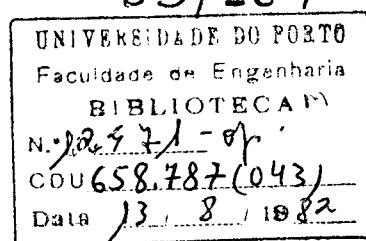# IMPROVEMENTS TO STOCHASTIC

STOCK CONTROL DECISIONS

Bernardo Jose Calafate de Vasconcelos

Thesis submitted in accordance with the regulations concerning admission to the degree of Doctor of Philosophy.

University of Warwick

School of Industrial and Business Studies

April 1982

CONTENTS

ILLUSTRATIONS

ACKNOWLEDGEMENTS

SUMMARY

Recent computer developments have allowed a new dimension to the scientific inventory control. An example is the multi-item inventory model described in Johnston (1980) which represents the departure from the traditional optimization towards decision aid models which the manager can use interactively in the examination of the trade-offs which could improve his current policy.

The Johnston model has however limitations. The scope of this thesis is to overcome some of these limitations, namely, the treatment of the non-captive demand, the determination of the reorder frequency and the extension to a two level inventory system, together with a critical examination of the variables involved, in order to improve the decision making.

In relation to non-captive demand, the problem can become relevant when more than one order overlaps. Then, service levels and average stocks are normally higher than the predictions from formulas derived for captive demand. The main result now achieved is the introduction of the notional control level which relates to the conventional reorder or top up levels and to the lost demand. The notional level allows the extension of established formulations, including the Johnston model, from non-captive to captive demand.

Johnston leaves the reorder frequency to be decided on a practical basis. Here, the same criteria adopted by Johnston have been used to derive consistent expressions for the number of orders. Empirical functions have been incorporated to reach formulas ready for use.

The two level system comprises one main warehouse and its satellites. The analysis covers, basically, the rules to decide the allocations, the theoretical prediction of service levels and the extension of the initial Johnston formulation to this system. The allocation rule derived says that quantities should be allocated so as to have the same probability of depletion. For the prediction of service levels, the depletion time distribution rather than the demand distribution has been used in the formulations, because the conventional approach, based on the latter, does not produce the desired results. Implementable formulas are given for situations in which satellites are of the same order of magnitude.

The results in the three areas mentioned above are accompanied by considerations about the economic meaning of the variables and a method is suggested to cross-check the consistency of the decisions. They are new contributions for the inventory control and constitute an important complement to the initial Johnston model.

Chapter ONE

INTRODUCTION

1.1 Motivation for the research

Inventory control has received much attention in the literature and yet it still is a most challenging topic for the operational researchers. The interest it raises is not purely academic, but stems from a real economic issue. In fact, the average capital tied up in finished goods can be higher than 10% of the Gross National Product (see Waller, 1978), and if these figures give the magnitude of the problem to a nation, they also reflect its importance to each organisation. From a different viewpoint, typical values for the stock turn of finished goods are around 5 and this is equivalent to say that capital tied up is about 20% of the annual revenue. Thus, even small improvements in percentage would bring significant benefits.

Computer packages based on scientific methods for inventory control are widespread. Whether they are suitable or well used has been questioned, for example the following quotation from Wagner (1980, p.447): "One might be inclined to conclude that today companies are faced only with the limited prospect of fine tuning their (inventory) systems to obtain modest improvements. In my experience as well as that of other practitioners who commented on a draft version of this article, major reductions in inventory investment, frequently with an accompanying improvement in systemwide service, can be attained in businesses that already have in place a scientific inventory management system."

Continuing developments in small computers have popularized their use for the control of inventories whose size is large enough to have required a main frame a few years ago. Small computers spread quickly in today organisations and managers become more and more acquainted to them. This favours the use of interactive computer packages for decision purposes and allows a new dimension to the scientific inventory control. Although computing power and storage potential is available at a cost effectiveness many times better than ten or more years ago, most of the current computer systems have not changed in this basic outlook. An opportunity for major improvement comes from integrating the stock control system into the total company management control and from ensuring that the stock control units are internally consistent.

This thesis has followed those two main aims. The first being to relate the controls traditionally associated with stocks (for example, reorder level or reorder interval) with the measures and concepts of more general management planning. It is an attempt to move away from the simplistic criterion of cost minimisation so often behind the traditional systems.

One interactive stock control system which started to incorporate the higher management aims was described by Johnston (1980) and that, together with the situation in which it was used, was taken as a basis for this work. This necessitated the second aim, namely a critical examination

of this model and its inconsistencies. However, a brief description of its situation needs to be made.

The Johnston approach had been implemented in a group of builders and plumbers merchants. The present author contacted them and spoke to the people in charge of inventory management. These meetings gave an opportunity to obtain information about the market environment, the group corporate policy, the physical distribution operations, the inventory structure, etc. and have indeed given the theme for the analysis carried through.

The merchant group is an aggregation of a dozen of merchandising companies whose global turn-over is some hundred million pounds a year. It has more than one hundred branches scattered over Great-Britain. These branches are grouped in four virtually independent distribution areas, each one having a main warehouse. Products are bought out and received at the main warehouse and the branches are supplied from there, i.e., they are satellites to the main warehouse. The inventory in a satellite has an average of 3,000 lines and a main warehouse deals with 7,000 to 10,000 lines.

Each main warehouse is controlled independently of its satellites. A computer based "real time" information system is kept up to date with the stock levels, and external replenishments are triggered on a reorder level policy. The control of the service levels and reorder levels, is based

on the Johnston approach, mentioned earlier.

The stock control in satellites follows a reorder cycle
policy whose decision variables are the reorder level, the
top up level and the cycle length. The reorder level is
calculated from the selected service. For this purpose,
items are classified into hierarchical ranks that are
denoted by 'vital', 'key', 'standard' and so on; each rank
is associated with a service level, typically, 95% for
vital, 93% for key and 88% for standard items where the
figures relate to the percentage of cycles which do not run
out of stock. The cycle length for reviewing the stock
levels is currently one week for all items; but if required
stocks at the end of the day can easily be calculated from
the transaction records. The top up (maximum) stock level is
set so that orders are of a 'convenient' quantity,
considering the item demand and the pack size. Practical
considerations, therefore, are influential in controlling
the stocks.

There is no intention of describing the merchants group in
detail but only to give an idea about the structure, size
and control of the inventory system that inspired the
theoretical framework for the research. Other particulars
will be mentioned later as the analysis goes on. This
analysis, on the other hand, is developed according to the
Johnston approach. His basic criteria are used to extend
previous results to areas which had not been considered but
are relevant to the merchants group.

1.2   Scope for analysis

Inventory control appears in a large variety of situations, uses quite a number of theories and techniques and has been the subject of an enormous amount of specialised literature. Therefore, delineating the province for the problem we are interested in is, more than ever, an essential pre-requisite in undertaking the analysis.

The real situation which inspires the analytical model is the merchants group. In that organisation, products are bought out ready for sale. The inventory system, therefore, relates to finished goods and can be decoupled from the production. Thus, integrated approaches such as the recent developments in "materials requirements planning", can be left aside. Further comments on this topic can be found in Brown (1977) and Wagner (1980).

The length of the decision horizon helps the drawing of another borderline. The research, here, will concentrate on tactical decisions associated with the daily management of inventories. Physical distribution decisions such as the number, location and capacity of warehouses, trunking methods and fixed assets, may have implications on inventory control: however, they are strategic in relation to inventory decisions as their effects last much longer. Therefore, physical distribution problems, including the scheduling of deliveries, will not be considered here. Basic

approaches to the physical distribution are found in Eilon et al (1970), while Mole (1979) gives a survey of more recent work on vehicle routing and scheduling.

Having removed production and distribution functions from the main area of concern, we remain with inventory control in the strict sense. The mathematical treatment of decisions in inventory control has been tackled in different ways. A brief reference is given below in order to introduce and localise the approach and the concepts that are going to be adopted in the analysis.

Consider, firstly, the model that expresses the relationships among the variables. The most frequent approach uses a stationary model, i.e., a model which comprises a single decision period, only. Subsequent periods are ignored for decision, an attitude called sometimes "myopic policy". Conversely, dynamic models contemplate a number of decision periods: the decision is taken only for the first period, but all of them are considered, according to a methodology borrowed from dynamic programming.

Dynamic models are theoretically powerful and might be of interest when stocks are held for speculative purposes or when the market is unstable but predictable. For instance, speculating with price fluctuations and capital investment; or dealing with fashionable goods, demand seasonality, finance availability and taxation. For most of the current

situations, however, the benefits from the use of dynamic models are irrelevant and hardly justify the heavy computational burden those models bring. Full details in dynamic models are given in Veinott(1966).

The myopic policy will be used here because it is also the one adopted by Johnston and will be most convenient for the attempt at extending his approach. Furthermore, single period models are much easier to manipulate and this should not be overlooked.

Most of the inventory control decision models treat demand as a random variable. This assumption normally involves two inter-related aspects: the choice of a probability distribution and the estimate of the distribution parameters.

Many of the traditional probability distributions have been indicated as suitable to fit demand data at one or other occasion. The main requirements of demand distributions are that they should be defined in the positive domain only (or almost) and that they should be easy to handle. Fortuin (1980) compares five of the most popular ones, namely, Normal, Logistic, Gamma, Log-normal and Weibull, in a particular application and concludes that differences in control are negligible. This conclusion might apply especially in relation to fast moving items and high service policies.

The Gamma distribution will be used through this thesis whenever a demand distribution has to be specified. The reasons for this choice are three fold. First, the arguments in Burgin (1967), namely, that the Gamma family is easy to handle, it fits the data very well and covers a wide range of distribution shapes. Second, because the good fit to demand data was confirmed during the contacts within the merchants group. Finally, because that distribution was also used by Johnston.

The distribution, as said before, is only one side of the demand characterisation. The other is the forecasting of the demand parameters. These are normally two, the mean and the variance. A comprehensive survey of the current forecasting techniques can be found in Fildes (1979) and details about the methods commonly used for inventory control are given in Lewis (1981a). Special forecasting techniques may be required for slow moving items, particularly if they have intermittent or lumpy demands. Croston (1972 and 1974), Johnston (1975) and Ward (1978) tackle this problem.

Slow moving items normally have to be dealt with separately not only because they require special forecasting techniques but also because they are often associated with specific policies, e.g. safety in relation to spare parts, trade image in relation to luxurious products, etc. Furthermore, the inclusion of slow movers may prevent some convenient simplifications usually adopted in the course of the analysis. These reasons prompted the exclusion of slow

moving items from our study .

The decision making implies an objective. In the most traditional approach, the objective for inventory control is minimising the annual cost of running inventories. Formally, such an objective could be set as

$$\text{Min} \left[C = Co + Ch + Cs + Cp \right] \qquad (1.1)$$

where 'Co','Ch', 'Cs, and 'Cp' denote the so called ordering, holding, shortage and product costs, respectively. Due to the nature of the model, only variable costs (i.e., those affected by the decision) need to be considered.

The ordering costs 'Co' stand for all the expenses involved in making the product available at the warehouse. Often these costs are assumed to be proportional to the annual number of orders: then, the proportionality constant is denoted by procurement or reordering cost rate.

The holding costs 'Ch' reflect the cost of the capital invested in stock and additional expenses in storing, maintenance, depreciation, insurance, etc. For convenience, the holding cost is usually made proportional to the average investment in stock. The proportionality constant is normally called holding or carrying charge.

The shortage costs, 'Cs' stand for damages caused by the item demanded being out of stock. Such damages may be

diverse. In a merchandising company they might include loss of profit, expediting expenses, contractual compensations, loss of customers good will, etc. Shortage costs are difficult to quantify mainly because they depend largely on the customer reaction. Usually it is assumed that the shortage costs increase linearly with one or more measures for the run out.

The measures for service and run out may sometimes give rise to confusions. Here, the following terminology will be adopted:

'Stockout' refers to having no stock to meet demand. The 'stockout rate' is then the proportion of replenishment cycles in which stockout occurs. This rate will be denoted by 'P'.

'Shortage' refers to each unit short, i.e. to each unit demanded when the item is out of stock. 'Shortage rate' is then the proportion of demand not met ex-stock. This rate will be denoted by 'V'.

The proportion of time out of stock is sometimes mentioned in relation to the service; actually, the expected value for this proportion is the same as for the shortage rate 'V'. Usually, the service levels are taken as the complements to one of the rates defined above, i.e. (1-P) and (1-V).

The product cost 'Cp', sometimes called 'annual usage value' is the buying cost corresponding to the annual throughput.

This throughput may depend on the service offered by the company, and the buying price may depend on the quantities ordered each time. In both cases 'Cp' would be affected by the final decision, so it should appear in expression (1.1). Most often, however, this term is not included in the first approach to a solution: it is considered for the final decision only and if the effects on 'Cp' of quantity discounts or other factors are expected to be relevant.

Hadley et al (1963, section 2.6) have shown that the objective of minimising the annual cost is strategically equivalent to the maximisation of the annual profit. The expression (1.1) is, on the other hand, a generalization of the function proposed in Wilson (1934) which led to the classical economic order quantity. These objectives have still a large acceptance, though they are sometimes opposed: Eilon (1962, 1964), for instance, recommends instead the maximisation of the profit per replenishment; and Burgin (1967) suggests the return on capital employed, a criterion which will be examined in section 2.3.

The approach which is being discussed, whatever the objective function considered, intends to reach an 'optimal' solution. The optimal solution would be the right target if the terms in the objective function could be precisely calculated. This is rarely the case: none of the costs included in expression (1.1) is a simple function; but even if those functions were linear there was still a large margin of uncertainty because of the difficulties in

estimating the cost rates (see Hadley et al, 1963, chs 1,9).

A different approach to decision problems is gaining popularity. Taking advantage of the present computing facilities, models are designed to be used on a "what-if" basis. Rather than pretending to give an optimal solution, those models illustrate the consequences of alternative decisions in order to aid the final decision making. This has been the attitude adopted by Johnston. Chapter 2 will examine this topic in detail as an introduction to the Johnston model and to the analysis in the subsequent chapters.

1.3   Guide to the reader

The broad objective of this thesis is to extend the theoretical results in Johnston (1980) to areas which have not been contemplated in his model.

In chapter 2, the transition from traditional optimising models to decision aid models is examined through some examples, and eventually, the Johnston model is introduced as a decision aid model. The characteristics of the model are discussed in detail and possible extensions indicated.

In chapter 3, the Johnston model is extended to situations of overlapping orders, i.e. when more than one replenishment order can be outstanding. The formulas traditionally used to estimate the service and the stock levels from the values set for the decision variables, assume captive demand. If those formulas are used when demand not met ex-stock is lost and orders overlap, they tend to underestimate the levels mentioned above. New expressions are derived there which cater for that effect.

In chapter 4, the analysis covers the reordering frequency, a variable which Johnston left to be set on a practical basis. Formulas are derived to calculate a value for that frequency, consistent with criteria established before.

In chapter 5, a typical centrally controlled 2-level

multi-item inventory system is examined and eventually, formulas are derived to predict the service levels from the values set for the decision variables. Then, the earlier criteria are used to complement the Johnston model and enable its use for the control of the system.

In chapter 6, a summary of the analysis along the thesis is provided, the main achievements listed and shortcomings examined. Finally, a mention is made to topics which require further investigation, and to neighbouring areas which might be scope for subsequent research.

The contents of each chapter is divided into headed sections and some of the latter end with a subsection. A subsection is an appendix to the respective section and contains comments, mathematical derivations and other details which have been withdrawn from the main body of the section for sake of clarity. These subsections are not essential for the understanding of the main argument of the chapter; so, they might conveniently be left in a first reading in order to avoid being distracted by minutiae.

A computer-like notation borrowed from Basic and Fortran is used for the analytical expressions. An attempt was made to keep that notation consistent and use the same symbols for the same variables. The notation symbols are repeatedly explained at convenient places, nevertheless a list is given

below for general reference.

1.3.1   Notation details

Variables are denoted by an alphanumeric string starting by a letter, and indices of indexed variables are given between brackets. The multiplication symbol is always explicit and operations follow the normal hierarchy: exponentiation (**), multiplication or division (*,/) and addition or subtraction (+,-), subordinated to the parenthesis rules. For instance:

F2(i,j)/A**2 reads as F2(i,j)/(A**2)

Note the indices 'i,j' in the variable 'F2'

The following symbols are consistently used:

A= stocking factor

a= sales loss fraction, i.e., the proportion of demand not met ex-stock which is lost for sale

B0= n*FI/F , n=avge. no. intermediate cycles (see 5.139)

B1= F/F1 : shortage penalty ratio

B2= (F2+λs)/F1 : holding ratio

B3= (F0+λn)/(d*L*F1) : ordering charge ratio

B4= (L/52)*(B2/B1)

B5= B3/B1

B6= B3/B2

c?= constant value for the variable '?'.

D= demand in a relevant period of time, e.g., the lead time. It may stand for 'Dm' when there is no risk of confusion.

Dm= mean of 'D'

Ds= standard deviation of 'D'

Dv= variance of 'D'

Dc= (Ds/Dm) : coefficient of variation

d= demand per unit of time. It may stand for 'dm' when there is no risk of confusion.

dm,ds,dv,dc: the same as for 'D.

DOQ= D/Q

$E\left[?\right]$ = expected value for '?'.

EXP(?)= exponential of '?'.

F=(a*F1+F3)

FI= as 'F' but in relation to intermediate cycles.

FO= cost per replenishment order.

F1= profit per unit of cost.

F2= holding cost per year and per unit of cost.

F3= penalty per unit of cost of a unit short.

f(x)= probability density function.

G= (Dm/Ds)**2 : modulus.

g: the same as 'G' but relating to 'd'.

I= total number of items.

i= index for item.

INT(?)= integer part of '?'.

J= total number of families of items.

j= index for family.

K= total number of satellites.

k= index for satellite.

L= reorder lead time. It may stand for 'Lm' when there is no risk of confusion.

Lm,Ls,Lv,Lc: the same as for D.

LOT= L/T

M(?)= maximum value between '?' and zero.

ML= $(52*F1*d-\pi)/(FZ+F0+\lambda n)$: margin loss.

mt= average in time of the number of orders outstanding.

mo= number of orders outstanding just before a new order being raised. It may stand for 'mom' when there is no risk of confusion.

mom= expected 'mo'.

m= (a*mom) : in reorder level policies;

= a*(mom+1) : in periodic review policies.

N= number of orders per year.

OU= 1/U

P= stockout rate, i.e., proportion of cycles that run out of stock.

p= probability of depletion in the allocation to satellites.

Q= reorder or delivery quantities.

R= notional nominal control level (reorder level or top up level, according to replenishment policy). 'Nominal' stands for on hand plus on order minus backlog.

RO= actual nominal control level (as for 'R').

RZ= aggregate shortage reorder level.

ri= return on investment.

rmi= return on marginal investment.

S= average stock.

T= replenishment period. It also stands for 'Tm' when there is no risk of confusion.

Tm,Ts,Tv,Tc : as for 'D'.

U= stock turn, i.e., the number of times per year the stock of the average year is renewed.

V= shortage rate, i.e., the proportion of demand not met ex-stock.

Y= nominal stock level.

Yh= on hand stock level.

YO= on hand stock in MWH.

YOf= stock in MWH relating to the clearing level.

Z= shortage quantity per cycle. It may stand for 'Zm' when there is no risk of confusion.

Zm= expected 'Z'.

Zt= number of days of stock out per cycle. It may stand for 'Ztm' when there is no risk of confusion.

Ztm= expected 'Zt'.

∂?= symbol for partial derivative of '?'.

Δ?= symbol for finite difference of '?'.

λ = Lagrangean multiplier.

π = profit.

Chapter TWO

COMMENTS ON THE BASIC MODELS

2.1   Optimizing and decision aid models

The inventory models examined in the previous chapter
attempt to optimise an objective function. The solution
depends on the objective aimed at, and importantly, the
position of the optimum will depend on the values of several
parameters or variables, exogenous to the model, and whose
estimation often presents a lot of uncertainty. These
parameters include the cost rates per order, per pound
invested in stock and per shortage. Such costs can, in part,
be assessed objectively, the remaining being an attempt to
quantify fuzzy relationships of preference between
conflicting objectives. This last component makes those
parameters intangible in a large measure.

The analysis in Boothroyd et al (1963) suggests that
accuracy in estimating the exogenous cost variables is not
crucial. It is shown that if estimates of the run out cost
charge are twice as much as the actual value, the total
system cost does not suffer more than a 2 or 3% increase. A
similar increase would result from a 30% overestimate on the
carrying charges. We may question the reasons for and
the meaning of such an insensitivity. The model explains
that, in the first case, the benefits of a higher customer
service compensate the cost of carrying extra stocks; in the
second, stocks would decrease and the benefits of it would
provide the trade-off for the higher shortages. Therefore,
incorrect estimates to the degree mentioned above, important

though they may appear, would not significantly affect the final perfomance or, at least, the perfomance as measured by the model. Graphically, it means that the total cost curve is flat near the optimum.

Arguing that it does not matter very much if some of the parameters estimates are not so good is not entirely satisfactory. The cases mentioned above may lead to comparable situations in terms of computed costs, yet be quite distinct in the stock levels and service to the customers resulting in a very different stock policy. Often, the sensitivity of the mathematically formulated objective function does not match entirely the sensitivity of the decision maker. This is because the model is a simplified and distorted vision of the reality, thus reducing the dimensionality of the problem and the scope of the different balancings that constitute the 'management art'.

We can expect disadjustments, sometimes profound, between the perfomance measured by the model and the perfomance felt by the manager. Borrowing concepts from Keen et al (1978,Ch.1), if the decision procedure were completely structured, i.e., if all possible perfomance variables (objectives) were identified, if the preferences relating each other were defined and if the logic sequence of decision stages were established, then the process could be automated to find 'the' solution... and the decision maker could eventually be dismissed. However, structured tasks do not occur so often in business. And this is not only because

the decision process is not very clear, but also because of the difficulty in establishing, once for all, the preference relationships between perfomance variables. In fact, the relative merits of different outcomes are continuously changing with the market environment, the corporate policy, the constraints upon the stock sub-system, etc. Being so, this sort of decision processes may eventually be pushed to a higher level of structuring, but they can never become structured altogether.

It is not intended to discuss the structuring of the decision process. However, if the analyst examining a task were let to question the manager about the sequence of decisions implied in the task, and the judgements backing each of the decisions, he might be able to extract underlying preference criteria. These, just as the preferences themselves, are likely to be time and place dependent; yet, some of them may appear more robust than others. Further, from the analysis of the criteria themselves, we can eventually ascend to higher rank criteria, necessarily more robust, walking in this way towards a higher level structure. Arriving at the complete structuring would mean having identified that essence which backs the decision criteria.

Decision aid models are erected from (relatively) robust criteria. Their function is to reduce the whole set of feasible solutions into a subset of consistent alternatives, by eliminating those which definitely have no interest. In

this way, the decision maker can concentrate his attention on a limited range of solutions in order to choose, among them, the most preferred. Note, the selection of the most preferred solution is valid for a time and a place as, a priori, no descrimination could be made inside the range of solutions; otherwise, such a descrimination would have been included in the model. Consistently, any solution inside the range shall be preferred to any feasible solution outside.

Obviously, the level at which the decision process has been structured, the robustness of the preference criteria and the effectiveness of the decision aid model are interrelated. If the structuring had reached completeness, criteria would give place to law(s) and we could talk about an optimizing model in the strict sense. Conversely, a model for a poorly structured process, if it is consistent, cannot be very discriminating and, therefore the range of solutions produced is large.

Returning to an earlier point, the cost parameters mirror, in part, the relations of preference, amongst perfomance variables, of the decision maker. Often, such relationships are very 'soft' in the sense that no robust criteria have yet been found to explain them. In such circumstances, it is naive to look for optima. We cannot do better than creating a decision aid model based on criteria of 'recognized robustness' in order to produce that range of solutions which would deserve to be looked at. The phrase 'recognized robustness' means that such criteria have been stable and

are expected to remain stable through the life of the model; moreover and not less important, it means that users of the model are acquainted and agree with them.

Choosing 'the' solution among the range of solutions may not be an easy task. The decision maker has to ponder a variety of combinations, each one a multiple trade-off involving the whole set of perfomance variables. The decision aid model may help such a choice by providing cross checks for consistency. For instance, when the manager 'feels' that a service level of 92%, say, is preferred to one of 89%, he should be provided with the implications of it, such as the extra capital involved, the expected extra revenue on sales, the profits, the assumed penalties for shortages, the marginal cost of the capital invested, etc., i.e. a set of yardsticks which he may like to look at. Then, if he still feels that 92% is preferred, he is probably right or at least consistent.

It is not the mathematical formulation that distinguishes optimizing from decision aid models, rather, it is their philosophies, the attitudes of their users and the problems they are directed to. The decision aid models bridge the gap between the theorist so concerned with the exactitude that he misses the action; and the manager who has to act anyway, but would like to have a good theory backing his intuition. Such characteristics are the subject of this chapter.

2.2   Earlier trade-off approaches

Many models include a parameter designated the carrying (holding) charge to reflect the inconvenience of tying the company resources in stocks. The problem in evaluating this charge starts with the concept itself and, in consequence, the sort of charges which shall be included. Most people agree that carrying charges shall reflect stock maintenance costs, borrowing interest rates, opportunity costs, but finding a figure is often controversial.

Brown (1967,tm3) uses the carrying charge in a quite different way. He says (p.31) "the carrying charge is a kind of scale factor that expresses one's relative reluctance to invest money in stocks". In this perspective, he suggests that stock levels be computed for different values of the carrying charge, according to the particular model being used. In his case, the Wilson formula is adopted to compute the economic order quantity 'Q' for each item 'i',

$$Q(i) = \sqrt{2*F0*D(i)/(F2*C(i))} \qquad (2.1a)$$

where

    F0= reoder charge;

    F2= holding charge;

    C(i)= buying price;

    D(i)= anual demand.

The average investment in stock, 'S' is given by

$$S = h * \sum C(i) * Q(i) \qquad (2.1b)$$

where 'h' is an averaging constant. 'S' is a function of the value used for the carrying charge 'F2'. Thus, an exchange curve relating carrying charges to investments can be pictured, then, letting the management "decide what value of the carrying charge to use" (p.39) in order to get a convenient level of investment in stock.

The point to emphasize is that Brown's proposal brought different views about the meaning of carrying charge. Rather than reflecting a real cost, as traditionally was supposed to do, it becomes purely the scanning knob used in the search for the best compromise. Thus, the stock level becomes, explicitly, a variable to be decided upon instead of being the mere effect of pursuing an overall objective as, say, the minimum total cost. And the carrying charge, traditionally supposed to be a relatively objective parameter, turned out as something intangible and not very precise, being the consequence of an intended stock policy. A simple reinterpretation of the model has enabled to pass from a 'optimum' solution to a range of solutions within which only the very decision maker is able to discriminate.

A different approach is described in Thomson (1967, ch5) and in Lewis (1970, ch8). In this approach, a restriction is added to the usual minimum cost model, that imposes a maximum investment in stock. Formally, the restriction could

be written as

$$h * \sum C(i) * Q(i) \leq cS \qquad (2.2a)$$

where the l.h.s. stands for the investment in stock and 'cS' is a constant that fixes the maximum stock investment. The inclusion of this restriction in the Wilson model would lead to a reorder quantity defined by

$$Q(i) = \sqrt{2*F0*D(i)/((F2-2*h*\lambda)*C(i))} \qquad (2.2b)$$

where 'λ' is the Lagrangean multiplier associated with the restriction.

The value of 'λ' compatible with (2.2a) could be easily calculated as shown in Lewis. Alternatively, 'λ' might be manipulated to produce sets of values for the 'Q(i)'s of all the items and hence for the investment in stock as calculated from (2.1b). An exchange curve relating 'S' to 'λ' could eventually be drawn and used in the way Brown suggested for the exchange curve between stock and carrying charge. The two approaches are, in fact, strategically equivalent: a simple generalisation of the carrying charge concept would make formulas (2.1a) and (2.2b) alike.

2.3   The Burgin-Wild model

Stock decisions based on exchange curves have been contemplated also, in Burgin et al (1967), though the approach in itself did not receive much emphasis. The paper deals especially with finished goods and looks for "a logical basis for the amount of money to invest in inventories". According to the authors such a basis could be borrowed from the accountancy concept of 'return on capital employed', or 'return on investment', here denoted by 'ri', which is defined as

$$ri = \pi 1/(J+S) \qquad\qquad (2.3)$$

where '$\pi 1$' is the profit, 'S' is the investment in stocks and 'J' stands for other assets.

The main objective of having stocks of finished goods is to meet immediatly a customer demand whenever it occurs. Being out of stock has, potentially at least, damaging effects on sales and, therefore, on revenues and profits. If more capital is invested in stocks, the average stockouts should be lower, assuming the investment is fairly distributed by the items in stock. Hence, some sort of relationship between profit and average investment in stock could be established.

The concept of profit in that paper is not totally clear. It refers to a profit per item which is assumed constant, no matter the level of stock. This makes sense only if it

relates to some gross profit before holding costs having been deducted. Such an interpretation will be assumed.
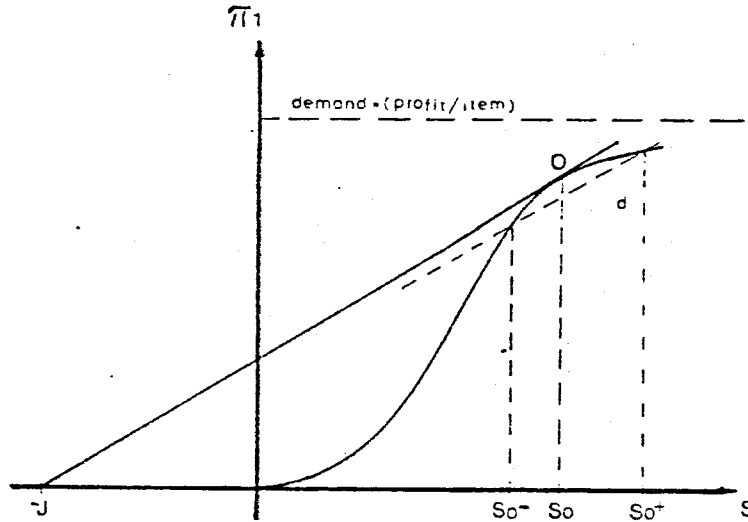
Fig. 2.1 - The profit-stock curve

Fig 2.1 depicts a typical profit-stock curve. The return on investment as from (2.1) would be the tangent of the angle formed by the straight line drawn from (-J,0) to the curve, and the horizontal axis. The maximum angle would be for the straight line to the point 'O' which, thus, would determine the optimum stock 'So'. The picture makes apparent that we can easily fall in a wide range of indifference, (So-;So+) say, for which the return on investment is roughly the same. Recognizing this, the paper moves on and modifies the criterion into "maximize return on capital employed by optimal deployment of stock subject to a given total stock investment". In a condensed form, it coud be written:

$$\text{Max } (ri) \quad \text{subject to} \quad S \leqq cS \qquad\qquad (2.4)$$

where 'cS' stands for a constant value of stock.
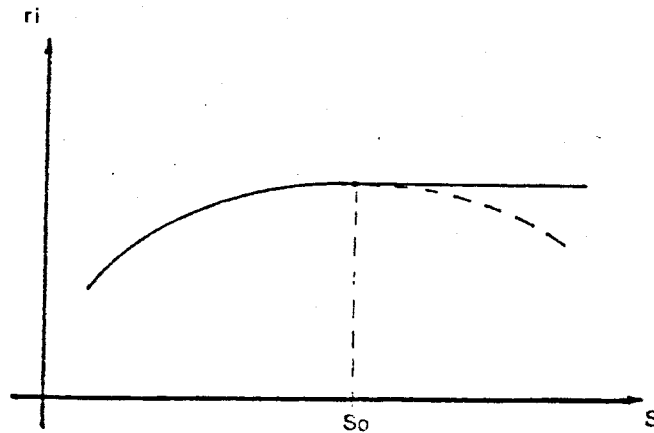
Fig. 2.2 - Return on capital as a function of stock

Actually, in B-W model there is an equality instead of the inequality constraint in (2.4). That, in fact, would correspond to follow the dashed line in fig. 2.2 instead of the solid one, when the availability in stock is greater than 'So'. This, however, is more a theoretical detail without any practical effect; if 'ri' were to be maximised, decisions based in either of the curves would, surely, be the same.

B-W model contemplates multi-item inventory situations. The article presents expressions to enable the computation of the reorder level and the replenishment quantities for each

item. The criteria of maximizing return on investment governs the allocation of 'S' to inventory lines. The method is further particularized to situations in which the lead time demand is Gamma distributed. The model assumes that demand not met at once is entirely lost. Later, in Burgin (1970), an empirical relationship between lost sales and time out of stock is proposed which would extend the applicability of the model. In Johnston (1974) procedures are suggested which facilitate the generation of consistent solutions. More precisely, this author introduces the idea of "knobs" to control the system trade-offs between perfomance variables; and, at the same time, to overcome the difficulties in assessing some of the cost components.

The basic philosophy to retain in the B-W model is the production of an exchange curve for the perfomance variables, namely, the return on capital versus investment on stock. An exchange curve is the range of solutions out of which the final decision is made. The criterion imbedded in the model is to seek the highest return on capital for any level of investment, from which a judicious allocation of capital to lines in stock is derived.

B-W model had a strong impact on the way of thinking about inventory control, nonetheless, for it clearly calls the attention to the necessity of measures of perfomance other than profit or cost. However, the authors put too much emphasis on the return on capital, ignoring others in which the manager might be and in general is interested.

Furthermore, the return on capital ' can raise some controversy as a measure of perfomance.

The return on capital would depend, in fact, on the assets 'J' about which the article is laconic. It may be difficult to decide the values which should be considered and this deteriorates the meaning attached to 'ri' as a perfomance variable. Actually, the only assets that should be considered for the return on capital are those which may be altered by the decision. The assets which are already allocated and, whatever the decision might be, are not to be disposed of, do not matter for that decision. Only the extra investments associated with each of the options should be examined. The extra investments involved in short term decisions about inventory control relate to stocks only, and so, the term 'J' is redundant in the expression of 'ri'.

The return on capital invested as measure of perfomance can be contested because it by-passes an important point that has just been touched: the manager is supposed to ponder the worth of the extra money to be invested. This leads directly to the notion of 'return on marginal investment', 'rmi', formally defined as:

$$rmi = \Delta \pi_1 / \Delta s \qquad (2.5a)$$
$$= \partial \pi_1 / \partial s \qquad (2.5b)$$

the first for discrete increments and the latter for a continuous function. (Note that return on marginal

investment is different from marginal return on investment).
The relationship between 'rmi' and 'ri' is depicted in
fig.2.3.
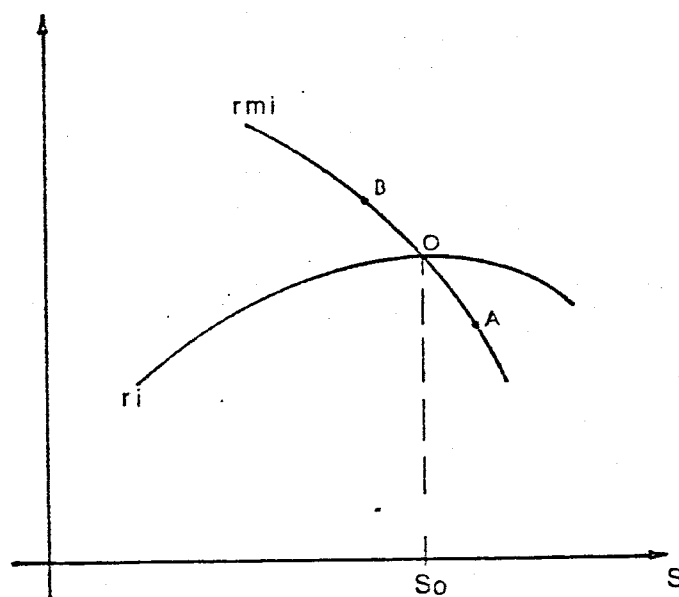
Fig. 2.3 - Return on investment and on marginal investment.

