $AV_i(t,t+w)$ be the total number of units that becomes available for issue at the base$_i$ spares pool in $(t,t+w]$. Then $W_i(t) > w \geq 0$ if and only if

$$X_i(t) \geq s_i(t) + AV_i(t,t+w).$$

Using (4.40) to substitute for $X_i(t)$ we have $W_i(t) > w \geq 0$ if and only if

$$D_i(t) + E_i(t) + R_i(t) + [N_i^0(t) - N_i^0(t-OST_i^*)] + Q_i(t-OST_i^*) \geq s_i(t) + AV_i(t). \quad (4.41)$$

$D_i(t)$ consists of: units that will still be in the diagnostic facility at t+w $[=D_i^D(t,t+w)]$; units that will leave the diagnostic facility in (t,t+w] after being condemned $[=D_i^E(t,t+w)]$; units that will leave the diagnostic facility in (t,t+w] and enter the base$_i$ repair facility $[=D_i^R(t,t+w)]$; and units that will leave the diagnostic facility in (t,t+w] and enter the depot pipeline $[=D_i^0(t,t+w)]$. Therefore, the inequality (4.41) can be written as

$$D_i^D(t,t+w) + D_i^E(t,t+w) + D_i^R(t,t+w) + D_i^0(t,t+w) + E_i(t) + R_i(t) + [N_i^0(t) - N_i^0(t-OST_i^*)] + Q_i(t-OST_i^*) \geq s_i(t) + AV_i(t). \quad (4.42)$$

Define:

$I_i^E(t,t+w)$ = 1 if the unit that failed at t^+ was condemned in (t,t+w] and a replacement from the external supplier arrived in $(t,t+w]$. $I_i^E(t,t+w) = 0$ otherwise.

$I_i^R(t,t+w)$ = 1 if the unit that failed at t^+ was sent to the base_i repair facility in $(t,t+w]$ and completed repair in $(t,t+w]$. $I_i^R(t,t+w) = 0$ otherwise.

$\#E_i(t_1,t_2,t,t+w)$ = number of units that arrive from the external supplier in $(t,t+w]$ to replace failures at base_i in $(t_1,t_2]$ that were condemned in $(t_1,t_2]$.

$\#E_i^D(t,t+w)$ = number of units that arrive from the external supplier in $(t,t+w]$ to replace failures at base_i in $(0,t]$ that were condemned in $(t,t+w]$. Note that $\#E_i^D(t,t+w) \leq D_i^E(t,t+w)$.

$\#R_i(t_1,t_2,t,t+w)$ = number of failures in $(t_1,t_2]$ that were sent to the base_i repair facility in $(t_1,t_2]$ and completed repair in $(t,t+w]$.

$\#R_i^D(t,t+w)$ = number of failures in $(0,t]$ that were sent to the base_i repair facility in $(t,t+w]$ and completed repair in $(t,t+w]$. $[\#R_i^D(t,t+w) \leq D_i^R(t,t+w)]$.

$\#DE_i(t,t+w)$ = number of units that arrive at base_i as resupply from the depot in $(t,t+w]$. (Possibly including resupply for the demand at t^+).

$M_i^+(t,t+w)$ = number of units management has sent to $base_i$ that arrived in $(t,t+w]$.

Then,

$$AV_i(t,t+w) = M_i^+(t,t+w) + I_i^E(t,t+w) + I_i^R(t,t+w) + \#DE_i(t,t+w)$$
$$+ \#E_i(0,t,t,t+w) + \#E_i^D(t,t+w) + \#E_i(t,t+w,t,t+w)$$
$$+ \#R_i(0,t,t,t+w) + \#R_i^D(t,t+w) + \#R_i(t,t+w,t,t+w).$$

After some rearrangement the inequality (4.42) can now be rewritten as

$$[D_i^D(t,t+w)] + [D_i^E(t,t+w) - \#E_i^D(t,t+w)]$$
$$+ [E_i(t) - \#E_i(0,t,t,t+w)] - [\#E_i(t,t+w,t,t+w)]$$
$$+ [D_i^R(t,t+w) - \#R_i^D(t,t+w)] + [R_i(t) - \#R_i(0,t,t,t+w)]$$
$$- [\#R_i(t,t+w,t,t+w)] - [I_i^E(t,t+w) + I_i^R(t,t+w)]$$
$$+ [N_i^0(t) - N_i^0(t-OST_i^*) + Q_i(t-OST_i^*) + D_i^0(t,t+w) - \#DE_i(t,t+w)]$$
$$\geq s_i(t) + M_i^+(t,t+w). \tag{4.43}$$

A term by term analysis of the left side of (4.43) yields:

(1) $D_i^D(t,t+w)$ has a Poisson distribution with mean

$$\int_0^t \lambda_i(s)[1-G_i(s,t+w)]ds.$$

Applying the Splitting Property P10, it is straightforward to show (see Chapter III) that $D_i^D(t,t+w)$ is independent of the number of units that left the $base_i$ diagnostic facility in $(0,t+w]$. Hence, $D_i^D(t,t+w)$ is independent of every other term in (4.43).

(2) $D_i^E(t,t+w) - \#E_i^D(t,t+w)$ is the number of units that failed in $(0,t]$, were condemned in $(t,t+w)$ and for which replacements from the external supplier have not arrived at $base_i$ by $t+w$. This quantity has a Poisson distribution with mean

$$\int_0^t \lambda_i(s) \int_t^{t+w} dG_i(s,y)[1-P_i(s,y)-P_R^i(s,y)][1-e_i(s,y,t+w)]ds.$$

P10 assures us that this term is independent of the other terms in (4.43).

(3) $E_i(t) - \#E_i(0,t,t,t+w)$ is the number of units that failed in (0,t], were condemned in (0,t] and for which replacements from the external supplier have not arrived at base$_i$ by t+w. This quantity has a Poisson distribution with mean

$$\int_0^t \lambda_i(s) \int_s^t dG_i(s,y)[1-P_i(s,y)-P_R^i(s,y)][1-e_i(s,y,t+w)]ds.$$

Applying P10 we can establish the independence of this term and the other terms in (4.43).

(4) $\#E_i(t,t+w,t,t+w)$ has a Poisson distribution with mean

$$\int_t^{t+w} \lambda_i(s) \int_s^{t+w} dG_i(s,y)[1-P_i(s,y)-P_R^i(s,y)]e_i(s,y,t+w)ds$$

and by P10 can be shown to be independent of the other terms in (4.43).

(5) $D_i^R(t,t+w) - \#R_i^D(t,t+w)$ is the number of units that failed in (0,t], entered the base$_i$ repair facility in (t,t+w] and are still in the repair facility at t+w. This quantity has a Poisson distribution with mean

$$\int_0^t \lambda_i(s) \int_t^{t+w} dG_i(s,y)P_R^i(s,y)[1-r_i(s,y,t+w)]ds.$$

and (by P10) is also independent of every other term in (4.43).

(6) $R_i(t) - \#R_i(0,t,t,t+w)$ is the number of units that failed in (0,t], entered the base$_i$ repair facility in (0,t] and are still in the repair facility at t+w. This quantity has a Poisson distribution with mean

$$\int_0^t \lambda_i(s)\int_s^t dG_i(s,y)P_R^i(s,y)[1-r_i(s,y,t+w)]ds.$$

and is independent of the other terms in (4.43).

(7) $\#R_i(t,t+w,t,t+w)$ has a Poisson distribution with mean

$$\int_t^{t+w} \lambda_i(s) \int_s^{t+w} dG_i(s,y)P_R^i(s,y)r_i(s,y,t+w)ds$$

and is independent of the other terms in (4.43).

(8) Since $I_i^E(t,t+w) + I_i^R(t,t+w) \leq 1$ these two random variables are not independent. However, the distribution of their sum is easily obtained after noting that

$$\Pr[I_i^E(t,t+w) + I_i^R(t,t+w) = 1) =$$

$$\int_t^{t+w} \{P_R^i(t,y)r_i(t,y,t+w) + [1-P_R^i(t,y)-P_i(t,y)]e_i(t,y,t+w)\}\, dG_i(t,y).$$

$[I_i^E(t,t+w)+I_i^R(t,t+w)]$ and $\#DE_i(t,t+w)$ may be correlated because $\#DE_i(t,t+w)$ may provide information on whether the failure at t^+ was or was not sent to the depot in $(t,t+w]$. We will return to this point momentarily.

(9) Since OST_i^* is deterministic, units that arrive from the depot in $(t,t+w]$ must have been shipped in $(t-OST_i^*,t+w-OST_i^*]$. First, consider the case where $w \leq OST_i^*$. If the arrival at t^+ is sent to the depot in $(t,t+w]$, a serviceable replacement for this unit could not have arrived at base$_i$ by $t+w$ even if $H_0(t^+) > 0$. Therefore,

$$\#DE_i(t,t+w) = [N_i^0(t+w-OST_i^*) - Q_i(t+w-OST_i^*)] - [N_i^0(t-OST_i^*) - Q_i(t-OST_i^*)]$$

and

$$\begin{aligned} &N_i^0(t) - N_i^0(t-OST_i^*) + Q_i(t-OST_i^*) + D_i^0(t,t+w) - \#DE_i(t,t+w) \\ &\quad = [N_i^0(t) - N_i^0(t+w-OST_i^*)] + D_i^0(t,t+w) + Q_i(t+w-OST_i^*). \quad (4.44) \end{aligned}$$

$Q_i(t+w-OST_i^*)$ is not affected by demands on the depot after $t+w-OST_i^*$. Hence, for $w \leq OST_i^*$, $Q_i(t+w-OST_i^*)$ is independent of $[N_i^0(t) - N_i^0(t+w-OST_i^*)]$ and $D_i^0(t,t+w)$. Furthermore, using P10 it is easy to show that $D_i^0(t,t+w)$ is independent of the number of units that left the base$_i$ diagnostic facility in $(0,t]$ and therefore $D_i^0(t,t+w)$ is independent of

$[N_i^0(t) - N_i^0(t+w-OST_i^*)]$. Therefore, the sum on the right of (4.44) is the sum of independent random variables. The distribution of this sum is readily obtained using (3.1), (4.30), (4.32) and the fact that from P10, $D_i^0(t,t+w)$ has a Poisson distribution with mean $pd(0,t,t,t+w)$ where

$$pd(t_1,t_2,t_3,t_4) = \int_{t_1}^{t_2} \lambda_i(s) \int_{t_3}^{t_4} dG_i(s,y)P_i(s,y)ds.$$

For $w \leq OST_i^*$, $\#DE_i(t,t+w)$ provides no information about the failure at t^+ and is independent of $[I_i^E(t,t+w) + I_i^R(t,t+w)]$. Hence, all nine terms on the left in (4.43) are mutually statistically independent. $Pr[OST_i^* \leq w < W_i(t)]$ can now be obtained in a straightforward manner.

Now consider the case where $w > OST_i^*$. The known failure at $base_i$ at t^+ results in a non-continuous MVF for the NHPP describing the number of $base_i$ resupply requests placed on the depot. Therefore,

$$\begin{aligned}\#DE_i(t,t+w) = {} & [I_i^0(t,t+w-OST_i^*) + N_i^0(t+w-OST_i^*) - Q_i^t(t+w-OST_i^*)] \\ & - [N_i^0(t-OST_i^*) - Q_i(t-OST_i^*)]\end{aligned}$$

where

$I_i^0(t,t+w-OST_i^*)$ = 1 if the failure at t^+ was sent to the depot no later than $t+w-OST_i^*$. $I_i^0(t,t+w-OST_i^*) = 0$ otherwise.

$Q_i^t(t+w-OST_i^*)$ = number of depot backorders at $t+w-OST_i^*$ that belong to $base_i$ given that there was a failure at $base_i$ at t^+.

Hence,

$$\begin{aligned}& N_i^0(t) - N_i^0(t-OST_i^*) + Q_i(t-OST_i^*) + D_i^0(t,t+w) - \#DE_i(t,t+w) \\ & = [N_i^0(t) - N_i^0(t+w-OST_i^*)] + D_i^0(t,t+w) + Q_i^t(t+w-OST_i^*) - I_i^0(t,t+w-OST_i^*) \\ & = Q_i^t(t+w-OST_i^*) - I_i^0(t,t+w-OST_i^*) + [D_i^0(t,t+w) - D_i^0(t,t+w-OST_i^*)] \\ & \quad - [N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]. \qquad (4.45)\end{aligned}$$

$D_i^0(t,t+w) - D_i^0(t,t+w-OST_i^*)$ is the number of failures at $base_i$ in $(0,t]$ that were sent to the depot in $(t+w-OST_i^*,t+w]$. This quantity has a Poisson distribution with mean $pd(0,t,t+w-OST_i^*,t+w)$ and is independent of $I_i^0(t,t+w-OST_i^*)$ [Assumptions 2 through 6] and $Q_i^t(t+w-OST_i^*)$ [since these demands were placed on the depot after $t+w-OST_i^*$].

$[N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]$ is the number of failures at $base_i$ in $(t,t+w-OST_i^*]$ that were sent to the depot in $(t,t+w-OST_i^*]$. This quantity has a Poisson distribution with mean $pd(t,t+w-OST_i^*,t,t+w-OST_i^*)$ and is independent of $I_i^0(t,t+w-OST_i^*)$ [Assumptions 2 through 6] and $[D_i^0(t,t+w) - D_i^0(t,t+w-OST_i^*)]$ [Assumptions 2 through 6 and the fact that NHPP have independent increments].

$Q_i^t(t+w-OST_i^*)$ is not independent of $I_i^0(t,t+w-OST_i^*)$ and $[N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]$. The known failure at $base_i$ at t^+ and the number of failures at $base_i$ in $(t,t+w-OST_i^*]$ that are sent to the depot in $(t,t+w-OST_i^*]$ certainly affect $Q_i^t(t+w-OST_i^*)$. However, the distribution of

$$[Q_i(t+w-OST_i^*) \mid I_i^0(t,t+w-OST_i^*),\ N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]$$

can be obtained straightforwardly, albeit tediously, by extending the arguments that led to (4.30) and (4.32). Details can be found in Kotkin (1985). Using the fact that $I_i^0(t,t+w-OST_i^*)$ and $[N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]$ are independent, the joint distribution of $Q_i(t+w-OST_i^*)$, $I_i^0(t,t+w-OST_i^*)$ and $[N_i^0(t+w-OST_i^*) - N_i^0(t) - D_i^0(t,t+w-OST_i^*)]$ can be obtained straightforwardly. The distribution of the sum in (4.45) can then be obtained directly (Kotkin [1985]).

For $w > OST_i^*$, $I_i^0(t,t+w-OST_i^*)$, $I_i^E(t,t+w)$ and $I_i^R(t,t+w)$ are not independent since at most one of these random variables can be positive. Assumptions 2 through 6 guarantee, however, that $I_i^E(t,t+w)$ and $I_i^R(t,t+w)$

are independent of the other random variables in (4.45). Since the first seven terms on the left in (4.43) are independent of the latter two terms, all that remains to be done in order to find $\Pr[W_i(t) > w > OST_i^*]$ is to find the distribution of $[I_i^E(t,t+w) + I_i^R(t,t+w)|I_i^0(t,t+w-OST_i^*]$. This is easily done. Clearly,

$$\Pr[I_i^E(t,t+w) + I_i^R(t,t+w) = 1|I_i^0(t,t+t+w-OST_i^*) = 1] = 0$$

so that

$$\Pr[I_i^E(t+w) + I_i^R(t,t+w) = 1|I_i^0(t,t+w-OST_i^*) = 0]$$
$$= \Pr[I_i^E(t+w) + I_i^R(t,t+w) = 1]/\Pr[I_i^0(t,t+w-OST_i^*) = 0]$$

where the numerator on the right is given in paragraph (8) above and

$$\Pr[I_i^0(t,t+w-OST_t^*) = 0] = \int_t^{t+w-OST_i^*} [1-P_i(t,y)]dG_i(t,y).$$

Similarly,

$$\Pr[I_i^E(t,t+w) + I_i^R(t,t+w) = 0|I_i^0(t,t+w-OST_i^*) = 1] = 1,$$
$$\Pr[I_i^E(t,t+w) + I_i^R(t,t+w) = 0|I_i^0(t,t+w-OST_i^*) = 0]$$
$$= \frac{\Pr[I_i^0(t,t+w-OST_i^*) = 0] - \Pr[I_i^E(t,t+w) + I_i^R(t,t+w) = 1]}{\Pr[I_i^0(t,t+w-OST_i^*) = 0]}.$$

Summary

The supply interactions between the depot and bases manifest themselves through the impact of stock policies at the depot on the bases' pipelines. In Chapter IV we explicitly examined this impact by deriving the distribution of a bases's pipeline as a function of the depot asset level. Therefore, inventory performance at the bases can be improved either by increasing stock at the bases or by increasing stock at the depot, thereby reducing the bases' pipelines. This tradeoff between depot and base stock forms the basis of the optimization problem that we formulate in the next chapter.

Next page is blank.

CHAPTER V

THE FIXED ASSET VECTOR PROBLEM

The emphasis of the previous two chapters has been on describing important stochastic processes that arise at the depot and the bases. In particular, in Chapters III and IV we derived the time-dependent probability distributions of the number of units in the pipeline at the depot and bases, respectively. Using these results, (2.2) and (2.3), we can obtain the probability distributions of the number of units on-hand and the number of backorders outstanding at any time during the horizon at every location in the inventory system. We can then develop measures of inventory performance which aid in evaluating different stock policies and asset levels during the horizon. It then becomes natural to formulate an optimization problem that allocates a valuable resource (money, weight, volume, etc.) over a catalog of items (for example, the items that comprise a weapon system) in order to maximize the inventory performance of the catalog over the horizon. The Fixed Asset Vector Problem (FAVP), introduced in this chapter, represents the first step towards formulating such an optimization problem. The FAVP formulation assumes that all asset levels remain unchanged during the horizon: there are no management directives to change the asset levels at any location during the horizon. In Chapter VIII we examine the optimization problem that arises when this assumption is removed.