It is more important to guarantee that any pound of capital
gets a return which compares favourably with any alternative
application than working near the maximum return on
investment. In other words, the return on marginal
investment shall not be less than the opportunity cost of
capital (on theoretical grounds, they should be equated). In
fact, having capital available there is no apparent reason
to refrain investing more than 'So', unless there were more
favourable alternatives. If, for instance, the opportunity
cost is at level A (fig.2.3), why should we stop at O?
Conversely, there may be more profitable alternatives which
raise the opportunity cost to B; then it would be wise to

stop the investment about that level.

Thus, the return on marginal investment has advantages over 'ri' , as a basis for decisions. The most profitable allocation of capital to simultaneous investments is obtained by equating the 'rmi', and not the 'ri', for all of them. Note, however, that unlike 'ri', the return on marginal investment is not supposed to be maximized. Its maximum occurs at 'S=0' or close, and it can be shown that 'rmi=ri' for 'S=So'. Therefore, the maximum return on marginal investment is unlikely to be a reasonable policy.

2.4   The Johnston model.

The model introduced in Johnston (1980) has some points in common with the approach in B-W model, namely: penalties for shortages are estimated on the basis of profit loss resulting from the lost sales; capital available for stocks is constrained; a criterion is embodied to allocate capital to product lines and to help generate the range of solutions; the final choice is left to the decision maker. The merits of J-model reside mainly on the refinement of the concepts and on their development for practical use.

The built-in criterion in J-model is the maximization of a notional profit (which in the article is called return). It will be seen, later, that its meaning is not very distinct from that profit referred to in the previous section. For the J-model, the annual profit associated with each product line is:

Notional profit, $\overline{\pi} =$

gross profit	$52*F1*d$
-shortage costs	$-N*(F1+F3)*Z$
-holding costs	$-F2*S$
-ordering costs	$-N*F0$

$$(2.7)$$

where

$F0 =$ cost per order

$F1 =$ gross profit rate

$F2 =$ holding cost rate $\Big\}$ (per money unit)

F3= lost sale cost rate

d= weekly demand-by-value rate

S= average on hand stock-by-value

Z= average shortage-by-value per cycle

N= no. of orders per year

The expression 'by-value' means that physical quantities associated with the variables concerned are multiplied by the respective unit costs. This is merely a way of keeping the formulation shorter.

First consider that inventories are operated according to the conventional reorder level system where 'R' stands for the nominal reorder level and 'Q' for the replenishment quantity. 'Nominal', here, relates to stock on hand plus on order. For simplicity, continuous review and unit withdrawals are assumed, so that, any new order is raised exactly when the level 'R' is reached. This assumption is usual and will be commented in sub-section 2.4.1.

Johnston formulation implies that all demand not met at once is lost; that replenishment orders do not overlap; and that the situations to be dealt with by the model concern only high service levels, so 'Z' is much lower than 'Q' (this might be controversial if the slow movers had not been excluded). Under these circumstances,

$$Z= \int_{R}^{\infty}(x-R)*f(x)*dx \qquad (2.9)$$

where 'f(x)' is the density function of the lead time

demand. The probability of stockout is then:

$$P= \int_{R}^{\infty} f(x)*dx \qquad (2.11a)$$

$$= -(\delta Z/\delta R) \qquad (2.11b)$$

For high service levels and as it is shown in sub-section 2.4.1, the following approximations can be used:

safety stock, $Ys= R-D$ (2.13a)

average stock, $S=Ys+Q/2$ (2.13b)

and $\delta S/\delta R= 1$ (2.15)

For a periodic review control procedure, the expressions (2.9) to (2.15) still hold provided that a different, but consistent, interpretation is given to the symbols. In this case, 'R' is the top up level (maximum order cover); and 'Q' is the average reorder quantity and should be the same as the average depletion in the reorder period 'T'. More details are given in sub-section 2.4.1.

After these introductory remarks we resume the discussions of Johnston's proposals. Though it has not been yet explicitly mentioned, the J-model is intended to deal with a multi-item inventory. The paper considers the possibility of defining 'buying families', each one consisting of items obtained from the same supplier, in order to take advantage of quantity discounts and to reduce distribution costs. According to the author, operations would follow a periodic

review control. At intervals of length 'T', a joint order would be placed raising the nominal stock of each item to its top-up level 'R(i)'.

The problem is then to calculate the 'R(i)'s so that the capital allocated to the whole buying family be restricted to some maximum 'cS'. The investment on each line as well as the expected services come then imposed by the respective 'R(i)'. In fact, with 'R(i)' and using expressions (2.9) and (2.11), 'Z(i)' and 'P(i)' could be calculated. On the other hand, from the reorder period 'T' and from the demand rate 'd(i)', the average reorder quantity would follow as 'Q(i)= T*d(i)'; and the average demand in the relevant decision period is 'D(i)=(L+T)*d(i)', where 'L' is the average lead time. Then having 'R(i)', 'Q(i)', and 'D(i)' the average stock 'S(i)' would follow from expressions (2.13).

The expression for the profit has a structure as in (2.7) but now summed up through the whole family. Formally, it can be writen:

$$\text{Max } \left(\sum \pi(i) \right), \quad \text{subject to} \quad \sum S(i) \leq cS \qquad (2.17)$$

where each '$\pi(i)$' is defined as in (2.7). By introducing a Lagrangean multiplier 'λ' in relation to the constraint, the objective function turns out:

$$\sum \pi(i) + \lambda * \left[cS - \sum S(i) \right] \qquad (2.19a)$$

Then, by taking derivatives to each 'R(i)' and equating to zero, it follows:

$$-N*F(i)*(\delta Z(i)/\delta R(i))-(F2+\lambda)*(\delta S(i)/\delta R(i)) =0 \quad (2.19b)$$

where

$$F(i)= F1(i)+F3(i) \quad\quad (2.19c)$$

and $i= 1,2,...,I$, with 'I' being the number of items of the buying family. Hence, after recalling (2.11) and (2.15), then:

$$P(i)= 1/N*(F2+\lambda)/F(i) \quad\quad (2.21)$$

which, by inversion of (2.11a) determines the 'R(i)'s.

The value of 'F2', the holding cost rate, is assumed not to have significant variation from item to item. 'N' is taken as exogenous to the model. 'F(i)/F1(i)' is implied to be roughly constant across the items of a same family. Then, after rearranging we get:

$$P(i)= T*A/F1(i) \quad\quad (2.23)$$

where 'T' is the average replenishment interval and the new symbol 'A' stands for what the paper calls the 'stocking factor', a constant for each family.

The latter expression above drives the J-model based system.

The stocking factor is simply a device, a 'knob' as it has been labelled by Johnston, used by the decision maker to generate potential solutions. Each one can be assessed by a vector of attributes for the perfomance variables. The perfomance variables referred to in the paper are the average stock, the stock turn, the service level and the expected loss in sales. Others can easily been added if required by the user.

The allocation criterion used in J-model does not differ radically from the one in B-W's. In the latter, the objective is to maximize 'ri= $\pi1/(J+S)$' for each available 'S' which, in fact, is the same as maximizing ' $\pi1$' if the constraint on 'S' is active. ' $\pi1$' differs from ' π', the profit in J-model, insofar as 'F2' and 'F3' are significant in relation to ' λ' and 'F1', respectively. If 'F2' and 'F3' were ignored (see further discussion in the next sections), both the criteria would be the same.

There are many features in J-model which are not mentioned here because they are not so important for the analysis to be carried out through this thesis. They relate to practical points developed to facilitate the user interaction. The whole package is much in line with the concept of the decision aid model. The user is provided with an extensive board of variables and, in this way, perfomances can be appreciated from several angles and in relation to different targets. With practice, the decision maker can learn how these variables interact with each other and develop, at the

same time, a consistent strategy. Better solutions might also be helped by working out logical relationships between variables as it will be discussed in sections 2.5 and 2.6.

2.4.1 Analytical details

The derivations above implied some approximations which will be commented here. First, in relation with the assumption of continuous review of stock and unit withdrawals. When the review of the stock levels is not strictly continuous, but instead, made at short intervals (for instance, at the end of each day), it is often assumed, as in Burgin et al (1967), that the actual inventory position at which the order is placed fluctuates uniformly in the interval (R ; R-X) where 'X' is the mean demand in the review period. Therefore, the actual reorder level would have 'mean= R-X/2' and 'variance= (X**2)/12'. If the review is continuous but demand is lumpy, an approximation can also be obtained by replacing the average quantity issued for 'X' in the above expressions. In this way, we can always relate to the basic assumption, in approximate terms.

Expressions (2.9) and (2.11) imply that any demand backlogged during a shortage is lower than the replenishment quantity 'Q', so all backorders can be supplied after a new replenishment. This is a reasonable assumption when service

levels are high and items are fast movers. Formulas for more general situations are given in Taylor et al (1976).

The expression (2.13a) for the safety stock neglects the shortage 'Z'. In fact, the expression should be:

$$Ys = \int_0^R (R-x)*f(x)*dx \qquad (2.25a)$$

$$= R-(D-Z) \qquad (2.25b)$$

Therefore

$$\delta Ys/\delta R = 1-P \qquad (2.25c)$$

Then, for the average stock:

$$S = Ys+Q/2 \qquad (2.27a)$$

$$\delta S/\delta R = 1-P \qquad (2.27b)$$

The results derived for the reorder level policy also hold for the periodic review policy, but in this case 'R' is the top up level and the relevant period to be considered is 'L+T', where 'L' is the lead time: thus, 'f(x)','D' and 'x' relate to that period.

Consider fig.2.4. If, over a given period 'Lb+T', the demand has been 'Db' and any intermediate shortage 'Za' were entirely backlogged and supplied, then, the shortage at the end would be

$$Zb = M(Db-R) \qquad (2.29)$$

where 'M(x)' means the maximum between zero and 'x'. As
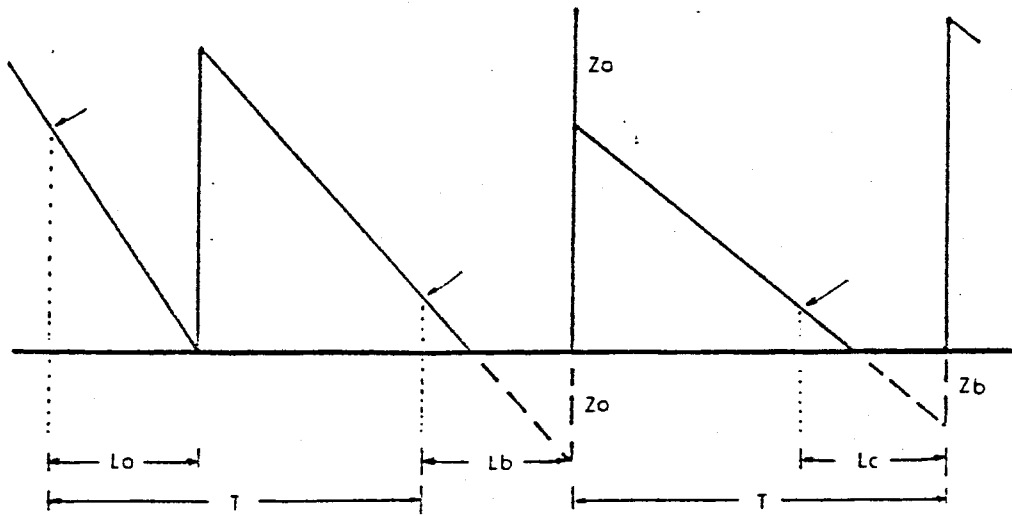
Fig.2.4 - Shortage effects with P.R. control

shortages cause a loss in sales, any intermediate 'Za' will not be issued later and 'Zb', actually, might be lower than the value obtained from (2.29). However, by a convenient re-interpretation of 'R', as it will be shown in Ch.3, expressions (2.9) to (2.13) can still be used. For the time being this should be taken on trust in order to make the following remarks: if the shortage 'Z' is entirely lost for sale, we can expect that

$$Q+Z= d*T \tag{2.31a}$$

So, $\partial Q/\partial R= P$ (2.31b)

and $\partial S/\partial R= 1-P/2$ (2.31c)

This last expression compares with (2.27b): for high services, the r.h.s is in both cases, close to one, hence the approximation considered in (2.15).

2.5   Interpretations of the holding costs.

Neither B-W nor J-model pays attention to the economical meaning of the Lagrangean multiplier. It is known that, in the optimum solution, the value assumed by that multiplier gives the marginal worth of the constrained resource. Therefore, for J-model formulation (2.17), the optimum 'λ' would give the increase in profit if one extra unit of money were made available for investment in stock. Or, in other words, 'λ' is the return on the marginal investment, formerly denoted by 'rmi'.

The return on marginal investment, as already emphasised in section 2.3, should compare favourably with the opportunity cost of the capital and, therefore, with the current borrowing interest rate. For this comparison being meaningful, the interest rate should not have been included in the holding charge 'F2', otherwise, it would be counted twice.

Expression (2.21) shows that the split between 'F2' and 'λ' is quite irrelevant as both can be interpreted as holding costs, complementing each other. In fact, the constraint in the formulation (2.17) could be ignored if in (2.7) 'F2+λ' substitutes for 'F2'; this would lead to the same expression (2.21). Additionally, 'λ' could also be ignored and then 'F2' would stand for the 'true' (i.e., the whole) holding cost. In the latter case, 'F2' might then be used to tune

the compromise between stock and service, very much in the
same way as the Brown proposal referred to in section 2.2.
Alternatively, 'F2' could be ignored and the tuning be made
based on 'λ', only. This illustrates that 'F2' and 'λ' are
not distinct from each other and that one of them is,
actually, redundant.

As the inventory includes several buying families, the
criterion used to share the money within a family should be
the same to share the money amongst the families. Then, for
the item 'i' of the family 'j' we could write:

$$P(i,j)= 1/N(j)*(F2(i,j)+\lambda)/F(i,j) \qquad (2.33)$$

Note that 'λ', being the return, should be the same for all
items. 'F2(i,j)' might stand for some discriminating holding
cost, specific for the item. For example, it may be wise
keeping fashionable or perishable goods with stocks lower
than for more durable items. This could be associated with a
higher holding cost; but, in most of cases it is easier to
fix a maximum stock than estimate the 'F2' required to
produce the same effect. To make it clear, suppose we are
dealing with a product more perishable than the rest of the
inventory; normally it would be more convenient to set a
maximum stock, say 2 weeks demand, rather than finding some
figure for 'F2' in order to push down the stock to that 2
week level.

For most of the inventory situations holding cost rates can

be approximated as being the same for all the goods or, at least, for the goods within each family. As mentioned above, the model reacts to the sum 'F2+λ' and not to each of this terms, individually; therefore, 'λ' can be assumed to include the whole of the holding costs, and in this case, 'F2' could be ignored. Items which require a special policy might be associated with special holding costs; but, as argued above, controlling directly the stock levels is normally easier.

2.6 Interpretations of the shortage costs

The other cost rate appearing in expression (2.33) is 'F(i,j)'. This could be unfolded to

$$F(i,j)= a*F1(i,j)+F3(i,j) \qquad (2.35)$$

Both 'F1' and 'F3' are shortage cost rates. As mentioned in section 2.4, 'F1' is the profit rate. It also corresponds to a shortage rate insofar as the 'Z' units short are not sold anymore. Thus, the remainder shortage costs 'F3' are costs other than direct loss in profit.

In J-model, it is implied that a unit short is a unit lost. Under this assumption, 'Z' units short would represent a loss in profit of 'F1*Z'. Actually, a shortage does not mean necessarily a one-to-one sales loss as some of the customers may wait for the arrival of the next shipment. For a more general statement we write for the expected loss in profit per cycle, 'a*F1*Z', where 'a' can assume values between 0 and 1. Thus, 'a' is the 'sales loss fraction', that proportion of 'Z' which is really lost for sale. The extreme 'a=0' corresponds to the situation of captive demand in which orders not met are backlogged and wait for the coming of the goods. At the other end of the range, 'a=1', the entire shortage is lost for sale. In the real world,'a' must be located somewhere in between.

The value of 'a' may depend on mahy factors: type of business, competition, market segment, customers good-will, existence of substitutes, etc. Across an inventory, 'a' may have   different values from one commodity to another. Moreover, for the same item and for the same environment, 'a' depends on how long the shortage lasts. Burgin (1970) suggests an exponential relationship between loss and time out of stock, as depicted in fig.2.5a. This could be converted into a more tractable form by noting that the shortage quantity and days out of stock have a direct relationship through the demand rate, thus arriving at

$$a = a0 - a1*Exp(-a2*Z) \qquad\qquad (2.37)$$

where 'a0', 'a1' and 'a2' are parameters. Fig.2.5b illustates the role of these constants.
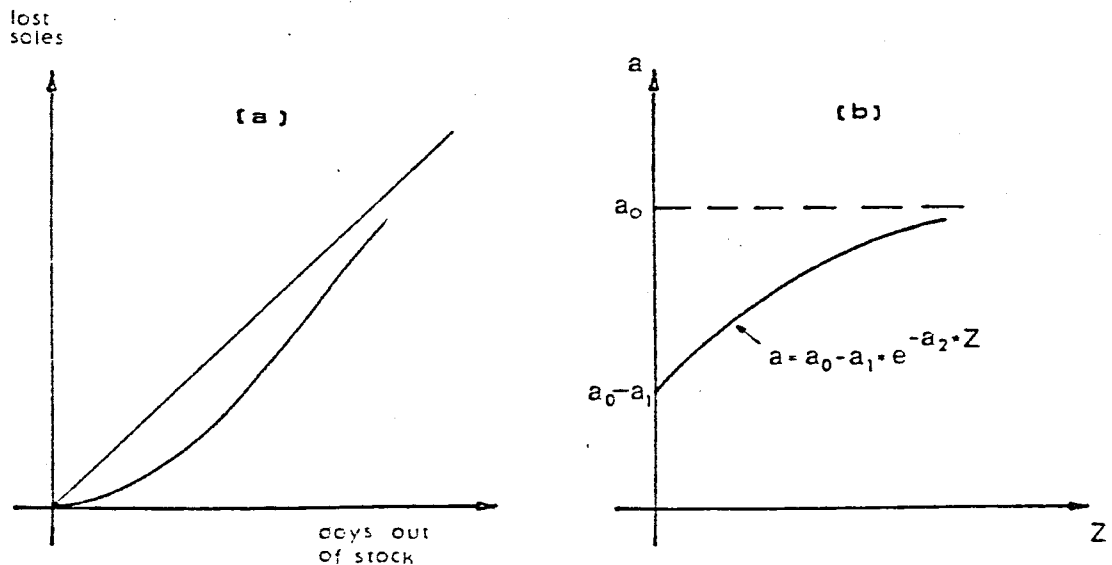
Fig.2.5 - (a): The Burgin curve

- (b): Shortage effect on loss fraction

For most of the situations, 'a0' will be close to 1; it is an asymtotic value for high 'Z' and, in the limit we can expect that customers do not wait if stocks are consistently out. There are, of course, some branches of business where stocks are kept very low or where are item lines not kept at all but such cases are not being discussed here; we are concerned with models for inventory policies of relatively low 'Z'. From the fig.2.5b above, this would lead us to the area of higher sensitivity of 'a' to 'Z'. Therefore, 'F=a*F1+F3' would be highly correlated to 'Z' and, hence, to 'R'. Then, in establishing the allocation rule, an expression different from (2.21) would be reached as a consequence of '$\partial F/\partial R$' having a value significantly different from zero.

Burgin (1970) did not give empirical evidence that his curve fitted the reality. Trade environments are so distinct from each other that customer behaviour patterns in respect to backlogging are likely to vary a lot. Further, these behaviours may change appreciably from one time to another. If we think of an inventory involving several thousands of items, determining the individual parameters for (2.37) may become quite unjustifiable.

Estimating the loss factor on a broad basis, for the whole inventory or by groups of items may be accurate enough for the purpose of control. Practical procedures to perform this will depend on the specific situation. Burgin (1970) has

suggested a way, based on sales records; other procedures could easily be thought of. Estimates on aggregate data, both across the inventory and across successive periods of time, will 'average' the resulting values for the loss factor. Thus, we could talk about an 'aggregate loss factor' as an average value for 'a' within a group of items and for the service levels currently adopted. Trade characteristics would dictate how those groupings should be made. The buying family, because of the inherent homogeneity of its components, appears to be a good candidate. Hopefully, one single aggregate loss factor for the whole inventory will satisfy in many situations.

The aggregate loss factor is, then, an average value for the current shortage range. With that aggregation, 'a' becomes practically insensitive to the variations in the 'Z's for the individual items and acquires robustness through time. By taking the aggregate value for 'a' in (2.35), the derivative '$\partial F/\partial R$' can be approximated as zero and, hence, (2.33) maintained for the allocation rule. Such an assumption will be implied hereafter whenever the loss factor is referred to. The letter 'a' will be used, without any subscript, but it will relate to the aggregation group used to estimate its value.

We have been concentrating mainly on the direct loss in profit caused by the shortage, 'a*F1*Z'. This loss is due to a non-sale of the item short. The shortage of an item may cause giving up a joint order and, thus, lead also to losses

on the sale of other items; also, it may affect future demands, jeopardize promotion campaigns and, if consistent, contribute to undermine the image of the firm. 'F3', the last component of the shortage cost rate in (2.35) relates to these indirect losses whose magnitude, generally speaking, depends on customers good will, on market structure and environment, and on the company intended policy. Quantifying that magnitude is, however, a very subjective matter as it should mirror the manager stance on the environment he perceives.

The allocation criterion as expressed in (2.33) reacts to the value of 'F', no matter how it is split into its components, 'F1', 'a' and 'F3'. 'F1' can be calculated easily and precisely from the profit margin, and 'a' could be estimated as an aggregate factor, in the way described above. 'F3', however, is highly subjective and this makes 'F' equally subjective.

The problem is again the ability to find a quantifiable expression to the manager judgement. As 'F1' is the tangible component in 'F', it may constitute a convenient basis to quantify 'F3' and, eventually, 'F' by means of a relationship 'F=function(F1)' common to all the items in inventory or, at least, in each family. Such a relationship would be fundamental in dictating stock allocations through (2.33): this is the criterion built-in in the J-model under the more compact expression (2.23). But as that relationship should match the views of the decision maker about the

consequences of the shortages, he should be aware of it and be able to intervene and modify the function.

Johnston (1980) simplifies the problem by assuming that 'F=c*F1', where 'c' is a constant in each family. Then, (2.33) would reduce to

$$P(i,j) = 1/N(j)*A(j)/F(i,j) \qquad (2.39)$$

similar to (2.23). On the other hand, Oral et al (1972) claimed to have found a method to compute shortage costs, based on a probabilistic model for customer reactions to shortages. In relation to the specific case they were considering, they obtained a regression of costs to the profit margins and have found an expression with a correlation coefficient better than 90% This, after conversion to our notation becomes:

$$(F*C) \simeq 0.2*(F1*C)**0.9 \qquad (2.41)$$

where 'C' is the cost of the item, so 'F*C' and 'F1*C' are the shortage cost and the gross profit for the item, respectively. Other approaches, either analytical or empirical could be and have been proposed (see, for instance, Walters (1968)). The point to emphasize is that the final choice about the function that is going to be used, shall be left to the decision maker.

The notation 'F' will be kept from now on, standing for

shortage costs. It will be assumed that in each specific case it is possible to make explicit some relationship 'F=function(F1)', the same for the whole inventory (or, at least, for a buying family) representing satisfactorily the managerial views about shortage costs. In a extreme situation where such views do not exist or cannot be expressed in a more satisfactory way, the relationship 'F=F1' can always be taken.

2.7 Extensions to the Johnston model

There are two visible weaknesses in the analytical treatment developed by Johnston. The first is associated with the lost sales assumption according to which, demand not met ex-stock is lost and, in consequence, a shortage causes a loss in profit. The lost sales assumption, that is basic for the allocation criterion, is ignored in the subsequent evaluation of the service levels. In fact, the formulas (2.9) and (2.11) which have been used for calculating the shortage per cycle 'Z' and the stockout 'P' hold when the demand is captive and could be used only as an approximation for non-captive demand situations. During the contacts with the merchants group referred in section 1.1, the errors of these approximations were reported to be significant when the orders overlap.

The second weakness relates to the replenishment period which Johnston assumes to be set on a practical basis. However, the practical sense may be misleading: for instance, one might be tempted to think, as with the classical economic ordering quantity approach, that the reorder costs are the only factor that prevents frequent reordering. This does not apply to the J-model, necessarily, because frequent reordering causes also frequent risks of shortage.

The situation of replenishment orders with overlap and the problem of determining the reordering frequence will be

examined in chapters 3 and 4.

Another point to be considered relates to the fact of the J-model having been devised to the control of inventories located at one place, only. The inventory in the merchants group spreads over several locations, but at present each one is controlled independently. An integrated control, including the main warehouse and the respective branches, might be appealing but the existing theoretical models for that control are very complicate to handle. An attempt will be made in chapter 5 to extend the approach introduced by Johnston into the 2-level inventory system found in the merchants group.

These extensions to the J-model constitute the centre of the research undertaken and will be the subject for the remainder of this thesis.

Chapter THREE

IMPLICATIONS OF NON-CAPTIVE DEMAND

3.1   An introduction to the problem

Customer service is a main reason for stock holding. The lack of service or requests not satisfied immediately may result in lost sales. Lost sales are distinct from shortages. The latter refers to the whole quantity requested and not immediately satisfied, a part of which might be backlogged and supplied when stock becomes available; the other part is never issued, because customers give up, and constitutes the lost sales. This has led to the introduction of the loss fraction 'a' such that, for an expected shortage 'Z' per cycle, the expected loss in sales is 'a*Z'. The extremes 'a=1' and 'a=0' correspond to the total lost sales and backlog cases, respectively. The complete backlog situation is also termed of captive demand and opposes to the non-captive situations in which lost sales occur.

Situations of non-captive demand raise analytical problems when decision periods overlap: the decision period is the supply lead time in the reorder level policy and the lead time plus the review period in the periodical review policy. That overlapping occurs whenever more than one order is outstanding, but in the periodical review policy it happens even if only one order outstands. These difficulties have been referred to by Hadley et al (1963, section 5.13): "it (is) not possible to treat rigourously the lost sales case when more than a single order (is) outstanding. For periodic review systems, things are even more difficult and trouble

is encountered even in treating the case where only a single order is outstanding". The loss in sales is a central point in the approaches followed by Burgin and Wild and by Johnston. These authors have avoided the problem by assuming that no more than one order is outstanding, and yet this restriction is often unreal.

Practitioners know that, when orders overlap and shortages cause a direct loss in sales, then stocks tend to be higher and shortages tend to be lower than the values expected from the reorder levels being adopted. These aspects have not received much attention in the literature and yet differences are significant when the degree of overlapping is high. This will be discussed in the next sections in relation to the reorder level and periodic review operating policies, however, the approach followed is likely to be valid for other models. Importantly, the derivations lead to a new interpretation of the reorder and top up levels for replenishment so that the current formulations can be extended to situations of lost sales and overlapping orders without further complication. This overcomes the difficulty mentioned in section 2.4 in connection with the periodical review operating control and eliminates any need for restricting the number of outstanding orders.

Some usual assumptions are implied in the course of the derivations. For instance, the independence of the demand in non-overlapping periods, the independence of the lead times and the non-overtaking of orders. Admittedly, the last two

assumptions are contradictory. The non-overtaking means that orders arrive in the same sequence as they are raised. This restricts, in theory, the length of successive lead times. In practice we can assume that the first quantity arrived relates to the first ordered. The distortion this introduces will be small relative to other assumptions (e.g. demand independence). Another assumption is that the whole quantity associated with any order arrives at one time and is larger than the total of backorders then outstanding. Therefore, all backorders will be supplied on arrival of a fresh replenishment. This is perhaps acceptable with the low shortage levels assumed in J-model. The above approximations are so standard that they are often not made explicit. They will be assumed in this work. Other approximations will be used on one or two occasions. They will be commented upon as introduced. They are often needed to achieve tractable results though such approximations proved to be reasonable after having been checked by simulation. This point will be further discussed at the end of the chapter.

3.2   The effects on service levels.

Let us consider the classical reoder level operating policy.
With this policy, a fixed quantity 'Q' is ordered when the
nominal stock falls to the so called reorder level. The
nominal stock is the stock on hand plus on outstanding
orders minus the quantity backlogged. It is always assumed
that quantities backlogged are going to be delivered later,
i.e., backorders are firm. Continuous review and unit
withdrawals are considered on the grounds introduced in
section 2.4. This implies that the nominal stock is actually
at the reorder level when a new order is raised. Let us
denote by 'RO' that reorder level. Initially, reorder
decisions are supposed to be independent from item to item,
so, only one item is considered in the analysis. Later, the
results will be extended to group replenishments.

For illustration purposes, consider deterministic processes
with shortages as those represented in fig.3.1. Here, the
replenishment lead time is 'L=5 weeks', the demand rate 'd=2
units', the lead time demand 'D=d*L=10', the replenishment
quantity 'Q=4', the reorder level 'RO=9'.

The item runs out of stock because 'RO' is lower than the
lead time demand. Denote, as usual, by 'a' the lost sales
fraction. Fig.3.1(a) relates to the total backlog case
'a=o'; and fig.3.1(b) to the total lost sales case, 'a=1'.
We can see that, on average, the shortage is higher in the

(a)

nominal
stock

O₄

O₃
O₂
O₁

a = 0
L = 5
d = 2
D = 10
RO = 9

stock
on hand

time

(b)

nominal
stock

O₅

O₄

O₃
O₂
O₁

a = 1
L = 5
d = 2
D = 10
RO = 9

stock
on hand

time
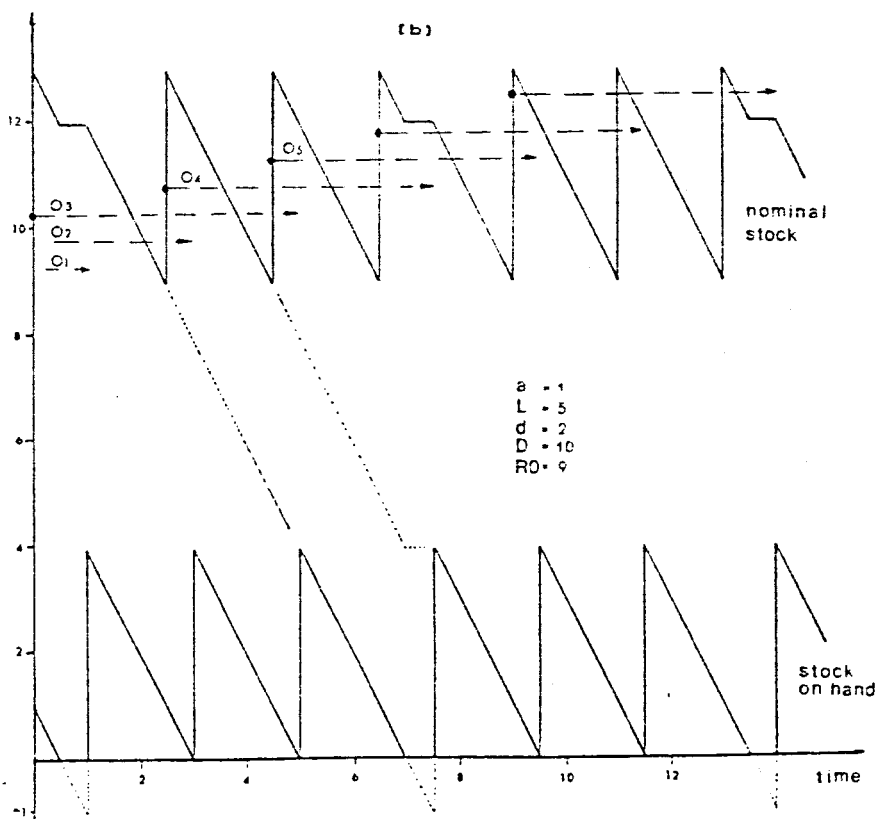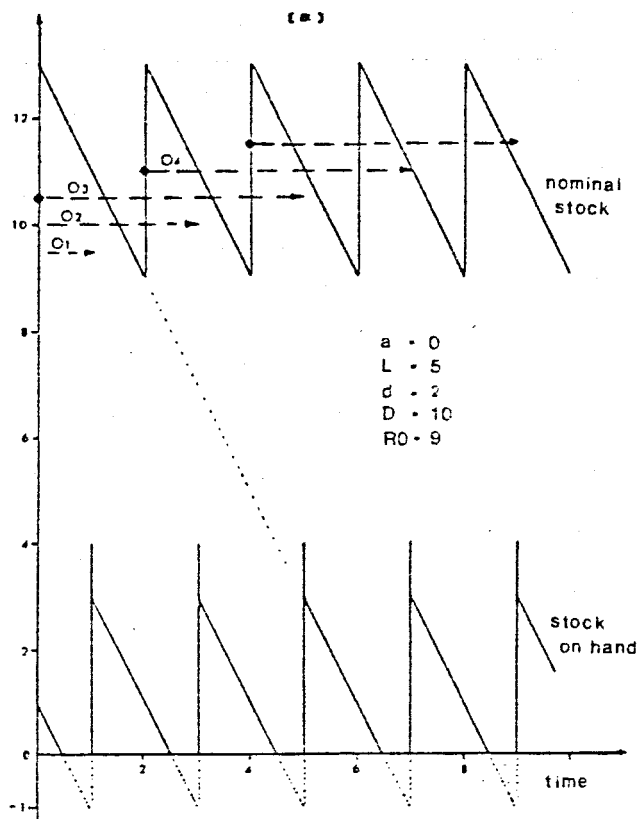
Fig 3.1 - Shortages with reorder level policy

(a) backlog case; (b) lost sales case

former than in the latter in spite of demand rates and operating policies being the same. The reasons become apparent after comparing the two diagrams.

In the diagram (a), the shortage associated with each replenishment period, shown just before a new receipt, is consistently 1 unit: exactly the difference from 'R0' to the lead time demand. In figure (b), shortages are also of 1 unit but take place every 3 periods, only. Examining the latter, it can be seen that, when the shortage of 1 unit occurs during the first week, there are 3 orders outstanding, those indicated by 'O1', 'O2' and 'O3'. Now, we need to assign a given stockout to one, and only to one, lead time period in order to avoid multiple counting. The shortage of 1 unit took place in each of the lead times 'O1', 'O2' and 'O3'. That unit is, however, 1 unit for all of them. By convention the shortage will be assigned to the oldest order outstanding (the next to arrive) which is the order 'O1'. All the shortages in Fig.3.1(b) correspond to lost demand; if the loss is assigned to order 'O1', the demand left for the lead time 'O2' is, in fact, reduced by that loss. The same applies to the remaining demand over lead time 'O3'. Therefore, the lead time demand to be covered by 'R0' is the remaining demand, i.e., after deducting any eventual losses which have occurred in that lead time and assigned to earlier orders.