However, the FAVP is extremely interesting in its own right. It can be used for scenarios during which management directed changes can not be implemented for any reason including, of course, lack of time and/or lack of asset visibility and control.

Inventory Performance Measures

There are several useful measures of inventory performance including criteria based on an item's fill rate, ready rate, and expected number of outstanding backorders. At $t \geq 0$, we define the fill rate at $location_j$ for a particular item, $FR_j(t)$, as the probability that there is at least one unit on-hand at $location_j$ at time t. Therefore, $FR_j(t) = Pr\ [X_j(t) < s_j(t)]$. Note that if $s_j(t) = 0$ then $FR_j(t) = 0$. The ready rate at $location_j$ at time t for a particular item, $RR_j(t)$, is defined as the probability there are no backorders outstanding at $location_j$ at time t. Therefore, $RR_j(t) = Pr\ [X_j(t) \leq s_j(t)]$. Finally, from (2.2), the expected number of backorders of a particular item outstanding at $location_j$ at time t is given by

$$EBO(s_j(t), s_0(t), t) = \sum_{k > s_j(t)} Pr(X_j(t) \geq k). \tag{5.1}$$

All of these performance measures depend upon $s_j(t)$ and also upon $s_0(t)$ since $X_j(t)$ depends upon $s_0(t)$.

Brooks, Gillen, and Lu (1969) discussed the relative merits of using the different performance measures in steady state models. They also presented some computational experience comparing the asset allocations from single echelon steady state optimization problems using the different performance measures. Since a backorder outstanding at a base implies a customer is waiting, the measure most directly related to customer satisfaction is the expected number of backorders outstanding at the bases. Fill rate and ready rate are measures more of immediate supply accommodation than of customer satisfaction. For this reason, the expected number of backorders outstanding at the bases is the performance measure used in most steady state multi-echelon models. (See, for example, Sherbrooke [1968], Clark [1978], Kaplan [1980], Vincent [1980], and O'Malley [1983]). For the remainder of this study we assume that the performance criteria are based on the expected number of backorders outstanding at the bases. However, any of the other measures could be used in developing optimization problems. The results presented here have analogs for each of the different measures.

Performance Criteria in Non-Stationary Systems

Definition 5.1: An asset vector is a vector containing an asset level for every item at every location. As asset policy specifies an asset vector at every point in time during the horizon.

Given an asset policy, inventory performance can be measured at any point in time during the horizon. While time plots of inventory performance may be useful management aids, they usually do not, in themselves, provide an objective way to evaluate alternative asset policies. The major exception is when one policy provides better performance at every point in time. Two important objective performance criteria that can be obtained from the time plots are the average performance over the horizon and the worst performance during the horizon. Usually, management has an interest in both these criteria. For obvious reasons, management is concerned with the worst performance predicted during the horizon. Regardless of the average performance over the horizon, management may not want to tolerate extended periods of severely poor performance or even poor performance during short critical periods.

However, the intervals of poor performance may be during non-critical time periods and may be mild enough and/or short enough so that the average performance over the horizon is at a satisfactory level. In this case, management may not want to perturb the inventory system and incur the extra expense of added assets because of small, non-severe, non-critical periods of poor inventory performance.

By controlling the average performance over the horizon, management assures that the inventory system provides satisfactory service during the horizon. The average performance also provides a way of distinguishing between two equal cost asset policies that have the same worst performance over the horizon.

Therefore, inventory managers are faced with a multi-criteria optimization problem. They wish to choose an asset policy that makes the most economical use of available resources while controlling the average and worst inventory performance over the horizon.

The major advantage of ergodic theory is that in steady state all points in time are stochastically identical. With one constraint on inventory performance it is possible to control both the average and worst behavior for some time interval in steady-state. In non-stationary systems, where there is no passage to steady state and each point in time may be stochastically different, one constraint may be insufficient to satisfactorily control both the average and worst performance. Different asset policies may have the same average expected number of backorders outstanding but vastly different values for the maximum expected number of backorders outstanding over the horizon. Similarly, different asset policies with the same maximum expected number of backorders outstanding over the horizon may have a significantly different average expected number of outstanding backorders. For these reasons, the FAVP formulation includes constraints on both the average and maximum expected number of outstanding backorders over the horizon.

Formulation of the FAVP

The major assumption of the FAVP is that the asset levels for every item at every location are fixed at time 0. There are no management directives to change the asset position of any item at any location during the horizon. This restricts the set of feasible asset policies to policies for which the asset vector does not change over the horizon. In Chapter VIII we discuss the optimization problem that results from dropping this restriction. For notational convenience, while discussing the FAVP, we shall drop the notation indicating the dependence of the asset level on time.

The other assumptions of the FAVP are:

a. Inventory performance is measured at the bases. Depot performance is included only insofar as it affects performance at the bases. This is reasonable since primary customer demands occur only at the bases (Assumption 1). However, the formulation could easily be modified to include depot performance explicitly.

b. Performance constraints are expressed in terms of the expected total number of backorders for all items at all the bases. The FAVP formulation can easily be modified to include constraints on the performance at each base and/or on the performance of each item.

c. The objective is to minimize, over a catalog of items, the total cost of procuring or holding the required assets. The FAVP could easily be modified to minimize the weight or volume of the optimal asset vector instead of the cost. Adding constraints on the weight and/or volume will increase the computational burden of obtaining optimal solutions.

Define

I = number of items in the catalog;

T = horizon length;

$X_{ij}(t)$ = pipeline for item_i at location_j at t, $i=1,2,\ldots,I$; $j = 0,1,\ldots,N$; $0 \leq t \leq T$;

s_{ij} = asset level of item_i at location_j, $i = 1,2..,I$; $j = 0,1,..N$;

s_{i+} = $(s_{i0}, s_{i1},\ldots,s_{iN})$, $i = 1,2,\ldots I$;

s_{+j} = $(s_{1j}, s_{2j},\ldots,s_{Ij})$, $j = 0,1,\ldots,N$;

s_{++} = $(s_{1+}, s_{2+},\ldots, s_{I+})$;

C_i = procurement or holding cost of item_i, $i = 1,2,..I$;

1_N = a vector of ones with dimension N+1;

$$AEBO(s_{ij},s_{i0}) = \frac{1}{T}\int_0^T EBO(s_{ij},s_{i0},t)dt$$

= the average expected number of backorders of item_i outstanding at base_j over the horizon. $i = 1,2,\ldots I$; $j = 1,2,\ldots,N$;

$$\text{MEBO}(s_{++}) = \max_{0 \le t \le T} \{ \sum_{i=1}^{I} \sum_{j=1}^{N} \text{EBO}(s_{ij}, s_{i0}, t) \}$$

= the maximum expected total number of backorders for all items outstanding at the bases.

The FAVP can now be written as

$$\text{Min} \sum_{I} C_i (\sum_{j=0}^{N} s_{ij})$$

$$\sum_{i=1}^{I} \sum_{j=1}^{N} \text{AEBO}(s_{ij}, s_{i0}) \le \text{aeb} \tag{5.2}$$

$$\text{MEBO}(s_{++}) \le \text{meb}$$

$$s_{ij} = 0,1,2,\ldots \quad i = 1,2,..I; \; j = 0,1,..,N.$$

If $\text{AEBO}(s_{ij}, s_{i0})$ has a limit as T goes to infinity for each item in the catalog and the constraint on MEBO (s_{++}) is removed, the FAVP reduces to the steady state two-echelon model first introduced by Sherbrooke (1968).

Solution of the FAVP

For aeb, meb ≥ 0 (5.2) always has a feasible solution since as all s_{ij} go to infinity, $\text{MEBO}(s_{++})$ and all the $\text{AEBO}(s_{ij}, s_{i0})$ go to zero. If we introduce a Generalized Lagrange Multiplier (GLM), $u_a \ge 0$ (Everett [1963]), we can rewrite (5.2) as

$$\text{Min} \sum_{i=1}^{I} \{ C_i s_{i0} + \sum_{j=1}^{N} K(s_{ij}; s_{i0}) \}$$

$$\text{MEBO}(s_{++}) \le \text{meb} \tag{5.3}$$

where $K(s_{ij}; s_{i0}) = C_i s_{ij} + u_a \text{AEBO}(s_{ij}, s_{i0})$. Here, and in the sequel, it is understood that all decision variables are non-negative integers. We can use a result of Everett (1963) to relate optimal solutions of (5.3) to solutions of (5.2).

<u>Theorem 5.2:</u> Let $s_{++}(u_a)$ be an optimal solution to (5.3) for a particular GLM $u_a \geq 0$. Then $s_{++}(u_a)$ is optimal in (5.2) with the right hand sides of the constraints replaced, respectively, with

$$aeb[s_{++}(u_a)] = \sum_{i=1}^{I} \sum_{j=1}^{N} AEBO[s_{ij}(u_a), s_{i0}(u_a)]$$

$$meb[s_{++}(u_a)] = MEBO[s_{++}(u_a)] \leq meb.$$

<u>Proof</u>: Consider (5.2) with aeb and meb replaced with $aeb[s_{++}(u_a)]$ and $meb[s_{++}(u_a)]$ respectively. Clearly, $s_{++}(u_a)$ is feasible. For all asset vectors y_{++} such that $MEBO(y_{++}) \leq meb$ we have that

$$\sum_{i=1}^{I} C_i[s_{i+}(u_a) - y_{i+}] \cdot 1_N \leq u_a \sum_{i=1}^{I} \sum_{j=1}^{N} AEBO[y_{ij}, y_{i0}] - AEBO[s_{ij}(u_a), s_{i0}(u_a)] \; .$$

since $s_{++}(u_a)$ is optimal in (5.3). If y_{++} is feasible in the augmented (5.2) we have that

$$\sum_{i=1}^{I} \sum_{j=1}^{N} AEBO[y_{ij}, y_{i0}] \leq \sum_{i=1}^{I} \sum_{j=1}^{N} AEBO[s_{ij}(u_a), s_{i0}(u_a)]$$

which implies that y_{++} has an objective function value in (5.2) which is no smaller than the value of $s_{++}(u_a)$. //

By varying u_a one can use (5.3) to obtain solutions to (5.2) with different values for the right hand sides of the constraints. Solutions to (5.3) are undominated (efficient) solutions of (5.2) in the sense that any asset vector that has lower procurement/holding costs than $s_{++}(u_a)$ must have either higher average expected total backorders, higher maximum expected total backorders, or both. Usually, one can obtain an undominated solution to (5.2) with $aeb[s_{++}(u_a)]$ and $meb[s_{++}(u_a)]$ sufficiently close to the desired levels aeb and meb. For two reasons, however, this can not be guaranteed. First, since all asset levels are non-negative integers, there are infinite values of aeb and meb for which the constraints in (5.3) will never hold at equality. Secondly, when using a GLM, a "duality gap" often arises (see Everett [1963]). A duality gap occurs when there is a solution to (5.2) for particular aeb and meb but there is no value of u_a that can obtain this solution. Everett showed that a gap arises when the objective function in (5.2) is not a strictly convex function of the right hand sides of the constraints in (5.2) (which, as we shall see later, is the case here). Everett's GLM procedure can only generate points of strict convexity in the three dimensional space representing the optimal objective function value in (5.2) as a function of aeb and meb.

In practice aeb and meb are usually soft management parameters and acceptably close values of the constraints are sufficient. Therefore, we will obtain undominated solutions of (5.2) by using Everett's (1963) GLM technique. For the moment, we shall concentrate on solving the relaxation of (5.3) obtained by removing the constraint on the maximum expected total number of backorders outstanding at the bases. Without this constraint, (5.3) is separable by item. Unfortunately, one can construct examples that show that $K(s_{ij};s_{i0})$ is not a convex function either of the 2-tuple (s_{ij},s_{i0}) or of the quantity $s_{i0} + s_{ij} = 0,1,2....$ Since $K(s_{ij};s_{i0})$ is not convex, the objective function of the subproblem for $item_i$ and the objective function in (5.3) are not convex.

However, using Theorem 5.3, it is straightforward to show that when s_{i0} is held fixed at some non-negative integer value, $K(s_{ij};s_{i0})$

is a convex function of s_{ij}, $j = 1,2,..,N$.

Theorem 5.3: For fixed s_{i0}, $AEBO(s_{ij},s_{i0})$ is a discretely convex decreasing function of s_{ij}, $i = 1,2,...,I$; $j = 1,2,...,N$.

Proof: From Theorem 2.2 we have that for $s_{ij} = 0,1,2...$ and $t \geq 0$,

$$EBO(s_{ij}+1,s_{i0},t) - EBO(s_{ij},s_{i0},t) = -\Pr(X_{ij}(t) \geq s_{ij}+1)$$

and therefore, for fixed s_{i0}, $EBO(s_{ij},s_{i0},t)$ is a discretely convex decreasing function of s_{ij}. The theorem now follows by applying two elementary properties of convex functions. //

We will exploit the convexity of $K(s_{ij};s_{i0})$ for fixed s_{i0} in an implicit enumeration scheme to solve the subproblem for item_i. The item_i subproblem, $i = 1,...,I$, can be written as

$$\min_{s_{i0} = 0,1,2,...} TC_i(s_{i0}) = C_i s_{i0} + TCB_i^*(s_{i0})$$

where

$$TCB_i^*(s_{i0}) = \min \sum_{j=1}^{N} K(s_{ij};s_{i0})$$

is the total optimal contribution by the bases to the objective function of the item_i subproblem when the depot asset level is s_{i0}. For fixed s_{i0}, the item_i subproblem is separable by base since we need only find $TCB_i^*(s_{i0})$. Therefore, for fixed s_{i0}, we can solve the item_i subproblem by minimizing $K(s_{ij};s_{i0})$ for each base.

Since $K(s_{ij};s_{i0})$ is convex for fixed s_{i0}, $\underline{s}_{ij}(s_{i0})$, the base_i asset level that minimizes $K(s_{ij};s_{i0})$, is the smallest non-negative integer for which $K(s_{ij}+1;s_{i0}) > K(s_{ij};s_{i0})$. Therefore, $\underline{s}_{ij}(s_{i0})$ is zero if and only if

$$\int_0^T \Pr(X_{ij}(t) \geq 1)dt < \frac{C_i T}{u_a} . \tag{5.4}$$

Otherwise, $\underline{s}_{ij}(s_{i0})$ is the unique positive integer satisfying

$$\int_0^T \Pr(X_{ij}(t) \geq \underline{s}_{ij}(s_{i0})+1)dt < \frac{C_i T}{u_a} \leq \int_0^T \Pr(X_{ij}(t) \geq \underline{s}_{ij}(s_{i0}))dt. \quad (5.5)$$

Say T is measured in days. Then u_a is a backorder cost in the sense that for every T backorder-days accumulated over the horizon, there is a "charge" to the inventory system of u_a dollars in $K(s_{ij};s_{i0})$. Intuitively, (5.4) and (5.5) state that $\underline{s}_{ij}(s_{i0})$ is such that the marginal reduction in backorder costs over the horizon from adding the $\underline{s}_{ij}(s_{i0})^{th}$ asset at base$_j$ must be greater than the marginal increase in procurement/holding costs from adding that asset. Furthermore, the reduction in backorder costs from adding the $(\underline{s}_{ij}(s_{i0}) + 1)^{th}$ asset (or any assets after that) must be less than the corresponding increase in procurement/holding costs.

The optimality conditions (5.4) and (5.5) allow for straightforward determination of $\underline{s}_{ij}(s_{i0})$, and thereby $TC_i(s_{i0})$, for any value of s_{i0}. However, empirical evidence from tests on the items in our data base (see Appendix B) has shown that not only is $TC_i(s_{i0})$ not convex, it is not unimodal. Therefore, in order to solve the item$_i$ subproblem, it is necessary to determine $TC_i(s_{i0})$ for all $s_{i0} = 0,1,2,\ldots$. The optimal depot asset level for item$_i$, $\underline{s}_{i0}$, is the non-negative integer that yields $TC_i^*(\underline{s}_{i0})$, the minimum value of $TC_i(s_{i0})$. Fortunately, it is possible to a priori determine an upper bound on $\underline{s}_{i0}$. Before establishing this upper bound we need to obtain some intermediate results.

<u>Lemma 5.4</u>: If the random variable G is stochastically larger than the random variable Z, then $E[G] \geq E[Z]$.

<u>Proof</u>: See Lehmann (1959). //

<u>Lemma 5.5</u>: For fixed s_{ij} and all $t > 0$, $EBO(s_{ij},s_{i0},t)$ is a decreasing function of $s_{i0} = 0,1,2,\ldots$.

Proof: $X_{ij}(t)$, $t \geq 0$, is given by (4.40). The only component of $X_{ij}(t)$ that depends on s_{i0} is $Q_{ij}(t-OST^*_{ij})$. By Theorem 4.2, $Q_{ij}(t-OST^*_{ij})$ is stochastically decreasing with respect to s_{i0}. Therefore, $X_{ij}(t)$ and $[X_{ij}(t)-s_{ij}]^+$ are stochastically decreasing with respect to s_{i0}. It now follows directly from (2.2) and Lemma 5.4 that for all $s_{i0} = 0,1,2,...$ and $t > 0$, $EBO(s_{ij},s_{i0}+1,t) \leq EBO(s_{ij},s_{i0},t)$.