Denote by 'Z(1)' the shortage assigned to order 'O1'. The

remaining lead time demand in relation to orders 'O2' and 'O3' is 'D-Z(1)=10-1=9', which is the same as 'RO'. Therefore Z(2) and Z(3) are zero. In relation to 'O4', however, the shortage is 1 again, because the older orders co-standing with 'O4' have no lost sales assigned to them.

Let us denote:

by 'Z0(k)' and 'Z1(k)' the shortage quantities assigned to order 'k' for the total backlogging (a=0) and total lost sales (a=1) cases respectively; and by 'Z(k), simply, the shortages for non-specified values of 'a';

by 'mo', the number of outstanding orders just before a new order being raised.

Under a deterministic process, shortages can only occur if 'RO' is less than 'D'. We write, then, formally:

$$Z1(k)= (D-\sum Z1(k-j))-RO \qquad j=1,2...,mo \qquad (3.1a)$$

So,
$$\sum Z1(k-j)= D-R\theta \qquad j=0,1,...,mo \qquad (3.1b)$$

Denote by 'Z1' the average of the 'mo+1' terms of the l.h.s. summation in (3.1b). Then,

$$(mo+1)*Z1= D-RO \qquad (3.3a)$$

and,
$$Z1= (D-RO)/(mo+1) \qquad (3.3b)$$

This last expression gives, for the data of fig.3.1(b), 'Z1=1/3'. Note that, as the process is deterministic, 'mo' is constant and so is 'Z1'; 'mo+1' is the number of terms in 'Z1' and is also the number of reorder periods between

consecutive shortages.

For the total backlogging case we would have:

$$Z0 = D - R0 \qquad\qquad (3.5)$$

A deterministic process with a constant lost sales fraction 'a' would be written, in general:

$$Z(k) = (D - \sum a * Z(k-j)) - R0 \qquad j=0,1,\ldots mo \qquad (3.7)$$

The shortage pattern, after stabilization, would follow a cycle of 'mo+1' reorder periods such that '$Z(k)=Z(k+mo+1)$'. With this in mind, add up the 'mo+1' shortages of a cycle from '$Z(k)$' to '$Z(k+mo)$' and use (3.7) for each '$Z(.)$'. This, after some rearrangements leads to:

$$(mo+1)*(D-R0) = (1+a*mo)* \sum Z(k+j) \qquad (3.9a)$$
$$j=0,1,\ldots mo$$

The shortage accumulated in 'mo+1' periods is

$$\sum Z(k+j) = (mo+1)*(D-R0)/(1+a*mo) \qquad (3.9b)$$
$$j= 0,1,\ldots mo$$

Hence, the average shortage per period 'Z' is

$$Z = \sum Z(k+j)/(mo+1) \qquad j=0,1,\ldots,mo \qquad (3.9c)$$
$$= (D-R0)/(1+a*mo) \qquad\qquad (3.9d)$$
$$= Z0/(1+m) \qquad\qquad (3.9e)$$

where 'm=a*mo'. It will be seen in section 3.5 that for the deterministic case 'mo= INT(RO/Q)' where INT(y) stands for 'the integer of y'

Now, consider the probabilistic situation. The lead time demand is a stochastic variable with mean 'D', and density function 'f(x)'. Denote by 'x(k)' the actual demand during the lead time of a particular order 'k'. The shortage associated with that order is:

$$Z(k) = M\left[x(k) - RO - a*\sum Z(k-j)\right] \qquad j=1,\ldots,mo(k) \qquad (3.11)$$

where '$M\left[x\right]$' is the maximum between zero and 'x'; 'mo(k)' denotes the number of orders that were outstanding just before raising the order 'k' and this number can now fluctuate from order to order.

In the case of total backlogging, 'a=o' then (3.11) above simplifies to

$$ZO(k) = M\left[x(k) - RO\right] \qquad (3.13)$$

Hence, the expected shortage would be given by

$$ZO = \int_{RO}^{\infty} (x - RO)*f(x)*dx \qquad (3.15)$$

And the probability of stockout would follow the usual expression

$$PO= \int_{RO}^{\infty} f(x)*dx \qquad (3.17)$$

The formulae for 'ZO' and 'PO' are independent of the degree of overlapping; that is, under the reorder level operating system with captive demand, overlapping does not affect service levels.

To deal with non-captive demand it is convenient to introduce 'R(k)' defined by

$$R(k)= RO+a*\sum Z(k-j) \quad j=1,\ldots,mo(k) \qquad (3.19)$$

Denote, then, by 'R', 'Z' and 'mo' the expected values for 'R(k)', 'Z(k)' and 'mo(k)', respectively, after the system has reached the statistical equilibrium. In the expression above, the elements 'Z(k-j)' are not independent variables. But it can be shown, see for instance Ray(1981), that the expected value

$$E\ (\sum Z(k-j),\ j=1,\ldots,mo(k)\ ) = mo*Z \qquad (3.21)$$

provided that 'mo(k)' is independent of the 'Z(k-j)'s. This condition is not met entirely but, as it will be seen later in section 3.5, correlation is likely to be weak. Expression (3.21) will be adopted and thence

$$R= RO+a*mo*Z \qquad (3.23a)$$
$$= RO+m*Z \qquad (3.23b)$$

From (3.11), the expected value is

$$Z = E\left[\, M(x(k) - R(k))\,\right] \tag{3.25a}$$

Hence

$$Z = \int_{RO}\int_{y}^{\infty} (x-y)*f(x\,|\,y)*g(y)*dx*dy \tag{3.25b}$$

where 'x' stands for the values of 'x(k)' and 'y' for the values of 'R(k)' for any order 'k'; 'g(y)' is the probability density function for 'R(k)'; and 'f(x|y) is the conditional probability density function for the lead time demand, given 'y'.

The exact evaluation of (3.25) is not easy. To overcome the difficulty the following approximation will be introduced: 'R(k)', in expression above, will be replaced by its expected value 'R'. By doing so, (3.25) becomes:

$$Z = E\left[\, M(x(k) - R)\,\right] \tag{3.27a}$$

or

$$Z = \int_{R}^{\infty} (x-R)*f(x)*dx \tag{3.27b}$$

Consistently, it will be adopted for the probability of stockout

$$P = \text{prob}(\, x(k) \gtrsim R \,) \tag{3.29a}$$

or

$$P = \int_{R}^{\infty} f(x)*dx \tag{3.29b}$$

instead of

$$P = \text{prob}(\ x(k) \geq R(k)\ ) \qquad\qquad (3.31a)$$

or

$$P = \int_{RO}^{\infty} \int_{y}^{\infty} f(x\,|\,y) * g(y) * dx * dy \qquad\qquad (3.31b)$$

The size of the errors involved in this approximation seems quite acceptable in simulation checks reported in section 3.6 .

Expressions (3.15) and (3.17) are particular cases of (3.27b) and (3.29b), respectively. In fact, for 'a=0' then 'R=RO', so, the latter reduces to the former. Furthermore, the relationships

$$\partial Z / \partial R = -P(R) \qquad\qquad (3.33a)$$

and $\qquad \partial P / \partial R = -f(R) \qquad\qquad (3.33b)$

apply in the same way as for situations of captive demand.

The variable 'R' has been introduced for convenience of the analysis and was defined earlier as

$$R = RO + m * Z \qquad\qquad (3.35)$$

Derivations above show that the role performed by 'R' in a situation with lost sales and decision period overlapping is the same as the role of 'RO' with captive demand. Moreover, 'R' reduces to 'RO' for captive demand or for non-overlapping cases, because 'a=0' or 'mo=0',

respectively, and hence 'm=0' in expression (3.35). Therefore, 'R' is a mathematical notion which generalizes the concept of reorder level and, because of this, will be called 'notional reorder level'. Virtually, the inventory system reacts directly to 'R' and indirectly only to 'R0', which is a finding of major importance for the analysis hereafter.

The service levels with lost sales can be approximated from the service levels which would be obtained if demand were captive, as follows:

$$Z \sim Z0/(1+m*P0) \qquad\qquad (3.37a)$$

$$V \sim V0/(1+m*P0+a*V0) \qquad\qquad (3.37b)$$

where 'V= Z/(Q+a*Z)' and 'V0=Z0/Q' are the shortage rates for non-captive and captive demand situations. Derivations are shown in subsection 3.2.1. Note that the formula (3.37a) fits the deterministic situation formerly discussed. For this case, in fact, if 'Z0' is not zero, then 'P0=1' and (3.37a) converts to (3.9e).

Expressions (3.27) to (3.37) derived for the reorder level control also hold with the periodical review policy as it is shown in sub-section 3.2.1, but symbols have to be interpreted differently. Differences are summarized below:

	R.O.L.	P.R.
R	Notional reorder level	Notional top up level
RO	Set " "	Set " " "
m=	a*mo	a*(mo+1)
f(x)	pdf for L demand	pdf for L+T demand

3.2.1   Analytical details

For the purpose of deriving the expressions (3.37) which relate the service levels with captive and non-captive demand situations, consider the general expressions

$$Z(X)= \int_{X}^{\infty} (x-X)*f(x)*dx \qquad (3.39a)$$

and

$$P(X)= \int_{X}^{\infty} f(x)*dx \qquad (3.39b)$$

By developing in Taylor's series, we can write

$$Z(X+h)= Z(X)+(\partial Z/\partial X)*h+1/2*(\partial^2 Z/\partial X^2)*h^2+... \qquad (3.41)$$

Substituting 'RO' for 'X', 'm*Z' for 'h'; and having in mind (3.33) and the expression (3.23b) defining 'R=RO+m*Z', then

$$Z(R)= Z(RO)-m*Z*P(RO)+1/2*(m*Z)**2*f(RO)+... \qquad (3.43)$$

Following the earlier notation, we would write:

$$Z(R) = \int_R^\infty (x-R)*f(x)*dx = Z \qquad (3.45a)$$

$$Z(R0) = \int_{R0}^\infty (x-R0)*f(x)*dx = Z0 \qquad (3.45b)$$

$$P(R0) = \int_{R0}^\infty f(x)*dx = P0 \qquad (3.45c)$$

thence

$$Z = Z0 - m*Z*P0 + 1/2*(m*Z)**2*f(R0)+... \qquad (3.45d)$$

The last expression would simplify further to

$$Z = Z0/(1+m*P0)+... \qquad (3.47a)$$

The shortage rate is $'V = Z/(Q+a*Z)'$. So, extending the usual notation structure, $'V0 = Z0/Q'$. If $'Z'$ is approximated by the first term of the series in (3.47a), it follows using the sa_me steps

$$V \sim V0/(1+m*P0+a*V0) \qquad (3.47b)$$

The periodical review policy could be analised on the same steps followed earlier. With this policy, stock levels are verified at fixed intervals $'T'$. The quantity ordered $'Q'$ is the difference from the present nominal stock to a top up level $'R0'$. Thus, the operating variables are now $'T'$ and $'R0'$. Note that $'R0'$ relates here to the top up level while previously it meant the reorder level.

The decision horizon is $'L+T'$ which implies that decision

periods always overlap, even when no more than one order is
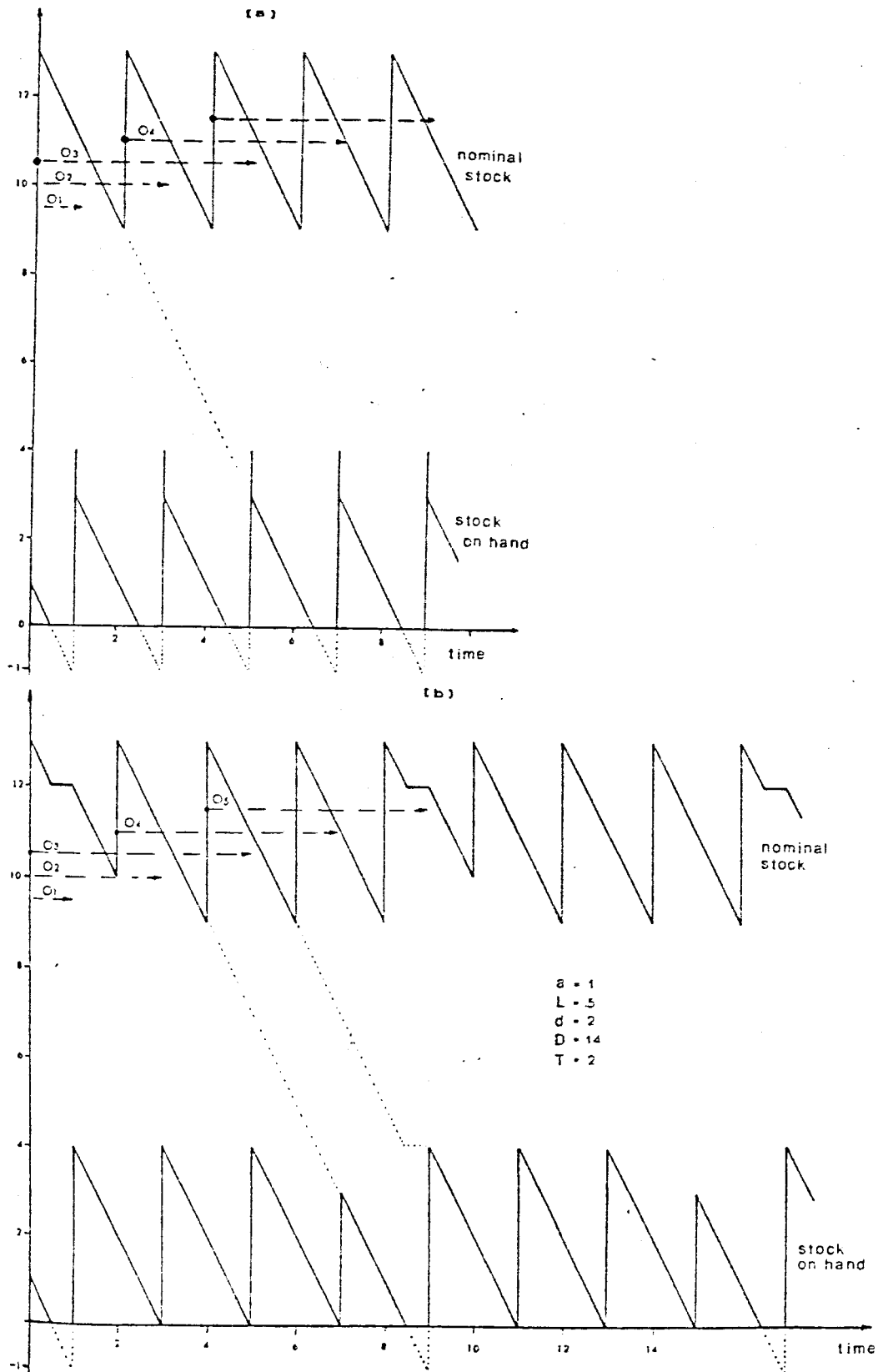
Fig. 3.2 - Shortages with periodical review policy

(a) backlog case;   (b) lost sales case

outstanding at any time. For this reason, the lost sales effect is more pronounced with this model than with the earlier one. Fig.3.2 helps make it clearer. The diagram relates to the same item of fig.3.1, but now, operated by periodical review.

Lead time and demand rate are the same in both figs. 3.1 and 3.2, namely, 'L=5' and 'd=2'. The review period is now set at 'T=2', so the quantity demanded in 'T' is 4, the same as the reorder quantity in fig.3.1. As the decision period is 'L+T=7', the demand in this period is 'D=14' (note that, now, 'D=(L+T)*d' while in the previous model, 'D=L*d'). The top up level is set as 'RO=13'. The provision for the decision period, 'RO', is 1 unit lower than the demand 'D', exactly the same deficit as in fig.3.1. Both situations, therefore, are quite comparable and the treatments are equivalent. When demand is captive, the results, in terms of shortage, are also the same (figs.3.1a and 3.2a). For the lost sales case, however, this is not true: in fig.3.1b the average shortage is 1/3 while in fig.3.2b it is 1/4.

Each order 'k' can be associated to a 'L+T' decision period. The shortage 'Z(k)' in relation to order 'k' occurs between the arrivals of the orders 'k' and 'k+1'. As it can be seen in fig.3.3, there were only 2 orders outstanding just before the order 'k' being raised, yet 3 decision periods were overlapping. In general, we have 'mo(k)' orders and 'mo(k)+1' periods outstanding. Expression (3.7) still

applies for the deterministic case, having in mind the present meaning of 'RO' and 'D', since the summation is extended from 'j=1' to 'j=mo+1'. If, now, we make 'm=a*(mo+1)', instead of 'm=a*mo', expression (3.9e) is also valid for the periodic review case. In the example of fig.3.2b, 'ZO=1', 'mo=2'; then, for 'a=1', 'm=3', so 'Z1=1/4'.
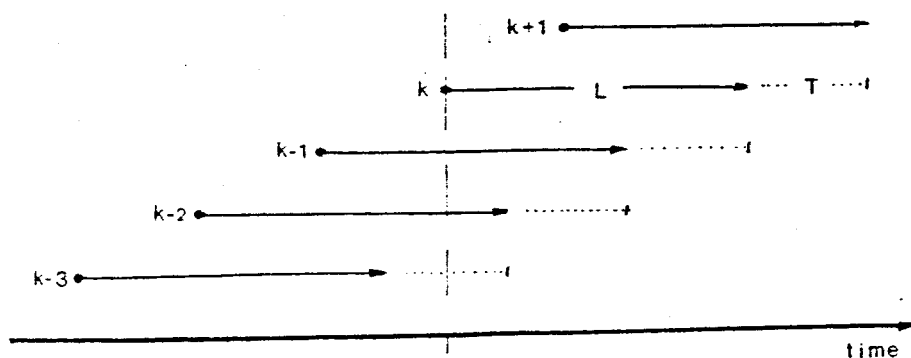
Fig. 3.3 - Overlapping of orders and decision periods

The approach used in the previous section to deal with the probabilistic case applies also here. Essentially, we would arrive at the same formulas though the symbols have a different meaning.

3.3 The effect on stock levels

The stock on hand, 'Ys(k)', at the end of the decision period associated to order 'k', is given by

$$Ys(k) = M\left[R(k) - x(k) \right] \qquad (3.61)$$

where 'R' and 'x' relate to the notional level and to the decision period demand, respectively. 'D' will denote the expected value for the latter. That result is based on arguments similar to those used in the preceding section. Adopting the same sort of approximations, the expected safety stock would be

$$Ys = E\left[Ys(k) \right] \qquad (3.63a)$$

$$= \int_0^R (R-x) * f(x) * dx \qquad (3.63b)$$

$$= R - D + Z \qquad (3.63c)$$

$$= R0 - D + (m+1) * Z \qquad (3.63d)$$

For the total backlogging, 'Ys0=R0-D+Z0'. For most of the cases 'P0' is low, so it follows from (3.47a) that 'Z>Z0/(1+m)'. Hence and from (3.63d)

$$Ys > R0 - D + Z0 = Ys0 \qquad (3.65)$$

Whenever 'm' is greater than zero. (Obviously, equalities

shall substitute the inequalities above when 'm=0'). Therefore, situations of non-captive demand coupled with overlapping decision periods result in higher actual safety stocks than for the equivalent captive demand and/or non-overlapping situations.
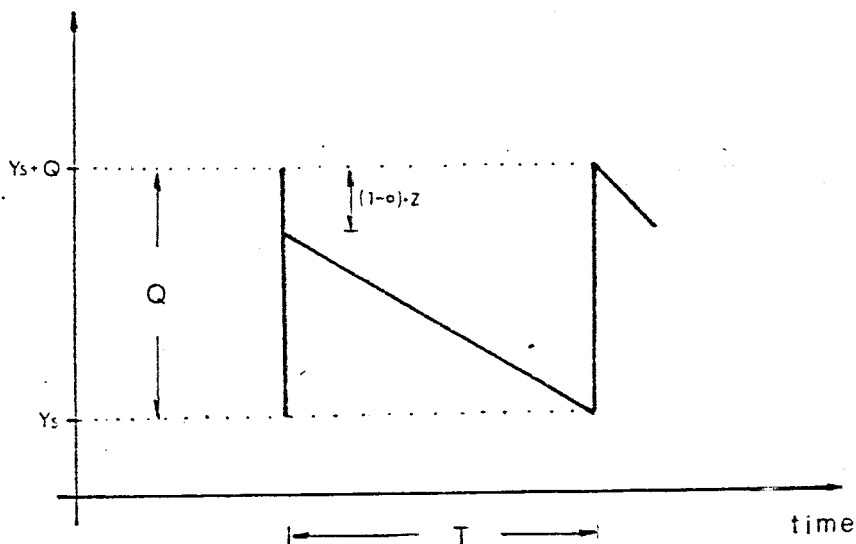
Fig.3.4 - Deterministic approximation for stocks

An approximation for the average stock can be obtained from the deterministic equivalent for the stock movement, as it is shown in fig.3.4. The deterministic equivalent substitutes the expected values, before and after the replenishment, for the actual levels and assumes a constant demand rate equal to the average of the actual demand. The expected stock level before each replenishment is the expected safety stock as from (3.63). Note, it is the expected value through all cycles including those which run out of stock. Therefore, 'Ys' is the lower stock level of the deterministic equivalent, as it is seen in the diagram.

'Q' is the expected replenishment quantity (a constant for the R.O.L control). The quantity backlogged is '(1-a)*Z', as 'Z' is the expected shortage. Such a quantity is captive and, theoretically, it is delivered to the customer or put at his disposal as soon as the replenishment arrives. Thus, it should not account for the average stock of the company (admittedly, internal conventions may superimpose on these views). In fig.3.4, it is assumed that the backlogged demand is issued immediately and so, it is not included in the average stock. This latter would be, then

$$S = Y_s + 1/2*(Q-(1-a)*Z) \qquad (3.67a)$$
$$= R-D+1/2*(Q+(1+a)*Z) \qquad (3.67b)$$

For 'Z' much lower than 'Q', the current assumption, we can approximate

$$S = R-D+Q/2 \qquad (3.69)$$

Taylor et al (1976), for captive demand situations, present exact formulas for the average stock (of difficult evaluation, however) as well as, expressions to compute the upper and lower bounds. The upper bound corresponds to (3.67) with 'a=0', i.e. captive demand, and the lower bound to (3.69).

'Q' is constant in the case of the inventory being operated in a reorder level policy. Therefore, it would follow from (3.67)

$$\partial S/\partial R = 1 - 1/2 * (1+a) * P \qquad (3.71)$$

In the case of a P.R. control, on average, 'Q+a*Z=d*T'. This quantity is constant, therefore

$$\partial S/\partial R = 1 - 1/2 * P \qquad (3.73)$$

Expression (3.71) reduces to (2.27b) for 'a=1' (total lost sales) and (3.73) is identical to (2.31c). But now the expressions are more comprehensive for 'R' has now a broader meaning. With high service levels, 'P' is small and both expressions can be approximated as in Johnston (1980), by putting

$$\partial S/\partial R = 1 \qquad (3.75)$$

Thus, the use of a notional level enables a common analytical treatment for both the lost sales and captive demand situations.

3.4   Expansion of the Johnston model.

The formulation introduced in Johnston (1980), as has been already mentioned, implies that orders do not overlap. This may constitute a very inconvenient restriction especially in situations with long procurement lead times. In order to bring down average stocks to reasonable levels it may be necessary to use replenishment intervals much shorter than the lead times, thus accentuating the problem. The restriction in the model can be withdrawn with the introduction of the notional level.

We recall from section 2.4 the criterion used in J-model for investment allocation

$$\text{Max} \left(\sum \pi(i) \right), \text{ subject to } \sum S(i) \leqq cS \qquad (3.81)$$

and the derivations which followed to arrive at

$$-N*F(i)*(\partial Z(i)/\partial R(i))-(F2+\lambda)*(\partial S(i)/\partial R(i)) =0 \qquad (3.83)$$

The above derivatives '$\partial/\partial R$' stand, regardless of 'R' being the notional or the set level, and the formulation is valid for any operating procedure. Now, using the results in (3.33) and the approximation (3.75) that has already been adopted earlier, we get exactly the same formal results. 'R(i)' should be such that

$$P(i) = 1/N*(F2+\lambda)/F(i) \qquad (3.85)$$

where P(i) is given by (3.29) for each specific 'i'.

The computational procedures are basically the same as before. 'P(i)' is fixed from (3.85); then, R(i) is computed by inversion of (3.29). From 'P(i)' we can also calculate 'Z(i)'. Now, it follows from (3.35) that the level to be set is

$$RO(i) = R(i) - m*Z(i) \qquad (3.87)$$

Computations for 'Z' and 'R' can be based on empirical relationships as those suggested by Johnston. The procedures to generate values for the perfomance variables can follow the same steps as for the non-overlapping case. The extra computations required are only those related to (3.87).

3.5 The estimate of the number of outstanding orders

The suggested approach to deal with situations of overlapping relies on the assumption that 'm' can be evaluated.This implies that we are able to estimate 'mo', the average number of orders outstanding when a new order is raised, as well as, the loss fraction 'a'. The loss fraction was discussed in section 2.6. Now, we will look at 'mo'.

First of all, it shall be stressed the distinction between average number of orders outstanding when a new order is raised, 'mo'; and the average number of orders outstanding throughout the time, hereafter, denoted by 'mt'. The latter can be easily evaluated with the help of fig.3.5.
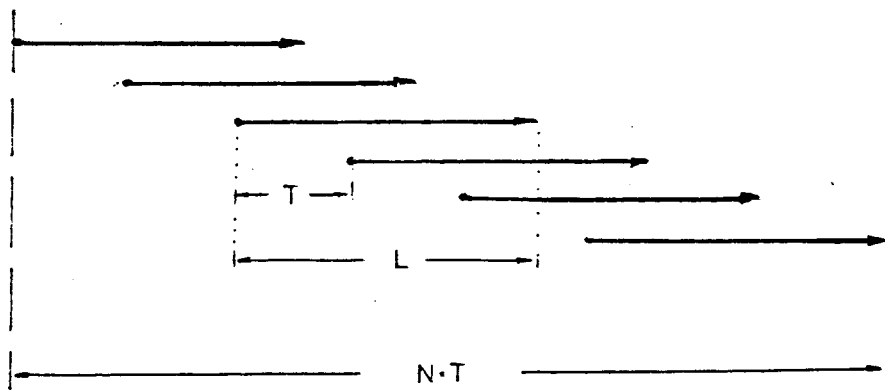
Fig.3.5 - Overlapping degree

Consider some length of time, one year, say, during which 'N' orders are placed at intervals of 'T', on average. Each order has a lead time of length 'L'; thus, the 'N' lead

time vectors add up to the length of 'N*L'. This length has now to be compacted into 'N*T' by sliding back the L-vectors such that they will overlap each other. Then, 'mt', the average number of vectors overlapping, would be such that

$$(N*L)/mt = N*T \qquad (3.91a)$$

so $\qquad mt = L/T \qquad (3.91b)$

$$= D/(Q+a*Z) \qquad (3.91c)$$

'mt' is the average number of orders outstanding throughout. It is also the expected number of orders outstanding at any random moment in time for both deterministic and probabilistic processes. 'mt' will be adopted as the measure for the overlapping degree.

The evaluation concerning 'mo' is not so simple. The following bounds can, however, be established

$$mt-1 < mo < mt \qquad (3.93)$$

In fact if, on average, there were 'mo=mt' orders outstanding 'just before' the new order raising, it would make 'mt+1' orders 'just after' it. This latter figure would fall again to 'mt' after the next arrival; and would remain so from then on to the subsequent order raising. Thus, the number of orders outstanding would vary in the range (mt;mt+1). The average through time would be, then, higher and not equal to 'mt'. Therefore, 'mo' has to be lower than 'mt'. By using similar arguments we could conclude that 'mo'

is greater than 'mt-1'.

Deterministic situations as those exemplified earlier in figs. 3.1 and 3.2 have the same 'mo(k)' for all 'k's; and as 'mo(k)' is an integer, so is the average 'mo'. Hence and from (3.93) we can conclude that for the deterministic case

$$mo= INT(mt) \qquad\qquad (3.95)$$

where INT(.) stands for the 'integer of'.

Expression (3.93) is general for both the deterministic and probabilistic processes. For the latter, however, the estimation of 'mo' is not so simple. Hopefully, for many applications, a rough estimate for 'mo' is enough. Then, the following might be reasonable:

$$mo= M(mt-1/2) \qquad\qquad (3.96)$$

where 'M(x)' is the maximum between zero and 'x'. A much more elaborate process of estimating 'mo' is examined in sub-section 3.5.1.

The rational behind expression (3.96) is better explained with the help of fig.3.6. Let be 'I=INT(mt)'. When 'mt~I+1/2', orders arrive more or less at the mid-point between two consecutive order raising (see figs. 3.1 and 3.2). So, a relatively high degree of variation in demand and lead times can occur before the order arrives out of the

T-period in which it was expected. It is likely that for most of the orders 'mo(.)=I'. A small fraction may have 'mo(.)=I+1' and higher. Another fraction of identical size may have 'mo(.)=I-1' and lower. In a random situation these should occur with equal frequence so that, on average, we can expect that 'mo~I'. The point 'A' in fig 3.6 would represent this situation.
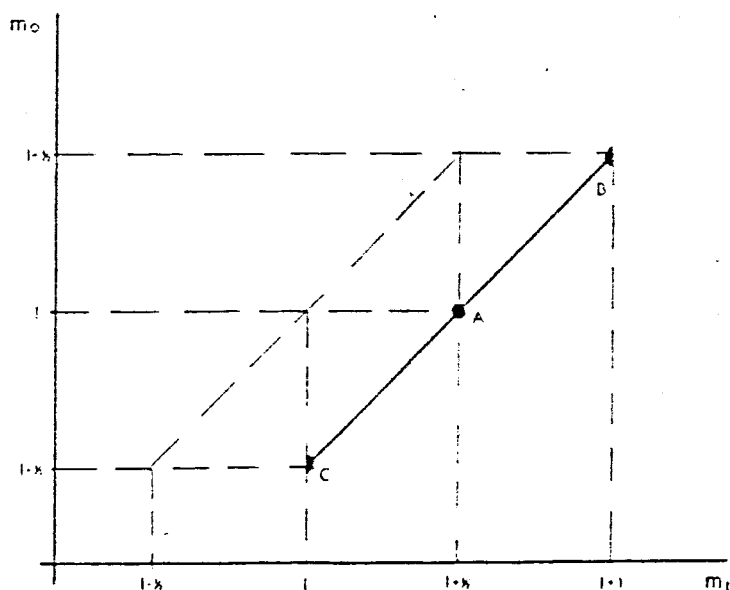
Fig. 3.6 - A rough estimate for 'mo'

Imagine, now, that 'mt' is near to an extreme, for instance, lower but close to 'I+1' (point 'B' in the diagram). Therefore, there should be 'I' orders outstanding when a new order is raised, but an order is due to arrive just before a new order raising. However, the process fluctuations can easily, either delay the arrival or antecipate the raising, so that the latter precedes the former and hence 'I+1'

orders are outstanding. From the position of point 'B' it is plausible that about half the orders have mo(.) equal to 'I' or lower and another half equal to 'I+1' or higher. On average, we can expect that 'mo~I-1/2'. On similar grounds, we can also expect that 'mo~I-1/2', if 'mt' is higher but close to 'I' (point 'C').

The three points, 'A', 'B' and 'C' are on a straight line. If we assume that a linear relationship gives a fair approximation, then (3.96) follows. Note that this latter expression gives for 'mo' the mid-point of the interval defined in (3.93).

3.5.1 Analytical details

The boundaries for 'mo' stated in (3.93) above imply that 'mt' is not an integer. Theoretically, we could imagine a limiting case where any new raising coincides with the time of a new arrival, such as is shown in fig.3.7. In such circumstances 'mt' is an integer and, in expression (3.93), 'mo' might take the value of one of the extremes, either 'mt-1' or 'mt', depending on the sequence in which the simultaneous events arrival/raising are looked at. Such a coincidence, however, can only occur consistently under deterministic situations.

In deterministic situations, the expression (3.95) holds for
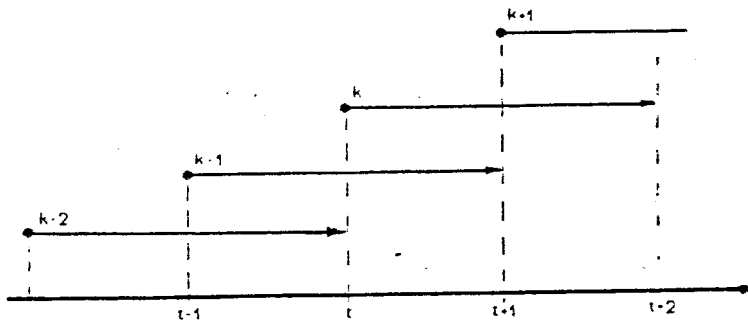
Fig. 3.7 - Simultaneous raising and arrival

mt=2; mo=1 or mo=2?

non-integer 'mt', but when 'mt' is an integer 'mo' could be again either 'mt' or 'mt-1'. The latter, however, shall be the figure to be used in the formulas because only the 'mt-1' orders preceding 'k' have their risk periods located in the k-lead-time. For instance, in fig.3.7 only the order 'k-1' has its risk period inside the k-lead-time, so 'mo' shall be taken as 1. This become more apparent with the example in fig.3.8 which relates to a periodical review control with 'mt=2', loss fraction 'a=1'. Note that this situation is very close to the one in fig.3.2; the differences come from reducing the lead time from 5 to 4, thus, pushing the overlapping degree 'mt' from 2.5 to the border line 'mt=2.0'. In both cases, 'R=D-1', 'T=2' and 'Z0=1'. However, in fig.3.2, 'Z1=1/4' while in fig.3.8, 'Z1=1/3'. Recalling from (3.9e) that 'Z1=Z0/(1+m)' and that 'm=mo+1' for the periodical review, the figures for 'Z1'

correspond to 'mo=2' and 'mo=1', respectively; thus, when 'mt' is an integer and the process deterministic, 'mo=mt-1'.
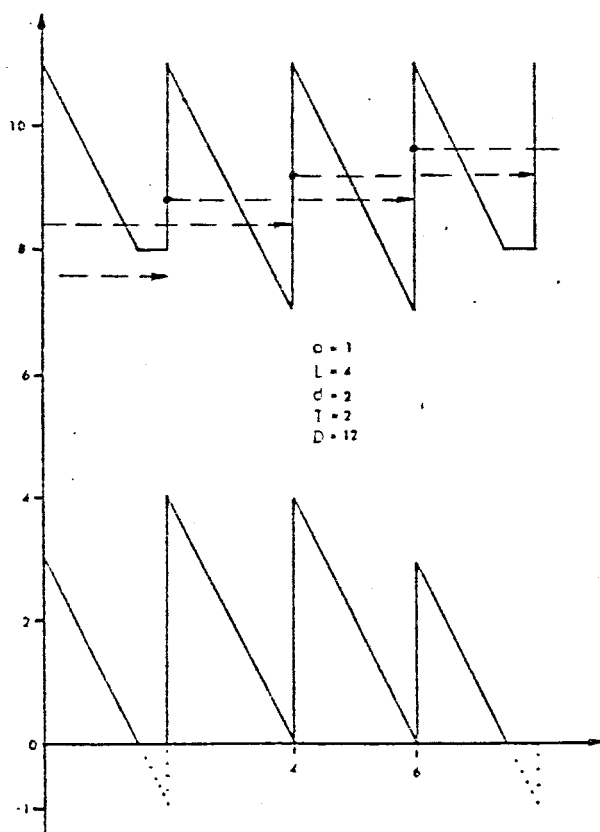
Fig. 3.8 - PR control with simultaneous arrival and
raising of orders

The value of 'mo' for the more general situation of probabilistic demand can be estimated, roughly, from expression (5.96). A more precise method is described below.