Theorem 5.6: $TCB^*_i(s_{i0})$ is a decreasing function of $s_{i0} = 0,1,2,....$

Proof: Let $k > h$ be non-negative integers. Then

$$\begin{aligned} TCB^*_i(h) &= \sum_{j=1}^{N} K(\underline{s}_{ij}(h);h) \\ &= \sum_{j=1}^{N} [C_i\underline{s}_{ij}(h) + u_a AEBO(\underline{s}_{ij}(h);h) \\ &> \sum_{j=1}^{N} [C_i\underline{s}_{ij}(h) + u_a AEBO(\underline{s}_{ij}(h);k) \\ &\geq TCB^*_i(k) \end{aligned}$$

where the next to the last inequality is a direct consequence of Lemma 5.5 and the last inequality follows from the optimality of $\underline{s}_{ij}(k)$.

Theorem 5.7: For all $s_{i0} = 0,1,2,...$

$$\underline{s}_{i0} \leq s_{i0} + \text{INT}\ \{[TCB^*_i(s_{i0}) - TCB^*_i(\infty)]/C_i\}. \quad (5.6)$$

Proof: Since $\underline{s}_{i0}$ is optimal in the item$_i$ subproblem we have for all $s_{i0} = 0,1,2,...$ that $TC_i(\underline{s}_{i0}) \leq TC_i(s_{i0})$ and therefore,

$$\begin{aligned} \underline{s}_{i0} &\leq s_{i0} + [TCB^*_i(s_{i0}) - TCB^*_i(\underline{s}_{i0})]/C_i \\ &\leq s_{i0} + [TCB^*_i(s_{i0}) - TCB^*_i(\infty)]/C_i \end{aligned}$$

by Theorem 5.6. Since $\underline{s}_{i0}$ must be an integer, the theorem follows. //

In particular, applying Theorem 5.7 with $s_{i0} = 0$ we have that

$$\underline{s}_{i0} < \text{INT}\ \{(TCB^*_i(0) - TCB^*_i(\infty)]/C_i\}. \quad (5.7)$$

$TCB^*_i(\infty)$ is easily calculated since for $s_{i0} =\infty$, $Q_{ij}(t) = 0$, $t > 0$, and therefore, $X_{ij}(t)$, $t > 0$, has a Poisson distribution with mean obtained from (4.40).

We can use Theorem 5.6 to develop an iterative procedure to try to improve the upper bound on $\underline{s}_{i0}$ obtained from (5.7). Let ub_1 be the upper bound obtained from (5.7). For all $s_{i0} \geq \underline{s}_{i0}$, we know that $TCB_i^*(s_{i0}) \leq TCB_i^*(\underline{s}_{i0})$. Therefore, given $ub_m \geq \underline{s}_{i0}$, $m \geq 1$ we can obtain

$$ub_{m+1} = INT\{[TCB_i^*(0) - TCB_i^*(ub_m)]/C_i\} \geq \underline{s}_{i0}.$$

We continue until $ub_{m+1} = ub_m$. Using Theorem 5.6, it is easy to show that this procedure does indeed terminate. Unfortunately, it does not necessarily terminate at $\underline{s}_{i0}$. Call the final upper bound obtained ub_f.

Initially, in order to obtain an optimal solution to the $item_i$ subproblem, we expect to have to calculate $TC_i(s_{i0})$ for all values of s_{i0} from 0 to ub_f. The first incumbent solution to the $item_i$ subproblem is $s_{i0} = 0$ and $\underline{s}_{ij}(0)$, $j = 1,\ldots,N$, with objective function value $TCB_i^*(0)$. Corollary 5.8 shows how the upper bound on $\underline{s}_{i0}$ can be updated every time a new incumbent solution is obtained.

<u>Corollary 5.8</u>: Let ub_c be the upper bound on $\underline{s}_{i0}$ based on the current incumbent solution with depot asset level $s_{i0}^1 \geq 0$. Let $s_{i0}^2 > s_{i0}^1$ be such that $ub_c \geq s_{i0}^2$ and $TC_i(s_{i0}^2) < TC_i(s_{i0}^1)$. Then,

$$\underline{s}_{i0} \leq s_{i0}^2 + INT\{[TCB_i^*(s_{i0}^2) - TCB_i^*(ub_c)]/C_i\} \leq ub_c.$$

<u>Proof</u>: The first inequality follows, as in the proof of Theorem 5.7, from the optimality of $\underline{s}_{i0}$ and from Theorem 5.6 (since $ub_c \geq \underline{s}_{i0}$). The second inequality is obtained in a similar manner after using the fact that $TC_i(s_{i0}^2) < TC_i(s_{i0}^1)$ and the fact that from Theorem 5.6, $TCB_i^*(ub_c)$ provides the best available lower bound on $TCB_i^*(\underline{s}_{i0})$. //

Originally, ub_f is obtained either from (5.7) or from applying the iterative procedure to improve (5.7). Since $s_{i0} = 0$ is the first incumbent solution, $ub_c = ub_f$. When the next incumbent solution is obtained for some $s_{i0} > 0$, ub_c is updated using Corollary 5.8. In principle, an iterative procedure similar to the one described above can be employed every time ub_c is updated. Computational experience indicates, however, that very little improvement in the upper bound on $\underline{s}_{i0}$ is obtained by applying the iterative procedure every time a new incumbent solution is obtained. The computational burden of applying the iterative procedure appears to outweigh the benefits obtained. Most of the improvement in the upper bound on $\underline{s}_{i0}$ comes about from using Corollary 5.8 without an iterative procedure. In fact, Corollary 5.8 (without the iterative procedure) is so effective in improving ub_c that we have found that it is not worthwhile even to apply the iterative procedure to improve the initial upper bound obtained from (5.7).

It is also possible to establish a "static" upper bound on $\underline{s}_{i0}$.

<u>Theorem 5.9</u>: Let ub_0 be the optimal solution to

$$\underset{s_{i0}=0,1,\ldots}{\text{Min}} \quad C_i s_{i0} + \sum_{j=1}^{N} K_j(0{:}s_{i0}). \tag{5.8}$$

Then, $\underline{s}_{i0} \leq ub_0$.

<u>Proof</u>: Since ub_0 is optimal in (5.8) we have for $w = 1,2,\ldots$

$$C_i w \geq \sum_{j=1}^{N} [K_j(0;ub_0) - K_j(0;ub_0+w)]$$

$$\geq \sum_{j=1}^{N} [K(\underline{s}_{ij}(ub_0+w);\ ub_0) - K(\underline{s}_{ij}(ub_0+w);\ ub_0+w)]$$

from Theorem 2.2 and the fact that $X_{ij}(t)$, $t > 0$, is stochastically decreasing with respect to s_{i0}. Since $K(\underline{s}_{ij}(ub_0+w);ub_0) \geq K(\underline{s}_{ij}(ub_0);ub_0)$, we have that $TC_i(ub_0+w) \geq TC_i(ub_0)$ and the theorem follows. //

We can further reduce the computational burden of obtaining the optimal solution to the item_i subproblem. Theorem 5.10 shows that no base asset level will increase when the depot asset level increases.

Theorem 5.10: $\underline{s}_{ij}(s_{i0})$ is a non-increasing function of $s_{i0} = 0,1,2,\ldots.$

Proof: For each base_j and $t \geq 0$, $X_{ij}(t)$ is stochastically decreasing with respect to s_{i0} (see the proof of Lemma 5.5). Along with the optimality conditions (5.4) and (5.5), this implies that $\underline{s}_{ij}(s_{i0}) \geq \underline{s}_{ij}(s_{i0}+1)$. //

Summarizing the above, we present Algorithm A1, an implicit enumeration scheme for finding optimal solutions to the item_i subproblem.

Algorithm A1: To find an optimal solution to the item_i subproblem:

Step 1: Use (5.4) and (5.5) to obtain $TCB_i^*(\infty)$.

Step 2: Use (5.4) and (5.5) to obtain $TCB_i^*(0)$.
Set $TCB_i^*(0)$ as the value of the incumbent solution and store the depot and base asset levels.

Step 3: Use (5.7) and Theorem (5.9) to find UB, an upper bound on $\underline{s}_{i0}$. Set $j = 1$.

Step 4: If $j > UB$, stop. The incumbent solution is optimal. Otherwise, use (5.4) and (5.5) to obtain $TCB_i^*(j)$.

Step 5: If $jC_i + TCB_i^*(j) <$ the value of the incumbent solution then:
(a) Set $jC_i + TCB_i^*(j)$ as the value of the incumbent solution.
(b) Store the depot and base asset levels.
(c) Update UB using Corollary 5.8.

Step 6: Set $j = j + 1$ and go to Step 4.

After using Algorithm A1 to solve every item subproblem, we obtain $\underline{s}_{++}$, the asset vector that is the optimal solution to the relaxation of (5.3). If $MEBO[\underline{s}_{++}] \leq meb$, $\underline{s}_{++}$ is optimal in (5.3) and is also an undominated solution of (5.2). In order to complete our discussion of the solution of the FAVP, let us assume that $MEBO[\underline{s}_{++}] > meb$.

We can introduce another GLM, $u_m \geq 0$, in order to bring the constraint in (5.3) into the objective function. By a theorem of Everett (1963), the solution to

$$\text{Min} \sum_{i=1}^{I} \{C_i s_{i0} + \sum_{j=1}^{N} K(s_{ij};s_{i0})\} + u_m MEBO(s_{++}) \tag{5.9}$$

is an undominated solution of (5.3). Empirical evidence from tests on the weapon systems in Appendix B indicate that $MEBO[s_{++}]$ is not a convex function either of s_{++} or of total system assets. We can show (see Theorem 5.11 below) that for fixed s_{+0}, $MEBO[s_{++}]$ is a convex decreasing function of $(s_{+1},\ldots,s_{+N})$. The fact that (5.9) is not separable by item, though, makes it extremely difficult and impractical to obtain solutions to (5.9). Therefore, we approximate $MEBO[s_{++}]$ by

$$MEBO(s_{++}) \simeq \sum_{i=1}^{I} MEBO(s_{i+})$$

$$MEBO_i(s_{i+}) = \max_{0<t\leq T} \{ \sum_{j=1}^{N} EBO(s_{ij},s_{i0},t).\} \tag{5.10}$$

Let $0 < TM_i(s_{i+}) \leq T$ be the time at which the expected total number of base backorders for item_i reaches its maximum value, $MEBO_i[s_{i+}]$. If $TM_i(s_{i+})$ were the same for all items, (5.10) would be an exact equality. (5.10) is conservative in the sense that it may result in overstocking, but will ensure that the performance constraints are satisfied.

Using the approximation (5.10), (5.9) is separable by item and the item_i subproblem is

$$\min_{s_{i0}=0,1,2,\ldots} TCM_i(s_{i0}) = C_i s_{i0} + \sum_{j=1}^{N} K(s_{ij};s_{i0}) + u_m MEBO_i(s_{i+})$$

Before solving the item_i subproblem, we need some intermediate results. Theorem 5.11 establishes the convexity of $\text{MEBO}[s_{i+}]$ for fixed s_{i0} while Theorems 5.12 and 5.13 establish, for fixed s_{i0}, upper and lower bounds on the optimal base asset levels.

<u>Theorem 5.11</u>: For fixed s_{i0}, $\text{MEBO}_i(s_{i+})$ is a discretely convex decreasing function of $(s_{i1}, s_{i2}, \ldots, s_{iN})$.

<u>Proof</u>: In the proof of Theorem 5.3 we saw that for all $t > 0$ and fixed s_{i0}, $\text{EBO}(s_{ij}, s_{i0}, t)$ is a discretely convex decreasing function of s_{ij}, $j = 1, \ldots, N$. Therefore, for all $t > 0$, the sum over all bases of $\text{EBO}(s_{ij}, s_0, t)$ is a discretely convex decreasing function of the total base asset vector $(s_{i1}, s_{i2}, \ldots, s_{iN})$. Since the maximum of convex functions is itself a convex function, the theorem follows. //

Let $s^*_{ij}(s_{i0})$ be the optimal base_j asset level for item_i when the depot asset level for item_i is s_{i0}.

<u>Theorem 5.12</u>: For $j = 1, \ldots, N$, $s^*_{ij}(s_{i0}) \geq \underline{s}_{ij}(s_{i0})$.

<u>Proof</u>: Assume there is at least one base_k such that $s^*_{ik}(s_{i0}) < \underline{s}_{ik}(s_{i0})$. Let y_{i+} be an asset vector such that $y_{i0} = s_{i0}$ and for $j \neq 0$,

$$y_{ij} = s^*_{ij}(s_{i0}) \text{ if } s^*_{ij}(s_{i0}) \geq \underline{s}_{ij}(s_{i0})$$

$$y_{ij} = \underline{s}_{ij}(s_{i0}) \text{ otherwise.}$$

Clearly, $\text{MEBO}[y_{i+}] \leq \text{MEBO}[s^*_{i+}(s_{i0})]$. From the definition of $\underline{s}_{ij}(s_{i0})$, we have that $K(y_{ij}; s_{i0}) \leq K(s^*_{ij}(s_{i0}); s_{i0})$ for all j with strict inequality holding for at least base_k. Hence, y_{++} is a better solution to the item_i subproblem for fixed s_{i0} than the asset vector $(s_{i0}, s^*_{i+}(s_{i0}))$. This contradicts the optimality of $s^*_{i+}(s_{i0})$. //

Theorem 5.12 assures us that we can use Algorithm A1 not only to solve the relaxation of (5.3) but also to provide starting values in the search for $s^*_{ij}(s_{i0})$, $j = 1, \ldots, N$.

<u>Theorem 5.13</u>: Let $ub_{ij}(s_{i0}) \geq \underline{s}_{ij}(s_{i0})$ be the smallest non-negative integer such that for all $0 < t \leq T$,

$$u_m[EBO(ub_{ij}(s_{i0}),s_{i0},t) - EBO(ub_{ij}(s_{i0})+1,s_{i0},t)] < [K(ub_{ij}(s_{i0})+1;s_{i0}) - K(ub_{ij}(s_{i0});s_{i0})].$$

Then, $s^*_{ij}(s_{i0}) \leq ub_{ij}(s_{i0})$, $j = 1,\ldots,N$.

<u>Proof</u>: The theorem follows directly from (5.10), Theorem 5.11 (and its proof) and the convexity of $K(s_{ij};s_{i0})$ for fixed s_{i0}.

For any asset vector s_{i+} we define the marginal benefit of putting an extra asset at base$_j$ as

$$MB_{ij}(s_{i+}) = u_m[MEBO_i(s_{i+}) - MEBO_i(s_{i+}+e_j)] - [K(s_{ij}+1;s_{i0}) - K(s_{ij};s_{i0})]$$

where e_j is the j^{th} unit vector in R^N. It is interesting to note that for base$_j$, $MB_{ij}(s_{i+})$ is not monotonically decreasing with respect to s_{i+}. The benefit of an extra asset at base$_j$ may actually increase as more stock is placed at the other bases. In fact, it is quite conceivable that for a particular asset vector, base$_j$ could be "blocked" in the sense that $MB_{ij}(s_{i+}) < 0$ even though $s_{ij} < ub_{ij}(s_{i0})$. Precisely, base$_j$ is blocked if there is some t^*, $0 < t^* \leq T$, such that

$$u_m[EBO(s_{ij}(s_{i0}),s_{i0},t^*) - EBO(s_{ij}(s_{i0})+1,s_{i0},t^*)] \geq [K(s_{ij}(s_{i0})+1;s_{i0}) - K(s_{ij}(s_{i0});s_{i0})]$$

but $MB_{ij}(s_{i+}(s_{i0})) < 0$. Base$_j$ can be blocked in many ways. For example, say there is another base with no assets but with an extremely high pipeline at a time at which base$_j$ has a very small pipeline. Giving assets to base$_j$ without giving assets to the other base would be pointless since the maximum could not be reduced. However, after allocating some assets to this other base, base$_j$ may be "unblocked". If not for blocking, we know that $s^*_{ij}(s_{i0}) = ub_{ij}(s_{i0})$, $j = 1,\ldots,N$. Blocking can not occur when, for all possible asset vectors, all bases attain their maximum expected number of $item_i$ backorders at the same time.

Using Theorem 5.11 and the bounds in Theorems 5.12 and 5.13, we could construct a branch and bound algorithm for finding $s^*_{ij}(s_{i0})$, $j = 1,\ldots,N$. However, since we may have to find $s^*_{ij}(s_{i0})$, $j = 1,\ldots,N$ for many values of s_{i0} we attempt to reduce the computational burden by using a greedy heuristic.

<u>Algorithm A2</u>: A greedy heuristic for finding $s^*_{ij}(s_{i0})$, $j = 1,\ldots,N$:

<u>Step 1</u>: Set the current asset vector to $\underline{s}_{i+}(s_{i0})$.

<u>Step 2</u>: For the current asset vector s^c_{i+}, the set of eligible bases consists of each $base_j$ for which $s^c_{ij} < ub_{ij}(s_{i0})$. If the set of eligible bases is empty, stop. $ub_{i+}(s_{i0})$ is the optimal solution.

<u>Step 3</u>: Find $MB_{ij}(s^c_{i+})$ for each eligible $base_j$. If $MB_{ij}(s^c_{i+}) < 0$ for each eligible base, stop. Set $s^*_{i+} = s^c_{i+}$. Otherwise, increase by 1 the asset level of the base with the largest value of $MB_{ij}(s^c_{i+})$, $j = 1,\ldots,N$.

<u>Step 4</u>: Update the current asset vector and go to step 2.

We note that if there is no blocking, the greedy heuristic will not stop until $s^*_{ij}(s_{i0}) = ub_{ij}(s_{i0})$ for each base. The greedy heuristic may not find the true optimal solution if the greedy allocation induces blocking that might not otherwise have occurred.

We can now use an implicit enumeration scheme to find the optimal solution of the $item_i$ subproblem. Let

$$TCMB^*_i(s_{i0}) = \min \{ \sum_{j=1}^{N} K(s_{ij};s_{i0}) + u_m MEBO[s_{i+}]\}$$

be the optimal contribution by the bases to $TCM_i(s_{i0})$. Similar to the proof of Theorem 5.6, we can prove that $TCMB_i^*(s_{i0})$ is a decreasing function of s_{i0}. Therefore, we can establish analogs of Theorem 5.7, the bound (5.7), Corollary 5.8 and Theorem 5.9. No analog of Theorem 5.10 is possible since adding an asset to the depot may unblock a base_j. If that happens, $s_{ij}^*(s_{i0})$ will increase.