Consider first, the reorder level policy. In this case, orders are raised at 'Q-intervals', that is, the sales between two consecutive orders amount to 'Q'. (Sales in this context include the backorders). In fig.3.9, the cumulative

sales are plotted against time. By choosing appropriate origins, we can say that order 1 is raised when the accumulated sales reach 'Q', order 2 when they reach '2*Q',..., order 'k' when they reach 'k*Q'. When order 4 was raised the orders 2 and 3 were still outstanding. We see that 'x(4)' is greater than 'x(3)-Q' and greater than 'x(2)-2*Q'. When order 3 was raised, only order 2 was standing and 'x(3)=x(2)-Q'. In fact, as a consequence of the non-overtaking condition, the following relationship follows:

$$x(k) \geq x(k-j) - j*Q \quad , \quad j=1,\ldots,mo(k) \qquad (3.97)$$

Note that orders 'k' and 'k-j' overlap.
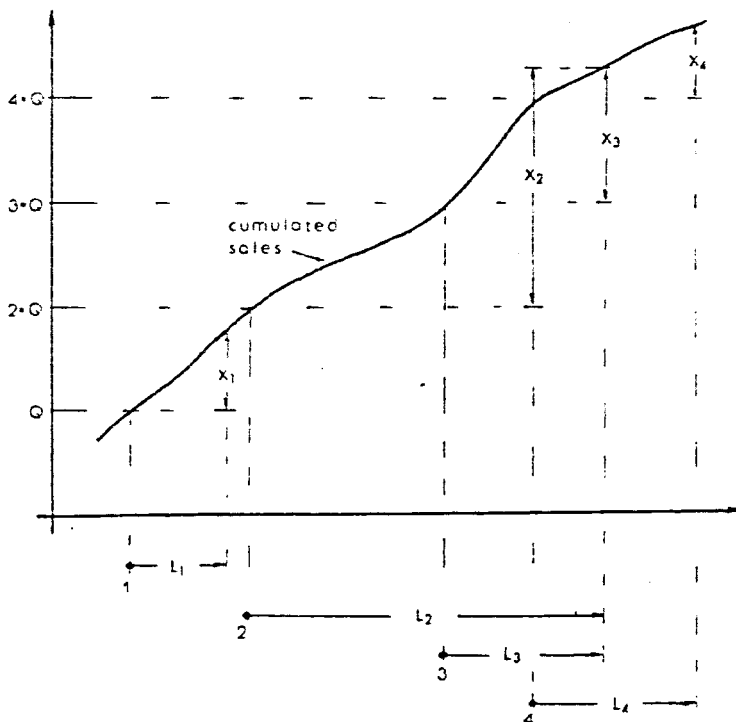
Fig. 3.9 - Cumulative sales

Now, we see from the picture that no orders were outstanding when order 2 was raised because 'x(1)' had been less than 'Q'. There was the order 2 outstanding when 3 was raised because 'x(2)' had been greater than 'Q'. The reason why order 2 still stands after two new orders being raised is because 'x(2)' is greater than '2*Q'. Formally, the condition for the order 'k-j' being standing when order 'k' is raised is

$$x(k-j) \geqq j*Q \qquad\qquad (3.99)$$

This relation could be derived from (3.99) as a necessary condition ('x(k)' is non-negative). It is now shown to be a sufficient condition for order 'k-j' being standing. If 'k-j' is standing so are all the subsequent orders up to 'k', because of the non-overtaking. Therefore, (3.99) is also the necessary and sufficient condition for at least 'j' orders being outstanding when the new order 'k' is raised. Then, introducing the notation

$$Pm(j) = prob(mo(k) \geqq j) \qquad\qquad (3.100a)$$

and

$$pm(j) = prob(mo(k) = j) \, , \qquad\qquad (3.100b)$$

for any 'k', it follows from above that

-3.5.1-

$$Pm(j) = \int_{j \cdot Q}^{\infty} f(x) * dx \qquad (3.101a)$$

$$pm(j) = Pm(j) - Pm(j+1) \qquad (3.101b)$$

$$= \int_{j \cdot Q}^{(j+1) \cdot Q} f(x) * dx \qquad (3.101c)$$

Note that 'x' and 'f(x)' relate to actual sales which, at most, are the same as the demand. With low shortages we can approximate these as being the same.

The average number of orders outstanding would be given by

$$mo = (\sum j*p(j))/(Pm(b)-Pm(u+1)) \ , \ j=b,\ldots,u \qquad (3.103a)$$

$$= (b*Pm(b)+\sum Pm(j)-u*Pm(u+1))/(Pm(b)-Pm(u+1)) \qquad (3.103b)$$
$$j=b+1,\ldots,u$$

where 'b' and 'u' are, respectively, bottom and upper sensible limits to be considered for 'mo(k)'. 'Sensible' means that, 'Pm(b)~1' and 'Pm(u+1)~0'. So, for practical purposes, (3.103) may become much simpler.

Note that, under a reorder level policy, when a new order is raised, the stock on hand plus on order minus backorders sum up to 'R'. The stock on hand is non-negative and the stock on order is 'mo(k)*Q'. The quantity actually backlogged, '(1-a)*Z(.)', is lower than '(1-a)*Q' since 'Q' is greater than 'Z(.)'. Therefore

$$mo(k)*Q \lesssim R+(1-a)*Q \qquad (3.105a)$$

or, $\qquad mo(k) \lesssim INT(R/Q+1-a) \qquad (3.105b)$

then, $\qquad mo(k) \lesssim INT\left[(R/D)*mt+1-a\right] \qquad (3.105c)$

as 'mt~D/Q'. The r.h.s. of the last inequality could be taken as the upper limit 'u'.

A sort of symmetrical reasoning could be followed in relation to the bottom limit. Saying that 'Z(.)' is less than 'Q' means that the actual lead time demand 'x(.)' does not exceed the average 'D' in more than 'R+Q-D'. In practical terms, the reverse is also likely, i.e., the actual lead time demand does not fall below 'D' in more than that 'R+Q-D'. Therefore

$$x(k) > D-(R+Q-D) \qquad (3.107a)$$

So

$$x(k)/Q > (2-R/D)*mt-1 \qquad (3.107b)$$

Note that the stock on hand just before the arrival of the order 'k' is 'Ys(k)=R-x(k)'. 'Ys(k)' is positive as we are considering, now, situations where the demand is less than the average, therefore 'x(k)' is less than 'R'. Consequently, the stock on order at that moment must exceed the difference between 'Ys(k) and 'R', i.e. has to be at least equal to 'x(k)'. Hence, the number of orders outstanding just before the arrival of order 'k', 'nb(k)' say, has to be an integer greater or equal to 'x(k)/Q'. Thus

$$nb(k) \gtrsim INT(x(k)/Q)+1) \qquad (3.109)$$

The number of orders outstanding falls to 'nb(k)-1'after the k-arrival. This last figure gives the number of orders outstanding when the next order (after the k-arrival) is raised. Refer to that order as 'k+i'. Then, it follows:

$$mo(k+i)= nb(k)-1 \qquad (3.111a)$$
$$\gtrsim INT(x(k)/Q) \qquad (3.111b)$$
$$\gtrsim INT((2-R/D)*mt)- 1 \qquad (3.111c)$$

The r.h.s. of the latter expression could be taken as a sensible value for the lower boundary 'b'. Both boundaries are close integers in practice, so the evaluation of 'mo' through (3.103) is not a lengthy task.

For the periodical review procedure, the estimation of 'mo' could be done in a similar way, substituting T-intervals for Q-intervals and lead time for lead time demand. Then, from the non-overtaking condition and through the same sort of arguments as before, the following necessary and sufficient condition could be stated:

$$mo(k) \gtrsim j \quad \text{if and only if} \quad L(k-j) > j*T \qquad (3.113)$$

In parallel with the previous case, we could use (3.103) to evaluate 'mo', as

$$Pm(j) = \int_{j \cdot T}^{\infty} f(x) * dx \qquad\qquad (3.115a)$$

$$pm(j) = Pm(j) - Pm(j+1) \qquad\qquad (3.115b)$$

where 'f(x)' is the density function for the lead time. Practical limits for 'b' and 'u' could be found as in the earlier approach.

3.6   Simulation checks.

In this chapter, a number of approximations have been introduced to overcome analytical difficulties. The individual approximations were accompanied by comments explaining why each was considered acceptable. In this section, the overall effect of the errors involved is to be looked at by comparing the analytically expected results to
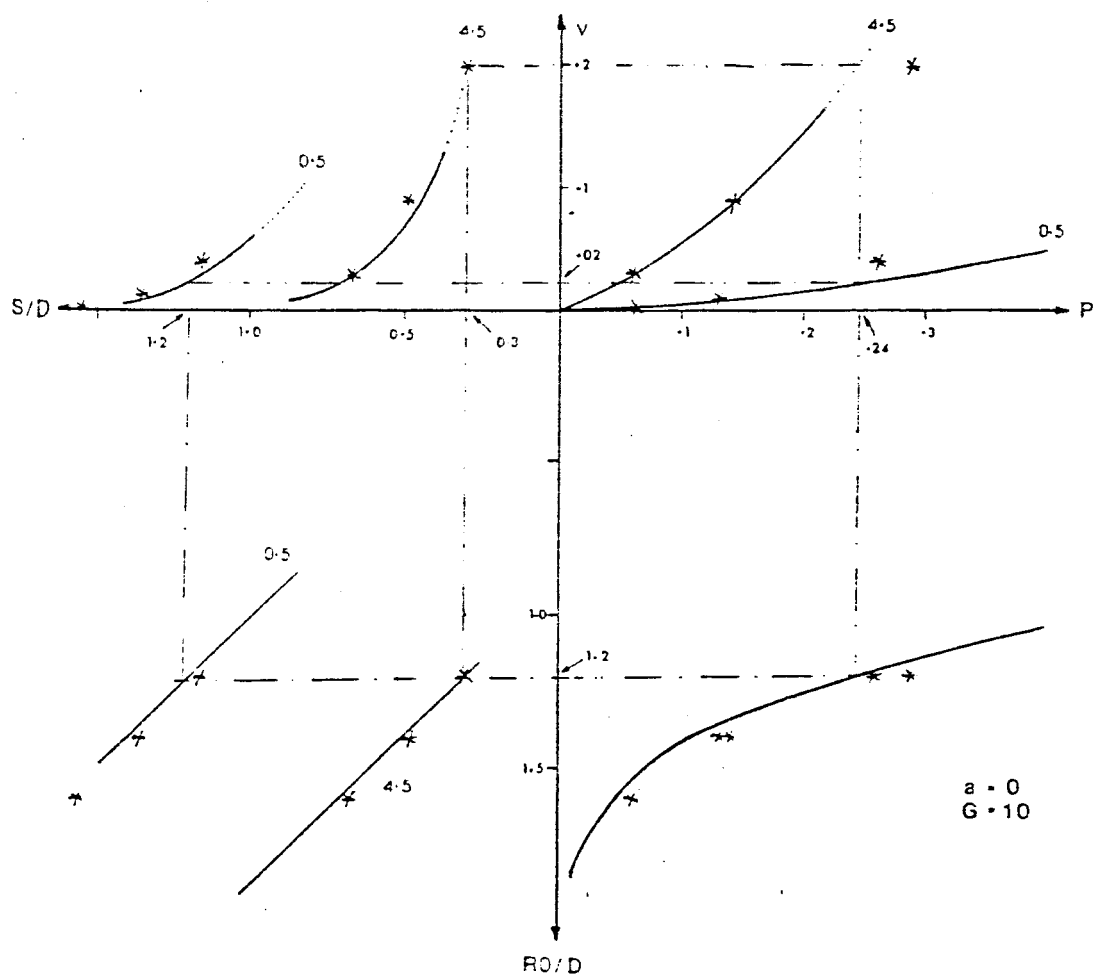
Fig. 3.10 - Analytical predictions and simulation
            with captive demand, a=0

the values obtained from simulation. Comparisons are

depicted in figs. 3.10 to 3.12 and show that differences are small and that the new formulations substantially improve the predictions.

The solid curves in fig.3.10 represent the analytical expectations in relation to a reorder level policy when the lead time demand is gamma distributed with modulus equal to 10 and demand is captive (a=0). 'P' and 'V' stand for the stockout and shortage rates, respectively; 'RO' for the reorder level as set and 'S' for the average stock. 'D' is the average demand in the lead time and 'RO/D' and 'S/D' were introduced to standardize the results. The numbers 0.5 and 4.5, assigned to each curve, are the values set for the overlapping degree 'mt' (see 3.91). Dots and stars mark typical values obtained from simulation (actually, only two sets are plotted but the results of all runs were very similar). More details can be found in subsection 3.6.1.

The diagram should read as follows. Consider that the reorder level is set so that 'RO/D=1.2'. This value at the lower axis determines, to the right, 'P~0.24'; to the left, 'S/D~1.2' for 'mt=0.5', and 'S/D~0.3' for 'mt=4.5'. The corresponding values for the shortage rate are 'V~0.02' for 'mt=0.5' and 'V~0.20' for 'mt=4.5'.

In the case considered in fig 3.10, notional and set reorder levels coincide because 'a=0', so 'R=RO+a*mo*Z' (see 3.23) converts into 'R=RO'. As a consequence of this, the expected stockout rate 'P' is independent of the degree of

overlapping and so is the expected shortage per cycle 'Z' (see expressions 3.27 and 3.29). On the other hand, when the overlapping 'mt' increases, the reorder quantity 'Q' decreases, then the shortage rate 'V' increases for 'V~Z/Q'. The decrease in 'Q' leads also to a lower average stock 'S'.
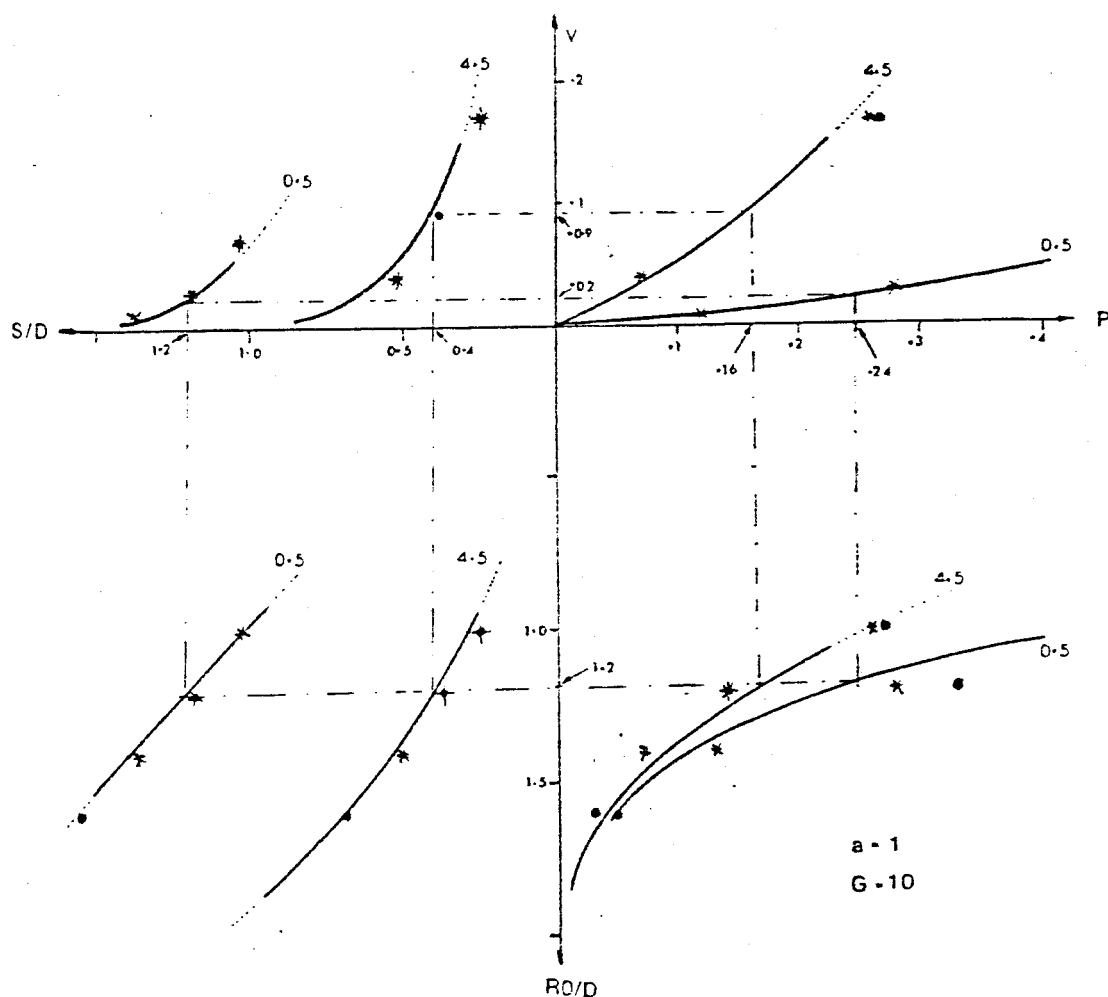
Fig. 3.11 - Analytical predictions and simulation
with total lost sales, a=1

In fig. 3.11 any shortages are lost for sale (a=1). In this case, 'P' depends on the overlapping degree. For the same value 'RO/D=1.2' as in the example above, the stockout rate is now 'P~0.24' for 'mt=0.5' and 'P~0.16' for 'mt=4.5'. 'V'

and 'S/D' also depend on the overlapping degree but the variation is less pronounced than with the captive demand: for 'mt=4.5' they are now 'V~0.09' and 'S/D~0.4' while in fig.3.10 they were 'V~0.20' and 'S/D~0.3'.
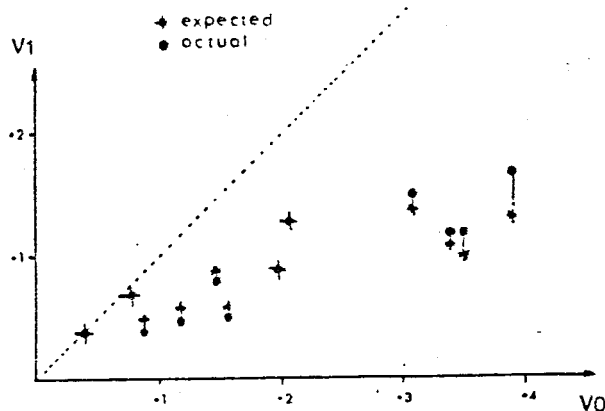
Fig. 3.12 - Lost sales effect on services

Fig.3.12 checks the formulas (3.37). 'V1' and 'V0' are the shortage rates for 'a=1' and 'a=0', respectively. The dots are results obtained from simulation runs and the stars are the values calculated from (3.37b) for several combinations of demand and control parameters. The differences are small for the low shortage levels currently adopted. Note that 'V0' would be the estimate for 'V1', if the lost sales effect were ignored (the estimates would be located on the bisecting line). Therefore, the improvement brought in by (3.37) is appreciable.

3.6.1 Further details

The predictions represented by the curves in figs. 3.10 and 3.11 refer to a reorder level policy and to a demand in the lead time Gamma distributed with modulus 'G=10'. The points to draw the curves were calculated from the values assigned to the ratio 'D/Q' and to the stockout 'P'

The overlapping degree was approximated as

$$mt \sim D/Q \tag{3.121a}$$

which follows from (3.91c), after neglecting 'Z'; and the number of outstanding orders, having then (3.96) in mind, was estimated as

$$mo \sim D/Q - 0.5 \tag{3.121b}$$

Therefore, the values '0.5' and '4.5' assigned to the curves correspond to 'mo=0' and 'mo=4', respectively.

The computing routine follows, basically, the steps mentioned in section 3.4 for the extended Johnston model which are detailed below:

> (i) Set a value for 'P'.
>
> (ii) Invert (3.29) and obtain 'R/Ds', where 'Ds' is

the standard deviation for the lead time demand, by using the "function 1" in Johnston (1980, p.1083). Then, compute

$$R/D= (R/Ds)*(Ds/D) \qquad (3.123a)$$
$$= (R/Ds)/\sqrt{G} \qquad (3.123b)$$

(iii) Compute 'Z/D' by using "function 3" in Johnston (1980. p.1083)

(iv) From (3.87), then

$$RO/D= R/D-a*mo*Z/D \qquad (3.125)$$

where 'mo' is determined by (3.121b) and 'a' is the value set for the lost sales fraction.

(v) Compute 'V' as follows:

$$V= Z/(Q+a*Z) \qquad (3.127a)$$
$$= (Z/D)/(Q/D+a*Z/D) \qquad (3.127b)$$

(vi) The stock ratio can then be derived from (3.67) as

$$S/D= R/D-1+1/2*(Q/D+(1+a)*Z/D) \qquad (3.129)$$

The simulation values in figs. 3.10 and 3.11 were obtained from a discrete time simulation program written to run inventory operations under both reorder level and periodical review controls. A sample of a printout is shown in fig.3.13. The input conditions are the loss factor 'a', 'D/Q' (as an approximation for the overlapping level) and 'RO/Q'; the demand and supply characteristics; the simulation length, run-in time and the random generator (seed).

Lead times were deterministic. Demands are generated daily,

then inventory status updated and all possible courses of actions analysed. The variables 'P','V', 'S/D' and 'mo' have the usual meaning. 'Vt' is the shortage rate on a time basis, i.e., time out of stock over total time. 'NTC' stands for the total number of cycles in the run.

```
**INVENTORY CONTROL SIMULATION**
 * Re-Order Level (R,Q) MODEL *


    PARAMETERS
    a       0.00    1.00
    D/Q     0.50    1.50    4.50
    RO/D    1.00    1.30    1.60

DAILY DEMAND DISTRIBUTION: GAMMA MOD= 1.0
DAILY DEMAND RATE,d= 30.0
LEAD TIME (DAYS),L= 10.
LEAD TIME DEMAND: GAMMA MOD= 10.0

SIMUL.LENGTH= 15000
RUN IN TIME=   500
RANDOM GEN=   1111
```

MOD	a	D/Q	RO/D	D	Q	P	V	Vt	S/D	mo	NTC
A	0.00	0.50	1.00	300	600	0.59	0.09	0.09	0.92	0.00	725
B	0.00	0.50	1.30	300	600	0.26	0.03	0.03	1.24	0.01	724
C	0.00	0.50	1.60	300	600	0.07	0.01	0.01	1.50	0.01	724
D	0.00	1.50	1.00	300	200	0.59	0.25	0.25	0.30	1.15	2174
E	0.00	1.50	1.30	300	200	0.24	0.08	0.08	0.55	1.15	2173
F	0.00	1.50	1.60	300	200	0.08	0.02	0.02	0.84	1.15	2174
G	0.00	4.50	1.00	300	67	0.60	0.46	0.46	0.14	4.42	6487
H	0.00	4.50	1.30	300	67	0.26	0.18	0.18	0.34	4.42	6488
I	0.00	4.50	1.60	300	67	0.08	0.05	0.05	0.61	4.42	6487
J	1.00	0.50	1.00	300	600	0.59	0.08	0.08	0.99	0.00	665
K	1.00	0.50	1.30	300	600	0.24	0.03	0.03	1.23	0.00	706
L	1.00	0.50	1.60	300	600	0.06	0.01	0.01	1.50	0.00	720
M	1.00	1.50	1.00	300	200	0.43	0.15	0.15	0.40	0.84	1847
N	1.00	1.50	1.30	300	200	0.19	0.06	0.06	0.58	0.92	2052
O	1.00	1.50	1.60	300	200	0.06	0.01	0.02	0.85	1.11	2141
P	1.00	4.50	1.00	300	67	0.29	0.19	0.19	0.21	3.29	5276
Q	1.00	4.50	1.30	300	67	0.12	0.08	0.08	0.38	3.89	5993
R	1.00	4.50	1.60	300	67	0.04	0.02	0.02	0.63	4.23	6328

```
ACTUAL DEMAND CHARACTERISTICS:
    RATE=    30.0
    ST.DEV=  29.9
*ET=12:02.6 PT=2:59.8 IO=0.2
```

Fig. 3.13 - Printout of inventory simulation

3.7   Conclusion

The formulations of the reorder level and periodical review operating procedures have been extended to situations in which orders overlap and demand is not captive. The approach used, based on a notional level, may be applicable to other inventory models whose replenishment decisions involve the setting of some sort of control stock level.

The notional level is the key concept that enables a common treatment for both lost sales and captive demand cases. The notional level is obtained by adding to the usual control level, be it the reorder level, the top up or other control level, the expected lost sales during the decision horizon. Referring to (3.23) in section 3.2, it reads

$$R = RO + m*Z \qquad\qquad (3.135)$$

where 'R' and 'RO' are the notional and the set reorder levels and 'm*Z' stands for the expected lost sales.

This new concept overcame the lost sales barrier, thus extending the applicability of the established inventory decision models. Service levels are given by expressions (3.27) and (3.29) in section 3.2. The same concept is applied to reach the formula for the average stock, expression (3.69) in section 3.3, and to expand the J-model in section 3.4. The formulas are the same as for captive

demand if the notional level 'R' substitutes for the set physical level 'RO'. The extra computational load is thus quite insignificant.

The traditional formulas to predict the perfomance variables with captive demand give poor predictions with non-captive situations if several orders are currently outstanding. The magnitude of the errors is clearily depicted in the diagrams included in section 3.6. If the overlapping degree is not high, those errors might be not as significant as others provenient from the estimates of the demand or supply characteristics: even so, the use of the notional level may be worthwhile because it brings no significant complexity.

The analytical way which led to the formulation of the notional level concept passed through some approximations in order to arrive at tractable and usable formulas. Each of these approximations seems reasonable but the chain of simplifications might have led to distorted and useless results. This is not the case, apparently, for the comparison between the predictions from the formulas so derived and results obtained by simulation have shown that they are close. Even when differences are larger, the predictions are better than if the notional level were ignored.

Chapter FOUR

REORDER FREQUENCY IN THE JOHNSTON MODEL

4.1   Sensitivity to the order frequency

The allocation criterion in J-model which has lead to the formula

$$P= 1/N*(F2+\lambda)/F$$

assumes that 'N', the number of orders per year, is an exogenous variable. Johnston mentions, in his paper, the 'mean inter-order interval' which is another way to fix the frequency. But he gives no help to choose a value for it. Apparently, that frequency is left to be decided on a basis of good sense.

It is obvious that the smaller are the orders the lower are the stocks. Decisions based on the Wilson classical approach should increase the number of orders if, by any reason, the ordering costs had become lower. This is apparent in formulas associated with· the so called economic reorder quantities, e.g. expression (4.61a) to be seen in section 4.3. With the J-model, however, a decrease in ordering costs does not lead necessarily to a more frequent reordering, as it is shown below.

The effect of the reorder frequency is exemplified in fig.4.1. The average stock 'S' and the shortage rate 'V' are variables being used as the two most important to assess the system perfomance. The figures plot the values of 'S/D' ('D'
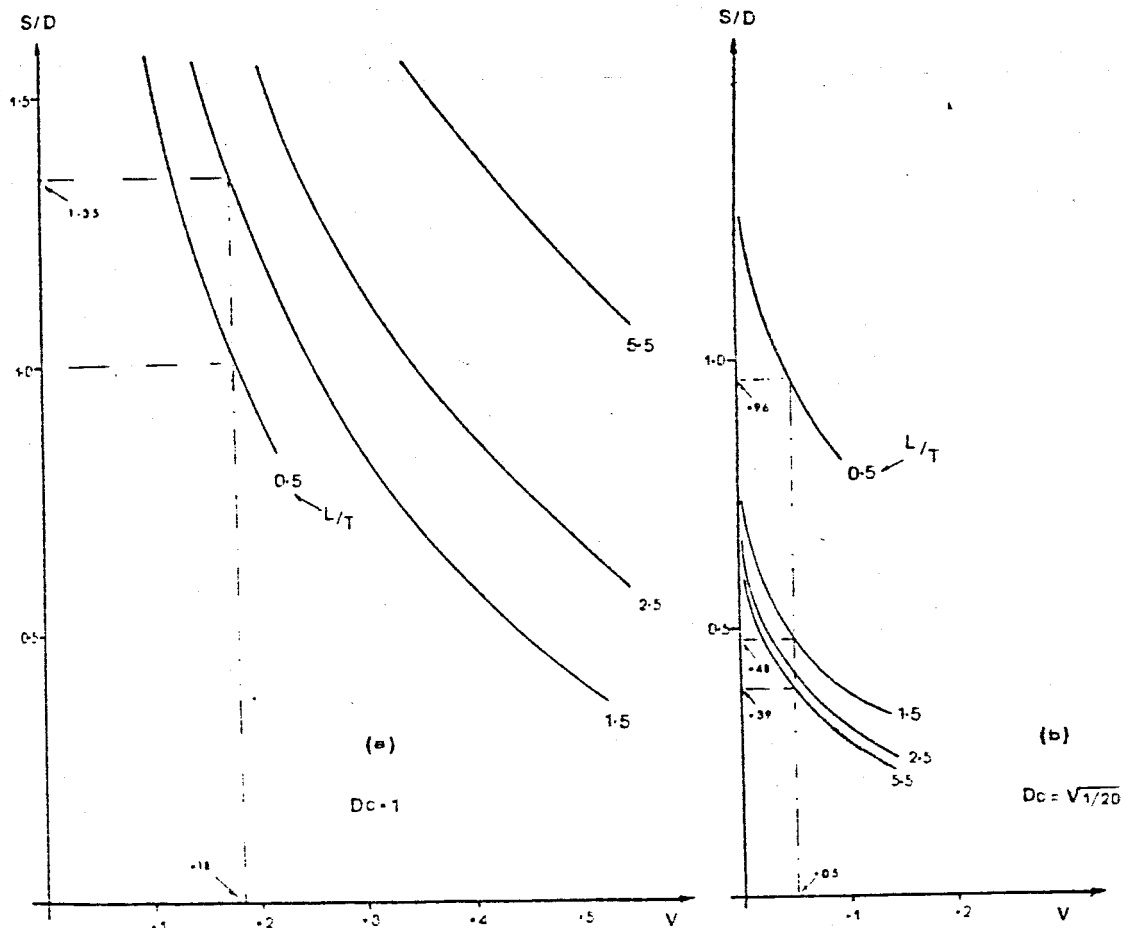
Fig. 4.1 - Reorder frequency effect on perfomance

is the lead time demand) against 'V' for different degrees
of overlapping 'L/T'. The curves, therefore, show the 'S,V'
perfomances the decision maker can expect for different
ordering frequencies. Values relate to the continuous review
reoder level control with captive demand. Captive demand has
been assumed in this example in order to isolate the present
problem from the lost sales case discussed in the previous
chapter. The overall pattern of the curves, however, is
common to other situations.

In fig.4.1a, the lead time demand is assumed to follow a

negative exponential. The coefficient of variation then is one, i.e. 'Dc=1'. We see that for 'L/T=0.5', i.e. 'N=26/L' orders a year (L in weeks), we can expect to have a shortage 'V' of about 18% for 'S/D=1'. If we increase the number of orders per year and retain the same shortage rate 'V', then the investment in stock must be increased. For instance, if 'L/T' changes from 0.5 to 1.5, the 'S/D' has to be raised from 1.0 to about 1.35 to keep the shortage at the same rate of 18%. That is, ordering more often brings no perfomance improvement, on the contrary, the result is to get either larger stocks or worse service. This may constitute a surprise.

Increasing the order frequency decreases the average stock if 'P', the probability of stockout, is kept constant. Increasing 'N', however, also increases the number of times the stockout may occur. i.e. 'P*N'. Thus, the shortage rate becomes higher. To recover the previous 'V', safety stocks need to be higher, too. This last adjustment may lead to higher average stocks than those we have begun with. In a case similar to the one depicted in fig.4.1a, increasing 'L/T' above 0.5 would lead to a worse situation: higher ordering costs and higher holding costs for the same service.

The figure (a) does not illustrate the effects in the region where 'L/T' is under 0.5, but it could be shown that when the overlapping degree raises from very low values there is, initially, a clear improvement in the 'S,V' perfomance. Such

an effect, however, becomes weaker and weaker·for higher overlapping degrees and ends being negative. The turning point depends on the demand distribution. For instance, if we look now to the picture (b) on the right side of the fig. 4.1, we see that 5% of shortage could be get either with 'S/D=0.96' and 'L/T=0.5; or 'S/D=0.48' and 'L/T=1.5'. There was a very signficant decrease of stock when the number of orders per year was trebled. However, the marginal effect for higher frequences decreases sharply. The decrease in 'S/D' when 'L/T' changes from 1.5 to 2.5 is already small; and the effect would be reversed at some stage if we kept on increasing the overlapping degree. The turning point in this example would be reached for a value of 'L/T' necessarily lower than 10, according to expression (5.57) to be seen in section 4.3.

The choice of a consistent value for 'N' may not be easy. In the Wilson approach it depends on estimates for holding and ordering costs. Such estimates may raise practical difficulties, yet the rationale is straight forward. The ralationship between 'N' and those costs is monotonic, the higher the ratio between holding and ordering costs the higher the 'N' and the lower the average stocks. It is not so simple with the J-model. A higher ordering frequency decreases the stock but increases the shortage quantities if the replenishment control levels are kept the same. The choice of a value for 'N', consistent with the model, is complicated by the fact that the marginal effect of 'N' on 'S' may be negative or positive (for the same 'V'). It is

not always obvious whether the extra cost of ordering more often is compensated by less capital tied up'or, on the contrary, increases that capital. The problem gets worse when groups of items are considered.

An analytical approach will be presented in the course of the next sections. Basically, we are interested in finding a way to get values for 'N' consistent with the objectives in J-model. First, the global formulation of that model will be developed to contemplate the 'optimization' of 'N'. Then, the single item case will be considered, i.e. orders are supposed to be raised individually, one for each item, at independent moments in time. This will help understand the nature of the parameters and variables involved. Subsequently, the analysis will be extended to group ordering. Finally, suggestions will be made to deal with the problem in practical terms.

4.2   Extension to the J-model

We recall once more from section 2.4 the formulation of the allocation criterion

$$\text{Max } \pi \text{ , subject to } S \leqq cS \qquad (4.1)$$

Now, 'π' and 'S' comprehend all the items in the inventory, grouped in 'J' buying families, each family containing 'I' items. The buying family has been introduced in section 2.4 as a group of items obtained from the same supplier and ordered simultaneously. Items can be identified as the elements of a matrix 'I*J'. The subscript (i,j) reads the item 'i' of the family 'j'. Families have, obviously, different numbers of items. We can make 'I' the same for all the families with a convenient introduction of dummies. This will simplify the notation. Another simplification is made through the collapsing of indices to indicate aggregation. Thus, for instance:

$$\pi = \sum \pi(j) = \sum \sum \pi(i,j) \qquad (4.3a)$$
$$S = \sum S(j) = \sum \sum S(i,j) \qquad (4.3b)$$

The profit associated with the family 'j' can be derived from expression (2.7) as:

$$\pi(j) = 52 * \sum F1(i,j)*d(i,j) - N(j)*FZ(j) -$$
$$F2(j)*S(j) - F0(j)*N(j) \qquad (4.5)$$

'N(j)', now, is an endogenous variable. Its value shall be in line with the objective stated in (4.1). The reoder cost 'FO' often is not very precise. So, it is convenient to stipulate a maximum number of orders per year 'cN'; this is similar to the limit placed on the stock investment. We have seen earlier that cost rates can be converted into constraints and vice-versa, so 'FO' and 'cN' really complement each other.