An implicit enumeration scheme very similar to Algorithm A1 can now be used for each item in order to find s_{++}^*, an undominated solution of (5.3). We leave the details to the reader. s_{++}^* is also an undominated solution of (5.2). Using standard techniques, we can now search for the Generalized Lagrange Multipliers that produce an asset vector(s) whose inventory performance is acceptably close to the targets aeb and meb.

Summary

The FAVP assumes that management will not and/or cannot change the asset vector over the horizon. The problem then is to find the least cost asset vector at time 0 that achieves management specified targets on the average and maximum expected total number of base backorders over the horizon. The FAVP (5.2) is a non-linear non-convex integer programming problem. By using Generalized Lagrange Multipliers we were able to show that the FAVP is separable by item. We developed an efficient implicit enumeration scheme to find the optimal levels for each item and thereby find an undominated solution of the FAVP. In the next two chapters we discuss some of the computational issues involved in using the FAVP to obtain cost effective asset vectors in non-stationary two echelon systems.

CHAPTER VI

APPROXIMATIONS TO THE BASE PIPELINE DISTRIBUTION

In using Algorithms A1 and A2 to solve the FAVP we must find the average and maximum expected total number of backorders at the bases for many different candidate asset policies. Therefore, for each item we must find the time dependent distributions of the base pipelines for various depot asset levels. We see from (4.40) that for a particular item the distribution of $X_j(t)$, $t > 0$, can be obtained from the convolution of a Poisson distribution and the distribution of $Q_j(t)$ obtained from (4.30) and (4.32). Most of the effort and cost incurred in obtaining the distribution of $X_j(t)$ lies in obtaining the distribution of $Q_j(t)$. To evaluate (4.30) and (4.32) we must perform many numerical integrations of a function that is the product of probability terms that contain MVF. Depending on the behavior of this function in $(0,t]$ it may be necessary to make many function evaluations in order to obtain accurate results from a numerical integration routine (Conte and de Boor [1980]). Furthermore, the MVF themselves may have to be numerically integrated. Therefore, the evaluation of (4.30) and (4.32) at many times during the horizon can be extremely time consuming and costly. Furthermore, it can also be time consuming to perform the actual convolution of the components of $X_j(t)$ for many different times during the horizon and many different depot asset levels. For these reasons, in this chapter we study approximations to the distribution of $X_j(t)$, $t > 0$, that reduce the burden and cost of solving the FAVP.

We obtained data on three current Army weapon systems in order to form a realistic data base for the tests we report upon in this chapter and in Chapter VII. A summary of the data for each weapon system can be found in Appendix B. Rather than test an approximation on an item by item basis, we felt it was more useful to test the approximation over representative catalogs of items. As long as an approximation consistently performs well over representative catalogs, the approximation provides a valuable tool even when one can find particular items and/or points in time during the horizon at which it fails.

All run times are for the CDC CYBER 700 with a cost of $800 per CPU hour. The weapon system used in a particular run will be identified by the abbreviations used in Appendix B. All numerical integrations were done using the International Mathematical and Statistical Libraries subroutine DCADRE (IMSL [1979]) which uses adaptive Romberg integration (Conte and de Boor [1980]). The upper bound on relative error was 10^{-8}. (Run times and costs did not significantly change when this bound was lowered to 10^{-6}). Finally, for consistency among the different weapon systems, the targets aeb and meb in (5.2) were expressed as a percentage of the number of each type of weapon system deployed. Therefore, a 10% aeb target actually implied a target of 32.7 backorders for the average expected total backorders on the AAH, 10 on the BHAWK, and 25 on the M60A3.

We studied two approximations to the distribution of $X_j(t)$, $j = 1,2,\ldots,N$ for any $t > 0$. The first one, which we call NEGBI, approximates the distribution of $X_j(t)$ with a negative binomial distribution with mean $E[X_j(t)]$ and variance $VAR[X_j(t)]$ obtained from (4.40). A priori, there are two reasons to believe that NEGBI may be a satisfactory approximation. First, as we mentioned in Chapter IV, empirical evidence from runs on the data base indicated that the probability mass function (pmf) of $X_j(t)$ was unimodal and that $VMR[X_j(t)] \geq 1$. The pmf of a negative binomial distribution is unimodal. The negative binomial is also a two parameter distribution with $VMR \geq 1$ so we can ensure that NEGBI uses the correct mean and variance and, hence, VMR. Secondly, NEGBI has been shown to be an excellent approximation in stationary systems (Slay [1980], Graves [1983]). The limiting distribution of $Q_j(t)$ in the stationary case resembles (4.6). Therefore, there is strong evidence to believe that at least for the proportionate base case, NEGBI will be an excellent approximation.

The second approximation, called POISSON, approximates the distribution of $X_j(t)$ with a Poisson distribution with mean $E[X_j(t)]$. Based on our empirical evidence, this underestimates the variance of $X_j(t)$. The advantage that POISSON has over NEGBI is that we do not have to calculate $VAR[X_j(t)]$. Therefore, we would normally expect that POISSON would run in less than one-half the time of NEGBI. However, from (4.35) we see that the integrands for $E[Q_j(t)]$ and $E[Q_j^2(t)]$ are very similar. We used DCADRE (when necessary) to compute $E[Q_j(t)]$ and stored the points at which the integrand was evaluated, along with the terms that were common to the integrand of $E[Q_j^2(t)]$. We then used the trapezoid rule (Kitchen [1968]) with the points and values saved from the computation of $E[Q_j(t)]$. This significantly reduced the time needed to compute $E[Q_j^2(t)]$. As we shall see, POISSON was, on average, approximately 15% faster than NEGBI. Although this approach does not guarantee that the accuracy achieved in calculating $E[Q_j(t)]$ will be matched in the calculation of $E[Q_j^2(t)]$, our tests indicated very little loss in accuracy. In fact the same approach was used to reduce the run time for obtaining the entire exact distribution of $X_j(t)$.

Table 1 is a comparison of the exact average expected total number of backorders at the bases with the projections made by NEGBI and POISSON. The FAVP was run using each approximation to determine asset policies for three target aeb percentages (meb was set to infinity). The final asset vector for each approximation supposedly achieved the target aeb percentage. These final asset vectors were then correctly evaluated using the exact distributions of the base pipelines. This yielded the true backorder ratio for those asset vectors where the backorder ratio is defined as the average expected total number of backorders at the bases divided by the total number of weapon systems deployed.

From Table 1 we see that NEGBI does significantly better than POISSON in projecting the true backorder ratio. In fact, using NEGBI instead of the exact distributions resulted in very little loss of accuracy. Notice that in all cases NEGBI and POISSON underestimated the backorder ratio. This is consistent with observations made by Slay (1980) and Graves (1983) for stationary systems. Graves (1983) reports on rare instances where NEGBI overstated the backorder ratio in a stationary system. We did not observe such aberrant behavior on any of the items in our data base. We also note NEGBI performed best on the M60A3 which has a support structure consisting of proportionate bases.

Ultimately, we are concerned not with the error in projecting backorder ratios but with the cost of the assets required to achieve a target backorder ratio. Table 2 is a comparison of the inventory

Table 1: Exact Backorder Ratios For Approximate FAVP Solutions.

System: AAH

TARGET	NEGBI	POISSON
15%	15.238%	16.64%
10%	10.209%	11.31%
5%	5.146%	6.98%
1%	1.040%	1.42%

System: BHAWK

TARGET	NEGBI	POISSON
15%	15.573%	20.53%
10%	10.612%	13.51%
5%	5.406%	8.64%
1%	1.094%	2.60%

System: M60A3

TARGET	NEGBI	POISSON
15%	15.1381%	15.94%
10%	10.0834%	11.12%
5%	5.0442%	6.58%
1%	1.0097%	1.39%

investment needed to achieve various target backorder ratios when assets are determined using the FAVP with either the exact, NEGBI or POISSON evaluation methods. In constructing Table 2 we proceeded as follows. First, for each target backorder ratio we ran the FAVP with the exact base pipeline calculations in order to obtain the true optimal asset vector and true optimal inventory investment. Then, we ran the FAVP with each of the approximations over a range of values for the Generalized Lagrange Multiplier. We evaluated the asset vectors so obtained using the exact base pipeline distributions. We then searched among these final asset vectors (using the exact evaluations) for the asset vector that achieved the target backorder ratio. For example, using POISSON it cost $11.52 million to achieve a target backorder ratio of 5% for the AAH. POISSON actually projected a lower backorder ratio for that money. However, when the asset vector was evaluated using the exact base pipeline distributions, a 5% backorder ratio was achieved.

In summary, we note that NEGBI performed very well with little loss in accuracy or increase in inventory investment when compared with the exact solution. NEGBI always performed considerably better than POISSON with only a small increase in run times.

As expected, these results were duplicated when we activated the constraint in (5.2) on the maximum expected total number of backorders at the bases by setting meb to a finite value. NEGBI provides such a good approximation to the base pipeline distributions that there was little loss in accuracy in using NEGBI to project both the average and maximum expected total number of backorders at the bases.

Once we accept the fact that the negative binomial is a satisfactory approximation to a base's pipeline distribution, it seems

Table 2: Inventory Cost Comparisons (Millions $)

System: AAH

TARGET	NEGBI	POISSON	Exact
10%	10.17	10.43	10.17
5%	11.12	11.52	11.08
1%	12.62	13.34	12.51
(RUN TIME (sec):	274	240	2327)

System: BHAWK

TARGET	NEGBI	POISSON	Exact
10%	18.43	18.53	18.43
5%	21.37	21.41	21.35
1%	27.66	27.71	27.56
(RUN TIME (sec):	118	89	1123)

System: M60A3

TARGET	NEGBI	POISSON	Exact
10%	25.89	26.57	25.84
5%	33.62	34.30	33.41
1%	40.91	43.05	40.74
(RUN TIME (sec):	337	301	1612)

that further effort toward reducing run times should be directed toward approximating $E[X_j(t)]$ and $VAR[X_j(t)]$ and avoiding the cumbersome expressions in (4.33) for $E[Q_j(t)]$ and $E[Q_j^2(t)]$. Approximations that significantly reduce the number of numerical integrations that need to be performed will significantly reduce the run times of the FAVP. This, in our opinion, is the next logical extension to the research and results presented in this chapter.

CHAPTER VII

COMPUTATIONAL EXPERIENCE

In this chapter we briefly discuss the sensitivity of the FAVP solutions to changes in the input data. In doing so, we shall briefly discuss the efficacy of using the more convenient stationary models to approximate the FAVP results.

Most of the parameters of operation (PO) that are input to the FAVP are used to determine the time-dependent distributions of the pipelines at the depot and bases. The demand intensities at the bases stand out as the most important input elements since they drive the distributions of the bases' pipelines. Changes in the order and ship times, base repair times and the inventory system's maintenance concept also impact upon the base pipelines. Changes in the depot input parameters impact directly on the depot pipeline and indirectly on the base pipelines through the distribution of $Q_j(t)$, $t \geq 0$.

Table 3 illustrates the effect on total inventory investment for the AAH helicopter when the demand intensity at each base for each item was multiplied by a common scaling factor but the weapon system's performance target was not changed. As we see, the change in inventory investment was far less than proportional to the change in scaling factor. Furthermore, for bigger changes in scaling factor, the change in inventory investment seemed less sensitive to the change in the bases' pipelines. This implies that if a 100% increase in the bases′ pipelines causes a 50% increase in inventory investment, an additional 100% increase in the bases′ pipelines will result in less than an additional 50% increase in inventory investment.

Table 3

Impact of Changing Intensities on Inventory Investment for the AAH

Scaling Factor	Inventory investment (Millions $)
.5	2.37
1	3.89
2	5.20
4	7.33
8	8.87

In Table 4, we again changed the intensities of the demand for items on the AAH by a scaling factor, but this time we held all asset levels fixed at the values obtained when the scaling factor was 1. Note that the decrement in backorder ratio (Chapter VI) appears to be more than proportional to the change in scaling factor but seems to be less sensitive to larger changes in the scaling factor.

The FAVP selects the least cost asset vector that achieves the performance targets. Unit prices have absolutely no impact upon the pipeline calculations for a particular asset vector. By examining Algorithms A1 and A2, we can see that if every unit price changed by the same percentage, the FAVP would produce exactly the same stock list to meet the performance targets. This is encouraging as economic tradeoffs should be based on relative and not absolute costs.

Table 4

Impact of Changing Intensities on Backorder Ratios for the AAH

Scaling Factor	Backorder Ratio
.5	2.38%
1	5.01%
2	10.45%
4	21.32%
8	43.02%

If only some unit prices change, the FAVP will try to substitute cheaper items for the more expensive ones.

There are economies of scale to be had by consolidating bases. Two bases each with a deployment of 50 AAH helicopters will generally require a higher total inventory investment to achieve the same performance as one base with 100 helicopters. This is a special case of the results in Table 3.

Finally, the FAVP results can be extremely sensitive to the initial system condition and the length of the scenario. Table 5 shows the budget requirements in order to achieve approximately a 5% backorder ratio for the AAH for various scenario lengths. For this run the AAH usage modifiers (Appendix B) were set to 1 so that the system would eventually reach steady state. All assets were assumed on-hand and ready for issue at time 0.

We note from Table 5 that the steady state budget was considerably more than the $2.42 million required for the original 30 day scenario. This was simply due to the fact that stationary models ignore the initial conditions and horizon length. We could have changed the

Table 5

AAH Budgets for Various Scenario Lengths

Scenario Length (Days)	Budget (Millions $)
5	1.04
10	1.77
30	2.42
60	4.61
Steady State	6.97

(Run time: 382 sec)

initial system condition so that the stationary model would understate the budget requirements. It is clear from Table 5 that one must proceed cautiously when using stationary models to approximate solutions of non-stationary models.

There are, of course, many heuristics one could try to improve upon the results of the stationary models. Table 6 shows the results of one such heuristic. We used SESAME (U.S. Army [1983]) to optimize

Table 6

Evaluation of Stationary Model's Recommended Stock

Scenario Length (Days)	Budget (Millions)
5	1.39
10	2.08
30	2.96
60	5.34
Steady State	6.97

(Run time: 23 sec)

but evaluated the final solutions using the FAVP methodology. The SESAME search routine with the FAVP evaluator found the budget required to achieve approximately a 5% backorder rate for various horizon lengths. Note that run times were reduced by a factor of 18.

Once the usage modifier was set to 1, the AAH had no non-stationary PO. It is certainly worthwhile to investigate further the efficacy of using a stationary model to approximate the FAVP results when all the PO are non-stationary. (We made no attempt to do so). This would require adjustment of the data (possibly averaging) to meet the input capabilities of the stationary models. Besides run time savings, stationary models are also convenient in that they require less data on the behavior of the PO than the FAVP does. Being able to approximate FAVP results without the associated data collection and input effort is a major reason for using stationary models to approximate the results of non-stationary models. This is one area we feel deserves a considerable amount of further research and study. As multi-echelon systems grow both in the number of locations and the number of items, judicious use of stationary models (including, possibly, coordinated single echelon policies) may be the only practical way to obtain cost effective operating policies.

Next page is blank.

CHAPTER VIII

THE ASSET VECTOR TRANSFORMATION PROBLEM

The two fundamental questions that arise in controlling inventories are when and how much to order. The FAVP answered these questions by determining the initial asset vector at time 0 and by following a strict one-for-one resupply policy thereafter. This resulted in a constant asset vector and, consequently, constant total system stock over the horizon. For many non-stationary systems it is infeasible to change the asset vector during the horizon. This could be due to any and all of the following:

a. A short time horizon which prevents external procurement and expeditious redistribution of assets.

b. Lack of asset visibility and control (e.g., a combat unit).

c. High cost of effecting desired changes in the asset vector during the horizon.

For these systems, the FAVP is a satisfactory method for determining an optimal asset policy. However, in many non-stationary systems we might reasonably expect that not only may the optimal asset vector change over time but also that the total system stock may change during the horizon. In order to accomplish these changes, management will intervene in the normal one-for-one resupply operations of the system by directing procurements, disposals and/or redistributions of assets.

These management interventions along with the initial asset vector and the one-for-one resupply policy, completely answer the when and how much to order questions for non-stationary inventory systems. The methods and costs involved in effecting these management decisions play an important part in determining an optimal operating policy for the system.

In this chapter we explore some of the issues involved in characterizing and specifying an optimal inventory control policy for non-stationary systems. We introduce the Asset Vector Transformation Problem (AVTP), a stochastic transshipment problem which begins to tackle the question of if, when, and how management should intervene to effect asset vector changes during the horizon. The AVTP and the FAVP together form a tool for analysis and determination of operating policies for non-stationary multi-echelon inventory systems.

Characterization of Optimal Operating Policies

An operating policy must specify: an asset policy; an ordered set of management intervention times, MT, at which procurement, disposal and/or redistribution decisions are to be made; and the method for accomplishing any planned procurements, disposals, and redistributions. Between times in MT, the system follows normal one-for-one resupply rules. At times in MT, disposals and new procurements always result in changes in the asset vector and the system "condition" (see Chapter I). Redistributions may change the asset vector but they always change the system condition. The choice of an asset vector for intervals between management intervention times should therefore reflect the changes in system condition brought about by management decisions. However, the choice of an asset vector for each interval should also reflect the cost involved in transforming one asset vector to another. These costs can be significant when scenarios dictate extensive management directed procurements, disposals, and redistributions.

We wish to choose an operating policy from among all feasible operating policies so that operating costs (costs to review, procure, dispose, hold, and redistribute) are minimized subject to constraints on the average and worst performance over the horizon.

Define

$h_i(t_1,t_2)$ = holding cost for a unit of item$_i$ between times t_1 and t_2;

$R(t)$ = cost of a management review at time t;

M = cardinality of MT;

$MC(s_{++}(t^-),s_{++}(t),t)$ = expected cost of effecting a management decision at t which will result in a change in the asset vector from $s_{++}(t^-)$ to $s_{++}(t^+)$ and/or in a change in the system condition;

$s_{++}(\cdot)$ = an asset policy;

$AEBO[s_{++}(\cdot)]$ = the average expected total number of backorders at the bases over the horizon given the asset policy $s_{++}(\cdot)$;

$MEBO[s_{++}(\cdot)]$ = maximum value over the horizon of the expected total number of backorders at the bases given the asset policy $s_{++}(\cdot)$;

$t_0 = 0$.

Then, the problem of determining an optimal operating policy can be formulated as finding MT and $s_{++}(\cdot)$ that solve:

$$\begin{aligned}
\text{Min} \sum_{m=1}^{M} \{ \sum_{i=1}^{I} & h_i(t_{m-1},t_m)[s_{i+}(t_{m-1}) \cdot 1_N] \\
& + R(t_m) + MC(s_{++}(t_{m-1}),\ s_{++}(t_m),t_m)\} \\
& + \sum_{i=1}^{I} h_i(t_M,T)[s_{i+}(t_M) \cdot 1_N] \\
& AEBO[s_{++}(\cdot)] \leq aeb \qquad\qquad (8.1) \\
& MEBO[s_{++}(\cdot)] \leq meb \\
& t_1 \leq t_2 \leq \cdots \leq t_M \\
& s_{++}(t) \geq 0 \text{ and integer for all } t.
\end{aligned}$$

The FAVP (5.2) is a special case of (8.1) obtained by assuming either $R(t)$ or $MC(s_{++}(t^-),s_{++}(t),t)$ is infinite for all $t > 0$. This implies that it is infeasible for management to intervene during the horizon. Therefore, M is zero, $h_i(0,T)$ is simply C_i and there is no need to worry about finding the most cost effective ways to implement management decisions. It is important to note that the solution to the FAVP (5.2), along with setting MT equal to the null set, is a feasible solution to (8.1).

(8.1) is solved at time 0 to determine the optimal operating policy for a system using an HCP. MT is therefore fixed at time 0 and operating policies may be expressed as functions of the system condition at the times in MT. The system condition at any future time can only be described stochastically. The cost of effecting any management decisions at times in MT clearly depends upon the system condition at these times. Therefore, this cost is a random variable and is included in the objective function via its expected value.

Under a RTCP we have complete knowledge of the system condition at the current decision time. However, (8.1) must still be solved at each decision time since our current decision may depend upon what we expect to do in the future. Thus, under a RTCP, MT and operating policies may be changed each time (8.1) is solved. The cost of effecting any management decision at the current decision time is deterministic but the cost of any future decisions is again a random variable.

The optimization problem (8.1) is very difficult to solve. Later, we shall briefly discuss a heuristic for solving (8.1) when MT is fixed at time 0. Before doing so, however, we need to introduce the AVTP, a stochastic transshipment model that determines $MC(s_{++}(t^-), s_{++}(t), t)$ for any candidate time t and any candidate asset policy.

Formulation of the AVTP

Let t_m, $1 \leq m \leq M$, be a management intervention time in MT and consider the problem of transforming the system asset vector from $s_{++}(t_{m-1})$ to $s_{++}(t_m)$. Let $AS_i(t) = s_{i+}(t) \cdot 1_N$ and define for each item $i = 1,2,..,I$,

$$DC_i(t_{m-1},t_m) = \{k;\ 0 \leq k \leq N,\ s_{ik}(t_{m-1}) > s_{ik}(t_m)\}$$

$$IN_i(t_{m-1},t_m) = \{k;\ 0 \leq k \leq N,\ s_{ik}(t_{m-1}) < s_{ik}(t_m)\}$$

$$EQ_i(t_{m-1},t_m) = \{k;\ 0 \leq k \leq N,\ s_{ik}(t_{m-1}) = s_{ik}(t_m)\}$$

as the set of locations for which the item i asset level decreases, increases or remains the same, respectively. For each item we can identify the following cases:

<u>Case 1a</u>: $AS_i(t_{m-1}) = AS_i(t_m)$; $EQ_i(t_{m-1},t_m) = \{0,1,\ldots,N\}$.

No procurements, disposals or redistributions are necessary unless management wishes to change the system condition by forcing locations either to exchange on-hand assets for on-order assets or to exchange one type of on-order asset for another type of on-order asset.

<u>Case 1b</u>: $AS_i(t_{m-1}) = AS_i(t_m)$; $EQ_i(t_{m-1},t_m) \notin \{0,1,\ldots,N\}$.

Procurements and disposals of item i are necessary only if the system condition at t_m prevents a redistribution of assets to attain $s_{++}(t_m)$.

<u>Case 2a</u>: $AS_i(t_{m-1}) < AS_i(t_m)$; $DC_i(t_{m-1},t_m)$ empty.

Procurements are necessary to raise the asset positions of the locations in $IN_i(t_{m-1},t_m)$ to their target values.

<u>Case 2b</u>: $AS_i(t_{m-1}) < AS_i(t_m)$; $DC_i(t_{m-1},t_m)$ not empty.

Along with external procurements there must be a redistribution of current assets between locations in $DC_i(t_{m-1},t_m)$ and $IN_i(t_{m-1},t_m)$.

<u>Case 3a</u>: $AS_i(t_{m-1}) > AS_i(t_m)$; $IN_i(t_{m-1},t_m)$ empty.

Disposals must be made to lower the asset positions of the locations in $DC_i(t_{m-1},t_m)$ to their target values.

<u>Case 3b</u>: $AS_i(t_{m-1}) > AS_i(t_m)$; $IN_i(t_{m-1},t_m)$ not empty.

Along with disposals, there must be a redistribution of assets between locations in $DC_i(t_{m-1},t_m)$ and $IN_i(t_{m-1},t_m)$.

The depot and $base_j$ may exchange an asset either by direct shipment or by the creation/cancellation of a backorder at the depot belonging to $base_j$. $Base_j$ and $base_k$ may exchange an asset either by direct shipment or by the reassignment to one base of a depot backorder belonging to the other base. (By assumption, all backorders outstanding at a base belong to primary customers and not to other bases). Any redistribution of assets consists of a composition of these basic exchange methods. We see that the ability to change from one asset vector to another depends upon the system condition at the time the change is planned. For example, if base_j is in $DC_i(t_{m-1},t_m)$ and the depot is in $IN_i(t_{m-1},t_m)$ there can be no direct exchange of an asset unless at t_m either base_j has a unit on-hand or there is a depot backorder belonging to base_j. If this is not the case, the only way to attain the new asset vector is by a simultaneous procurement at depot and disposal at base_j. It is highly unlikely this will ever be desirable. Rather, it would probably be better to seek an adjustment to a different asset vector.

Define, for $i = 1,2,..,I$ and $j = 0,1,...,N$:

$OH_{ij}(t)$ = number of units of item_i on-hand at location_j at time t;

$Z_{ij}(t)$ = number of depot backorders of item_i at time t belonging to location_j;

x_{ijk} = number of on-hand units of item_i directly shipped from location_j to location_k;

ds_{ij} = number of disposals of item_i at location_j;

pr_{ij} = number of new procurements of item_i at location_j;

cb_{ij} = number of item_i depot backorders belonging to location_j that are cancelled;

nb_{ij} = number of item_i depot backorders created that belong to location_j;

$CX_{ijk}(x_{ijk})$ = cost of directly shipping x_{ijk} units;

$CCB_{ij}(cb_{ij})$ = cost of cancelling cb_{ij} depot backorders;

$CNB_{ij}(nb_{ij})$ = cost of creating nb_{ij} depot backorders;

$CPR_{ij}(pr_{ij})$ = cost of procuring pr_{ij} units;

$CDS_{ij}(ds_{ij})$ = cost of disposing ds_{ij} units.

The AVTP at time t_m can now be formulated as a stochastic transshipment problem:

$$\text{Min} \sum_{i=1}^{I} \sum_{j=0}^{N} E\{CCB_{ij}(cb_{ij}) + CNB_{ij}(nb_{ij}) + CPR_{ij}(pr_{ij}) + CDS(ds_{ij}) + \sum_{k\neq j} CX_{ijk}(x_{ijk})\} \tag{8.2}$$

subject to:

(1a) For $i = 1,\ldots,I$ and $j = 1,\ldots,N$:

$$\sum_{k\neq j} x_{ijk} + cb_{ij} + ds_{ij} - \sum_{k\neq j} x_{ikj} - nb_{ij} - pr_{ij} = s_{ij}(t_{m-1}) - s_{ij}(t_m);$$

(1b) For $i = 1,\ldots,I$:

$$\sum_{k\neq 0} [x_{i0k} + nb_{ik}] + ds_{i0} - \sum_{k\neq 0} [x_{ik0} + cb_{ik}] - pr_{i0} = s_{i0}(t_{m-1}) - s_{i0}(t_m);$$

(2) $ds_{ij} \leq s_{ij}(t_m) - Z_{ij}(t_m)$ $\quad i = 1,\ldots,I;\ j = 0,1,\ldots,N;$

(3) $\sum_{k\neq j} x_{ijk} \leq OH_{ij}(t_m)$ $\quad i = 1,\ldots,I;\ j = 0,1,\ldots,N;$

(4) $cb_{ij} \leq Z_{ij}(t_m)$ $\quad i = 1,\ldots,I;\ j = 0,1,\ldots,N;$

(5) $\sum_{k\neq j} x_{ijk} \geq a_{ij}$ $\quad i = 1,\ldots,I;\ j = 0,1,\ldots,N;$

(6) $\sum_{k\neq j} x_{ikj} \geq u_{ij}$ $\quad i = 1,\ldots,I;\ j = 0,1,\ldots,N;$

where all variables are non-negative integers.

Constraint (1) is a balance constraint ensuring that the new asset vector is attained. Constraint (2) restricts the disposals at

$location_j$ to the number of units on-hand, in diagnosis, in repair and en route from the depot at t_m. Constraint (3) ensures that no location ships more units than are on-hand while constraint (4) requires that the number of depot backorders belonging to $location_j$ that are cancelled at t_m be less than the number outstanding. Constraints (5) and (6) are additional supply and demand constraints by which management can alter the system condition at t_m by requiring some locations to exchange on-hand assets for on-order assets. These constraints may be employed to change the system condition at t_m even though there may be no desire to change the asset vector. For example, (5) and (6) can be used to require a base in $EQ_i(t_{m-1},t_m)$ to send an on-hand unit to another base that is expected to have many customer demands backordered at t_m. Since the sending base is in $EQ_i(t_{m-1},t_m)$, this base must receive a replacement asset that will, perforce, be on-order. In fact, if management sets $a_{ij} = OH_{ij}(t_m)$, $location_j$ will be required to ship all its on-hand units. Some or all of these units may be replaced by on-order assets in the form of external procurements and/or the creation of new depot backorders belonging to $location_j$.

It is reasonable to assume that for all $n_{ij} \geq 0$,

$$CNB_{ij}(n_{ij}) \geq CX_{i0j}(n_{ij}) \qquad i=1,\ldots I;\ j=1,\ldots N,$$

since creating a depot backorder belonging to $base_j$ will eventually require shipping a serviceable unit from the depot to $base_j$ that otherwise would not have been shipped. Under this assumption the depot will prefer to ship units that are on-hand rather than to create new backorders. This ensures that

$$[\sum_{k\neq 0} n_{ik}][OH_{i0}(t_m) - \sum_{k\neq 0} x_{i0k}] = 0$$

and thereby prevents the depot from having any backorders outstanding while it also has on-hand stock.

The AVTP (8.2) can be augmented with additional constraints that affect the system condition by forcing a location to exchange one type of on-order asset for another type of on-order asset. These constraints can also be used to change the system condition without changing the asset vector. For example, at time t_m a base in $IN_i(t_{m-1},t_m)$ $[DC_i(t_{m-1},t_m)]$ can pass a procurement (disposal) action to a base in $EQ_i(t_{m-1},t_m)$ in exchange for the reassignment of an outstanding depot backorder belonging to the latter (former) base. In this case, the latter (former) base loses (gains) ownership of an outstanding depot backorder in exchange for a procurement (disposal) at that location. Constraints of this type are similar to (5) and (6) in that they are bounds on the number of procurements and disposals at a location.

If $a_{ij} = u_{ij} = 0$, (8.2) always has a feasible solution as $s_{++}(t_m)$ can be achieved via procurements, disposals and the creation and or cancellation of outstanding depot backorders. The AVTP is then a stochastic transshipment problem with full recourse. If either a_{ij} or u_{ij} is positive, (8.2) may not have a feasible solution indicating that the new asset vector cannot be achieved. In this case management must decide on an alternate asset vector until the next management intervention time.

For a discussion of solution techniques for stochastic transshipment problems, the reader is referred to Madansky (1960), Walkup and Wets (1966), Ziemba (1970a, 1970b), Huang et. al. (1977), Wets (1979) and Dempster (1980).

A Heuristic For Determining Optimal Operating Policies

In principle, specifying an optimal operating policy requires decision rules for the selection of an asset vector for every possible realization of the system condition at times in MT. However, these rules are computationally cumbersome to obtain and probably impractical to implement in systems with many items and bases. In this section we propose to restrict the form of the decision rules and to use the FAVP and AVTP to restrict the set of candidate asset policies. The FAVP and AVTP thereby form a basis for a heuristic for determining an effective operating policy.

For expository purposes we assume that MT is fixed at time 0 and consists of a lone element, t^*. The generalization to fixed MT with cardinality larger than one is straightforward. Say we have a set of candidate asset vectors, V_0, with the property that each asset vector in V_0 satisfies the performance constraints in (8.1) during $[0,t^*)$. For each v_0 in V_0 let $V_1(v_0)$ be the set of companion asset vectors for use in $[t^*,T]$. For each $v_1(v_0)$ in $V_1(v_0)$, the asset policy $[v_0,v_1(v_0)]$, along with the AVTP solution for transforming v_0 to $v_1(v_0)$, satisfies the performance constraints in (8.1) during $[0,T]$. We restrict ourselves to operating policies of the following form:

a. At time 0 an asset policy $[v_0^*, v_1^*(v_0^*)]$ is chosen that minimizes holding and expected asset vector transformation costs over the horizon;

b. At t^* the system <u>always</u> switches from v_0^* to $v_1^*(v_0^*)$ regardless of the system condition at t^*.

If $a_{ij} = u_{ij} = 0$ for all i and j, the change at t^* can always be accomplished (with a cost that depends on the actual system condition at t^*). Management can elect to have system condition constraints in the AVTP and still ensure that property b holds by replacing each a_{ij} with Min $[a_{ij}, OH_{ij}]$ and each u_{ij} with

$$\text{Min } [u_{ij}, \sum_{k \neq j} OH_{ik}].$$

A consequence of property b is that at t^* the system asset vector always changes to $v_1^*(v_0^*)$ even though for a particular realization of the system condition at t^*, $v_1^*(v_0^*)$ may no longer satisfy the performance constraints during $[t^*,T]$. For a system under an HCP this is reasonable as $v_1^*(v_0^*)$ was chosen with cognizance of this possibility. In fact, it is quite likely that under an HCP management would not have enough information on the system condition at t^* to realize that $v_1^*(v_0^*)$ is no longer feasible.

Restricting operating policies to the form above reduces the problem to choosing v_0^* in V_0 and $v_1^*(v_0^*)$ in $V_1(v_0^*)$ that minimizes

$$HC_0(v_0) + HC_1(v_1(v_0)) + MC(v_0, v_1(v_0), t^*) \tag{8.3}$$

where $HC_0(v)$ is the holding cost of asset vector v during $[0,t^*)$ and $HC_1(v)$ is the holding cost of asset vector v during $[t^*,T]$. (8.3) is formidable and cumbersome to solve unless the sets of candidate asset vectors are small. The second part of the heuristic uses the FAVP and AVTP to construct small sets of candidate asset vectors.

Say v_0 is fixed in (8.3). Let $f_1(v_0)$ be the solution to the FAVP for $[t^*,T]$ with initial system condition determined by v_0 and the methods for transforming from v_0 to $f_1(v_0)$ (obtained from the AVTP). We note that the asset policy $[v_0, f_1(v_0)]$ is feasible and has cost

$$HC_0(v) + HC_1(f_1(v_0)) + MC(v_0, f_1(v_0), t^*). \tag{8.4}$$

Since $f_1(v_0)$ is the FAVP solution, any other vector in $V_1(v_0)$ must necessarily have higher holding costs in $[t^*,T]$. Therefore, if $[v_0,v_1(v_0)]$ has lower total cost than (8.4) it must be true that

$$MC(v_0,f_1(v_0),t^*) - MC(v_0,v_1(v_0),t^*) \geq HC_1(v_1(v_0)) - HC_1(f_1(v_0)) \quad (8.5)$$

There are, of course, many asset vectors in $V_1(v_0)$ for which (8.5) may hold. However, we restrict attention to vectors $v_1(v_0) \geq f_1(v_0)$ so that holding costs increase since no asset position for any item decreases at any location. This ensures that the asset level for each item at each location is at least as large as the level that would have been obtained from solving the FAVP for $[t^*,T]$ with the proper initial system condition. (The Army (U.S. Army [1983]) uses a similar policy when changing peacetime asset levels at a management review time). If $f_1(v_0) \geq v_0$ there can be no larger companion vector for which (8.5) holds since the larger vector must necessarily incur more transformation costs. In this case, $v_1^*(v_0) = f_1(v_0)$.