The objective stated in (4.1) above would then turn out

$$\text{Max } \pi \text{ , subject to } S \leq cS \text{ and } N \leq cN \qquad (4.7)$$

Now, we introduce the Lagrangean multipliers 'λs' and 'λn' in relation to the 'S' and 'N' constraints, respectively, in order to form the associated Lagrangean expression. Then, taking partial derivatives in relation to each of the 'R(i,j)'s and equating them to zero we would arrive at

$$P(i,j) = 1/N(j)*(F2(j)+\lambda s)/F(i,j) \qquad (4.9a)$$

as seen in section 2.4. Similarly, but now taking the partial derivatives in relation to each of the 'N(j)'s, it follows

$$N(j)**2 = 26*d(j)*(F2(j)+\lambda s)/(FZ(j)+FO(j)+\lambda n) \qquad (4.9b)$$

for a reorder level policy.

With a reorder level policy, a new order is triggered when the stock falls to the reorder level. This raises no theoretical difficulty when the orders are placed for each item independently, but it is not clear how an order should be triggered when it involves all the items of a buying family. The problem can be overcome by the introduction of the aggregate reorder level which is defined below.

Let us suppose for a moment that the 'P(i,j)'s had been set for a given 'N(j)' according to (4.9a). Then the 'R(i,j)'s and 'Z(i,j)'s could be calculated as usual. The corresponding expected shortage penalty for the family would be given by the formula

$$FZ(j)= \sum F(i,j)*Z(i,j) \qquad (4.11)$$

Denote by 'RZ(j)' the value of 'FZ(j)' that results from the particular values set for the 'P(i,j)'s. This 'RZ(j)' will perform the role of the aggregate reorder level for the family 'j'.

The aggregate reorder level 'RZ(j)' is then compared with the shortage penalty 'FZ(j)' which could be expected if a new family order were raised immediately. If an order is raised now, the expected shortage 'Z(i;j)' can be computed for each item from the quantity in stock and the lead time distribution; and then, the associated 'FZ(j)' could be calculated from (4.11). The current shortage penalty 'FZ(j)' should be kept updated at all times and whenever it falls

down to the aggregate reorder level 'RZ(j)', a new order should be triggered.

Note that when a family order is placed, the actual stock position of each item is unlikely to be exactly the 'R(i,j)' which individually had been admitted in order to compute the 'RZ(j)' by means of (4.11). One can expect that mal-distributions take place and hence, when the aggregate reorder level is reached, some items are above and others are below the respective 'R(i,j)'. The actual stock investment for a given 'RZ(j)' will be increased as a result of this mal-distribution. This effect which will be examined in subsection 4.2.1, was neglected in the derivation of (4.9b).

The equations (4.9) constitute the basis for determining the combinations 'P-N' which should be used for the trade-offs involving service level, investment in stock and procurement effort, the latter measured by the annual number of orders for the whole inventory. Note that the ordering cost rate 'FO(j)' adds to 'λn' as well as the holding cost rate 'F2(j)' adds to 'λs'. The model reacts to each of these sums and not to how they are composed.

The complementarity of 'F2(j)' and 'λs' has been discussed in section 2.5 and most of the considerations developed there in relation to holding costs apply to the reordering costs. In fact, 'FO(j)' and 'λn' also complement each other. 'λn' is the cost of an extra order per year, on a common

basis for all the buying families. Such a value can be compared to the current accountancy figures for the variable costs of reordering. 'FO(j)' is the extra charge, on top of 'λn', specific of that family 'j'. It might stand for extraordinary costs in procurement, freight, reception control or whatsoever, but usually it can be ignored in the presence of 'λn' which broadly accounts for the common ordering costs.

4.2.1   Analytical details

A remark shall be made about the consistency of the aggregate reorder level. The 'actual' stock levels when a family order is raised, 'Ra(i,j)', say, are unlikely to coincide with the 'R(i,j)'s specified earlier for each item. There is a variety of possible combinations of 'Ra(i,j)'s which have that same global shortage penalty 'RZ(j)'. Obviously, it would be desirable to have the minimum stock compatible with that penalty. Or in other words, the minimum 'Ra(j)' compatible with the set 'RZ(j)'. This could be written formally as

$$\text{Min}(\sum Ra(i,j)) \text{ , subject to } FZ(j) = RZ(j) \qquad (4.19)$$

By equating to zero the derivatives of the Lagragean equivalent in order to each of the 'Ra(i,j)'s, we would get

$$1 - \lambda * F(i,j) * P(i,j) = 0 \qquad (4.21a)$$

$$\text{or} \quad P(i,j) = (1/\lambda)/F(i,j) \qquad (4.21b)$$

This result is in line with (4.9a). It means that the optimum combination would be for the 'Ra(i,j)'s coinciding with the 'R(i,j)'s. Denoting by 'Ra(j)' and 'R(j)' the summations for 'Ra(i,j)' and 'R(i,j)', respectively, we have

$$Ra(j) \gtrsim R(j) \qquad (4.23a)$$

$$= R(j) + \Delta R(j) \qquad (4.23b)$$

$$\text{where} \quad \Delta R(j) = \sum \Delta R(i,j) \qquad (4.23c)$$

$$Ra(i,j) = R(i,j) + \Delta R(i,j) \qquad (4.23d)$$

Note that '$\Delta R(i,j)$' can assume negative values, but '$\Delta R(j)$' cannot.

These arguments become clearer if we depict the trade-off between two items as in fig.4.2. The straight line (1) is the locus of all possible combinations of 'Z(1)' and 'Z(2)' leading to 'RZ', the reorder level. Curves (2) and (3) represent all possible combinations of 'Z(1)' and 'Z(2)' if the actual stock values, when reordering, summed up to 'R' and 'RA', respectively. Note, 'RA' is greater than 'R' because the latter gives higher shortages. Curve (2) is tangent to line (1) on 'o'. This is the point which represents the best combination of 'Z's as it corresponds to the lowest stock at the reorder moment. On that point 'o', 'Ra(1)=R(1)' and 'Ra(2)=R(2)'; and, naturally, 'Ra=R'. The

reorder level 'RZ', however, is likely to be reached with a non-optimum combination of 'Z's. Let point 'A' represent that combination. On this point, 'Ra(1)+Ra(2)=RA' and 'RA' is grater than 'R'.

Curves
1: F(1)*Z(1)+F(2)*Z(2)= RZ= const
2: Ra(1)+Ra(2)= R= const
3: Ra(1)+Ra(2)= RA= const
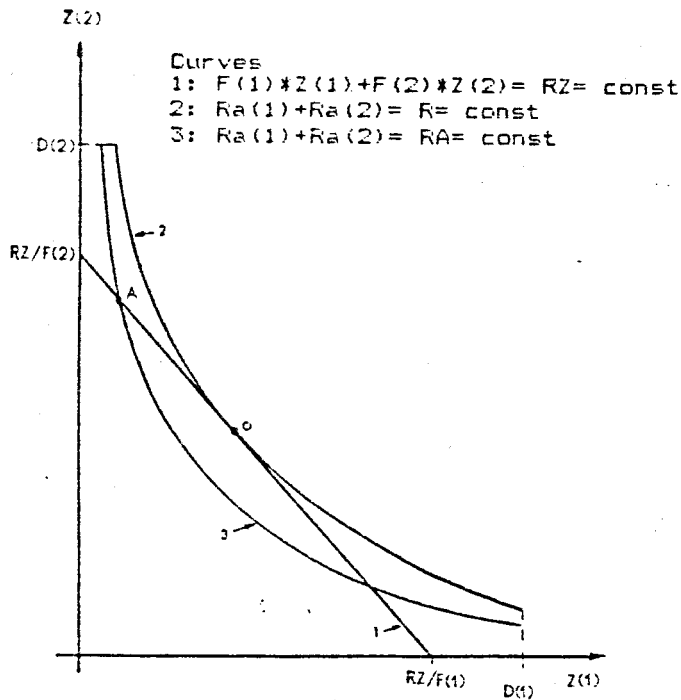
Fig.4.2 - Effect of mal-distribution

These considerations about the aggregate reorder level are required for the derivation of equation (4.9b). This was obtained by equating to zero the partial 'N(j)' derivative of the Lagrangean expression for (4.5). It would follow, then

$$-FZ(j)-N(j)*(\partial FZ(j)/\partial N(j))-$$
$$(F2(j)+\lambda s)*(\partial S(j)/\partial N(j))-(F0(j)+\lambda n)= 0 \qquad (4.25)$$

Then, it has been assumed that

$$\partial FZ(j)/\partial N(j)= 0 \qquad (4.26a)$$

-4.2.1-   -116-

$$\partial S(j)/\partial N(j) = -26*d(j)/N**2 \qquad (4.26b)$$

Consider, first, the single item case. Under a reorder level policy, the relevant decision period is the lead time 'L'. 'Z', therefore, does not depend on the reorder interval, so (4.26a) holds. On the other hand, from (2.13) in section 2.4

$$S = R-D+Q/2 \qquad (4.27a)$$
$$= R-D+26*d/N \qquad (4.27b)$$

As '$\partial R/\partial N = 0$', (4.26b) follows.

Consider, now, the multi-item situation. The value of 'FZ(j)' is fixed from (4.19) and does not depend on 'N(j)', therefore, (4.26a) holds. The expression (4.26b), however, implies that 'Ra' does not vary with 'N', when 'R' (or, equivalently, 'RZ') is kept constant. In fact, 'Ra' should replace 'R' in (4.27) because the average aggregate stock should take into account the actual aggregate reorder level.

Assume for a moment that the nominal stock of each item, when a new group order is placed, is raised to the level 'R(i,j)+Q(i,j)'. If the subsequent actual demands, for all items of the family, were exactly as expected, the next order would have taken place 'T' time units later. At this moment, all the items would be also at their respective 'R(i,j)'s, i.e. the inventory would be at its optimum reorder level combination (point 'o' in fig.4.2).

The actual issues between two successive orders, however, are not exactly $'Q(i,j)'$ but say, $'Q(i,j)-\Delta R(i,j)'$, in line with (4.23d). To deal with these differences, we can use the approach in (3.41) and write

$$\Delta Z \sim -Po*(\Delta R) + fo*(\Delta R)**2 \qquad (4.29)$$

where 'fo' stands for the value of the probability density function of the lead time demand in relation to a reference stand point 'o'. Now, aggregate the 'I' items of the family 'j' as follows

$$\Delta FZ(j) = -\sum F(i,j)*Po(i,j)*\Delta R(i,j) + \sum fo(i,j)*\Delta R(i,j)**2 \qquad (4.31)$$

The optimum combination is now taken as the reference stand point. $'\Delta FZ(j)'$ is zero, because the reordering always takes place when $'FZ(j)=RZ(j)'$. It follows, then, from (4.9a) and (4.21b) that

$$\Delta R(j) = N(j)/(F2(j) + \lambda s)*(\sum fo(i,j)*\Delta R(i,j)**2) \qquad (4.33)$$

The expected value for $'\Delta R(i,j)**2'$ is the variance of the demand between two successive reorder points. The period involved is, on average, $'T(j)=52/N(j)'$. If 'ds' is the standard deviation of the demand per unit of time, we could write

$$E(\Delta R(i,j)^{**}2) = C(I)^*T(j)^*ds(i,j)^{**}2 \qquad (4.35)$$

where 'C(I)' is a function of the number of the items of the family and 'E(.)' stands for the expected value.

The expected value for (4.33), having (4.35) in mind, converts into

$$E(\Delta R(j)) \sim 52^*C(I)/(F2(j)+\lambda s)^*$$
$$^*(\sum fo(i,j)^*ds(i,j)^{**}2) \qquad (4.37)$$

The expected value of 'ΔR' does not depend on 'N'...significantly, at least, if the above approximations are reasonable. Note, further, that 'fo(i,j)*ds(i,j)' might be easily converted into a standard density function.

The average stock 'S' is modified if '$E(\Delta R)$' is greater than zero. To keep, however, the same number of orders 'N(j)', the 'Q(i,j)'s shall remain the same. So, the top up level is given by

$$\text{Top up level} = Ra(i,j)+Q(i,j) \qquad (4.39a)$$
$$= R(i,j)+\Delta R(i,j)+Q(i,j) \qquad (4.39b)$$

The actual average stock, then, is increased to '$Sa=S+\Delta R$'. However, as 'ΔR' is independent of 'N', '$\partial Sa/\partial N=\partial S/\partial N$', therefore the expression (4.13b) holds. The $\Delta R(i,j)$ can be approximated empirically from the inventory running data.

4.3   Single item order frequency

Equations (4.9) can be used to generate 'P-N' combinations consistent with the allocation criterion in J-model. Some further refinements are convenient to ease their manipulation. They are introduced in this section for single item reordering and will be extended in the next section to buying families.

Assume then, that decisions are made individually for each item. So are the orders. The basic equations in (4.9) would simplify to:

$$P = (1/N)*(F2 + \lambda s) \qquad (4.41a)$$

$$N**2 = 26*d*(F2 + \lambda s)/(FZ + F0 + \lambda n) \qquad (4.41b)$$

The following notation is, now, introduced:

Margin loss, $ML = (52*F1*d - \pi)/(52*F1*d)$ (4.43a)

Shortage penalty ratio, $B1 = F/F1$ (4.43b)

Holding ratio, $B2 = (F2 + \lambda s)/F1$ (4.43c)

Ordering charge ratio, $B3 = (F0 + \lambda n)/(d*L*F1)$ (4.43d)

Stock turn, $U = 52*d/S$ (4.43e)

Years of stock, $OU = 1/U$ (4.43f)

Recall also that:

Shortage rate, $V = Z/(d*T)$ (4.45a)

Overlapping rate, $LOT = L/T$ (4.45b)

This notation is introduced for two main reasons. The first is to facilitate the derivations to be presented later. The second and more important is to make apparent relationships involving cost variables (B1, B2 and B3) and perfomance variables (V, OU and LOT). For instance, the formulation in (4.7) is strategically equivalent to

$$\text{Min } (ML = B1*V + B2*OU + B3*LOT) \qquad (4.47)$$

This equation can be a useful tool for decision making and will be examined in section 4.5.

Let us discuss the meaning of the new symbols before proceeding. The margin loss 'ML' gives the variable costs as a fraction of the potential profit ceiling. The variable costs referred to are the costs of ordering and holding stocks as well as of running out. 'B1' is the penalty for shortages as a fraction of the profit rate 'F1'. 'B1' mirrors the marketing policy, as mentioned in chapter 2 and can be expected to be close to one. 'B2' expresses the holding costs measured on that same basis. Holding cost rates for non-perishable and non-fashion products are normally in the range of 30 to 50% Typical figures for 'F1' are 50 to 100%, therefore, 'B2' is likely to stay between 0.5 and 1.0. Note that with "volatile" products (fashionable, perishable, pilferable, etc.), when holding costs are higher, so are, normally, the profit margins. Thus 'B2' is likely to remain in the same range. 'B3' is the reordering cost as a fraction of the profit of the lead time

demand. The value of 'B3' is likely to have a wide variation. Typical figures are well below 1. Note that 'B1', 'B2' and 'B3' are costs which have been expressed in relation to the profit in order to deal with standardized figures, easier to manipulate. These 'B's can be used as buttons or knobs to search the appropriate performance vector, as it will be discussed in section 4.5.

The stock turn 'U' and years of stock 'OU' are quite common yardsticks. The former gives the number of times per year the capital is reinvested (or, the whole stock is renewed). The latter, the time length of each investment (or, the time to exhaust stocks if fresh cargoes were not received). A lower 'OU' corresponds to a lower investment in stock. The shortage rate 'V' is the complement to one of the rate of fulfilment and thus, a lower 'V' gives a better service level to customers. The overlapping rate 'LOT' relates to the reorder frequency and the lower it is the lower is the number of orders. Therefore, decreasing 'OU', 'V' and 'LOT' is the aim, and this is in line with (4.47).

The equations (4.41) would turn out:

$$P= (1/N)*B2/B1 \qquad\qquad (4.49a)$$
$$N**2= 26*d*B2/(B1*Z+B3*d*L) \qquad\qquad (4.49b)$$

Rearranging (4.49) with (4.45) in mind, we get

$$N= (1/2)*B2/(B1*V+B3*LOT) \qquad\qquad (4.51a)$$

$$P/2 - V = (B3/B1)*LOT \qquad (4.51b)$$
$$= (L/52)*(B3/B1)*N \qquad (4.51c)$$

The formula (4.51a) shows that 'N' remains finite even if the reorder costs were zero (B3=0). Recall that with the Wilson model, 'N' would tend to infinity in this case. The last two expressions show that, for a consistent decision, 'V' shall be lower than 'P/2'.

The link just mentioned between 'P' and 'V' may become useful to write off, a priori, high values for the overlapping degree. Further insights may be derived from the particular demand distribution. As shown in subsection 4.3.2, if demand is gamma distributed, when 'P' tends to zero we have

$$\lim (V/P) = LOT*(Dc**2) \qquad (4.53)$$

where 'Dc' is the coefficient of variation of the lead time demand. It is also shown, for any value of 'P', that:

$$V/P \geq LOT*Dc**2 \ , \ for \ Dc \leq 1 \qquad (4.55a)$$
$$\leq LOT*Dc**2 \ , \ for \ Dc \geq 1 \qquad (4.55b)$$
$$= LOT \qquad , \ for \ Dc = 1 \qquad (4.55c)$$

From (4.51b), 'V/P<1/2, and then

$$LOT < 1/(2*Dc**2) \quad , \ for \ Dc \leq 1 \qquad (4.57)$$

This last expression is an important result. The condition 'Dc≤1' is generally met, if the very slow movers are excluded, and then, the boundary imposed by (4.57) overrides immediately the practice of high overlapping degrees. In particular, for the negative exponential distribution, 'Dc=1'. So, 'LOT' shall be lower than 1/2. This explains why, in fig. 4.1a, the performance goes worse when 'LOT', i.e. 'L/T' increases from 0.5.

The simultaneous equations (4.49) determine the P-N combination. An exact solution, however, is not easy to find. Empirical approximations for the probabilistic functions, as those in Johnston (1980), can simplify the task considerably. That paper proposes, for Gamma and for Normal distributed demands, approximations of the form

$$Z/D= A0+A1*P+A2*P**2 \qquad (4.58)$$

which are reproduced in subsection 4.3.1. Using such approximations, we would find that the overlapping degree should be

$$LOT= (-C1+\sqrt{C1**2+4*C0*C2}) \ / \ (2*C2) \qquad (4.59a)$$

where

$$C0= 1/2-A2*B4 \qquad (4.59b)$$

$$C1= A1 \qquad (4.59c)$$

$$C2= A0/B4+(52/L)*B6 \qquad (4.59d)$$

and

$$B4= (L/52)*(B2/B1) \qquad (4.59e)$$

$$B5= B3/B1 \qquad (4.59f)$$

$$B6= B3/B2 \qquad (4.59g)$$
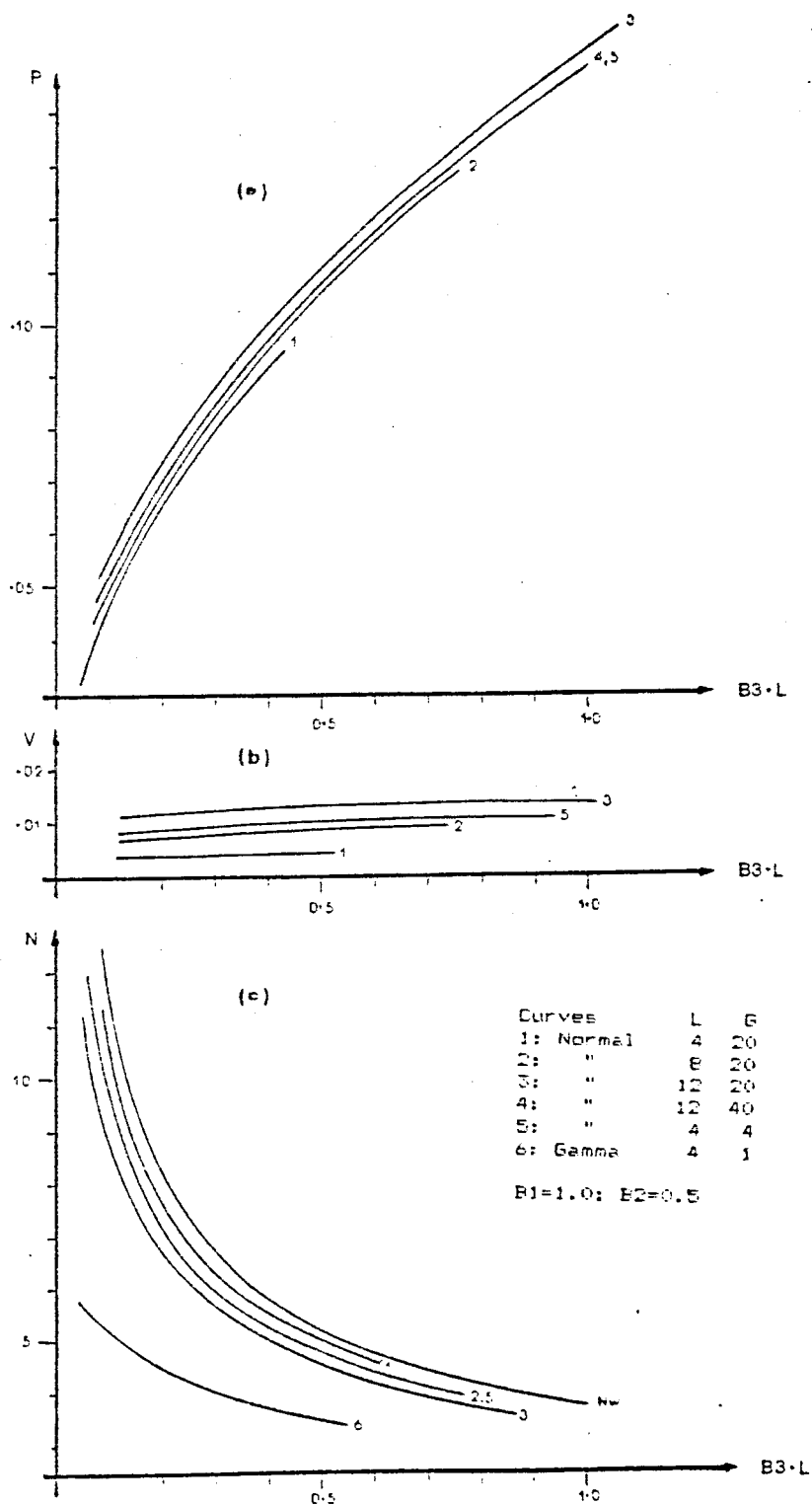
Fig 4.3 - Cost effects on the combination of attributes

The annual number of orders is 'N=(52/L)*LOT', 'P' is given by (4.49a) and the computation of other perfomance attributes is, now, straight forward. Fig.4.3 shows the variations of 'P','V' and 'N' with the reorder costs, for different demand distributions. Note the low value of the shortage rate.

It is interesting to compare the optimum frequency, as obtained from derivations above, to the values from the Wilson model. We recall that the latter assumes deterministic demand with no shortages. The choice of 'N' is such that the variable costs of ordering and holding stocks are minimized. Formally, we have

$$\text{Min (Cost= } (F2+\lambda s)*S+(F0+\lambda n)*Nw \text{)} \qquad (4.60)$$

where, 'Nw' stands for the Wilson 'N'. Equating to zero the derivatives in relation to 'Nw' and rearranging, it follows:

$$Nw**2= 26*d*(F2+\lambda s)/(F0+\lambda n) \qquad (4.61a)$$
$$= (26/L)*(B2/B3) \qquad (4.61b)$$

Comparing this result with the expression (4.49b) we see that 'Nw' is an upper bound for the corresponding value in J-model, thus it can be used to set a practical limit if one wants to avoid the computations involved in the method described above. 'Nw' is plotted in fig.4.3c, for comparison purposes.

4.3.1 Analytical determination of LOT

The Appendix III in Johnston (1980) gives approximate
expressions to compute service levels in relation to Gamma
and Normal distributed lead time demands. The expressions
below are drawn from there.

For the Gamma distribution, the percentage of lost sales
(PLS) is given by

$$PLS= A7*P+A8*P**2$$

where

$$PLS= 100*Z/D$$
$$A7= 9.4608205+101.30969/k-9.5595537/(k**2)$$
$$A8= 20.574471+9.9995001/k-27.350124/(k**2)$$

and 'k' is the modulus of the distribution, 'Z' is the
expected shortage quantity per cycle and 'D' is the mean
lead time demand.

For the Normal distribution

$$PLS= X6+X7*P+X8*P**2$$

where

$$X6= -0.0495939$$

X7= 40.16012

X8= 78.359788

but now

PLS= 100*Z/Ds= (100/Dc)*(Z/D)

Both cases could be standardized under the form

$$Z/D= A0+A1*P+A2*P**2 \qquad (4.62)$$

where, for the Gamma distribution,

A0= 0

A1= A7/100

A2= A8/100

And for the Normal

A0= X6*Dc/100

A1= X7*Dc/100

A2= X8*Dc/100

On the other hand, from (4.45)

$$V= (Z/D)*LOT \qquad (4.63a)$$

Then, it follows from (4.51)

$$N= (1/2)*B2/(B1*Z/D*LOT+B3*LOT) \qquad (4.63b)$$

As $$N= (52/L)*LOT \qquad (4.63c)$$

-4.3.1-

Then $(52/L)*(B1*Z/D+B3)*LOT-B2/2= 0$ (4.63d)

Denote

$$B4= (L/52)*B2/B1 \qquad (4.64a)$$

$$B5= B3/B1 \qquad (4.64b)$$

$$B6= B3/B2 \qquad (4.64c)$$

Then, from (4.61d)

$$(1/B4*Z/D+52/L*B6)*LOT**2-1/2= 0 \qquad (4.65a)$$

From (4.49), $P=B4/LOT$ (4.65b)

Combining, now, (4.62) and (4.65) we get

$$(1/B4)*\left[A0*LOT**2+A1*(P*LOT)+A2*(P*LOT)**2\right] +$$
$$+(52/L)*B6*LOT**2-1/2= 0 \qquad (4.67a)$$

or

$$(A0/B4+52/L*B6)*LOT**2+ A1*LOT- (1/2-A2*B4)=0 \qquad (4.67b)$$

Denoting

$$C0= 1/2-A2*B4 \qquad (4.69a)$$

$$C1= A1 \qquad (4.69b)$$

$$C2= A0/B4+52/L*B6 \qquad (4.69c)$$

the solution for (4.67) is

$$LOT= (-C1+\sqrt{C1**2+4*C0*C2}) / (2*C2*) \qquad (4.71)$$

-4.3.1-

Note that only the positive root is contemplated in (4.71).
In fact, 'C1' and 'C2' are positive. If 'C0' is also
positive, that root is the only one which gives positive
values for 'LOT. Currently, 'B4' is much lower than 1. So is
'A2' as it can be seen above. Thus, 'C0' is positive for
most of the cases. Otherwise, the other solution for (4.67)
would also have to be considered.

4.3.2   Details about the ratio V/P

Consider the expectation for the stockout and shortage rates
given by the expressions

$$P = \int_R^\infty f(x)*dx \qquad\qquad (4.76a)$$

$$= -\partial Z/\partial R \qquad\qquad (4.76b)$$

$$Z = \int_R^\infty (x-R)*f(x)*dx \qquad\qquad (4.76c)$$

$$= \int_R^\infty x*f(x)*dx \;-R*P \qquad\qquad (4.76d)$$

$$V = LOT*Z/D \qquad\qquad (4.76e)$$

The density function for the Gamma distribution could be
written

$$f(x)= (1/Gk)*(h**k)*(x**(k-1))*EXP(-h*x)$$

where 'k' is the modulus, 'h' the scale factor and 'Gk' the Gamma function. Denote by 'f1(x)' the first derivative of 'f(x)'. Then note that

$$f1(x)= (k-1)*(f(x)/x)-h*f(x) \qquad (4.76f)$$

Integrating both sides of (4.76f) as follows

$$\int_{R}^{\infty} f1(x)*dx= \int_{R}^{\infty}(k-1)*(f(x)/x)*dx-h*\int_{R}^{\infty}f(x)*dx \qquad (4.77a)$$

then

$$0-f(R)= \int_{R}^{\infty}(k-1)*(f(x)/x)*dx-h*P \qquad (4.77b)$$

hence

$$h*P-f(R)= \int_{R}^{\infty}(k-1)*(f(x)/x)*dx \qquad (4.77c)$$

On the other hand, the integration by parts of the r.h.s. of (4.76a) gives

$$P= \left[x*f(x) \right]_{R}^{\infty} -\int_{R}^{\infty}x*f1(x)*dx \qquad (4.78a)$$

and considering (4.76f)

$$P= 0-R*f(R)-\int_{R}^{\infty}(k-1)*f(x)*dx+h*\int_{R}^{\infty}x*f(x) \qquad (4.78b)$$

and considering (4.76d)

$$P= -R*f(R)-\int_{R}^{\infty}(k-1)*f(x)*dx+h*(Z+R*P) \qquad (4.78c)$$

Hence

$$h*Z= k*P-R*(h*P-f(R)) \qquad (4.78d)$$

and considering (4.77c)

$$h*Z= k*P-R*(k-1)*\int_{R}^{\infty}(f(x)/x)*dx \qquad (4.78e)$$

Now, as 'h=k/D', dividing by 'k' and 'P' it follows

$$Z/(D*P)= 1-R*(k-1)/(k*P)*\int_{R}^{\infty}(f(x)/x)*dx \qquad (4.78f)$$
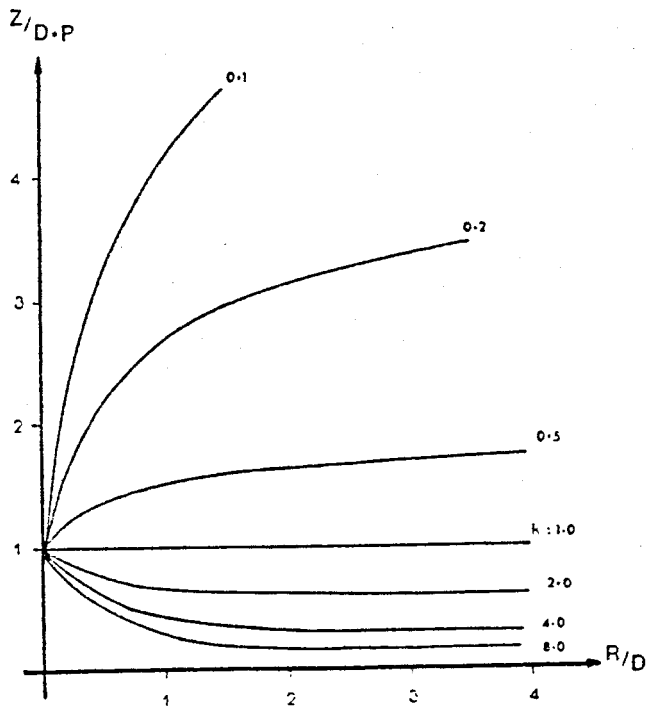
Fig. 4.4 - Ratio between shortages and stockouts

In (4.78f), we see that 'Z/(D*P)=1' for k=1. Moreover 'x' and

'f(x)' are non-negative (so is the integral on the r.h.s.); 'R' and 'k' are positive; therefore, 'Z/(D*P) is greater than one for 'k' lower than one and lower than one for 'k' higher than one. This is confirmed by fig.4.4 where each curve was drawn from a set of points obtained by computing 'Z' and 'P' by the formulas in Johnston (1980). The curves are monotonic and the limiting value is '1/k'. This can be seen by applying successively the rule of L'Hopital to 'Z/P' while making 'R' tending to infinity:

$$\lim (Z/P) = \lim \left[(\partial Z/\partial R)/(\partial P/\partial R) \right] \qquad (4.79a)$$
$$= \lim \left[P/f(R) \right] = -\lim \left[f(R)/f1(R) \right] \qquad (4.79b)$$

and considering (4.76f)

$$\lim (Z/P) = -\lim \left[1/((k-1)/R-h) \right] \qquad (4.79c)$$
$$= 1/h \qquad (4.79d)$$

Therefore,

$$\lim \left[Z/(D*P) \right] = 1/(D*h) = 1/k \qquad (4.79e)$$

The curves in fig.4.4 are monotonic and tend to '1/k'. Then, one can conclude that

$$1 < Z/(D*P) < 1/k \qquad \text{for } k<1 \qquad (4.80a)$$
$$Z/(D*P) = 1 \qquad \text{for } k=1 \qquad (4.80b)$$
$$1/k < Z/(D*P) < 1 \qquad \text{for } k>1 \qquad (4.80c)$$

The boundaries for 'V/P' as in (4.55), would then be derived from (4.76e), recalling that '1/k=Dc**2'.

4.4   The family order frequency

The manipulation of formulae (4.9) in relation to a buying family can follow the same steps as in the previous section. The results can be expressed by the same formulas though the symbols have now a more complex meaning to contemplate the multi-item situation. This will be seen below.

The subscript 'j' referring to the family will be omitted in order to simplify the notation. The subscript 'i', as usual, stands for the item. Then it reads

$$B1(i) \text{ for } B1(i,j) = F(i,j)/F1(i,j) \tag{4.81a}$$

$$B2(i) \text{ for } B2(i,j) = (F2(j)+\lambda s)/F1(i,j) \tag{4.81b}$$

$$B3(i) \text{ for } B3(i,j) = (F0(j)+\lambda n)/(d(j)*L(j)*F1(i,j)) \tag{4.81c}$$

$$B4(i) \text{ for } B4(i,j) = (L(j)/52)*(B2(i,j)/B1(i,j) \tag{4.81d}$$

$$B5(i) \text{ for } B5(i,j) = B3(i,j)/B1(i,j) \tag{4.81e}$$

$$B6 \quad \text{ for } B6(j) \quad = B3(i,j)/B2(i,j) \tag{4.81f}$$

Note that $B4(i)*B6 = L/52*B5(i)$ (4.81g)

The basic equations corresponding to (4.49) read now

$$P(i) = 1/N*B2(i)/B1(i) \tag{4.83a}$$

$$= B4(i)/LOT \tag{4.83b}$$

$$N**2 = 26*d/ \left(\sum B1(i)/B2(i)*Z(i) + B6*d*L \right) \tag{4.83c}$$

This last equation can be rearranged to give

$$N = (1/2)/ (\sum B1(i)/B2(i)*Z(i)/DT + B6*LOT) \qquad (4.85)$$

where 'DT=d*T' is the family mean demand value in a replenishment period 'T'. The formula (4.85) corresponds to (4.51a) for the single item.

Finally, adopting empirical relationships, we would arrive at the same formula (4.59)

$$LOT = (-C1 + \sqrt{C1**2 + 4*C0*C2}) / (2*C0) \qquad (4.86a)$$

But now,

$$C0 = 1/2 - a2*B4 \qquad (4.86b)$$
$$C1 = a1 \qquad (4.86c)$$
$$C2 = a0/B4 + (52/L)*B6 \qquad (4.86d)$$

and

$$a0 = (1/D)* \sum A0(i)/D(i) \qquad (4.86e)$$
$$a1 = (1/D)* \sum A1(i)/D(i) \qquad (4.86f)$$
$$a2 = (1/D)* \sum A2(i)/D(i) \qquad (4.86g)$$

where 'A0', 'A1' and 'A2' are the coefficients of the empirical expression (4.58) for each item, 'D(i)' is the mean lead time demand for the item and 'D' is the aggregate demand for the family (recall that demands are expressed in money units).