Now assume $f_1(v_0)$ is not $\geq v_0$. We restrict $V_1(v_0)$ to vectors $v_1(v_0)$ such that for every item_i, $i = 1,\ldots,I$ and location_j, $j = 0,1,\ldots N$,

$$\begin{aligned} v_{1ij}(v_0) &= f_{1ij}(v_0) \quad \text{if } f_{1ij}(v_0) \geq v_{0ij}; \\ f_{1ij}(v_0) &\leq v_{1ij}(v_0) \leq v_{0ij} \quad \text{otherwise.} \end{aligned} \quad (8.6)$$

(Note that we have extended the notation in the obvious manner). All such asset vectors are feasible since $f_1(v_0)$ satisfies the performance constraints in $(t^*,T]$. From (8.2) and (8.6) we see that the cost of transforming v_0 to any $v_1(v_0)$ in $V_1(v_0)$ is no larger then the cost of transforming v_0 to $f_1(v_0)$. The cost of transforming v_0 to any $v_1(v_0)$ in $V_1(v_0)$ can be obtained precisely by a parametric solution of the AVTP for v_0 and $f_1(v_0)$. The parameters are the right hand sides of the balance constraints for the items and locations for which $v_{0ij} > f_{1ij}$.

In this manner, the transformation costs for all vectors in $V_1(v_0)$ can be obtained by solving only one, albeit parametric, stochastic transshipment problem.

It is now a straightforward matter to select $v_1^*(v_0)$ as the asset vector in $V_1(v_0)$ that minimizes the sum of transformation and holding costs in $[t^*,T]$. Depending on the cost structure it may be possible to further reduce the computational burden. For example, say that it is possible to order $V_1(v_0)$ so that the holding cost is an increasing convex function on $V_1(v_0)$ and the transformation cost is a decreasing concave function on $V_1(v_0)$. We can then select $v_1^*(v_0)$ as the first asset vector for which the increase in holding cost exceeds the decrease in transformation cost.

All that remains is the selection of V_0. Let f_0 be the solution of the FAVP over $[0,t^*)$. The best asset policy using f_0 has cost

$$TC(f_0) = HC_0(f_0) + HC_1[v_1^*(f_0)] + MC[f_0,v_1^*(f_0),t^*].$$

V_0 should therefore contain all vectors v_0 that are feasible in $[0,t^*)$ and satisfy $HC_0(f_0) \leq HC_0(v_0) \leq TC(f_0)$. However, we shall assume that for all v_0 in V_0, $HC_1[v_1^*(v_0)] = HC_1[v_1^*(f_0)]$. V_0 can then be restricted to the v_0 such that

$$HC_0(v_0) \leq HC_0(f_0) + MC[f_0,v_1^*(f_0),t^*]. \tag{8.7}$$

There may be many asset vectors that satisfy (8.7). Some of these vectors are obtained during the implicit enumeration scheme (see Chapter V) used to determine f_0. The heuristic restricts V_0 to precisely these asset vectors. Hopefully, this will reduce V_0 to a manageable size. Otherwise, it may be necessary to further reduce the size of V_0 using heuristics based on the special cost structure of a particular problem.

In summary, the steps of the heuristic we propose are:

(1) Determine f_0 and $f_1(f_0)$. Set $v_0 = f_0$. Set the cost of the incumbent solution to infinity.

(2) If $f_1(v_0) \geq v_0$ set $v_1^*(v_0) = f_1(v_0)$ and go to step 5. Otherwise, construct $V_1(v_0)$ using (8.6).

(3) Parametrically solve the AVTP for v_0 and $f_1(v_0)$ thereby determining the transformation costs for all $v_1(v_0)$ in $V_1(v_0)$.

(4) Find $v_1^*(v_0)$ that minimizes

$$HC_1(v_1(v_0)) + MC(v_0,v_1(v_0),t^*).$$

(5) If the cost of $[v_0,v_1^*(v_0)]$ in (8.3) is less than the cost of the incumbent, set $[v_0,v_1^*(v_0)]$ as the incumbent asset policy.

(6) If $v_0 = f_0$ construct V_0 using (8.7). Otherwise, go to step 7.

(7) If all elements of V_0 have been examined, stop. Otherwise, choose a new v_0 and return to step 2.

This heuristic can be considered to be myopic in the sense that if transformation costs are assumed to be irrelevant and arbitrarily set to 0, the heuristic sets v_0^* to the FAVP solution over $[0,t^*)$. In fact the ultimate myopic heuristic would set $v_0^* = f_0$ and never bother with the construction of V_0. This would greatly reduce the computational burden as only one parametric AVTP would need to be solved to determine an asset policy.

Assume that conditions are such that we are certain that at t^* the asset position of every item at every location should increase. Furthermore, say for all v_0 in V_0, $f_1(v_0) \geq v_0$. The heuristic then chooses among asset policies of the form $[v_0,f_1(v_0)]$. It is interesting to note that the heuristic may not select and the optimal solution may not be $[f_0, f_1(f_0)]$. This is because transformations are not instantaneous so that some vector other than f_0 may leave the system better positioned at t^*. This can result in a significant reduction in holding costs over $[t^*, t)$ than would result if $[f_0, f_1(f_0)]$ were selected as the asset policy.

Summary

In this chapter we generalized the FAVP by allowing management to change the asset vector at any time during the horizon. We gave a brief overview of the issues and difficulties involved in obtaining solutions (operating policies) to this more general problem. We showed that an operating policy must consider the costs and methods of transforming one asset vector to another. For this reason we introduced the AVTP and showed that the FAVP solution was an integral part of the solution of the general inventory control problem. This general inventory control problem appears to be very difficult and cumbersome to solve. However, we outlined a heuristic using the FAVP and AVTP to obtain approximately optimal operating policies for non-stationary multi-echelon inventory systems.

CHAPTER IX

SUMMARY AND EXTENSIONS

Summary

The main goal of this dissertation was to develop a model and methodology for determining "cost effective" asset policies for a two echelon non-stationary one-for-one inventory system in which primary customer demands at the bases form mutually independent NHPP. In order to accomplish this, it was necessary to do three things. First, we had to obtain the time dependent distributions of the pipelines at the depot and bases. In Chapter III we derived the depot pipeline distribution by considering the depot as a single echelon inventory system unto itself. In Chapter IV we studied the supply interactions between the depot and bases and thereby derived the time dependent distributions of the bases' pipelines as functions of the depot's asset policy. We also studied and obtained results on other important stochastic processes that arise at the depot and bases. These results were used to derive the time dependent distributions of customer wait at the bases and the time dependent distribution of the delay at the depot before satisfying a base resupply request.

Secondly, we had to precisely define what we meant by a cost effective (efficient) asset policy. In Chapter V we discussed the problem of choosing meaningful inventory performance measures that adequately distinguished and ranked different asset policies for a non-stationary system. We decided that two performance measures were necessary: one to monitor average performance over the horizon and one to monitor the worst performance over the horizon. Specifically, we chose to measure the average and maximum expected total number of customer backorders outstanding at the bases over the horizon since at any point in time a customer backorder directly corresponded to an inoperable weapon system/end item. An efficient asset policy was then defined as an asset policy that achieved, at the least cost, management specified targets on the average and maximum expected total number of customer backorders over the horizon.

The FAVP introduced in Chapter V assumed that during the horizon management could not change the asset level of any item at any location. Therefore, the cost of an asset policy was simply the cost of procuring (holding) the assets that were placed at the locations in the system at the beginning of the horizon. Hence, an efficient asset policy could be obtained by solving (5.2).

In Chapter VIII we extended the FAVP by allowing management to change the asset levels of one or more items at one or more locations at one or more times during the horizon. In that chapter we considered the cost of an asset policy as: the costs to procure, hold and dispose assets during the horizon; the cost to review/observe the system condition at management review times; and the cost of redistributing assets among the locations in the system. Consideration of this latter cost led to the formulation of the AVTP, a stochastic transshipment problem for determining the least cost method of transforming one system asset vector to another. Under the above cost structure, an efficient asset policy could be obtained as a by-product of the optimal operating policy obtained by solving (8.1).

Lastly, we had to actually obtain efficient asset policies by solving the mathematical pr-gramming problems (5.2) and (8.1). In Chapter V we developed an implicit enumeration scheme to solve the non-linear integer program (5.2). We established in Chapter VI the efficacy of approximating the computationally cumbersome base pipeline distribution with a negative binomial distribution. Use of this approximation significantly reduced the effort and cost involved in solving (5.2). In Chapter VII we reported on some computational experience with the FAVP. In that chapter we also briefly addressed the issue of using the more facile steady state models such as SESAME (U.S. Army [1983]) to obtain approximate solutions to the FAVP. This is certainly an interesting area for further research.

It is very difficult, costly and burdensome to obtain the optimal solutions to (8.1). The reader can appreciate the enormity of the task by noting the effort involved in solving the FAVP which is itself a special case of (8.1). However, in Chapter VIII we briefly outlined a heuristic using the FAVP and AVTP to obtain close to optimal solutions to (8.1). Developing algorithms for obtaining exact and/or approximate solutions to (8.1) is a natural and fruitful continuation of the research presented in this dissertation.

Extensions

We mentioned above two areas of research that we felt were natural continuations of the research presented in this dissertation. In this section we will briefly present some of our (untested) thoughts on other extensions to some of the results we have presented.

The optimization scheme in Chapter V for solving the FAVP requires that for any specific asset policy we be able to calculate the average and maximum expected total number of backorders at the bases over the horizon. However, the optimization scheme itself and the other results in Chapter V are essentially independent of the actual distributions of the pipelines and the backorders at the locations. Therefore, the implicit enumeration scheme used to solve the FAVP is robust in the sense that it does not depend on the most of the major assumptions made in Chapter II. Therefore, the extensions we discuss below will generally only impact upon how we calculate the distributions of the pipelines and backorders at the locations in the inventory system.

(1) $\underline{N > 2 \text{ Echelons}}$: Because of our assumptions, the depot (echelon-N) pipeline will always have a Poisson distribution with mean given by (3.6). The time dependent distributions of the pipelines at the echelon-(N-1) locations are precisely the distributions that we have given for the base pipelines. In fact, all of the results obtained for the bases in this dissertation hold for the echelon-(N-1) locations. As long as the order and ship times between locations are constant, we can establish an analog to Theorem 4.8 for locations below echelon-(N-1). This allows us to write the pipeline at any time at location_j on echelon-K, $1 \leq K < N - 1$, as the sum of: a Poisson component representing the number of units in-repair at location_j and the number of units due-in from the external supplier; a Poisson component representing the number of failed units sent to higher echelon resuppliers for which serviceable

replacements could not yet have arrived at $location_j$ (even if the higher echelon resuppliers had infinite stock) because of the constant order and ship times; and a component representing the number of higher echelon backorders belonging to $location_j$. (4.30) and (4.32) can be used to find the distribution of the number of depot backorders belonging to $location_j$. The difficulty in extending our results to $N > 2$ echelons lies in finding the distribution of the number of backorders at echelons (K+1) through (N-1) that belong to $location_j$.

If $location_j$ is a "proportionate location" (see Chapter IV) then we can use (4.6) to find the distribution of the number of higher echelon backorders that belong to $location_j$. Otherwise, since our research has shown that locations below the depot will usually not have Poisson pipelines, we must modify the arguments we used to find the distribution of $Q_j(t)$, $t > 0$. However, the same underlying approach is still valid. Namely, for each higher echelon resupplier of $location_j$ we need to find the time interval over which backorders accumulated and the number of resupply requests made by $location_j$ during that interval. The fact that higher echelon locations usually do not have Poisson pipelines will probably lead to even more complex and cumbersome expressions than (4.30) and (4.32). However, it is reasonable to expect that the negative binomial will still be an effective approximating distribution. In fact, since we have shown the negative binomial distribution with proper mean and variance can be used to approximate (4.6), we know that the negative binomial approximation is adequate for N-echelon systems consisting of proportionate locations.

The implicit enumeration scheme given in Chapter V for solving the FAVP is easily extended to problems with $N > 2$ echelons. We can generalize Theorems 5.6 and 5.7 so that bounds can be established on the optimal asset levels at all the locations on echelons 2 through N. By fixing the asset levels at all the locations on echelons 2 through N we can calculate the distributions of the pipelines at the echelon-1 locations. We can then find the optimal asset levels at the echelon-1 locations in precisely the same manner that we found the optimal base asset levels in Chapter V. Now, however, to find the optimal asset vector we must not only search over all possible depot asset levels but also over all possible asset levels at the locations on echelons 2 through N - 1.

(2) Finite Repair Capacity/Batch Repair Policy: Assumptions 4, 5, and 6, by maintaining the statistical independence of the times different units spend in the repair facility, allowed us to use the Splitting Property of NHPP to establish that the number of units in the repair facility at a location had a Poisson distribution. This proved extremely convenient when we determined the distributions of the pipelines at the depot and bases. In principle, we could have removed Assumptions 5 and 6 at the bases. We would then have had to obtain the transient distribution of a finite server queue (possibly with a batch service policy). Such distributions are notoriously complex, difficult to obtain, and cumbersome to use. Removing Assumptions 5 and 6 at the depot would result in a non-Poisson depot pipeline. Computing the distribution of $Q_j(t)$, $j = 1,2,\ldots N$, would then require modification of the arguments used in Chapter IV (see paragraph (1) above).

The analysis becomes even more difficult when Assumptions 5 and 6 are removed at the diagnostic facilities. Our arguments relied heavily upon the fact that the output of a location's diagnostic facility was a NHPP. To establish independence among the components of a location's pipeline we also relied on the fact that the number of departures from a diagnostic facility during $(t_1,t_2]$ was independent of the number of units remaining in the same diagnostic facility at t_2. As far as we know, these results are valid only when the diagnostic facility acts as an $M(t)/G(t)/\infty$ queue. (In stationary systems we could deal with M/M/s queues at the diagnostic and repair facilities ([Kleinrock (1975)]). We believe that losing the independence of the different components of the pipeline at a location would lead to intractable expressions that would, in practice, be impossible to implement.

(3) Other Demand Processes: Most of the results in this dissertation can be extended to the case where customer demands at the bases form independent NHCPP. The problems involved in using other types of demand processes seem insurmountable. In fact, to the author's knowledge, there is no adequate model for stationary continuous review multi-echelon one-for-one inventory systems in which customer demands do not form a homogeneous (compound) Poisson Process.

(4) Uncertainty in the Intensity Function: The correct way to deal with a prior distribution on the demand intensity at a base is to calculate the performance of an asset policy for each possible value of the demand intensity and then to weight these performances by the prior. Attempting to account for uncertainty in the intensity function by developing a NHCPP model of demand ignores the correlation of demand at the depot and bases when there is uncertainty in the base demand intensity. At best this approach would be an approximation whose effectiveness would have to be investigated.

(5) Indentured Items: An indentured item is a module that

consists of several components that can be removed from the module. The basic idea behind this type of design is that when a module fails (thereby causing a weapon system/end item to fail), a failed component can be quickly removed and replaced. The module is then serviceable while repair proceeds on the failed component. Therefore, the actual downtime of the module will be less than if the repair had to be done on the whole module. Since serviceable components are not always available, a location's pipeline must include the number of modules awaiting serviceable components. (4.30), (4.32) and (2.2) can be used to find the number of backorders for each component at each location. This quantity is then added to the rest of the quantities in each location's pipeline. The FAVP can be straightforwardly extended to consider the tradeoffs between investments in modules and components.

The other assumptions in Chapter II (including the assumption of a one-for-one resupply policy) are identical to the assumptions in METRIC (Sherbrooke [1968]). Most of the stationary continuous review multi-echelon models in the literature are basically variants of METRIC and they have adopted the same set of assumptions. The difficulties and issues involved in relaxing these assumptions (along with some suggested research directions) have been discussed in the literature (see, for example, Kaplan [1980]) and will not be repeated here. Further research into developing tractable models and methodologies when some or all of these assumptions are removed would be useful in the analysis of both stationary and non-stationary multi-echelon inventory systems.

APPENDIX A

PROPERTIES OF NON-HOMOGENEOUS POISSON PROCESSES

This appendix catalogs some useful and important properties of non-homogeneous Poisson Processes (NHPP). Many of these correspond to results for homogeneous Poisson Processes and many can easily be generalized to non-homogeneous compound Poisson Processes. We assume, for ease of exposition, that all mean value functions (MVF) are differentiable.

P1: Let $\{\{N_i(t),\ t \geq 0\}\ \ i = 1,2,\ldots\}$ be a countable set of mutually statistically independent NHPP with the ith process having intensity $\lambda_i(t)$ and MVF $m_i(t)$ with $\sum m_i(t) < \infty$ for all $t \geq 0$. Define for $t \geq 0$, $N(t) = \sum N_i(t)$ as the superposition of the NHPP. Then $\{N(t),\ t \geq 0\}$ is a NHPP with intensity $\lambda(t) = \sum \lambda_i(t)$ and MVF $m(t) = \sum m_i(t)$ for all $t \geq 0$.

Proof: Define for $t \geq 0$ and $|z| < 1$

$$g(z,t) = E[z^{N(t)}]$$

$$g_i(z,t) = E[z^{N_i(t)}] = e^{-m_i(t)(1-z)} \qquad i = 1,2,\ldots.$$

By independence

$$g(z,t) = e^{-\sum m_i(t)(1-z)} \qquad t \geq 0,\ |z| < 1$$

so that $N(t)$ has a Poisson distribution with mean $\sum m_i(t)$. Clearly, $N(0) = 0$, and the fact that $N(t)$ has independent increments follows from the fact that each $N_i(t)$ has independent increments. //

For the remainder of this appendix let $\{N(t),\ t \geq 0\}$ be a NHPP with intensity $\lambda(t)$ and differentiable MVF $m(t)$, $t \geq 0$. Property 2 is given as a problem in Ross (1970) and Property 3 is given as a problem in both Parzen (1962) and Ross (1970).