The coefficients in (4.86) assume that the same value 'B4'

can be adopted for all the items in the family. The general situation of having distinct 'B(i)'s is dealt with in subsection 4.4.1. In this case determining 'N' would become much more complicate. The question now is to know how accurate 'B4' needs to be for the purpose of computing 'N'. If the required accuracy is not high, then, for practical purposes we could use a common 'B4'.


```
1: Gamma, (1/Dc)**2= 4, D=25
2: Normal,    "    = 20, D=625
12: Aggregation of 1 and 2
Nw: Wilson model, Nw=8.06

L=4; B2=0.5; B3=0.05; B6=0.1
```
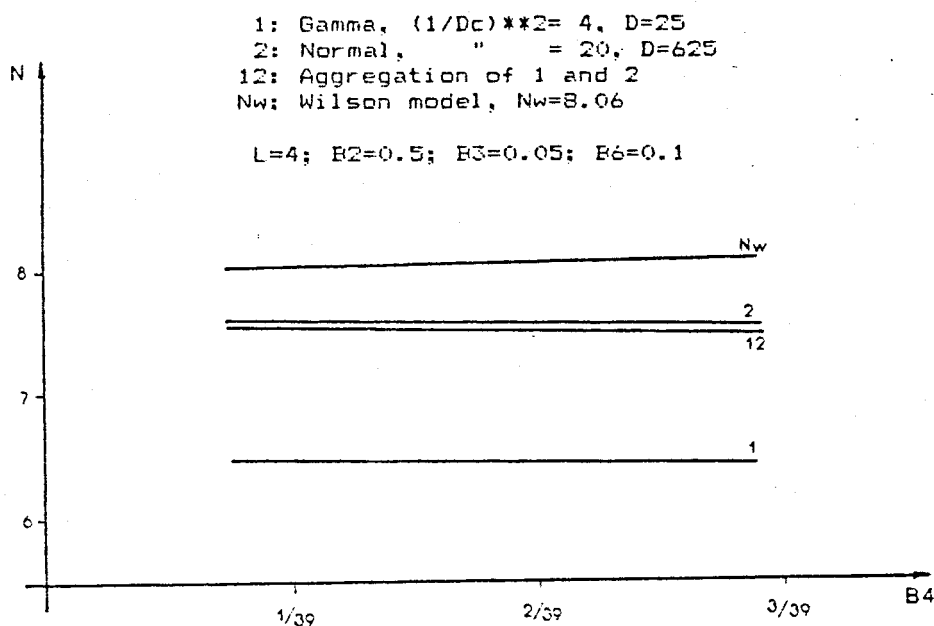
Fig. 4.5 - Sensitivity of 'N' to 'B4'

The sensitivity of 'N' to 'B4' is shown in fig.4.4, for a number of situations. In all of them, 'L=4' and 'B6=0.1'. Curve (1) relates to a lead time demand Gamma distributed with modulus equal to 4 (i.e.,'1/Dc**2=4'). Curve (2) corresponds to a Normal distribution with '1/Dc**2=20'. Note that as 'L=4', curve (1) corresponds to having a weekly demand with modulus of one, thus close to the limits of skewness, whilst curve (2) being Normal is at the opposite end.

The sensitivity to 'B4' is negligible when compared with the sensitivity to the distribution, though the latter is not very high, either. Curve (w) gives the value for the Wilson deterministic model (Nw=8.06, for any B4). The curve (12) corresponds to the family constituted by items of curves (1) and (2). Curve (12) lays between (1) and (2) as expected. The latter, which relates to the faster moving item, is shown to be highly dominant. Normally, the higher the demand the lower the coefficient of variation. Demands in cases (1) and (2) are in the ratio of 1:25. The relationships between 'D' and 'Dc' that have been used, agree with empirical data.

The example above reveals that having an accurate value for 'B4' to calculate the reorder frequency, is not very important. The low sensitivity to this parameter indicates that a common 'B4' for each family may well be satisfactory, since this value is chosen having in mind the dominance of the fast sellers. The problem is then reduced to compute the reorder frequency for as many single-like items as the number of families.

4.4.1 Analytical details

The derivations leading to the formulas (4.86) are as follows. After rearranging (4.85), then

$$\left(\sum 1/B4(i)*Z(i)/D + 52/L*B6\right)* LOT**2 - 1/2 = 0 \qquad (4.87)$$

The empirical relationships corresponding to (4.58) are now written as

$$Z(i)/D(i) = A0(i) + A1(i)*P(i) + A2(i)*P(i)**2 \qquad (4.89)$$

$$Or \quad Z(i)/D = a0(i) + a1(i)*P(i) + a2(i)*P(i)**2 \qquad (4.89b)$$

$$where \quad ak(i) = Ak(i)*D(i)/D , \quad k = 0,1,2 \qquad (4.89c)$$

Bringing these results into (4.87) with (4.43a) in mind we arrive easily at the equation

$$C2*LOT**2 + C1*LOT - C0 = 0 \qquad (4.91a)$$

$$where \quad C2 = \sum a0(i)/B4(i) + 52/L*B6 \qquad (4.91b)$$

$$C1 = \sum a1(i) \qquad (4.91c)$$

$$C0 = 1/2 - \sum a2(i)*B4(i) \qquad (4.91d)$$

The assumption that 'C0' and 'C1' are positive will be maintained on the grounds mentioned in the previous section. Then, the optimum 'LOT' for the model would be given as mentioned earlier by

$$LOT = (-C1 + \sqrt{C1**2 + 4*C0*C2}) / (2*C2) \qquad (4.93)$$

Therefore, we can determine analytically the optimum aggregate 'LOT' (and, hence, 'N') for the family, assuming that we know the demand distribution for each item and we are able to discriminate the 'B4(i)'s.

When the same 'B4' can be assigned to all the items in the family, then, from (4.81g) all items have the same 'B5'. And, from (4.83a) all have the same 'P'. Equation (4.85), then, could be rearranged to give

$$N = 26/L*B4/(V+B5*LOT) \tag{4.95a}$$

and $$P/2-V = B5*LOT \tag{4.95b}$$

where $$V = \sum Z(i)/DT \tag{4.95c}$$

These expressions are parallel to (4.51)

The aggregate 'Z/D' as a function of 'P' is easy to find in this case as the 'P(i)'s are the same for all the items. It follows from (4.89):

$$Z/D = a0 + a1*P + a2*P**2 \tag{4.97a}$$

where $$ak = \sum ak(i) , \quad k = 0,1,2 \tag{4.97b}$$

We would also find that:

$$C2 = a0/B4 + 52/L*B6 \tag{4.99a}$$

$$= (a0+B5)/B4 \tag{4.99b}$$

$$C1 = a1 \tag{4.99c}$$

$$C0 = 1/2 - a2*B4 \tag{4.99d}$$

4.5   Shadow prices for decisions

The preceding sections helped make clearer the relationships
between the input vector of costs and the output vector of per-
formance attributes. The components of those vectors have been,
basically, holding, shortage and reorder costs, on one hand;
services, stock levels and reorder frequency, on the other.
Different combinations could be chosen. Performance vectors
could include more variables, but only two would be independent.
This means that the model has two degrees of freedom. The user,
however, may like to manipulate three control knobs, for instance,
$(F0+\lambda n; F; F2+\lambda s)$ or $(B1; B2; B3)$ which relate to specific costs. The
model converts these parameters into ratios, e.g. $B4$, $B5$, $B6$
defined in (4.59). Note in the latter that each one is a combin-
ation of the other two. They can also be used directly for
control, but in this case with two knobs only.

The cost vector was initially considered as the input because
this is the conventional way of introducing these models.
However, the decision maker very often has a much clearer picture
about the attributes for the performance variables than about
those costs. He might hesitate about writing a figure for 'F'
while being quite positive about a maximum shortage rate of 2%.
The input, in such cases, could be in terms of $(V; B2; B3)$.

If the decision maker were absolutely sure that input
figures were right,   the decision process would terminate
just there.   This is not the common case. More often he is
not so certain. He may feel,   say, that a shortage   rate
higher than 2% is quite undesirable.   'V' might work as an

input, then 'B1' would be the dependent variable, an implied shortage penalty. The system should enable the input of any relevant vector of variables and calculate the values for the dependent ones. If the latter included costs, these could be used for consistency checks. For instance, suppose that for 'V=2%', the corresponding 'B1' would be 3.25: this would mean each unit short costs 3.25 times the profit per unit, which is unusually high. So, unless it relates to an exceptionally high sensitive item, such a figure might prompt to reconsider the targets set for 'V'.

The choice of the right trade-off, ultimately, depends on the 'feelings' or intuition of the decision maker. The reasons why those ultimate decisions cannot be modeled have been mentioned earlier. That 'intuition', however, can be improved by providing consistency checks. For example, when the manager decides to decrease 'V' from 4 to 2%, does he realize the cost of it? Is he, really, prepared to pay?

The margin loss 'ML' has been introduced earlier in (4.47). It measures the relative loss of profit. It converts into the absolute loss after multiplying by the potential annual profit (a figure which is easily reckoned). There is an 'ML' associated to each performance vector which represents the 'shadow price' for that performance.

Let us assume that the decision maker can rank the perfomance vectors and that a higher ordinal means a higher preference. Call that ordinal, the 'utility' of the

performance vector. Fig.4.6 depicts the margin loss against
utility. The decision maker is assumed to be consistent, so,
transitive properties apply to those utilities. We are
interested in situations like those represented by 'A' and
'B' lying on the so called 'effective frontier'. A point
like 'C' is dominated by 'A' (the same utility with lower
margin loss) and by 'B' (higher utility for the same margin
loss). The effective frontier in fig.4.6 gives the most
preferred performance vectors for each margin loss.
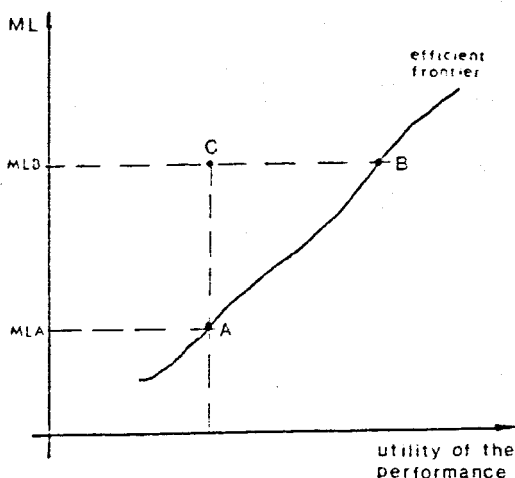
Fig. 4.6 - Margin loss and performance

The consistency assumption imposes that the curve in the
picture be monotonically increasing. If the manager wants to
move the perfomance from 'A' to 'B', the implied cost of
such a move is given by an additional margin loss of
'MLB-MLA'. This, multiplied by '52*F1*d' converts into
absolute values. The value of having 'B' for 'A', therefore,
can be easily checked before making a final decision.

Let us consider a numerical example. Suppose that the demand
and supply characteristics mentioned on the top of fig 4.7

```
**OPTIMAL P-N COMBINATION**

L  = LEAD TIME (WKS) = 4.00
G1 = SQ.MEAN/VARIANCE = 4.00
   NOTE: LEAD TIME DEMAND ASSUMED TO BE
         GAMMA DISTRIBUTED FOR  G1.LE.12
         AND NORMAL OTHERWISE

B1 = SHORTAGE PENALTY RATIO = (a.F1+F3)/F1
B2 = HOLDING RATIO = (F2+Ls)/F1
B3 = ORDER CHARGE RATIO = (K+Ln)/(D1.L.F1)
P  = STOCKOUT RATE
V  = SHORTAGE RATE
OU = YEARS OF STOCK
T  = REORDER INTERVAL (WKS)
N  = No. ORDERS PER YEAR
ML = MARGIN LOSS RATE = B1.V+B2.OU+B3.L/T
PARTIAL LOSS,  ML1 = MLR-B1.V
               ML2 = MLR-B2.YS
               ML3 = MLR-B3.L/T
```

B1=	B2=	B3=	P=	V=	OU=	L/T=	N=	ML =	ML1=	ML2=	ML3=
0.500	0.500	0.010	0.090	0.028	0.108	0.856	11.134	0.076	0.062	0.022	0.068
0.500	0.500	0.030	0.129	0.028	0.119	0.598	7.776	0.092	0.077	0.032	0.074
0.500	0.500	0.050	0.156	0.029	0.128	0.492	6.401	0.103	0.089	0.039	0.078
0.500	0.500	0.070	0.179	0.029	0.136	0.430	5.590	0.113	0.098	0.045	0.082
0.500	0.500	0.090	0.199	0.030	0.143	0.387	5.036	0.121	0.106	0.050	0.086
0.500	1.000	0.010	0.155	0.058	0.089	0.992	12.897	0.127	0.099	0.039	0.117
0.500	1.000	0.030	0.208	0.059	0.094	0.741	9.628	0.146	0.117	0.052	0.124
0.500	1.000	0.050	0.246	0.061	0.099	0.625	8.120	0.161	0.131	0.062	0.130
0.500	1.000	0.070	0.278	0.062	0.104	0.553	7.187	0.174	0.143	0.070	0.135
0.500	1.000	0.090	0.306	0.063	0.108	0.502	6.532	0.185	0.153	0.077	0.140
1.000	0.500	0.010	0.044	0.014	0.123	0.867	11.270	0.084	0.070	0.022	0.075
1.000	0.500	0.030	0.064	0.014	0.134	0.605	7.859	0.099	0.085	0.032	0.081
1.000	0.500	0.050	0.077	0.014	0.144	0.497	6.467	0.110	0.097	0.039	0.086
1.000	0.500	0.070	0.089	0.014	0.152	0.434	5.646	0.120	0.106	0.044	0.090
1.000	0.500	0.090	0.099	0.014	0.159	0.391	5.085	0.129	0.115	0.049	0.094
1.000	1.000	0.010	0.075	0.028	0.105	1.019	13.251	0.142	0.115	0.038	0.132
1.000	1.000	0.030	0.101	0.028	0.111	0.758	9.858	0.162	0.134	0.051	0.139
1.000	1.000	0.050	0.120	0.028	0.116	0.639	8.303	0.176	0.148	0.060	0.145
1.000	1.000	0.070	0.136	0.029	0.121	0.565	7.344	0.189	0.161	0.068	0.150
1.000	1.000	0.090	0.150	0.029	0.126	0.513	6.671	0.201	0.172	0.075	0.154
1.500	0.500	0.010	0.029	0.009	0.131	0.870	11.314	0.088	0.074	0.022	0.079
1.500	0.500	0.030	0.042	0.009	0.143	0.607	7.887	0.103	0.090	0.032	0.085
1.500	0.500	0.050	0.051	0.009	0.152	0.499	6.489	0.115	0.101	0.039	0.090
1.500	0.500	0.070	0.059	0.009	0.161	0.436	5.664	0.124	0.111	0.044	0.094
1.500	0.500	0.090	0.065	0.009	0.168	0.392	5.101	0.133	0.119	0.049	0.098
1.500	1.000	0.010	0.050	0.018	0.113	1.028	13.368	0.151	0.124	0.037	0.140
1.500	1.000	0.030	0.067	0.018	0.120	0.764	9.935	0.170	0.143	0.050	0.147
1.500	1.000	0.050	0.080	0.018	0.125	0.643	8.364	0.185	0.158	0.060	0.153
1.500	1.000	0.070	0.090	0.019	0.130	0.569	7.396	0.198	0.170	0.068	0.158
1.500	1.000	0.090	0.099	0.019	0.135	0.517	6.717	0.209	0.182	0.074	0.163

Fig. 4.7 - Control figures

apply. Each line of the table gives a set of values, consistent with the allocation criterion of the extended J-model. Use (B1;B2;B3) as the input vector. Assume that point 'A' in fig.4.6 results from the input (1.0;0.5;0.01) which leads to 'V=0.014', 'OU=0.123' and 'N=11.27'. Imagine that the manager was quite happy with the values of 'V' and 'OU' but 'N' was considered excessive. Then, he adjusted 'B3', the knob for 'N', from 0.01 to 0.09. The new perfomance vector, hereafter referred to as point 'B', is 'V=0.014', 'OU=0.159' and 'N=5.085'. Note that the other performance variables may also undergo an adjustment to the new balance. Assume that those figures are found satisfactory.

The consistency of moving from 'A' to 'B' can be tested in different ways. First, 'B1', 'B2' and 'B3' have concrete meanings and their values should be in line with other numbers the manager may have in his records. For instance, if 'F1=50%', the value 'B2=0.5' corresponds to 'F2+λs=25%'. This last figure shall be enough to cover the current borrowing interest plus other variable holding costs. Similar exercises could be done for 'B1' and 'B3' (the relevant equations being 4.43).

Another way of looking at the changes is through the margin loss. When changing from 'A' to 'B', the margin loss rises from 'MLA=0.084' to 'MLB=0.129', an increase of 4.5% of the potential profit. If the annual usage value were estimated in '52*d=$10,000' and 'F1=50%', the implied cost of

preferring 'B' to 'A' would be $225. Is that a good value?

The rationale subjacent to these last calculations, however, is not entirely correct. The increase in 'ML' was also due, in part, to the increase in 'B3', the reorder cost associated to 'N' (see 4.47). The other part was the price for the readjustments imposed on 'V' and 'OU'. The decrease in 'N', really, brings a saving in reorder costs. Therefore, it is a benefit not a cost. Such a benefit shall be compared to its counterpart, the costs due to the variation of 'V' and 'U'. Being so, what matters is the change in 'B1*V+B2*OU' (hereafter denoted by 'ML3') rather than in 'ML'. In the example, 'ML3' varied from 0.075 to 0.094, i.e., 1.9% of the potential profit that is $95. This is the net cost of reducing 'N' from 11 to 5, in other words, the shadow price for that decision. Note that shadow prices are evaluated at the implied costs.

We can assign a net margin loss to each perfomance variable. Thus, the effects of changing one attibute on the others can be easily evaluated. The printout reproduced in fig.4.7 mentions:

for V: ML1= ML-B1*V= B2*OU+B3*LOT   (4.101a)

for OU: ML2= ML-B2*OU= B1*V+B3*LOT   (4.101b)

for LOT: ML3= ML-B3*LOT= B1*V+B2*OU   (4.101c)

Many other combinations would be possible.

The utility of the performance vectors was introduced earlier as an ordinal. 'ML' in fig.4.6 could be seen as the quantification or 'cardinalization' of that utility. The effective frontier, as pictured, would be the utility function or the value function for the perfomance vector. The exercise of ranking perfomance vectors and mapping those preferences against their 'ML's may help extract the utility function of the decision maker. In a similar way, 'ML1', 'ML2' and 'ML3' at their efficient frontiers could be thought of as the utilities of 'V', 'OU' and 'N', respectively.

Each final choice, on the other side, represents a subjective trade-off with implied cost parameters, 'B1','B2' and 'B3'. The intangible 'B1' which so much reflects the managerial policy, might be empirically evaluated. This could help towards the structuring of the decision process. The drawback of such an approach is that the results, if any, may lack robustness.

4.6   Conclusion

The reorder frequency in the J-model is considered as an exogenous variable to be set on an empirical basis. By doing so, the interaction of the frequency with the other performance variables is missed, and yet, this interaction may become important for the objectives pursued in the model.

The reorder frequency consistent with the criteria in the J-model may be significantly lower than the frequency that resulted from adopting the economic order quantity given by the Wilson formula. According to those criteria, it was shown that the replenishment period 'T' should be such that

$$L/T < 1/(2*Dc**2) \qquad\qquad (4.105)$$

where 'Dc' is the coefficient of variation of the lead time demand and 'L' the lead time.

The relationship above has been earlier referred to as (4.57) and holds when the lead time demand can be approximated by a Gamma distribution with modulus greater or equal to one, so it holds in most of the situations, excluding the very slow movers. In particular, if the lead time demand follows a negative exponential (modulus=1, hence Dc=1), then 'T>2*L'. This means that no more than one order should outstand, no matter the length of the lead time, a

result that probably would not be evident.

The relationship (4.105) establishes an upper bound for the overlapping degree 'L/T'. The value that should be used can be computed in practice from the formula

$$L/T= (-C1+\sqrt{C1**2+4*C0*C2}) / (2*C2) \qquad (4.107)$$

earlier referred to as (4.86) and where 'C0', 'C1' and 'C2' depend on the carrying, shortage and reorder costs and on demand characteristics, as explained in section 4.4. This formula gives the answer to the initial question about the reorder frequency, since this relates to 'T' and 'L' is known.

Having derived in chapter 3 the formulas to deal with non-captive demand, and having solved in the present chapter the problem of the reorder frequency, the mathematical formulation of the J-model for inventories at one location has been thoroughly discussed. To complete this stage, considerations were made about an important procedural aspect, namely, the ability for checking the consistency of the decisions. The method proposed is based on what has been called in section 4.5, the net margin loss associated with each of the performance variables. The net margin losses for the shortage rate 'V', for the stock turn 'U' or for the duration of stock '1/U' and for the overlapping degree 'L/T' are reproduced below from (4.101):

for V: $ML1 = B2*(1/U) + B3*(L/T)$ (4.109a)

for 1/U: $ML2 = B1*V + B3*(L/T)$ (4.109b)

for L/T: $M13 = B1*V + B2*(1/U)$ (4.109c)

where 'B1', 'B2' and 'B3' have been defined in (4.43) as the shortage, the holding and the ordering costs, respectively, expressed as fractions of the profits.

The formulas (4.109) make apparent that each performance variable is associated with a cost, in this case, 'B1' associated with 'V', 'B2' with '1/U' and 'B3' with 'L/T'. These 'B's can be used as buttons to generate possible solutions in accordance with the allocation criteria of the model. If the decision maker wants to change one particular performance variable, he should then change the value for the respective 'B'. For instance, if the shortage rate 'V' has to be decreased, the previous value input for 'B1' shall now be increased. The decrease in 'V' has a price, the increase in stocks, and this price will be reflected on the respective 'ML'. That is, the increase of 'ML1' would reflect the price of decreasing 'V', and 'ML2' and 'ML3' would do the same in relation to decreases of '1/U' and 'L/T', respectively.

The 'ML's are expressed as a fraction of the annual profit and so, they are easily converted into a money amount. Therefore, the price of improving the performance can be calculated, that is, the 'shadow price' of a decision can be estimated.

In conclusion, if the decision maker were absolutely sure of the values which should be adopted for the shortage, holding and reorder cost rates, the problem would be confined to compute 'T' from (4.107) and then to calculate values for the other performance variables, this leading to the optimal solution. The decision maker, however, cannot be precise about those costs, so he has to proceed tentatively, by comparing different trade-offs for the variables. The method of the shadow prices can provide an important tool to weigh up these trade-offs.

Chapter FIVE

AN APPROACH TO A 2-LEVEL INVENTORY

5.1   Framework

The merchants group mentioned earlier is organised in four distribution areas and each of them comprises one main warehouse (MWH), their branches or satellites in the region and the related logistic facilities. These distribution areas are controlled independently and further integration is not considered for the time being. The system to be analysed in this chapter will be shaped from one of these distribution areas.

Presently, the MWH is considered as a single level inventory for the purpose of external procurement. A family order is placed to the supplier when the stock in the MWH reaches an aggregate reorder level as described in section 4.2. A delivery from the MWH to a satellite is triggered when the stock in the latter falls below a replenishment level. The control parameters for each satellite are set on the assumption of unlimited availability at the MWH, and satellite allocations are made independently of each other.

The MWH and the satellites are, thus, controlled on an independent basis. Improvements might be obtained with the introduction of central control, if replenishments and allocations were decided and controlled on a system-wide basis. This will be discussed through the present chapter. The rules and formulas achieved are consistent with the criteria in the Johnston model and thus, the model can be

extended to 2-level systems.

The external demand appears at the satellite level and external supplies enter the system through the MWH. There are exceptions. Some large customers are supplied directly from the MWH. These are perfectly identified firms, normally in the building industry. The marketing intelligence is often able to know in advance their requirements with reasonable certainty. When this is the case, the respective flows of materials can be dealt with quite apart from the remaining stochastic demand. Even if it were not so, a dummy satellite, adjacent to the MWH, could be 'created' for that trade. Thus, the external demand can be supposed to appear at the satellites, only. On the other hand, some items are supplied directly from external sources to the satellites. In relation to those goods, the branch acts on an independent basis, therefore, the problem reverts to the one warehouse case discussed earlier. In conclusion, the flow of materials to be considered relates only to those items coming from an external supplier to the MWH and from there to the branches where the external demand appears.

When an organisation holds several warehouses geographically dispersed, two initial questions might be whether they should be operated independently or under central control; and whether warehouses should be replenished directly from the external source or from intermadiate depots (i.e. main warehouses) run by the organisation. The system is termed "one-level" if the warehouses are supplied directly from

outside; "two-level" if the materials are sent firstly to a main warehouse and then, from here to the others (satellites); and in general, "multi-level" if there exists a hierarchy of warehouses through which the materials flow.

Note that multilevel systems as defined above might not mean central control: satellites could be considered as customers and place the orders to the MWH in the same way as to an external supplier; and the MWH would deal with the orders as with external demand and would control its own stocks in order to meet this demand to a convenient degree. A control like this would mean that the system had been decoupled into one-level inventories. In the context of this thesis, multilevel systems imply central control, i.e. control decisions are made having into account the status of the entire system. This will become clearer later in this chapter.

The central control of one-level system means, normally, the simultaneous replenishment of all the warehouses by placing a joint order to the supplier. The whole quantity ordered is allocated to the satellites and shipped to them. At a further stage, transhipments between warehouses may be required to keep the stocks in balance. The central control provides better service to customers than independent warehouses (for the same total stock investment in the system), but transhipment and information costs are higher. The latter may be partially offset by eventual economies to scale derived from joint procurement and replenishment. An

analysis of this problem is developed in Shakun (1962).

The 2-level system reduces the need for transhipments by retaining part of the stock in a MWH. Then, satellites are fed, basically, when their stocks become low. That means that the external replenishment quantity, when it arrives, is not allocated all at once to the satellites; a fraction remains in the MWH. Thus, on average, each external replenishment originates more than one internal replenishment to each satellite. When the subsequent allocation is made, compensation, in a part at least, can be given to the demand distortions experienced by the satellites. Naturally, the magnitude of such distortions increases for longer external lead times, as does the importance of the compensation effect. Therefore, a long external lead time favours the 2-level option.

Note that, in any case, the external lead time is supposed to be longer than the internal lead times, otherwise, retaining stock in the MWH might be useless. As recognised in Hadley et al (1961), if the external lead time was always less than the internal lead time, satellites would tend to be supplied from the external source.

The MWH, furthermore, reduces the capacity requirements in the satellites, as the deliveries are smaller, and gives opportunity for larger procurements. A convenient location of the MWH, having in mind the distribution network up and downstream, may lead to substantial savings in

transportation and holding costs. This is a matter which will not be detailed here (see Eilon et al, 1971, for basic approaches to this topic and references).

In a one-level system with joint replenishment, some criteria are needed to divide new supplies amongst the distribution outlets. In a multilevel system, a similar need arises each time the satellites are supplied from the common MWH. The rules to allocate a quantity 'Q' of stock to warehouses depend on the objectives pursued, on the distribution policy and also on the particularities of the system. An early study about this allocation is found in Simpson (1959). Here, emergency transhipments can be allowed, and the targets are either the minimisation of the transhipment costs or the minimisation of the shortage quantities. Different assumptions and operating policies have been considered by a number of authors as Shakun (1962), Gross (1963), Hadley (1963), Brown (1967,TM20), and Barrett (1969,ch6). The models vary from author to author, so do the conclusions. The most common objectives is, however, the minimisation of either the number of runout situations (stockouts) or the shortage quantities for the whole system. Both cases will be examined here.

Consider first the objective of minimising the stockouts. Stockouts and transhipments are directly related because the minimum number of stockouts leads to the minimum number of transhipments. Transhipments may be weighted by cost rates and a minimum transhipment cost sought. Assume at the

allocation time that the warehouses have no stock, then, the objective can be formulated as follows:

$$\text{Min} \left(\sum w(k) * P(k) \right), \text{ subject to } \sum Q(k) = Q \qquad (5.1)$$

where 'P(k)' is the probability of stockout in warehouse 'k', 'Q(k)' is the part of 'Q' allocated to warehouse 'k' and 'w(k)' is the weight associated with the stockout, for instance the transhipment cost. Formally,

$$P(k) = \int_{Q(k)}^{\infty} f(k,x) * dx \qquad (5.3)$$

with 'f(k,x)' standing for the p.d.f. for demand in warehouse 'k'. This demand relates to a given period, for instance, the expected time until the next replenishment of the warehouse.

The Lagrangian equivalent to (5.1) is

$$\sum w(k) * P(k) + \lambda * \left(\sum Q(k) - Q \right) \qquad (5.5)$$

Its minimum is met for:

$$w(k) * (\partial P(k) / \partial Q(k)) + \lambda = 0 \qquad (5.7a)$$
$$w(k) * f(k, Q(k)) = \lambda \qquad (5.7b)$$

Therefore, the 'Q(k)' shall be such that the left hand side of (5.7b) is the same for all warehouses. If the warehouses have some remaining stock, 'Yr(k)' say, when the allocation

is made, 'Yr(k)+Q(k)' substitutes for 'Q(k)' in (5.7b).

The other common objective is the minimization of the (weighted) shortage. The allocation rule is then to equate the (weighted) stockout probability. Formally,

$$\text{Min } (\textstyle\sum w(k)*Z(k)) \text{ , subject to } \textstyle\sum Q(k)=Q \qquad (5.9)$$

where 'Z', as usual, stands for the shortage quantity. The result would be an allocation such that

$$w(k)*P(k)= \text{constant} \qquad (5.11)$$

The weight 'w(k)' may stand for the shortage cost rate in each warehouse 'k'. That rate, earlier, has been denoted by 'F(k)' (only 1 item is being considered). Then

$$F(k)*P(k)= \text{constant} \qquad (5.13)$$

This allocation rule is the one which agrees with the criteria in J-model: getting the lowest shortage penalty for each level of investment. We will return to this point in section 5.3.

The allocation is only one of the decisions to be made in controlling the multilevel inventory. Others relate to: the mechanism that triggers the allocations; the stock balance between the MWH and the satellites; and the external replenishment control for the whole system. The integration

of these problems have been the subject of many studies. Literature surveys can be found in Scarf (1963), Veinott (1966), Clark (1972) and Hollier (1976). They comprehend long lists of references and detailed comments whose repetition, here, would be redundant. More recent published work on the topic did not attempt new ways. Most authors have followed a dynamic approach, i.e., consider a succession of periods of time with decisions being made at the beginning of each period. The objective, normally, is the minimum expected discounted cost through the multiperiod horizon.

Dynamic models are of two different types. One of them, pioneered by Clark et al (1960 and 1962) uses dynamic programming. Exact optimum solutions have been found only for the series structure (one receiver for each supplier). Upper and lower cost limits can be determined, however, for arborescent structures. The other type of model uses the so called dynamic process analysis. It was introduced by Veinott (1965) and extended by Bessler et al (1966) and Ignall et al (1969). The analysis considers a non-stationary stochastic process over N periods. Further, the conditions which make the optimum a myopic policy (where the horizon for the decision is only one period) were identified. When such conditions apply, the formulation can be simplified considerably.

The main difficulty with the dynamic approach is its analytical complexity and the consequent computational

burden. Such a drawback is magnified when the model is intended for decisions based on "what-if" analysis. Furthermore, those models work on a periodic review procedure and are not prepared to deal with multi-item situations. Simulation models by-pass those mathematical difficulties and shortcomings and may constitute in the future a real alternative to the mathematical approach. Relevant work in this area was developed by Cran (1966), Connors (1972) and Aggraval et al (1975). These models are essentially concerned with replenishment and allocation rules based on parameters whose best value is chosen empirically. It must be remembered that the parameters are optimum for the simulation model which may or may not represent the real world well.

Recalling the formulation of the criterion for capital investment and the way as J-model is used within the decision process, we realise that the ability to predict the service levels for each level of investment in stock is essential. None of the current multi-level inventory models gives or is directed to give tractable expressions to estimate those services. Therefore, they do not help to extend the Johnston approach.

To conclude this section, a special reference should be made to the control procedures proposed by Cran (1966) and by Brown (1967, TM27). Both deal, specifically, with a 2-level system comprising a MWH and several satellites. Brown's basic rules are as follows:

(i) The MWH replenishments are based on the aggregate demand forecasts for the whole system.

(ii) For the allocation purpose, stocks are broken down into: working stocks, corresponding to the expected demand; and the safety stocks. Working stocks are allocated to satellites as soon as received from the external supplier. Safety stocks are retained centrally.

(iii) Each satellite has a 'warning level' enough to cover, with some safety margin, the expected demand during the time for internal delivery. When such a level is reached at any one satellite, the remaining stock is allocated to all satellites but the only shipment is to that one in risk of stockout.

(iv) The record of the inventory situation at all locations is kept up-to-date. Replenishment orders for the main warehouse are raised when the aggregate stock on hand reaches a reorder point which has been calculated on the basis of the aggregate demand forecasts.

Cran's proposal is basically the same, except in relation to (ii), the allocation rule. When a new allocation is required, the "stock time", i.e., the time to deplete the stock existent in the system, is estimated from the system demand characteristics. The allocation is, then, given by

the expression

$$Q = d*T - A*s \qquad (5.14)$$

where:

d= satellite demand rate

T= mean stock time

s= standard deviation for the satellite demand
in the stock time

A= hold back factor, a constant whose best value
is found by simulation

The above rules are simple and easy to implement. They are the opposite to the dynamic models, highly sophisticated and difficult to use. The model which will be presented in the next sections possesses some of the that pragmatism. However, it is more elaborated and, furthermore, is consistent with the basic criteria in J-model.

5.2 Perspectives for the approach.

Distribution is the main activity in a merchandising company. That involves, basically, physical transportation and stocking. Warehouses are the nodes of the distribution network. The design of such a network, the location of the warehouses, the logistic means used, all affect the lower bounds for transportation and stock costs and impose restrictions on perfomances. The stock and transportation procedures interfere with each other. Therefore, a stock policy which ignores the other partners of the distribution complex is liable to be sub-optimal.

An integrated distribution analysis might involve: (i) the number and location of the warehouses, hierarchy and capacities; (ii) the means of delivery, delivery routes and delivery journeys; (iii) the items in the inventory for each warehouse; (iv) and the replenishment policies. These aspects are not independent, but, decoupling the system into subsystems to be worked separately is normally the only way to achieve practical results.

Another reason favours the split of the system. It is the quite different inertia of the variables involved. Decisions in relation to service levels may produce results in a few weeks; the reorganisation of the transports may take months; and, modifying warehouses location, their capacity or number is a matter for years. Therefore, those decisions have very

distinct time horizons. A stock policy usually needs to be flexible because it has to respond quickly to small fluctuations in the environment. Such fluctuations are not to be followed by immediate changes in the logistic means.

The analysis which will be developed here rests on the assumption that the distribution network will not undergo structural changes: the number, location and capacity of the warehouses; and the delivery means and routes will be kept the same through the horizon of analysis. Furthermore, slow moving items will not be considered for the reasons given in section 1.2; and in respect to the others, the current inventory lists for the satellites are supposed to be fixed. Thus, the following analysis will not intervene to modify the logistic means or the inventory bill of materials. Rather, it will concentrate exclusively on the replenishment policy. This reflects the short term character of the approach.

The replenishment policies contemplate both the external and internal replenishments. By external replenishment is meant the supply from external sources to the system. External orders are raised on a family basis when a system-wide reorder level is reached. For this purpose, a central data base is kept up to date with the stock levels at all locations. External supplies are received at the MWH.

Internal replenishments are those from the MWH to the satellites. This is the only direction for the internal flow

of materials allowed in the analysis. Transhipments among satellites and returns to the MWH are supposed not to be a regular flow. They will exist in practice as a result of spot decisions but, by assumption, they are not frequent and, more important, not planned for. Internal replenishments are decided individually for each item. Buying families, in the terms considered for external supplies, have no interest internally since the MWH is the common supplier.

There are lead times associated with external and internal replenishments. As mentioned earlier, external lead times are supposed to be much longer than the internal ones. A comment should be made in relation to the external lead time for a multilevel system. The period elapsed between the replenishment process being triggered and the product being available at the MWH is the external supplier lead time. It is necessary, however, to add an extra period of time for that product to become available also at the satellite. This extra period is the internal lead time. The external lead time for a satellite, therefore, is the external supplier lead time plus the internal lead time.

Two aspects should be noted in relation to the internal replenishments. The first is that, once the stock is in possession of the company, the capital is tied up. Capital cost is the same no matter where the stock is located. Therefore, ignoring small differences which may occur in maintenance and handling, the holding costs are the same for

MWH and for satellites, provided that space is available. It is true that a MWH is normally located on cheaper land than the satellites. However, capacities and locations are longer term decisions, therefore, the land price is a fixed cost from the present perspective and does not take part in the model. Then, as the holding costs are the same in any place of the system, there is no advantage in having the satellites half empty while the MWH is full.

The second aspect is a consequence of the first. If space is available in the satellites and stock is abundant in the MWH, savings in picking and transportation costs can be obtained if larger quantities are supplied each time. Similar reasons may justify the practice of 'route deliveries': when an item is required for a satellite, a joint shipment can be sent to others in the same delivery route. This practice may reduce the picking effort at the MWH, but tend to increase the number of shipments received at each warehouse, so the actual advantages have to be assessed locally. Both individual and group deliveries can be dealt with by the model to be introduced.

After these initial comments we can concentrate on the main problem which is the allocation policy that should rule the internal replenishments. In other words, when a fresh external supply arrives, a decision is needed about how much to retain in the MWH and how much to ship to the branches. There are two reasons for retaining stock in the MWH. One is

the capacity available in the satellites; the other is providing a better service to customers. Whenever the system stock is split into various locations there will be a risk of having a stockout in one location while others have excess. This imbalance could be corrected at the expense of transhipments. When these are not allowed the overall service becomes poorer. Therefore, quantities allocated should be small enough to give a relatively high chance of being consumed before the system is supplied again. This point will be discussed in section 5.3. For the time being it is enough to realise that, as we progress through the external replenishment cycle, the quantities allocated shall become smaller to keep a constant risk of stock excess. This agrees with common sense in that when the MWH is full, the allocation can be generous, whilst rationing is likely to take place towards the end of the cycle.

If satellite capacity restrictions are binding, maximum internal shipments need to be fixed for each satellite. For this purpose, the capacity might be partitioned by the items in the inventory, for instance, on a pro rata basis to demand-volume rate. Such limits may affect the number of internal replenishments, but not to a significant degree the service to customers. Capacity restrictions may prevail after a fresh replenishment, when there is plenty of stock in the system. The risk of excess is more important at the end of the cycle. This makes the successive allocations become increasingly smaller, as it will be seen in section 5.3. A limit is set, in practice, to avoid the sending of

too small quantities. This limit will take the form of a 'clearing level' for the MWH. Whenever a new allocation is required and the stock in the MWH is equal or lower than the clearing level, then all the remaining stock is dispatched.
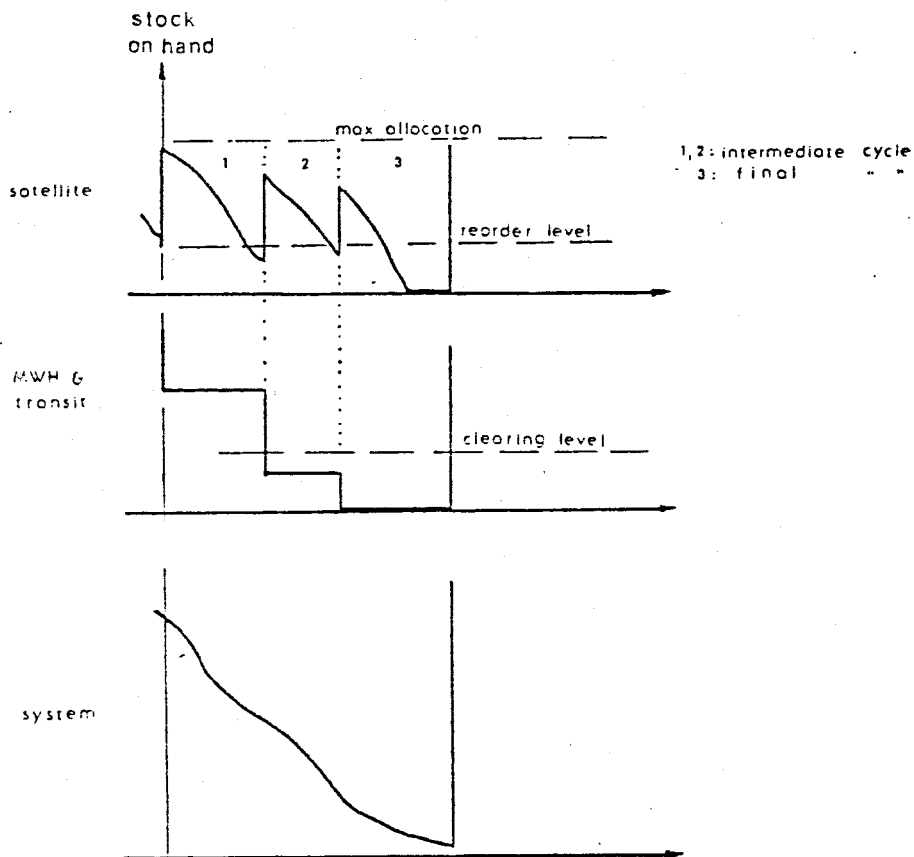
Fig. 5.1 - Types of replenishment cycles

As yet, the mechanism that triggers the internal replenishment has not been mentioned. It is presumed that there is a reorder level for each satellite. The expected shortages in the intermediate internal cycles will depend on these reorder levels. Intermediate cycles, as shown in fig.5.1, are those during which stock still remains in the

pipeline, i.e., in the MWH or in transit to the satellite. The final internal cycle (within an external cycle) is the one during which that pipeline becomes exhausted. The shortages will be called intermediate and final, within the given cycles.

The mathematical analysis of the internal reorder levels and intermediate shortages, consistent with the global model, will be presented in section 5.7. Meanwhile, it will be assumed that internal reorder levels are set empirically to give a reasonable low intermediate shortage, so they are exogenous to the model. Therefore, the final shortages are the only ones that are at the present stage included in the model and subjected to 'optimization'.

The characteristics of the system to be examined in this chapter can be summarized as follows:

(i) The distribution system is looked at on a short term basis. The whole distribution network will remain constant, as will, the inventory list of items. The only variables are those relating to the stock replenishment policies.

(ii) The system, for analysis purposes, comprises one MWH and its satellites. A storage capacity is allocated to each item at each satellite. There is no capacity constraint for the MWH.

(iii) The inventory is controlled centrally, the knowledge of the actual stock levels at any time and location is implied. In practice, this assumption can be applied to situations in which stocks are updated at short intervals where extrapolation is accurate enough.

(iv) Internal replenishments are tackled separately for each item. Flows of materials are from the MWH to the satellites and from these to the outside system. No transhipments or returns are considered.

(v) The holding costs for an item are the same, regardless of the storage location.

(vi) Picking, handling and transportation cost rates decrease as internal replenishments are made in larger quantities.

(vii) Internal lead times are small compared with the external cycle. Internal reorder levels are set exogenously. In consequence, intermediate shortages are ignored by the model.

(viii) Internal deliveries are restricted to the capacity allocated to the item in each satellite and to a maximum risk of stock excess. Minimum deliveries for each satellite and a clearing level for the MWH are fixed on a practical basis.

5.3   Criterion to successive allocations (1 satellite)

The first approach to the allocation problem has been mentioned in section 5.1. It involves two decisions. The partition of quantity 'Q' shall be such as to give to all satellites the same weighted probability of stockout till the time the next replenishment is expected. This rule, stated earlier in (5.13), solves only part of the allocation problem. The other part concerns the decision about the size of the quantity 'Q'. That is, when a new internal delivery is required what portion of the stock available in the MWH shall be sent out and what portion shall be retained.

Empirical rules for the quantity of stock "held back" have been mentioned earlier in section 5.1, namely those proposed by Brown and by Cran. Those proposals have intuitive appeal, however, they lack a solid rationale. The rationale which will be adopted in this thesis, stems from the notion of stock excess introduced in the previous section. The larger the quantity allocated and the later the allocation takes place in the external cycle, the higher the risk of stock excess. This risk should rule the successive allocations along the external cycle, as will be seen later.

Picking, handling and transportation costs are assumed to decrease if the quantities allocated each time are made greater. Therefore, the quantities allocated should be as large as possible, given the risk of stock excess the

manager is prepared to take. That means that successive allocations shall have the same maximum risk of stock excess. Other constraints, such as the stock availability or the capacity may superimpose themselves, but this does not invalidate the argument.

Note, now, that the expected risk of stock excess can be measured by the probability of the stock (just after the allocation) exceeding the demand (from that moment until the end of the external cycle). This probability is, in turn, the complement to one of the probability of that stock being depleted during the same period. Thus, the rule above is equivalent to:

"successive allocations shall have

the same probability of depletion"

This last enunciation is more convenient for the analytical treatment which follows. Note that the probability of stockout is the probability of depletion of the whole stock available; yet, the probability of depletion may relate to any quantity and so, they are distinguished here.

The remainder of this section will tackle mathematically the problem of successive allocations by considering a system with the MWH and one satellite. Admittedly, such a structure would raise no allocation problems, but it is used here to introduce the next section where structures with any number of satellites will be analysed.

Denote, then, by 'x' the external demand over the external

lead time, 'R' the system reorder level for the external replenishments and 'R1' the satellite reorder level for the internal replenishments. The system stock on hand is retained at the MWH except for the quantities shipped out to the satellite each time the latter reaches 'R1'. For the time being, 'R1' is assumed to be zero. We want to calculate the allocations 'Q(1)', ...,'Q(t)', ...,'Q(n)', so that they have the same probability of depletion 'p'. The time moments when those allocations take place are 'tm(1)', ...,'tm(t)', ...,'tm(n)'.
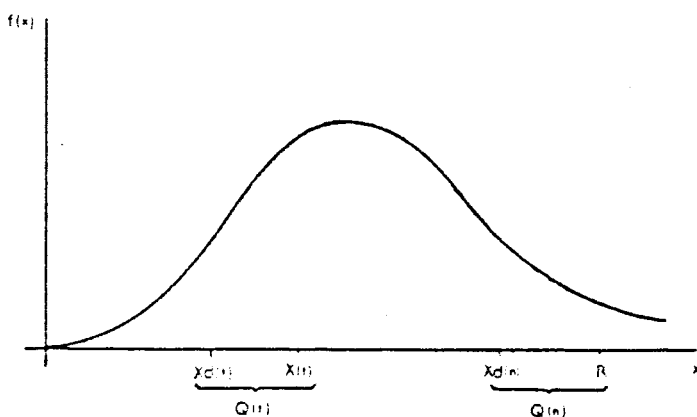
Fig. 5.2 - Probability associated with
the successive allocations

Let's start the analysis when the system reaches its reorder level 'R' and a new order is raised. That is, when a new external lead time begins. That moment is referred to as 'tm(1)'. Assume that the satellite has no stock on hand. That being so, an internal allocation is required. Denoting the quantity involved by 'Q(1)', then, according to the

criterion, 'Q(1)' should have the probability 'p' of being depleted. Thus,

$$pr(\ x > Q(1) \ ) = p \qquad (5.15)$$

Now, consider some time later in the cycle, say 'tm(t). Denote by 'Xd(t)', as in fig.5.2, the quantity so far depleted since the beginning of the current lead time period, i.e. since 'tm(1)', and assume that a new allocation is going to be made. This new allocation, 'Q(t)', shall be such that it has the probability 'p' of being depleted in the current period, having in mind that 'Xd(t)' has been used already. The new allocation involves the following conditional probability:

$$pr(\ x > Xd(t) + Q(t) \ | \ x > Xd(t) \ ) = p \qquad (5.17)$$

This is equivalent to

$$P(t)/Pd(t) = p \qquad (5.19a)$$

$$\text{where,} \quad X(t) = Xd(t) + Q(t) \qquad (5.19b)$$

$$P(t) = pr(\ x > X(t) \ ) \qquad (5.19c)$$

$$Pd(t) = pr(\ x > Xd(t) \ ) \qquad (5.19d)$$

The allocations ruled by expressions (5.19) decrease quickly and, eventually, they may become too small for delivery purposes. Usually, there is some limit imposed on the minimum delivery quantity. That limit fixes the size of the last allocation and therefore the depletion probability associated with it. As the depletion probability should be the same for all the allocations, 'p' should be set from that last delivery. That is

$$p= P(n)/Pd(n) \qquad (5.21)$$

The allocation conditions, in a real system, are normally slightly different from those described above. First, there is an internal reorder level, 'R1' say, which triggers the allocations. When the stock in the satellite falls to 'R1', a new allocation is made, provided that the MWH is not empty. This allocation is calculated by (5.19), but (5.19b) should now read

$$X(t)= Xd(t)+R1+Q(t) \qquad (5.23)$$

in order to take into account the quantity 'R1', the stock existing in the satellite.

Second, only the present allocation 't' is considered, rather than the whole series. 'Xd(t)' can be computed from the current stock on hand 'Yh(t)'. If 'Yh(t)' is greater than 'R', no order is outstanding. In such a case, we take for the calculations 'Xd(t)=0'. If 'Yh(t)' is less than 'R', there is an order outstanding, at least. Consider the order that will arrive next and the respective lead time vector. The quantity depleted since the beginning of that lead time is 'R-Yh(t)'. Thus

$$Xd(t)= M\left[R-Yh(t)\right] \qquad (5.24)$$

where 'M(x)' means the maximum between zero and 'x'.

The minimum delivery can be fixed in several ways. Here, it will be assumed that the management sets a 'clearing level' based on practical considerations. When a new allocation is required and the stock is below that clearing level, all the stock is delivered. If the clearing level is fixed at some value 'x', say, then the last delivery can vary from that value to zero and will be, on average, about 'x/2'. For convenience, the clearing level will be written as '2*YOf' so that 'YOf' represents the average quantity in the last delivery, i.e.

$$Q(n) = YOf \qquad\qquad (5.25)$$

The final allocation, in prospect, involves the quantities

$$X(n) = R \qquad\qquad (5.27a)$$
$$Xd(n) = R - R1 - YOf \qquad\qquad (5.27b)$$

from which we could have 'P(n)', 'Pd(n)' and hence

$$p = P(n)/Pd(n) \qquad\qquad (5.29)$$

The procedure for a new allocation 't' would be as follows:

$$Xd(t) = M\left[R - Yh(t)\right] \qquad\qquad (5.31a)$$

From $Xd(t)$, the current depletion, find

$$Pd(t) \text{ according to } (5.19d)$$

$$P(t)= p*Pd(t) \qquad (5.31b)$$

From P(t) find X(t) according (5.19c)

$$Q(t)= X(t)-Xd(t)-R1 \qquad (5.31c)$$

Obviously, 'Q(t)' may have to be restricted further by the satellite capacities or the MWH availability.

The following example will help clarify what has been said. Suppose a MWH with 1 satellite whose external demand in the external lead time is normally distributed with mean 'D=1000'and standard deviation 'Ds=200'. Let the reorder levels be 'R=1400' and 'R1=150', for the system and for the satellite, respectively. The clearing level is '100' and hence 'YOf=50'.

The attributes of the final partition can be calculated from (5.27) as

$$X(n)= R= 1400$$

$$Xd(n)= R-R1-YOf= 1400-150-50= 1200$$

The corresponding standard Normal values are

$$(1400-1000)/200= 2.0 \; ; \; and$$

$$(1200-1000)/200= 1.0 \; , \; respectively.$$

Hence, the depletion probabilities from (5.19c,d) are

$$P(n)= 0.02275; \; Pd(n)= 0.1587$$

Then, from (5.29), $\quad p= P(n)/Pd(n)= 0.143$

Now, suppose that the stock in the satellite is at the reorder level and that there are 1300 units in the MWH.

-5.3-

Hence, 'Yh(t)=1300+150=1450' which is greater than 'R'.
Then, from (5.31a), Xd(t)=0'. It follows from expressions
(5.31):

Pd(t)=1.0, because Xd(t)=0

P(t)= p*Pd(t)= 0.143. The coresponding standard
Normal value is x~1.065, then

X(t)= 1.065*200+1000= 1213

Q(t)= X(t)-Xd(t)-R1= 1213-0-150= 1063

Consider, now, a later stage in which the stock on hand is
'Yh(t)=500' split into 350 in the MWH and 150 in the
satellite. Then, from (5.31):

Xd(t)= R-Yh(t)= 1400-500= 900.

The associated depletion probability is Pd(t)=0.8085.

P(t)= p*Pd(t)= 0.143*0.8085= 0.1156.

The corresponding standard Normal value is x~1.2, then

X(t)= 1.2*200+1000= 1240

Q(t)= X(t)-Xd(t)-R1= 1240-900-150= 190

5.3.1 Analytical details

Consider again the successive allocations ruled by (5.19).
Under the circumstances there assumed and with the
convention 'X(0)=0', we have 'Xd(t)=X(t-1)' and
'Pd(t)=P(t-1)'. It follows then:

$$P(0) = 1 \qquad\qquad (5.33a)$$

$$P(1) = p*P(0) = p \qquad\qquad (5.33b)$$

$$P(2) = p*P(1) = p**2 \qquad\qquad (5.33c)$$

$$\dots\dots\dots$$

$$P(n) = \dots \quad = p**n \qquad\qquad (5.33d)$$

This gives 'n' simultaneous equations involving the 'n+2' unknowns: 'Q(1)', ...,'Q(n)','p', and 'n'. Furthermore

$$X(n) = \sum Q(t) = R , \quad t = 1, \dots, n \qquad\qquad (5.35)$$

which makes 'n+1' equations, one less than the number of unknowns.
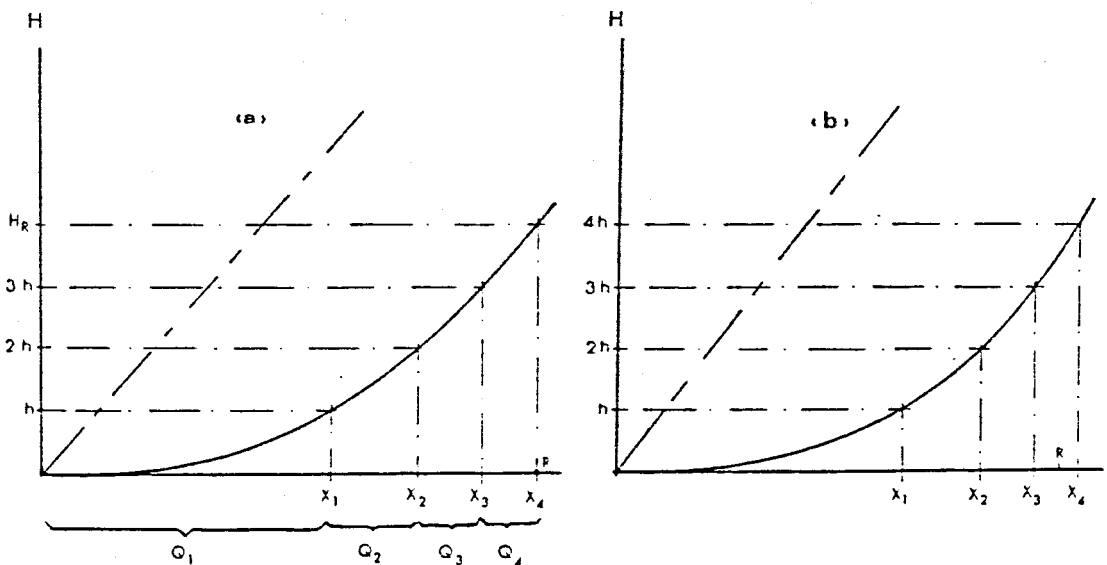
Fig. 5.3 - Allocations with constant depletion probability

Fig.5.3 gives a graphical method to solve the above equations. 'H(x)' is the log-transform of the probability

function for the external lead time demand. The equations (5.21) are equivalent to determine on 'H(x)' points spaced of 'h', where 'h=ln(1/p)'. Their projection on the x-axis gives the 'X(.)'s. Then, the 'Q(.)'s can be computed as

$$Q(t)=X(t+1)-X(t) \qquad (5.37)$$

As there is one more variable than equations, we can try to fix one of them. The diagram (a), in the figure, would give the solution when 'n' is preset. The procedure is straight forward. 'H(R)' can be computed and, then, we make 'h=H(R)/n'. The ceiling 'X(t)' is obtained by inversion of 'H(t*h)'. The diagram (b) relates to a situation where 'p' has been fixed. The value of 'h' is immediately calculated from 'p'. The procedure follows by marking on 'H(x)' the points at the levels 'h', '2*h', etc., and inverting the function. The value of 'X(4)', on purpose, has been represented beyond 'R'. It means that the last allocation would be smaller (or larger) than it should. 'X(n)' cannot be made equal to 'R' for that 'p', i.e., the choice of 'p' is restricted. This, however, has no relevant practical consequences.

When the lead time demand is gamma (modulus greater than 1) or normal distributed, 'H(x)' has a shape as in the figure. The slashed straight line in the diagrams correspond to the 'H(x)' for a negative exponential lead time demand. The 'Q(.)'s, in this case, would be equal. However, the negative exponential is unlikely to fit the lead time demand of a

fast mover. Within the range of interest, successive allocations decrease steadily. They decrease sharply, at first, so very small quantities are reached rapidly. Good sense dictates a minimum quantity, earlier referred to as the clearing level. This clearing level ties up the size of the last allocation and, therefore, the values for 'p' and 'h'. Denote by 'Yf0' the final allocation quantity. Then,

$$Q(n)= Yf0 \qquad\qquad (5.39a)$$

$$X(n-1)= R-Yf0 \qquad\qquad (5.39b)$$

$$h= H(R)- H(X(n-1)) \qquad\qquad (5.39c)$$

$$p= Exp(-h) \qquad\qquad (5.39d)$$

This value of 'p' can be used as the depletion probability for all the allocations.

5.4   Combined criteria for internal allocation

The preceding section considered the successive allocations to one satellite alone. This unreal situation was used to make the method of approach and the terminology clearer. The true allocation problem, however, arises only for multi-satellite situations. Recall that a new allocation is made whenever a satellite reaches its reorder point. The quantities allocated depend on the status of the whole system. This implies that the calculations are made as if quantities were allocated to all the satellites. Whether such quantities are sent to all of them, to a group or only to the satellite which has triggered the process, is a matter of distribution convenience. Thus, a simultaneous allocation is required each time, and simultaneous allocations are ruled by (5.13).

Note that the consequences of a new allocation may depend also on the recent history of the system. More precisely, they may depend on the depletion which has already taken place in the current lead time. Therefore, the probabilities in (5.13) shall be read as conditional probabilities. These latter have been denoted by 'p': say 'p(k,t)', to stand for satellite 'k' at time moment 'tm(t)'. Sucessive allocations, however, shall have equal depletion probabilities. That is, we can write 'p(k)' for 'p(k,t)', as these probabilities do not depend on 't'. The allocation criterion in (5.13) becomes, then

$$F(k)*p(k) = \text{constant} \qquad (5.47)$$

Having the 'p(k)'s, a new allocation can be calculated easily. The steps involved are as follows:

From $Xd(k,t)$ find $Pd(k,t)$

$$P(k,t) = Pd(k,t)*p(k) \qquad (5.49a)$$

From $P(k,t)$ find $X(k,t)$

$$Q(k,t) = X(k,t) - Xd(k,t) - Y(k,t) \qquad (5.49b)$$

$$\sum_k Q(k,t) \leq Y0(t) , \quad k = 1, 2, \ldots, K \qquad (5.49c)$$

These expressions correspond to those referred to as (5.31) in the preceding section. 'Y(k,t)' is the current nominal stock (i.e., on hand plus on order) in satellite 'k'. 'Y(k,t)=R(k)' for the satellite that triggered the allocation, but is not necessarily so for the others. 'Y0(t)' is the current stock available in the MWH. Obviously, it constitutes an upper bound to the sum of the quantities delivered, hence the relationship (5.49c). If 'Y0(t)' is exceeded in the first allocation attempt, the 'X(k,t)'s should be recalculated. Reducing the 'X(k,t)'s proportionally until (5.49c) is met is likely to be accurate enough.

The value 'Xd(k,t)' can be picked from the data sales for each satellite. Or, more conveniently, it can be roughly estimated from the sales for the whole system, in proportion to the demand rate. In this case, recalling (5.31a), we would have

$$Xd(t) = M\left[R - Yh(t)\right] \qquad (5.51a)$$

Then, $\quad Xd(k,t) = Xd(t) * d(k) / d \qquad (5.51b)$

where 'd', and 'd(k)' are the demand rates for the system and for the satellite, respectively, and 'M$\left[x\right]$' is the maximum between zero and 'x'.

The 'p(k)'s can be found from expressions (5.27) and (5.29) in relation to the final allocation. The process may involve many iterations if a precise result is required. However, the computational load is significantly reduced when the satellites can be considered 'twin'.

> The satellites are said to be 'twin' if
>
> the shortage penalties 'F(k)'s and
>
> the coefficients of variation 'Dc'
>
> are constant for all satellites

One procedure to calculate the 'p(k)'s for a situation with twin satellites and Normal demand is presented in subsection 5.4.1. Another will be discussed in section 5.5.

5.4.1   Analytical details

The calculation of the 'p(k)'s by applying to the final allocation the general criterion formalized in expression

-5.4.1-

-184-

(5.47), presents some computational difficulties.

When the final allocation takes place, the stocks in satellites are expected to be close to their reorder levels and the MWH to be about the average clearing level 'YOf'. In prospect,

$$Yhf= YOf+\sum R(k) \qquad (5.53a)$$

$$Xd(n)= R-Yhf \qquad (5.53b)$$

$$Xd(k,n)= Xd(n)*d(k)/d \qquad (5.53c)$$

Now, 'Yhf', the stock on hand in the whole system, has to be shared by all satellites. Denote by 'Yhf(k)' the share for the satellite 'k'. It is taken that 'Yhf(k)' is greater than 'R(k)'. The strict application of the criterion to the final allocation would involve the following set of simultaneous equations extended to all the satellites:

$$\sum Yhf(k)= Yhf \qquad (5.55a)$$

$$X(k,n)=Xd(k,n)+Yhf(k) \qquad (5.55b)$$

Find Pd(k,n) from Xd(k,n)

   "   P(k,n)   "   X(k,n)

$$p(k)= P(k,n)/Pd(k,n) \qquad (5.55c)$$

$$F(k)*p(k)= constant \qquad (5.55d)$$

$$k=1,2,..K$$

The solution of this set would have to be found iteratively. This might not be an easy task, but eventually, would determine the 'Yhf(k)'s and 'p(k)'s.

A considerable simplification is introduced with the assumptions of Normal distributed demand and twin satellites. The normality for demand seems reasonable for most of the cases since we are, now, especially interested in long lead times and fast moving items. The 'twinning' simplification implies that the satellites are balanced in terms of throughput and commercial importance. Analytically, if 'F(k)' is constant, the cost of one unit short is the same in all the satellites. Then, from (5.55c), 'p(k)' is constant. The other assumption, the constancy of the coefficient of variance implies that allocations are made in proportion to demand rates. This point is shown next.

A constant coefficient of variance is expressed as

$$s(k)/D(k)= Dc= \text{ constant} \qquad (5.57)$$

where 's' and 'D' are the standard deviation and the mean of the lead time demand. With Normal demand distribution, the depleted quantity 'Xd(k,n)' relates to a standard value

$$xd(k,n)= (Xd(k,n)-D(k))/s(k) \qquad (5.59a)$$
$$= (Xd(n)/D-1)/Dc \qquad (5.59b)$$

As 'xd(k,n)' does not depend on 'k', the same happens to 'Pd(k,n)'. Then, it follows from (5.55) that 'P(k,n)' and 'x(k,n)' are also independent of 'k'. We would arrive easily at

$$X(k,n) = R*d(k)/d \qquad \text{(5.61a)}$$

$$x(k,n) = (R/D-1)/Dc \qquad \text{(5.61b)}$$

'P(k,n)' and 'P(n)' are found from 'x(k,n)' and 'x(n)'. Thence, 'p' can be computed.

The assumption of twin satellites may look unrealistic particularly in view of the many non-linear variance laws which operate. For example, one form relates the variance and the mean as

$$Dv = a*(D**b) \qquad \text{(5.63)}$$

where 'a' and 'b' are constants. Johnston (1980) mentions that values for 'b' were found rather constant, actually 'b~1.5'. In this case,

$$Dc \propto (1/D)**0.25$$

$$\Delta Dc/Dc = -0.25*(\Delta D/D)$$

That is, the coefficient of variation increases when the demand increases but in a proportion about 4 times lower. The 'twin' assumption leads to compute a 'p' lower than it should be if the satellites are different. Recall that 'p', as computed from the last allocation shall be the minimum depletion probability (maximum risk of stock excess) to be considered for the intermediate allocations. Therefore, to be on the safe side, 'p' should be taken a bit higher than

-5.4.1-

the value calculated under the twin assumption.

Note that 'Pd(k,n)' and 'P(k,n)' as determined from
expressions (5.55) are only approximations. Those formulas
assume a given 'Xd(k,n)' which has been calculated from
(5.53). Actually, the 'expected value' has been taken for
the value of the 'stochastic variable Xd(k,n)'. The
distribution of 'Xd(k,n) given Xd(n)' should be considered
for an exact evaluation of 'p(k)'. This, however, is not
easy.

An alternative procedure stems from the obvious relationship

$$P = Pd(n) * p \qquad\qquad (5.65)$$

where 'P' is the system stockout rate and 'Pd(n)' is the
probability of clearing out the MWH in a cycle (i.e. of
'Xd(n)' being depleted). 'Pd(n)' can be evaluated easily,
known the demand distribution. 'P' can be estimated through
the depletion time distribution, a new approach that will be
introduced in the next section.

5.5   Depletion time

The service to customers is estimated, usually, from the lead time demand distribution defined by the density function 'f(x)'. This has been also the procedure followed so far to compute 'P' and 'Z'. We recall the expressions

$$P= \int_R^\infty f(x)*dx \qquad\qquad (5.70a)$$

$$Z= \int_R^\infty (R-x)*f(x)*dx \qquad\qquad (5.70b)$$

Another way to compute the services is based on the depletion time. This approach is more convenient in relation to multilevel inventory systems, as it will be seen later in this chapter.

In inventories operated on a reorder level, 'R' is the amount of stock available when a new order is raised. Consider the depletion time 't' as the time to completely deplete that quantity 'R'. A stockout will occur when 'R' is depleted before the arrival of the related order, i.e., when 't' is lower than the lead time 'L'. Formally, it can be written as:

$$P= \int_0^L q(t)*dt \qquad\qquad (5.71a)$$

where 'q(t)' stands for the distribution density of the time to deplete 'R'

Similarly, the expected time out of stock can be computed as

$$Zt = \int_0^l (L-t)*q(t)*dt \qquad (5.71b)$$

Then, if 'd' is the demand rate, the expected shortage is given by

$$Z = Zt*d \qquad (5.72a)$$

$$= d*\int_0^l (L-t)*q(t)*dt \qquad (5.72b)$$

Note now that from (5.70) we can derive the result met before

$$\partial Z/\partial R = -P \qquad (5.73a)$$

Similarly, from (5.71)

$$\partial Zt/\partial L = P \qquad (5.73b)$$

Thus, $\quad \partial Z/\partial R = -\partial Zt/\partial L \qquad (5.73c)$

That is, the marginal effect on the shortage quantity of an increase on the reorder level is the same as the marginal effect on the shortage time of a decrease on the lead time.

The shortage rate 'V' could be computed from 'Z' in the usual way. Or, directly from 'Zt' and the average period 'T' as

-5.5-

$$V = Zt/T \qquad (5.74)$$

The density function 'q(t)' can be estimated empirically from recorded data. Or, on some occasions, derived from the demand distribution 'f(x)'. Burgin (1969) gives some formulas to calculate the mean and the variance when the demand is Normal, and the mean when it is Gamma distributed. Furthermore, if the demand follows a Poisson distribution, the time to deplete 'R' is Gamma distributed with modulus 'R'. In this case, the mean and the variance would be

$$tm = R/d \qquad (5.75a)$$
$$tv = tm/g \qquad (5.75b)$$

where $g = (d/ds)^{**}2$; 'd' and 'ds' are the mean and the standard deviation of the demand per unit of time.

Further discussion about these points can be found in subsection 5.4.1. There, it will be also argued that the expressions (5.75) are likely to constitute sensible approximations for the first two moments of the depletion time distribution when demands are Gamma or Normal distributed. In that case, they would cover the demand situations which have been especially contemplated along this study.

Depletion time distributions tend to the Normal when 'R'

tends to infinity. If demand were Poisson, the depletion time would be Gamma. If demand were Gamma, the depletion time relates to a Poisson. Both Gamma and Poisson distributions can be approximated well by a Normal if the modulus or the mean, respectively, is high enough. This last condition is usualy met in the lead time demand distributions of a 2-level inventory system because the lead time (external) is long. Further details are given in the subsection 5.4.1. The point to be stressed here is that formulas (5.75) and the Normal distribution may provide approximations for a large range of demand situations.

Note, now, the 'additivity' of the depletion times. If the amount 'R' is split into 'R1' and 'R2', the times to deplete 'R1' and to deplete 'R2' (say 't1' and 't2' respectively) are independent variables. Therefore, we can expect that 't=t1+t2', the time to deplete 'R=R1+R2' has mean and variance equal to the sum of the means and the variances, respectively, of 't1' and 't2'. This is consistent with expressions (5.75).

The additivity property just mentioned applies straight to the 2-level system with 'K' satellites. Let 'Xd(n)=R-Yhf' be the quantity depleted in the system up to the final partition. The corresponding depletion time, according to the formulas (5.75) would have

mean= (R-Yhf)/d

variance= (R-Yhf)/(d*g)

At the final allocation, 'Yhf' would be split into the 'Yhf(k)'s. The additional depletion time would have

$$mean= Yhf(k)/d(k)$$

$$variance= Yhf(k)/(d(k)*g(k))$$

Therefore, the mean and the variance for the total depletion time would be

$$tm(k)= R/d-Yhf/d+Yhf(k)/d(k) \qquad (5.77a)$$

$$tv(k)= R/(d*g)-Yhf/(d*g)+Yhf/(d(k)*g(k)) \qquad (5.77b)$$

The problem is again to fix the 'Yhf(k)'s such that the criterion of equal weighted probability of depletion is maintained. A precise solution would involve many iterations. Under the assumption of twin satellites, 'Yhf(k)' is proportional to 'd(k)'. Then, those formulas simplify to give

$$tm= R/d \qquad (5.79a)$$

$$tv= tm/g+(g/g(k)-1)*Yfh/(d*g) \qquad (5.79b)$$

The evaluation of the service rates could be done according to formulas (5.71) and (5.72). The normal approximation can be used as it will be discussed in subsection 5.5.1. In this case, the value for the standard normal variable would be taken as

$$\eta = (Lm-tm)/s \qquad\qquad (5.81a)$$

where 'Lm' is the mean lead time, 'tm' the mean depletion time as from (5.79). The standard deviation 's' is such that

$$s**2 = tv+Lv \qquad\qquad (5.81b)$$

'Lv' is the lead time variance. The values for 'P', 'Z' and 'V' would then follow.