P2: For all $0 \leq s \leq t$, $Pr(N(s) = 1|N(t) = 1) = m(s)/m(t)$.

Proof:
$$Pr(N(s) = 1|N(t) = 1) = \frac{Pr(N(s) = 1,\ N(t) - N(s) = 0)}{Pr(N(t) = 1)}$$

$$= \frac{e^{-m(s)}m(s)\ e^{-[m(t) - m(s)]}}{e^{-m(t)}m(t)} = \frac{m(s)}{m(t)}. \qquad (A.1)$$

//

Given an event occurred in $(0,t]$, P2 shows that the time that the event occurred has the same distribution as a random variable with distribution function $m(s)/m(t)$, $0 \leq s \leq t$ and that the probability density function of the time of occurrence of that event is

$$f(s) = \frac{d}{ds} Pr(N(s) = 1|N(t) = 1) = \frac{\lambda(s)}{m(t)}. \qquad (A.2)$$

P3: Given $N(t) = n$, the joint distribution of the n epochs at which events occurred, $t_1 \leq t_2 \ \ldots \leq t_n \leq t$ is the same as if they were order statistics corresponding to n i.i.d. random variables $Y_1, Y_2, \ldots Y_n$ with common distribution

$$F(s) = \frac{m(s)}{m(t)} \qquad 0 \leq s \leq t.$$

Proof: Let $Y_{[1]} \leq Y_{[2]} \ \ldots \leq Y_{[n]}$ be the order statistics corresponding to $Y_1, Y_2, \ldots, Y_n$. The joint density function of these order statistics is given by

$$g(s_1, s_2, \ldots, s_n) = n! \prod_1^n f(s_i) \qquad 0 \leq s_1 \leq s_2 \cdots \leq s_n$$

where $f(s)$ is given by (A.2). Thus,

$$g(s_i,\ i = 1,2,\ldots,n) = \frac{n!}{m(t)^n} \prod_1^n \lambda(s_i). \qquad (A.3)$$

Let h_i be small enough so that $s_i + h_i < s_{i+1}$. Then

$$\Pr(s_i \le t_i \le s_i + h_i \ \ i=1,..,n|N(t)=n) = \frac{\Pr(s_i \le t_i \le s_i + h_i, \ i=1,..,n)}{\Pr(N(t)=n)}$$

$$= \Pr(N(s_i+h_i) - N(s_i) = 1 \qquad i = 1,..,n;$$

$$N(s_i) - N(s_{i-1} + h_{i-1}) = 0 \qquad i = 2,...,n;$$

$$N(s_1) - N(0) = 0;$$

$$N(t) - N(s_n+h_n) = 0)/\Pr(N(t) = n).$$

Since the intervals $[s_i, s_i + h_i]$, $i = 1,...,n$ do not overlap, and the number of events in each interval has a Poisson distribution with mean equal to the MVF over the interval, the above becomes after some reduction and rearrangement,

$$\frac{e^{-m(t)} \prod_1^n [m(s_i+h_i) - m(s_i)]}{e^{-m(t)} \ m(t)^n/n!} = \frac{n!}{m(t)^n} \prod_1^n [m(s_i+h_i) - m(s_i)].$$

Therefore,

$$\frac{\Pr(s_i \le t_i \le s_i + h_i, \ i = 1,...n|N(t) = n)}{\prod_1^n h_i} = \frac{n!}{m(t)^n} \prod_1^n \frac{m(s_i+h_i) - m(s_i)}{h_i}.$$

Taking the limit of both sides as the $h_i \to 0$ uniformly we find that the left side is just an ordinary probability density function and that this exists since the right hand side limit exists by the assumption that the MVF is differentiable.

Thus,

$$f(s_1,s_2,...,s_n|N(t) = n) = \frac{n!}{m(t)^n} \prod_1^n \lim_{h_i \to 0} \frac{m(s_i+h_i) - m(s_i)}{h_i}$$

$$= \frac{n!}{m(t)^n} \prod_1^n \lambda(s_i)$$

which is precisely (A.3) //

As a consequence of P3, if an event of the NHPP occurred in $(0,t]$ the time that this event occurred has distribution function $m(s)/m(t)$ and density $\lambda(s)/m(t)$, $0 \le s \le t$, independently of any other events that have occurred. The next two properties are simple corollaries of this powerful fact.

P4: For $0 \le s \le t$ and $0 \le k \le n$,

$$\Pr(N(s) = k \mid N(t) = n) = \binom{n}{k}\left(\frac{m(s)}{m(t)}\right)^{k}\left(1 - \frac{m(s)}{m(t)}\right)^{n-k} \qquad \text{(A.4)}$$

//

P5: Let W_k be the time of the kth event of the NHPP, $k = 1,2,\ldots$

For $1 \le k \le n$, the density of W_k given $N(t) = n$ is

$$\frac{n!}{(k-1)!\,(n-k)!}\left(\frac{m(s)}{m(t)}\right)^{k-1}\left(1 - \frac{m(s)}{m(t)}\right)^{n-k}\frac{\lambda(s)}{m(t)} \qquad \text{(A.5)}$$

Proof:

$$\Pr(W_k \le s \mid N(t) = n) = \sum_{j=k}^{n} \Pr(N(s) = j \mid N(t) = n)$$

Use (A.4) and differentiate to get (A.5). //

Alternatively, and as a direct consequence of P3, (A.5) is obtained by noting that given $N(t) = n$, $W_k \in [s, s+ds]$ iff $k-1$ of the n events occurred before s, $n-k$ occurred after $s+ds$ and one occurred in $[s, s+ds]$. Since the event times behave as independent random variables with distribution function $F(s)$, (A.5) follows. //

P6: Let $\{\{N_i(t),\ t \ge 0\}\ i = 1,2,\ldots\}$ be a countable collection of independent NHPP with $m(t) = \sum m_i(t) < \infty$ for all $t \ge 0$. Let $N(t) = \sum N_i(t)$ be their superposition. Then,

$$\Pr(N_i(t) = 1 \mid N(t) = 1) = m_i(t)/m(t).$$

Proof: Clearly, we have that

$$\Pr(N_i(t) = 1,\ N(t) = 1) = e^{-m_i(t)}\, m_i(t) \prod_{j \neq i} e^{-m_j(t)}$$
$$= m_i(t)\, e^{-m(t)}.$$

Dividing by $\Pr(N(t) = 1)$ (see P1) yields the desired result. //

P7: Let $\{\{N_i(t),\ t \geq 0\}\ i = 1,2\ldots\}$ and $\{N(t),\ t \geq 0\}$ be as in P6. Then for $t \geq 0$,

$$\Pr\left\{\begin{array}{l}\text{an event of the superposition}\\ \text{process that occurs at } t \text{ is}\\ \text{from the ith NHPP}\end{array}\right\} = \frac{\lambda_i(t)}{\sum_j \lambda_j(t)} \tag{A.6}$$

independently of the other events in $(0,t]$.

Proof: By P6, for $h > 0$,

$$\Pr(N_i(t+h) - N_i(t) = 1 \mid N(t+h) - N(t) = 1)$$
$$= \frac{m_i(t+h) - m_i(t)}{m(t+h) - m(t)} = \frac{[m_i(t+h) - m_i(t)]/h}{[m(t+h) - m(t)]/h}.$$

Taking the limit as $h \to 0$ yields (A.6). Independence follows from the independent increment property of the superposition process (P1).//

P8: Let $\{\{N_i(t),\ t \geq 0\}\ i = 1,2,\ldots\}$ and $\{N(t),\ t \geq 0\}$ be as in P6, and let $\gamma_i(t)$ be the probability that an event of the superposition process that occurred in $(0,t]$ was from the ith NHPP. Then, for $t \geq 0$,

$$\gamma_i(t) = m_i(t)/m(t) \tag{A.7}$$

independently of the other events in $(0,t]$.

Proof: By P3 and P1,

$$\gamma_i(t) = \int_0^t \frac{\lambda_i(x)}{\sum_j \lambda_j(x)} \frac{\sum_j \lambda_j(x)}{m(t)}\, dx = \frac{m_i(t)}{m(t)}.$$

Independence follows trivially as in P7. //

Therefore, if an event of the superposition process is known to have occurred in $(0,t]$, the probability it was an event of the ith NHPP is given by (A.7) and the fact that this event was from the ith NHPP has no effect on the probabilities that other events of the superposition process that occurred in (0,t] were or were not from the ith NHPP. This immediately leads to

<u>P9</u>: Let $\{\{N_i(t),\ t \geq 0\}\ i = 1,2,\ldots\}$ and $\{N(t),\ t \geq 0\}$ be as in <u>P6</u>. For $m \leq n$,

$$\Pr(N_i(t) = m \mid N(t) = n) = \binom{n}{m} \left(\frac{m_i(t)}{m(t)}\right)^m \left(1 - \frac{m_i(t)}{m(t)}\right)^{n-m} \qquad //$$

Property 1 ensures that the superposition of independent NHPP is itself a NHPP. Property 10, as a sort of dual to Property 1, ensures that a NHPP can be "split" into independent constituent NHPP.

<u>P10 (Splitting Property)</u>: Suppose an event of the NHPP $\{N(t), t \geq 0\}$ that occurs at time $t > 0$ is classified as a type i event with probability $p_i(t)$, $i = 1,2\ldots$ ($\sum p_i(t) = 1$ for all t) and that classification of each event is independent of the classification of other events. Let $N_i(t)$ be the number of type i events that have occurred by time t. Then the counting process $\{N_i(t),\ t \geq 0\}$ is a NHPP with intensity $\lambda_i(t) = \lambda(t)p_i(t)$ and MVF $m_i(t) = \int_0^t \lambda(s)p_i(s)ds$. Furthermore, $\{\{N_i(t),\ t \geq 0\}\ i = 1,2,\ldots\}$ are mutually statistically independent.

<u>Proof</u>: Given an event occurred in $(t_1,t_2]$, the probability it was classified as type i is

$$c_i(t_1,t_2) = \int_{t_1}^{t_2} p_i(x) \frac{\lambda(x)}{m(t_2) - m(t_1)} dx \qquad 1 = 1,2,\ldots$$

independently of the other events that occurred in $(t_1,t_2]$. Therefore, $[N_i(t_2) - N_i(t_1) \mid N(t_2) - N(t_1)]$ is a Binomial random variable with parameters $N(t_2) - N(t_1)$ and $c_i(t_1,t_2)$ so that for $|z| < 1$

$$E[z^{N_i(t_2) - N_i(t_1)} | N(t_2) - N(t_1)] = [1 - c_i(t_1,t_2)(1-z)]^{N(t_2) - N(t_1)}.$$

Since $N(t_2) - N(t_1)$ is Poisson with mean $m(t_2) - m(t_1)$ and $|1 - c_i(t_1,t_2)(1-z)| < 1$,

$$g_i(z,t_1,t_2) = E[z^{N_i(t_2)-N_i(t_1)}] = \exp[-c_i(t_1,t_2)(1-z)(m(t_2)-m(t_1))].$$

Therefore, $N_i(t_2) - N_i(t_1)$ has a Poisson distribution with mean $c_i(t_1,t_2)[m(t_2)-m(t_1)]$. In particular, $N_i(t)$ has a Poisson distribution with mean $c_i(0,t)m(t)$, $i = 1,2,\ldots$.

That each process has independent increments follows straightforwardly since $\{N(t), t \geq 0\}$ has independent increments and each event is classified independently of other events.

Since for all i, $N_i(0) = 0$, $\{N_i(t), t \geq 0\}$ has independent increments and the number of type i events in any interval has a Poisson distribution, $\{N_i(t), t \geq 0\}$ is a NHPP, $i = 1,2,\ldots$ All that remains is to show that the processes are mutually independent.

Let $Z = (Z_1, Z_2 \ldots)$ be a vector such that $|Z_i| < 1$ for all i and define the joint generating function $g(Z,t) = E[\Pi Z_i^{N_i(t)}]$. By the Law of Total Probability,

$$g(Z,t) = E_{N(t)} [E(\prod_i Z_i^{N_i(t)} | N(t))].$$

In light of the earlier discussion we know the joint distribution of $(N_1(t), N_2(t),\ldots | N(t))$ is multinomial with parameters $N(t)$ and $c_i(0,t)$, $i = 1,2,\ldots$ Therefore,

$$g(Z,t) = E_{N(t)}[(\sum_i c_i(0,t)Z_i)^{N(t)}]$$

Since $N(t)$ has a Poisson distribution and $|\sum c_i((0,t)Z_i| < 1$ we have that

$$g(Z,t) = \exp\left(-m(t)\left(1-\sum_i c_i(0,t)Z_i\right)\right)$$

$$= \exp\left(-\sum_i m(t)c_i(0,t)(1-Z_i)\right)$$

$$= \prod_i g_i(Z_i,0,t)$$

where in the next to the last equality we used the fact that $\sum_i c_i(0,t)=1$. Independence is thus established for $(0,t]$ and can be similarly established for any non-overlapping time intervals. //

The next properties are simple properties of NHPP given as problems in Parzen (1962) and included here for completeness. Proofs are straightforward and therefore omitted.

Let $T_0 \equiv 0$;

T_i = time between the (i-1)st and ith event of a NHPP;

$W_n = \sum_{i=0}^{n} T_i \qquad n = 0,1,2,\ldots;$

$B_n(t) = \Pr(T_n \le t) \qquad n = 0,1,2,\ldots;$

$b_n(t) = \frac{d}{dt} B_n(t) \qquad n = 0,1,2,\ldots;$

$G_n(t) = \Pr(W_n \le t) \qquad n = 0,1,2,\ldots;$

$g_n(t) = \frac{d}{dt} G_n(t) \qquad n = 0,1,2,\ldots.$

<u>P11</u>: $b_1(t) = \lambda(t)e^{-m(t)} \qquad t \ge 0$

<u>P12</u>: $b_2(t|t_1) = e^{-m(t+t_1) - m(t_1)} \lambda(t+t_1) \qquad 0 \le t_1 \le t.$

<u>P13</u>: $b_n(t|t_1,t_2,\ldots,t_{n-1}) = \lambda(t+W_{n-1})\, e^{-(m(t+W_{n-1}) - m(W_{n-1}))}$

$= b_n(t|W_{n-1}) \qquad t \ge 0$

By <u>P13</u>, the nth inter-arrival time depends on the first n-1 inter-arrival times only through their sum, W_{n-1}. While they are not independent, the inter-arrival times have conditionally exponential distributions. Of course, if $\lambda(t) = \lambda$ for all t, the nth inter-arrival time has an exponential distribution and does not depend on the waiting time, W_{n-1}.

<u>P14</u>: $$g_n(t) = e^{-m(t)} \frac{m(t)^{n-1}}{(n-1)!} \lambda(t) \qquad t \geq 0.$$

<u>P15</u>: $$\bar{B}_n(t|W_{n-1}) = 1 - B_n(t|W_{n-1}) = e^{-(m(t+W_{n-1}) - m(W_{n-1}))} \qquad t \geq 0.$$

<u>P16</u>: $$b_n(t) = \int_0^t \lambda(s)\, e^{-m(t+s)} \frac{m(s)^{n-2}}{(n-2)!} \lambda(t+s)\, ds \qquad \begin{array}{l} n \geq 2. \\ t \geq 0 \end{array}$$

<u>P17</u>: $$\bar{B}_n(t) = 1 - B_n(t) = \int_0^t e^{-m(t+s)} \frac{m(s)^{n-2}}{(n-2)!} \lambda(s)\, ds \qquad \begin{array}{l} n \geq 2. \\ t \geq 0 \end{array}$$

Next page is blank.

APPENDIX B

DESCRIPTION OF THE DATA BASE

We obtained data on three U. S. Army weapon systems in order to establish a realistic data base for the testing and validation reported upon in this dissertation. In this appendix, we give a brief description of this data. All data are unclassified and some scenario data were fictionalized but are typical of actual scenario data.

Each of the weapon systems described below actually consists of over 1000 items. However, we restricted our data collection efforts to the high cost items on each system since it is precisely these items that tend to drive inventory budgets and policies.

The Army believes (U. S. Army [1983]) that item failures are proportional to usage. Each weapon system (e.g. a tank) has a projected yearly usage during peacetime. This forms the basis for determining the maintenance factor for each item which is the expected number of failures of the item per application of the weapon system per year. Non-peacetime scenarios account for different failure rates by using a usage modifier (applied to all items on the weapon system) for each day in the scenario. The usage modifier on a particular day of the scenario is the ratio of scenario usage to daily peacetime usage. For example, an item on a tank may have a maintenance factor of .1 based on a usage profile for the tank of 2 miles driven per day during peacetime. If the scenario calls for the tank to drive 6 miles on a particular day, the usage modifier would be 3 and the failure rate for that day would be [(.1)(3)/365]. For each weapon system we give a typical scenario for the weapon system in terms of its daily usage modifiers.

Weapon System Name: Blackhawk Helicopter (BHAWK)

Number of Items in Data Base: 75

Distribution of Unit Prices and Maintenance Factors:

Maintenance Factors	Unit Price (Thousands) ≤ 1	1-5	5-10	10-25	25-50	50-100	100-250	> 250
< .01	0	0	1	4	0	0	0	0
.01-.05	0	2	5	4	0	0	1	0
.05-.1	0	1	4	6	1	0	1	1
.1-.25	0	3	2	8	0	1	1	1
.25-.5	0	0	1	2	2	0	0	1
.5-1	0	0	3	4	1	1	1	0
1-2.5	0	0	2	2	3	1	0	0
> 2.5	0	0	1	1	1	0	1	0

Support Structure: 3 bases: A Company: 30 helicopters; B Company: 50 helicopters; C Company: 20 helicopters.

Order and Ship Time: 7 days between bases and Depot.

Repair Policy: No base repair. Repair time at depot is log - normally distributed with a mean of 10 days and variance of 18 days. On average, 5% of all failures are condemned.

Scenario: 30 day horizon.

Usage modifier of 1.5 on each day at A Company;

Usage Modifier of 2 on each day at B Company;

At C Company, Usage Modifier of 2 on days 1-5; 5 on days 6-15; 2 on days 16-18; 4 on days 19-25; 3 on days 26-28; 2 on days 29-30.

Weapon System Name: Army Attack Helicopter (AAH)

Number of Items in Data Base: 155

Distribution of Unit Prices and Maintenance Factors:

	Unit Price (Thousands)							
Maintenance Factors	≤ 1	1-5	5-10	10-25	25-50	50-100	100-250	> 250
< .01	1	8	0	0	0	0	0	0
.01-.05	8	13	1	4	0	0	0	0
.05-.1	3	24	4	1	1	1	0	0
.1-.25	3	18	8	4	1	0	0	0
.25-.5	2	9	4	5	0	0	0	0
.5-1	0	3	2	4	5	1	1	1
1-2.5	0	1	1	3	5	0	1	0
> 2.5	0	2	1	0	0	1	0	0

Support Structure: 16 bases with a distribution of 24, 19, 11, 16, 58, 15, 16, 23, 13, 16, 24, 20, 16, 16, 16, and 24 helicopters.

Order and Ship Times: 2 days between bases and Depot.

Repair Policy: No base repair. Depot repair time 15 days. We assumed no condemnations although on average, 5% of all failures are condemned.

Scenario: 30 day horizon.

Usage modifier of 1.5 on each day at bases with < 20 helicopters.

For bases with > 20 helicopters:

Usage modifier of 3 on days 1-5; 2 on days 6-15;

3 on days 15-25; 1 on days 26-30.

Weapon System Name: M60A3 Tank (M60A3)

Number of Items in Data Base: 250

Distribution of Unit Prices and Maintenance Factors:

	Unit Price (Thousands)							
Maintenance Factors	≤ 1	1-5	5-10	10-25	25-50	50-100	100-250	> 250
< .01	3	6	1	1	0	0	0	0
.01-.05	16	12	5	3	0	0	0	0
.05-.1	9	4	5	4	0	0	0	0
.1-.25	14	9	2	2	0	0	0	0
.25-.5	36	8	1	1	0	0	0	0
.5-1	22	4	5	1	0	0	0	0
1-2.5	43	8	3	1	0	0	0	0
> 2.5	16	3	1	1	0	0	0	0

Support Structure: 10 identical bases each with 25 tanks.

Order and Ship Time: 2 days between bases and Depot.

Repair Policy: Very little base Repair. Depot repair time: 30 days. On average about 7% of the units that reach the depot are condemned.

Scenario: 180 day horizon.

Usage modifier on day t is $1 + 1.5 \sin^2 \pi t/2$.

Next page is blank.

BIBLIOGRAPHY

Bessler, Steven A. and Arthur F. Veinott, Jr. 1966. Optimal Policy for a Dynamic Multi-Echelon Inventory Model. Naval Logistics Research Quarterly 13, 355-390.

Bessler, Steven A. and Paul W. Zehna. 1968. An Application of Servo-Mechanisms to Inventory. Naval Logistics Research Quarterly 15, 157-168.

Brooks, Robin B.S., Collen A. Gillen, and John Y. Lu. 1969. Alternative Measures of Supply Performance: Fills, Backorders, Operational Rate and NORS. RM-6094-PR, Santa Monica, CA.

Brown, Mark and Sheldon M. Ross. 1969. Some Results For Infinite Server Poisson Queues. Journal of Applied Probability 6, 604-611.

Burns, John F. 1970. Derivation and Simulation of an Ordering Rule for the Multi-Echelon Supply Chain. TR #45, Department of Industrial & Systems Engineering, University of Florida, Gainesville, FL.

Burns, John F. and B. D. Sivazlian. 1978. Dynamic Analysis of a Multi- Echelon Supply System. Computers & Industrial Engineering 2, 181-193.

Cinlar, Erhan. 1975. Introduction to Stochastic Processes. Prentice Hall, Englewood Cliffs, New Jersey.

Clark, Andrew J. 1960. The Use of Simulation to Evaluate a Multi-Echelon Dynamic Inventory Model. Naval Logistics Research Quarterly 7, 429-446.

Clark, Andrew J. 1972. An Informal Survey of Multi-Echelon Inventory Theory. Naval Logistics Research Quarterly 19, 621-650.

Clark, Andrew J. 1978. Optimal Operational Availability Inventory Model. R-7806, CACI Inc., Arlington, VA.

Coggins, James. 1983. A Method to Incorporate the Dependent Parts Demand Due to Combat Damage Into Existing U. S. Army Multi-Echelon Provisioning Models. Unpublished MS Thesis. Georgia Institute of Technology, Atlanta, GA.

Conte, Samuel and Carl de Boor. 1980. Elementary Numerical Analysis. McGraw-Hill, New York.

Daley, D. J. 1973. Notes on Queueing Output Processes. Mathematical Methods in Queueing Theory: Proceedings of Conference at Western Michigan University, 10-12 May, 351-358.

Dempster, M.A.H. 1980. Stochastic Programming. Academic Press, New York.

Everett, Hugh III. 1963. Generalized Lagrange Multiplier Method For Solving Problems of Optimum Allocation of Resources. Operations Research 11, 399-417.

Fakinos, D. 1984. The Infinite Server Queue. Journal of the Operational Research Society 35, 439-445.

Fox, Bennet L. 1966. Discrete Optimization Via Marginal Analysis. Management Science 13, 210-216.

Frazza, Sally. 1982. Test of Poisson Failure Assumption. ADA137285, U. S. Army Inventory Research Office, Philadelphia, PA.

Galliher, Herbert P. and Richard C. Wilson. 1975. Aircraft Engines: Demand Forecasting and Inventory Redistribution. Aerospace Research Laboratories, Wright-Patterson AFB, OH.

Grassman, William. 1977. Transient Solutions in Markovian Queueing Systems. Computers and Operations Research 4, 47-56.

Graves, Steven C. and Leroy B. Schwarz. 1977. Single Cycle Continuous Review Policies for Arborescent Inventory Systems. Management Science 26, 529-540.

Graves, Steven C. 1983. A Multi-Echelon Model for a Reparable Item With One-For-One Replenishment. Management Science, To Appear.

Gross, Donald. 1982. Ample Service Assumption of Palm's Theorem in Inventory Modeling. Management Science 28, 1065-1079.

Gross, Donald and Douglas R. Miller. 1982. Multi-Echelon Repairable Item Provisioning In A Time-Varying Environment Using The Randomization Technique. Technical Paper Serial T-468, Institute for Management Science and Engineering, The George Washington University, Washington, DC.

Hadley, George and T. M. Whitin. 1963. Analysis of Inventory Systems, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Hillestad, Richard J. and Manuel J. Carillo. 1980. Models and Techniques for Recoverable Item Stockage When Demand and the Repair Process are Non-Stationary-Part I: Performance Measures. N-1482-AF, Rand Corp., Santa Monica, CA.

Hillestad, Richard J. 1982. Dyna-METRIC: Dynamic Multi-Echelon Technique for Recoverable Item Control. R-2785-AF, Rand Corporation, Santa Monica, CA.

Huang, G. C., I. Vertinsky and W. T. Ziemba. 1977. Sharp Bounds on the Value of Perfect Information. Operations Research 25, 128-139.

Iglehart, Donald L. and Samuel Karlin. 1962. Optimal Policies for Dynamic Inventory Processes with Non-Stationary Stochastic Demands. Studies in Applied Probability and Management Science, Arrow, Karlin, Scarf, ed., Stanford University Press, Stanford, CA.

Ignall, Edward and Arthur F. Veinott Jr. 1969. Optimality of Myopic Inventory Policies For Several Substitute Products. Management Science 15, 284-304.

IMSL. 1979. IMSL Library Reference Manual. IMSL, Houston, TX.

Kaplan, Alan. 1980. Mathematics for the SESAME Model. ADA081931, US Army Inventory Research Office, Philadelphia, PA.

Karlin, Samuel. 1960. Dynamic Inventory Policy with Varying Stochastic Demand. Management Science 6, 231-258.

Kelly, Frank P. 1979. Reversibility and Stochastic Networks. John Wiley & Sons, New York.

Khintchine, A. Y. 1960. Mathematical Methods in the Theory of Queueing. Griffin Publishing Co., London.

Kim, T. M. 1978. Two Echelon Reorder Point Periodic Inventory Control System with Performance Constraints. Unpublished PhD Thesis. University of Michigan, Ann Arbor, MI.

Kitchen, Joseph W. Jr. 1968. Calculus of One Variable. Addison Wesley Publishing, Reading, MA.

Kleinrock, Leonard. 1975. Queueing Systems, Volume I: Theory. John Wiley & Sons, New York.

Kotkin, Meyer and Donald P. Rhoads. 1977. An Analysis of a Three Echelon Inventory System. Unpublished Master of Engineering Thesis. Cornell University, Ithaca, N.Y.

Kotkin, Meyer. 1978. A Heuristic For Multi-Echelon Multi-Indentured Inventory Problems. ADA066590, U. S. Army Inventory Research Office, Philadelphia, PA.

Kotkin, Meyer. 1982. A Final Word on Simon's Two-Echelon Inventory Model. Unpublished Research Notes.

Kotkin, Meyer. 1983. The Expected Backorder Function in Inventory Theory. U.S. Army Inventory Research Office, Philadelphia, PA. To be published.

Kotkin, Meyer. 1985. Customer Wait In A Multi-Echelon Inventory System. U.S. Army Inventory Research Office, Philadelphia, PA. To be published.

Kruse, Karl. 1979. An Exact N Echelon Inventory Model: The Simple Simon Method. ADA067762, U. S. Army Inventory Research Office, Philadelphia, PA.

Kruse, Karl and Martin Cohen. 1983. Personal Communication.

Lehmann, Edward. 1959. Testing Statistical Hypothesis. Wiley and Sons, New York.

Madansky, Albert. 1960. Inequalities For Stochastic Linear Programming Problems. Management Science 6, 197-204.

Mason, David W. 1975. The FATSO (Father-Son) Sparing Model and Computer Program. D LOG A Working Paper No. 8/75, Ottawa, Canada.

Metzner, Howard E. 1981. Empirical Determination of WRSK Component Failure Distributions. ADA104497, USAF Business Research Management Center, Wright-Patterson AFB, OH.

Miller, Bruce L. 1968. Real Time METRIC: A Technique for the Distribution of Serviceable Assets. RM-5687-PR, Rand Corporation, Santa Monica, CA.

Miller, Bruce L. 1974. Dispatching from Depot Repair in a Recoverable Item Inventory System: On the Optimality of a Heuristic Rule. Management Science 21, 318-325.

Mitchell, Charles R., Robert A. Rappold and Wayne B. Faulkner. 1980. An Analysis of Air Force Economic Order Quantity Type Inventory Data With An Application to Reorder Point Calculations. TR80-6, U. S. Air Force Academy, Colorado Springs, CO.

Morton, Thomas E. 1978. Non-Stationary Infinite Horizon Inventory Problems. Management Science 24, 1474-1482.

Muckstadt, John A. 1973. A Model for a Multi-Item Multi-Indenture Inventory System. Management Science 20, 474-481.

Muckstadt, John A. 1980. Comparative Adequacy of Steady State Versus Dynamic Models for Calculating Stockage Requirements. R-2636-AF, Rand Corporation, Santa Monica, CA.

Nahmias, Steven. 1981. Managing Reparable Item Inventory Systems: A Review. Multi-Level Production/Inventory Control Systems: Theory and Practice, L. B. Schwarz, ed., North Holland Publishing Co., Amsterdam.

Nahmias, Steven and W. S. Demmy. 1982. The Logarithmic Poisson Gamma Distribution: A Model for Leadtime Demand. Naval Logistics Research Quarterly 29, 667-679.

Orr, Donald A. 1977. BARE BONES: A Method for Estimating Provisioning Budget Requirements in the Outyears. ADA044508, U.S. Army Inventory Research Office, Philadelphia, Pa.

O'Malley, T. J. 1983. The Aircraft Availability Model: Conceptual Framework and Mathematics. Logistics Management Institute, Washington, DC.

Parzen, Emanuel. 1962. Stochastic Processes. Holden-Day, San Francisco.

Petersen, Rein and Edward Silver. 1979. Decision Systems For Inventory Management and Production Planning. John Wiley and Sons, New York.

Petersen, J. W., R. M. Paulson and W. A. Steger. 1959. Supply and Depot Repair Interactions: A Case Study of Electronics Support. RM-2365, Rand Corporation, Santa Monica, CA.

Proschan, Frank. 1963. Theoretical Explanation of Observed Decrease in Failure Rate, Technometrics 5, 375-388.

Ross, Sheldon. 1970. Applied Probability Models With Optimization Applications. Holden-Day, San Francisco, CA.

Ross, Sheldon. 1972. Introduction To Probability Models. Holden-Day, Inc., San Francisco, CA.

Shanker, Kripa. 1977. An Analysis of a Two Echelon Inventory System for Recoverable Items. Technical Report #341, Cornell University, Ithaca, NY.

Sherbrooke, Craig C. 1968. METRIC: A Multi-Echelon Technique For Recoverable Item Control. Operations Research 16, 122-141.

Sherbrooke, Craig C. 1971. An Evaluator for the Number of Operationally Ready Aircraft in a Multi-Level Supply System. Operations Research 19, 618-635.

Silver, Edward A. 1972. Inventory Allocation Among an Assembly and Its Reparable Subassemblies. Naval Logistics Research Quarterly 19, 261-280.

Simon, Richard M. 1971. Stationary Properties of a Two Echelon Inventory Model for Low Demand Items. Operations Research 19, 761-777.

Slay, Michael. 1980. The Vari-Metric Inventory Model. Presented at Multi-Echelon Inventory Conference, Miami, FL.

U.S. Army. 1983. User's Guide for the Selected Essential Item Stockage for Availability Method (SESAME) Program. DARCOM PAM 700-18, HQ, U.S. Army Materiel Development and Readiness Command, Alexandria, VA.

Veinott, Arthur F. Jr. 1963. Optimal Stockage Policies with Non-Stationary Stochastic Demand. Multi-Stage Inventory Models and Techniques, Scarf, Gilford and Shelly, eds., Stanford University Press, Stanford, CA.

Vincent, Paul A. 1980. METRIC2: The Mathematical Theory of a Sparing Model for Reparable Items. ORAE Memo No. M102, Ottawa, Canada.

Walkup, David W. and Roger J. B. Wets. 1966. Stochastic Programs with Full Recourse. Document 01-82-0551, Boeing Scientific Research Laboratories, Seattle, Washington.

Wets, Roger J.B. 1979. Solving Stochastic Programs With Simple Recourse. WP#69, Department of Mathematics, University of Kentucky.

Ziemba, William T. 1970a. Computational Algorithms for Convex Stochastic Programs with Simple Recourse. Operations Research 18, 414-431.

Ziemba, William T. 1970b. Two Period Stochastic Programs with Simple Recourse. Working Paper No. 78, University of British Columbia, Vancouver, Canada.

Zmurkewycz, Maria. 1984. Personal Communication.

DISTRIBUTION

COPIES

COPIES	
2	Defense Technical Info Center, Cameron Station, Alexandria, VA 22314
1	Commander, US Air Force Logistics Cmd, WPAFB, ATTN: AFLC/XRS, Dayton, OH 45433
1	US Navy Fleet Materiel Support Office, Naval Support Depot, Operations Analysis Dept, Code 93, Mechanicsburg, PA 17055
1	George Washington University, Inst. of Management Science & Engr., 707 22nd St., N.W., Wash., DC 20006
1	Naval Postgraduate School, ATTN; Dept of Opns Anal, Monterey, CA 93940
1	Air Force Institute of Technology, ATTN: SLGQ Head Quantitative Studies Dept., Dayton, OH 43433
1	RAND Corp., ATTN: Library, 1700 Main St., Santa Monica, CA 90406-2138
1	Dr. Jack Muckstadt, Dept of Industrial Engr & Operations Research, Upson Hall, Cornell University, Ithaca, NY 14890
1	Dept of Industrial Engr, University of Michigan, Ann Arbor, MI 48104
1	Prof E. Gerald Hurst, Jr., Dept of Decision Science, The Wharton School, University of Penna., Phila., PA 19104
1	Prof Harvey M. Wagner, Dean, School of Business Adm, University of North Carolina, Chapel Hill, NC 27514
1	Pentagon Library, ATTN: ANR-AL-RS (Army Studies), Pentagon, Rm 1A518. Wash., DC 20310
1	Prof Leroy B. Schwarz, Dept of Management, Purdue University, Krannert Bldg., West Lafayette, IN 47904
1	US Army Research Office, ATTN: Robert Launer, Math. Div., P.O. Box 12211, Research Triangle Park, NC 27709
1	Prof William P. Pierskalla, 3641 Locust Walk CE, Phila., PA 19104
1	Air Force Logistics Mgt Center, Gunter Air Force Station, AL 36144 AFLMC/LGY
1	Prof Peter Kubat, Graduate School of Mgmt, University of Rochester, Rochester, NY 14627
1	Dr. Warren H. Hausman, Industrial Engr. & Engr. Management, Stanford University, Stanford, CA 94305
1	Dr. Paul Zipkin, Graduate School of Business, Columbia University, 416 Uris Hall, New York, NY 10027
1	Logistics Management Institute, ATTN: Librarian, 6400 Goldsboro Road, Bethesda, MD 20817-5886
1	ALMSA, ATTN: Dan Maksymowicz, AMXAL-L, St. Louis, MO 63188

END

12-86

DTIC