The above results can also be used to determine the conditional probability 'p', i.e. the parameter for intermediate allocations. The expression (5.29) in section 5.3 can be written

$$p = P/Pd(n) \qquad\qquad (5.83a)$$

where
$$Pd(n) = \int_{Xd(n)}^{\infty} f(x)*dx \qquad\qquad (5.83b)$$

$$Xd(n) = R-Yhf \qquad\qquad (5.83c)$$

'P' is the average stockout rate for the system and is calculated from (5.81). The density 'f(x)' relates to the system lead time demand. This procedure to compute 'p' is simpler than the one considered in subsection 5.4.1.

5.5.1   Analytical details.

The stockout rate and the expected time out of stock given by the expressions

$$P= \int_0^L q(t)*dt \qquad\qquad (5.85a)$$

$$Zt= \int_0^L (L-t)*q(t)*dt \qquad\qquad (5.85b)$$

assume that the lead time 'L' is constant. With variable lead times they should be replaced by

$$P= \int_0^\infty h(L)* \int_0^L q(t)*dt*dL \qquad\qquad (5.87a)$$

$$Zt= \int_0^\infty h(L)* \int_0^L (L-t)*q(t)*dt*dL \qquad\qquad (5.87b)$$

where 'h(L)' stands for the lead time density function.

The complexity brought about by a variable lead time can be eliminated by considering the time gap 'y=L-t'. Denote by 'f(y)' the time gap density function. Then 'P' and 'Zt' could also be calculated from

$$P= \int_0^\infty f(y)*dy \qquad\qquad (5.89a)$$

$$Zt= \int_0^\infty y*f(y)*dy \qquad\qquad (5.89b)$$

Note that 'y' may vary from minus to plus infinity.

The estimate of 'f(y)' can be done empirically from the actual time gaps recorded. Alternatively, it can be derived from the depletion time and lead time distributions. Assuming that depletion times and lead times are independent, we can estimate the mean and the variance of the time gap as

$$ym= Lm-tm \qquad (5.91a)$$
$$yv= Lv+tv \qquad (5.91b)$$

Later in this section, reasons are given to expect 'L' and 'R' to be high in the multilevel inventories we are dealing with, so that 'q(t)' and 'h(L)' are likely to be close to Normal densities. Then, 'f(y)' is likely to be also fit by the Normal.

Note that if 'q(t)' is Normal, the evaluation of 'P' and 'Zt' in (5.71) would involve the computation of the standard value

$$\xi = (L-tm)/s \qquad (5.92)$$

If 'f(y)' is Normal, the evaluation through (5.89) would require the standard value

$$\eta = ym/s \qquad (5.93a)$$
$$= (Lm-tm)/s \qquad (5.93b)$$

The standard deviation in (5.92) relates to the depletion time distribution while in (5.93) it relates to the time gap distribution. The latter expression reduces to the former when the lead time is constant. Therefore, the depletion time and the time gap approaches are basically the same if we take for the variance in both cases the depletion time variance plus the lead time variance.

The characterisation of the depletion time can be approximated, and sometimes exactly derived, from the demand distribution. In Burgin (1969) and Comments (1970), the following results are derived. Consider the demand per unit of time and let 'd', 'dv' be its mean and variance, respectively, and 'g=(d**2)/dv'. Then,

(i) When the demand is Gamma distributed, the mean time to deplete a quantity 'R' is given by

$$tm= R/d+(1+1/g)/2 \qquad\qquad (5.95)$$

(ii) When demand is Normal, the time to deplete 'R' has for mean and variance, respectively

$$tm= R/d+1/(2*g) \qquad\qquad (5.97a)$$
$$tv= (tm+3/(4*g))/g \qquad\qquad (5.97b)$$

A comment should be made in relation to these results. Burgin considered the depletion time as a discrete variable, rounded up. For instance, the depletion time is taken as 1 day if the whole quantity is sold in the first day, whether if in the morning or in the afternoon. In consequence, the

evaluation of the 'continuous' depletion time becomes overestimated when the expressions above are used. The error may become significant when 'R' is split into quantities 'R1', 'R2', ..., 'Rn'. Then the expected time to deplete 'R1', plus 'R2', plus...'Rn' should remain the same as the time to deplete 'R'. The use of Burgin's expressions would inflate the results. For instance, for a Gamma distributed demand, the expected time to deplete 'R' would be given by (5.95). However, the sum of the expected times to deplete 'R1', 'R2',..., 'Rn' would be

$$R/d+n*(1+1/g)/2$$

i.e., the last term comes multiplied by 'n'. Obviously, such a result lacks coherence.

For the analysis of multilevel inventories the depletion of 'R' has to be considered in stages. Burgin's formulas will be simplified to:

$$tm= R/d \qquad\qquad (5.99a)$$
$$tv= tm/g \qquad\qquad (5.99b)$$

This eliminates the inconsistency mentioned above. The legitimacy of above simplification requires to be investigated. Multilevel systems are associated to long lead times. This makes 'R' relatively high. On the other hand, slow movers are excluded from the analysis, and for a fast mover, '1/g' is likely to be much lower than 'R/d'.

Therefore, in the area of our interest, differences from (5.99) to the Burgin's formulas are not dramatic. The expressions (5.99), actually, are the exact formulas if demand were Poisson with a rate 'd'. In fact, it is known that Poisson and Gamma distributions relate to each other. Such a relationship can be introduced as follows. Assume that the Poisson variable relates to "failures" which take place with a rate 'a'. The distribution of the interval per failure (i.e. the interval between 2 consecutive failures) is a negative exponential with mean '1/a' and variance '1/a**2'. The distribution of the interval over 'G' failures is Gamma with modulus 'G', mean 'G/a' and variance 'G/a**2'. Therefore, if demand is Poisson with rate 'd', the time per unit demanded (i.e., per "failure") has mean '1/d' and variance '1/d**2'. The time per 'R' units demanded is Gamma distributed with mean 'tm=R/d' and variance 'tv=R/d**2=t/g'.

Let's consider, now, a Gamma distributed demand. Let 'd', 'g' and 'a=g/d' be the mean, the modulus and the scale factor, respectively, for the distribution of the demand per unit of time. The lead demand follows, then, a Gamma distribution with modulus 'G=L*g', mean 'D=L*d' and the same scale factor 'a'. Without loss of generality, take the 'day' as the current time unit; so, the demand rate is 'd' units per day, the lead time is 'L' days, etc. Denote by 'sut' a standard unit of time, such that '1 sut = 1/g day'. And consider 'x' a variable representing the integer number of 'suts' spent to sell each additional unit of product; i.e. 'x' is the number of 'suts' per unit demanded. Assume that

'x' follows a Poisson distribution with a rate 'a=g/d'. Then, the demand per sut is a negative exponential with mean 'd/g'. And the demand over 'G suts' follows a Gamma distribution with modulus 'G' and mean 'G*d/g=D'. The latter is the lead time demand distribution initially considered. Therefore, 'x' is the Poisson distributed variable associated to the Gamma distributed demand.

The variable 'x' represents the integer number of suts that elapse per unit demanded. It is Poisson distributed with mean 'a=g/d'. Hence, the variable 'X' representing the integer number of suts that elapse per 'R' units demanded is a Poisson with mean and variance 'm=g*R/d'. Now, if 't' represents the time in days that corresponds to 'X' in suts, then 't=X/g'. Therefore, the mean and the variance would be, respectively:

$$tm= m/g= R/d$$
$$tv= m/(g**2)= tm/g$$

This agrees with (5.99).

Note that 't' is Poisson distributed only for 'g=1'. However, for long lead times and fast moving items, 'm' is high enough to legitimate the Normal approximation for the 'X' distributed variable. Consequently, 't=X/g' can be approximated by the Normal, too.

Note also that 'X' is an integer, therefore, it

underestimates the actual time. The error, however, should be less than 1 sut.

In order to clarify this last approach, consider a lead time demand Gamma distributed with modulus 'G=30'. Denote the reorder level by 'R', the standard deviation for the lead time demand by 's' and the ratio 'R/s' by 'u'. Make 'u=6.0'. From the tables in Burgin et al (1976), we can read:

$$Z/D= 3.67\% \qquad \text{(L/D in the tables)}$$
$$P= 28.52\% \qquad (\text{F} \quad " \quad " \quad " \quad)$$

The approach by the associated Poisson would take the following steps:

$$E (X)= m= a*R= u*\sqrt{G}= 32.863$$
$$\text{Lead time} = L \text{ days}= (L*g) \text{ suts} = G \text{ suts}$$
$$P= pr(X<G)= 1-pr(X\geqslant G)= 28.86\%$$
$$Zt= m*pr(X\geqslant G)-G*pr(X\geqslant (G+1))-(m-G)$$
$$= 32.863*0.7114-30*0.6481-2.863= 1.07 \text{ suts}$$
$$Z/D= Zt/L= Zt(\text{in suts})/G= 1.07/30= 3.57\%$$

So, the values for 'P' and 'Z/D' agree with those found before.

The rate 'm' in the Poisson distribution is high enough to be approximated by the Normal. By doing so, we would get

$$yo= (30-32.863)/\sqrt{32.863}= -0.4994$$
$$P\sim 31\%$$

$$Zt= 0.198*\sqrt{32.863} \text{ suts}= 1.135 \text{ suts}$$

$$Z/D= 1.135/30= 3.78\%$$

These results are comparable with those just found above

The expressions (5.99) are theoretically exact for Poisson distributed demands and look to be reasonable approximations over the range of the Gamma family. The latter, as it is known, tends asymptotically to the Normal. Those formulas, therefore, apply over a spectrum of distributions large enough to fit the majority of the demands patterns of items in stock.

-5.5.1-

5.6 Extension to the Johnston model

The maximisation of the profit in J-model, as seen before, involves the solution of the following two simultaneous equations:

$$\partial FZ(i,j)/\partial R(i,j)= -1/N(j)*(F2(j)+\lambda s) \qquad (5.105a)$$

$$N(j)**2= 26*d(j)*(F2(j)+\lambda s)/(FZ(j)+F0(j)+\lambda n) \qquad (5.105b)$$

The indices refer to the item 'i' of the family 'j'. The above expressions can be extended to the 2-level inventories if the shortage and the shortage penalty are understood as follows:

$$Z(i,j)= \sum Z(i,j,k) \qquad (5.111a)$$

$$FZ(i,j)= \sum F(i,j,k)*Z(i,j,k) \qquad (5.111b)$$

where 'k', as usual, stands for the satellite.

The exact evaluation of the l.h.s. of the equation (5.105a) is not an easy matter if we want to contemplate systems with satellites having rather different demand structures and commercial strategies. Because of such a difficulty, the analysis will be restricted to structures of twin satellites. The main results are presented below. Further analytical details can be found in the subsection 5.6.1.

For twin satellites the 'F(i,j,k)'s are constant in 'k',

'F(i,j,k)=F(i,j)' say. In consequence, the allocation policy leads to 'P(i,j,k)=P(i,j)', i.e., the same shortage for each item (i,j), no matter which satellite. Under these circumstances

$$\partial FZ(i,j)/\partial R(i,j)= -F(i,j)*P(i,j) \qquad (5.113)$$

Applying the last result to equation (5.105a), it follows

$$P(i,j)= 1/N(j)*(F2(j)+\lambda s)/F(i,j) \qquad (5.115)$$

which is the equation found for the single level inventory. The other equation, (5.105b), remains unchanged.

The decision sequence can follow steps parallel to those in the single level case. Chapter 4 gives the way to choose a convenient value for 'N(j)' in the single level case. Those formulas apply also to the multilevel case. The value of 'N(j)' can so be computed.

The equation (5.115) generates a balanced set of 'P(i,j)'s. The depletion time approach and the approximation of the depletion time distribution through the Normal would lead to the values for 'R(i,j)'. The steps could be

From P(i,j), compute the standard value 'η'
From 'η' compute 'R(i,j)'. This can be done by means
of expressions (5.131) in the subsection 5.6.1

-5.6-

Once the set of 'R(i,j)'s has been calculated for all the items, the procedures are exactly the same as for the single case.

5.6.1   Analytical details

The main problem with the evaluation of the l.h.s. of the equation (5.105a), that is ' $\partial FZ/\partial R$' (the subscripts 'i,j' are implied), results from the difficulty of finding a simple mathematical expression to give the marginal effect of 'R' on the final partition to the satellites. In other words, if 'R' is increased by one unit, how would this fact change the quantities 'YHf(k)' in expressions (5.77) of section 5.5?

If we are dealing with twin satellites, the 'Yhf(k)'s remain the same for a constant global 'Yhf', whatever the value of 'R'. Furthermore, as the shortage penalties 'F(k)' are the equal to all the satellites, so is the expected shortage rate 'P(k)'. Then, 'P=P(k)' can be taken as the expected average stockout rate for the system. Note that when the twin assumption does not apply, the 'P(k)'s should be different from one satellite to another, so, the concept of system stockout rate becomes less clear.

For twin satellites, whenever some quantity is split, say

'R' into 'R1+R2+...', we have

$$\partial Z(k)/\partial R(k)= -P(k)= -P \qquad (5.121a)$$

As
$$Z=\sum Z(k) \qquad (5.121b)$$

then
$$\partial Z/\partial R= \sum (\partial Z(k)/\partial R(k))*(\partial R(k)/\partial R) \qquad (5.121c)$$

$$= -P*(\sum \partial R(k)/\partial R \qquad (5.121d)$$

Therefore

$$\partial Z/\partial R= -P \qquad (5.121e)$$

because $R= \sum R(k)$

The last relationship (5.121e), obviously, holds in two extremes: when 'R', the reorder level is split altogether into 'R1','R2',..., and sent to the satellites which, from then on, are run separately; and when the whole stock is kept common until being completely used. This last case is, actually, equivalent to have only one outlet with several counters.

The situations that are being considered here are located between these two extremes, therefore the expression (5.121e) should hold too. Anyway, equations (5.71) in section 5.5 apply to each satellite as follows:

$$P(k)= \int_o^L q(t)*dt \qquad (5.123a)$$

$$Zt(k)= \int_o^L (L-t)*q(t)*dt \qquad (5.123b)$$

Then, $\partial Zt(k)/\partial L= P(k)$ \qquad (5.123c)

Since 'P(k)=P', and as '$\partial Z/\partial R = -\partial Zt/\partial L$' from (5.73), then the relationship '$\partial Z/\partial R$' holds in general.

This result leads to equation (5.115) seen before. The latter generates a value for 'P' which, in turn, dictates the value for the reorder level 'R'. The depletion time approach can be used for the purpose of calculating the 'R' in relation to a given 'P'. The whole procedure is detailed below.

With the depletion time normally distributed, we can find the value for the standard variable that corresponds to the given 'P'. Denote by 'η' that value. Recalling, now, the expressions (5.79) in section 5.5:

$$\eta = (L-t)/s \qquad (5.125a)$$
$$s**2 = tv+Lv \qquad (5.125b)$$
$$t = R/d \qquad (5.125c)$$
$$tv = t/g+(g/g(k)-1)*Yhf/(d*g) \qquad (5.125d)$$

Now, denote

$$C1/\eta^2 = 1/(2*g) \qquad (5.127a)$$
$$C2/\eta^2 = (g/g(k)-1)* Yhf/(d*g)+Lv \qquad (5.127b)$$

By rearranging the set of equations (5.125) we obtain

$$t**2-2*(L+C1)*t+(L**2-C2) = 0 \qquad (5.129)$$

-5.6.1-

Then

$$t = L + C_1 + C_3 \qquad \text{(5.131a)}$$

Where $\quad C_3 = \sqrt{(C_1{**}2) + 2{*}L{*}C_1 + C_2} \qquad$ (5.131b)

Finally, $\qquad R = d{*}t \qquad$ (5.131c)

5.7 Determining the internal reorder and clearing levels

The internal reorder levels 'R(k)'s and the average clearing quantity 'YOf' have been considered to be set empirically. They add up to give the 'Yhf'. So, the latter has been a variable exogenous to the model. The results in (5.79) show that 'Yhf' increases the variance, and therefore, brings a penalty for the average investment in stock. The relative magnitude of this penalty increases with the number of satellites. With 1 satellite, 'g/g(k)=1', so the penalty is zero as expected. For 9 satellites the increment in the depletion time variance can be as large as '8**Yhf/(d*g)' which is likely to represent more than 50% of the whole variance. The size of this effect illustrates the convenience of making the 'R(k)'s and the 'YOf' endogenous variables.

The inclusion of such variables in the model requires that some sort of cost or profit rates are found for each of them. Though, fixing a charge in relation to 'YOf' may be more difficult and less tangible than fixing a value for 'YOf', directly. In respect to 'R(k)', we know that it governs the intermediate shortages. Shortage penalties have been considered already in relation to the final shortages and can be extended to the intermediate ones. Intermediate shortages, however, are likely to produce less damage. In fact, expediting is much easier, quicker and cheaper, so, the probable effect on sales and on 'good will' is much

weaker than in relation to the final shortages. Yet, intermediate shortages may occur more frequently.

The treatment given to the intermediate shortages depends largely on the attitude of the decision maker. He may have practical reasons to decide, for instance, that

$$PI = c*P \qquad\qquad (5.135)$$

i.e. the intermediate probability of stockout is taken as a fraction of the system stockout rate. When penalties for the intermediate shortages can be estimated, the 'optimum PI' can be found by following the basic approach in the J-model. This will be discussed in subsection 5.7.1. The calculation of a precise value for 'PI' would involve one or more iterations. To avoid a significant increase on the computational burden it is suggested that 'PI' should be set, roughly, as in (5.135). Occasionally, a 'PI' coherent with the current 'P' should be calculated in order to check the value that is being used for 'c' in (5.135). The computations required for such a checking are quite simple. An example is given in subsection 5.7.1.

So far, it has been assumed that the internal reorder level remains the same through all the internal cycles. However, it does not need to be so. There is plenty of stock in the MWH after an external supply, so, the risk of an intermediate shortage gives no pay-off. This is shown in

fig.5.4. The advantage to the final shortages of having low
'R(k)'s only appears in relation to the last partition.
Therefore, we can operate with two internal reorder levels.
A safe level, say 'R(k,1)' for the first internal cycles;
and a lower one, say 'R(k,n)' for the end, when the system
stock becomes scarce.
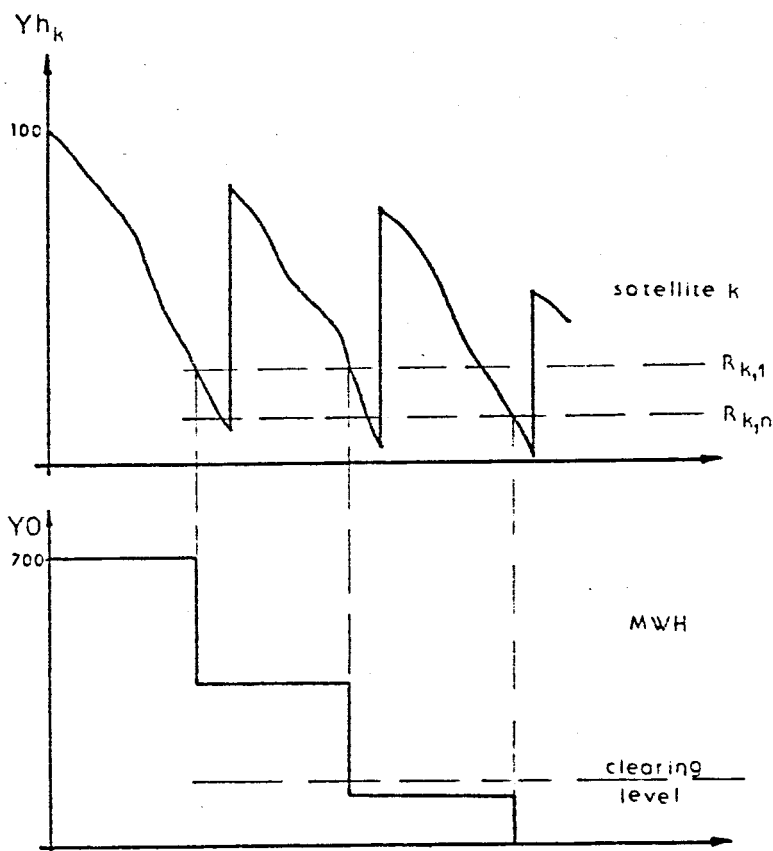
Fig. 5.4 - Operating on 2 reorder levels

The rule to change from 'R(k,1)' to 'R(k,n)' appears to be
quite obvious. The 'R(k,n)' holds only when the MWH is at or
below its clearing level, i.e., in the cycle which precedes
the last partition. 'R(k,1)' stands for the others. The
figure exemplifies this procedure. Reallocations are

-5.7-

triggered by 'R(k,1)' up to 'tm(n-1)'. When the 'R(k,1)' is reached next, the MWH is below the clearing level. The allocation is, then, postponed until the level 'R(k,n)'. Note that the final shortage is not affected by this procedure. Conversely, the intermediate shortage can be reduced drastically. In practical terms, this shortage becomes negligible in all but the last intermediate period.

5.7.1   Analytical details

The intermediate shortage can be treated analytically on the same grounds used for the final shortage. For this purpose some sort of shortage penalty needs to be assumed. Let us take that the intermediate shortage penalty is set in proportion to the penalty for the final shortage. Writing 'FI' for the intermediate shortage penalty, such a relationship could take the form 'FI=b0*F'. The latter will be adopted here, after multiplying by 'n', the number of intermediate cycles in each external one. Then, we get for each item

$$n*FI= BO*F \qquad\qquad (5.139)$$

These parameters can be included in the basic model. The general expression for the profit would become

$$\pi = 52*F1*d-N*(FZ+B0*FZI)-F2*S-N*F0 \qquad (5.141a)$$

$$\text{where} \quad FZ = \sum\sum\sum F(i,j,k)*Z(i,j,k) \qquad (5.141b)$$

$$FZI = \sum\sum\sum F(i,j,k)*ZI(i,j,k) \qquad (5.141b)$$

We take, as usual, the partial derivatives of the Lagrangian and equate them to zero. Without loss of generality, let 'YOf' be kept as an exogenous variable. Denote by 'RI' the sum of all 'R(k)'s. The decision variables can be chosen as 'R' and 'RI'. 'N' is not being considered, now. The rule to break down the 'RI' into the 'R(k)'s follows the criterion of equal weighted probability of stockout.

We have

$$R = Xd(n)+YOf+RI \qquad (5.143)$$

From the approximation in (3.75)

$$\partial S/\partial R = 1 \qquad (5.145a)$$

If 'R' remains constant

$$\partial S/\partial RI = 0 \qquad (5.145b)$$

From (5.121e)

$$\partial FZ/\partial R = -F*P \qquad (5.147)$$

We need to find, next, the partial derivative, '$\partial FZ/\partial RI$'. This will be attempted below for the twin satellite

-5.7.1-

situation with normally distributed depletion times.

Note first that, as 'R' is kept constant, we conclude from (5.79) in section (5.5) that

$$\partial tm/\partial RI = 0 \tag{5.151a}$$

$$\partial tv/\partial RI = (1/g(k)-1/g)/d \tag{5.151b}$$

Furthermore, from (5.81)

$$\partial \eta/\partial s = -\eta/s \tag{5.153a}$$

$$\partial s/\partial tv = 1/(2*s) \tag{5.153b}$$

As $F(k)=F$

$$FZ = F*Z \tag{5.155a}$$

$$= F*d*Zt \tag{5.155b}$$

And if 'ζ' is the shortage for the standard Normal distribution for the standardized value 'η', then

$$Zt = s*\zeta \tag{5.157}$$

Note $\quad \partial \zeta/\partial \eta = P$ $\tag{5.159}$

Then $\partial Zt/\partial s = \zeta + s*(\partial \zeta/\partial \eta)*(\partial \eta/\partial s)$ $\tag{5.161a}$

$$= \zeta - \eta*P \tag{5.161b}$$

Now

$$\partial FZ/\partial RI = F*d*(\partial Zt/\partial s)*(\partial s/\partial tv)*(\partial tv/\partial RI) \tag{5.163a}$$

-5.7.1-

$$=F*d*(\zeta-\eta*P)/(2*s)*(1/g(k)-1/g)/d \qquad (5.163b)$$

$$=F*((\zeta-\eta*P)/(2*s))*(1/g(k)-1/g) \qquad (5.163c)$$

Denote by

$$AO= 1/g(k)-1/g \qquad (5.165a)$$

$$Ax= (\zeta-\eta*P)/(2*s) \qquad (5.165b)$$

Then

$$\partial FZ/\partial RI= F*AO*Ax \qquad (5.165c)$$

The partial derivatives of 'FZI' are also required. Obvously

$$\partial FZI/\partial R= 0 \qquad (5.167a)$$

And, reasoning as for deriving the expressions (5.121),

$$\partial FZI/\partial RI= -F*PI \qquad (5.167b)$$

The optimalty conditions, finally, can be established. By equating to zero the partial derivatives of (5.141a) we get:

From $\partial/\partial R(i,j)=0$: $-N*(-F*P)-(F2+\lambda s)=0$ $\qquad (5.169a)$

and hence $P= (F2+\lambda s)/F$ $\qquad (5.169b)$

From $\partial/\partial RI(i,j)=0$: $-N*(F*AO*Ax-BO*F*PI)=0$ $\qquad (5.171a)$

and hence $PI=(AO/BO)*Ax$ $\qquad (5.171b)$

Note from (5.165) that 'AO~(K-1)/g' if demands for the 'K'

satellites are close to each other because, if demands had the same distribution, $g(k)=g/K$. When $K=1$, $A0=0$ and $PI=0$. The value of Ax interacts with the value of P and of PI itself, therefore, finding a precise figure for Ax requires the use of iterative methods. The example below will make it clearer.

Consider a system with $K=5$ identical satellites each with daily demand Gamma distributed with $d(k)=10$ and $g(k)=0.5$. The external and internal lead times are $LE=25$ and $LI=5$, both deterministic. The external reorder ratio is $R/D=1.2$ and the internal is $RI/DI=2.5$. The average clearing quantity is $Y0f=100$

Then:

system daily demand $\quad$ $d=K*d(k)=50$

$\qquad\qquad\qquad\qquad\qquad$ $g=K*g(k)=2.5$

total lead time $\qquad\qquad$ $L=LE+LI=30$

system lead time demand $D=L*d=1500$

external reorder level $\quad R=R/D*D=1800$

internal lead time demand $DI(k)=LI*d(k)=50$

internal reorder level $\quad RI(k)=RI/DI*DI(k)=125$

final partition $\qquad\qquad$ $Yhf=K*RI(k)+Y0f=125$

From (5.79), section 5.5

$\quad$ tm= $R/d=$ 36

$\quad$ tv= tm/g+(K-1)*Yhf/(d*g)= 37.6= 6.13**2

From (5.81) and admitting the normal approximation

$$s = 6.13$$

$$\eta = (30-36)/6.13 = -0.98$$

Then $P = 0.16$

$$\zeta = 0.09$$

From (5.165)

$$AO = (1/0.5)-(1/2.5) = 1.6$$

$$Ax = (0.09+0.98*0.16)/(2*6.13) = 0.02$$

Finally, from (5.171)

$$BO*PI = 1.6*0.02 = 0.03$$

The expected intermediate stockout for the internal reorder level ratio indicated above is 'PI=0.02', for a Gamma distribution. If we take 'BO=1' in the last result, the value set for 'RI' is about right.

The expression (5.171b) does not give, directly, a value for 'PI' because both 'PI' and 'Ax' depend on 'RI'. An approximate solution would have to be found iteratively. However, a rough estimate of the magnitude of 'PI' is easy to obtain from the current figures for 'P' and 'Yhf'. That magnitude can then be compared with the value which is being adopted for 'PI'.

5.8   Simulation checks

Formulas to compute stockout and shortage rates, 'P' and 'V', in a 2-level inventory system have been derived in section 5.5. Those derivations involved three sorts of approximations necessary to overcome analytical difficulties.

The first sort relates to the depletion time distribution. Considering that we are dealing with fast moving items and that a multilevel system implies relatively long lead times for the external supplies, it has been assumed that the Normal distribution would fit that depletion time with reasonable accuracy. Furthermore, that the mean and the variance for the depletion time could be calculated through the formulas derived for a Poisson distributed demand.

The second sort of approximations was introduced when mal-distributions of stock among the satellites were ignored. It was assumed that the last partition takes place when the stock in the MWH is at the level 'YOf'. This is an estimate of the real value which, actually, can vary from zero to '2*YOf'. Furthermore, when the last partition takes place, there is one satellite at its reorder level, but the others are likely to be above theirs. However, for analytical purposes, all of them have been considered at the respective reorder levels: so, the last partition actually takes place earlier in the cycle than the analysis

considers.

The last partition is supposed to equalise the probability of stockout for all the satellites. However, this may be impossible if some of the satellites are overstocked, because transhipments are not allowed. Furthermore, practical reasons often prevent the sharing to be made exactly as theoretically it should be, which adds new distortions.

The effects of these two sorts of approximations will be seen below for two cases of demand situations, by comparing the expectations from the formulas with the values obtained from simulation runs. The third sort of approximations relates to the assumption of twin satellites which was used to simplify the analysis. The use of the expressions so derived, when the satellites have distinct demands, is another source of errors. This, however, will not be investigated here.

The Poisson distribution has the variance equal to the mean whilst demands of fast moving items, usually, have the variance greater than the mean. Then, one can expect the variance of the depletion time to be greater than that given by the simplified formula in (5.75b) which is exact for Poisson demand, only. Hence, one can also expect the values of 'P' and 'V' computed on that basis to be lower than the actual ones. Furthermore, the approximation of the depletion time distribution by the Normal is another source of

-5.8-

distortion.

The mal-distributions mentioned in relation to the second sort of approximations cause again an under-estimate of the shortages. It has been shown in section 5.1 that shortages would be minimised if satellites kept equal probability of stockout. If such condition is not verified at the last partition, actual shortages are greater than the analytical expectation.
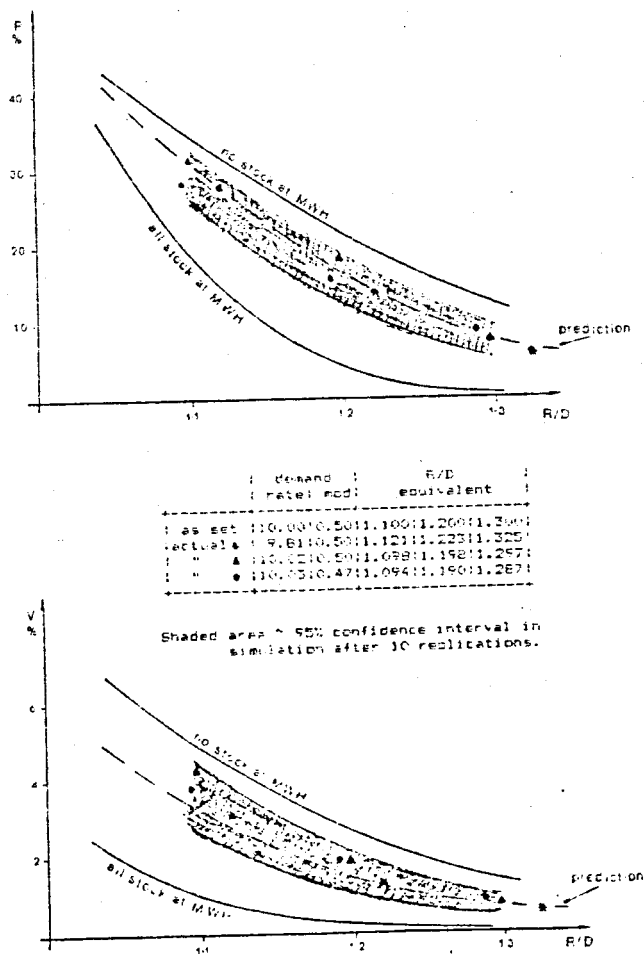
Fig. 5.5 - Simulation check, g=0.5

In order to get a feeling about the magnitude of these

errors, the analytical predictions were checked by simulation, under two demand situations: 'g=0.5' and 'g=1.0', where 'g' is the modulus of the Gamma distribution which has been assumed for the daily demand at each satellite. The system comprises 1 MWH and 5 equal satellites whose daily demand rate is 10 units for each. Other details are given in section 5.8.1.
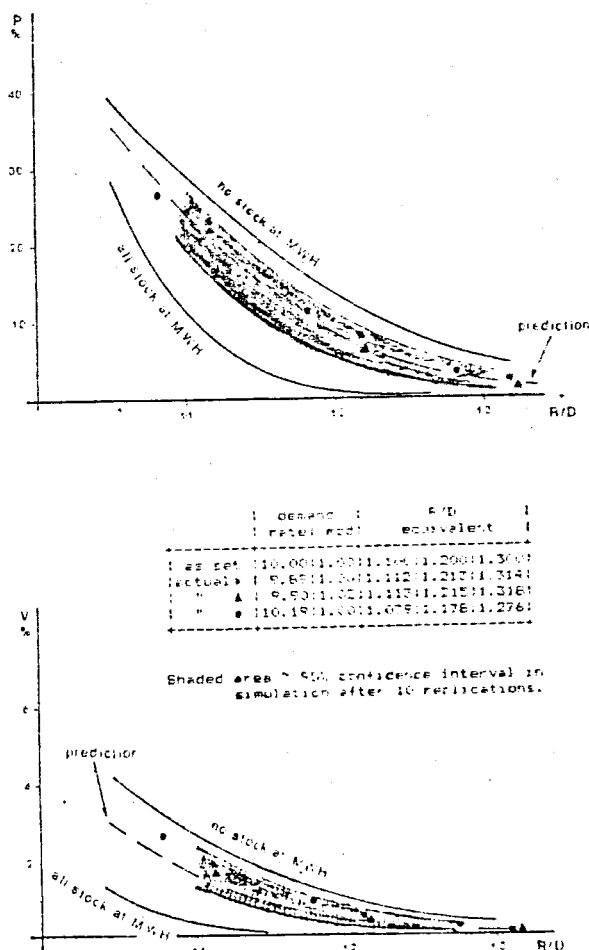
Fig. 5.6 - Simulation check, g=1.0

Figs. 5.5 and 5.6 depict the results of simulation in comparison with the predictions, at different reorder levels 'R/D'. The 3 marks, star, triangle and dot, are the values for 3 of the seeds used to generate the demand streams. The

simulation values are plotted against the 'equivalent R/D', not the value of 'R/D' that was set for the simulation. The 'equivalent R/D' relates to the actual mean and variance of the stream, as it will be explained in section 5.8.1. The dashed line represents the predictions. The solid lines relate to the limit situations. The upper line, "no stock at MWH" corresponds to the situation of all the stock being allocated and delivered to the satellites at once, when the procurement arrives at the system. The lower line, "all stock at MWH", relates to having the whole system demand concentrated on one outlet.

The graphs show that the simulation results are reasonably close to the prediction line, particularly in the region of high service levels. Inventory policies for fast moving items, usually, adopt stockout rates 'P' lower than 15% In this range, differences from simulated to analytical results look small; and, definitely, the dashed line constitutes a better estimate than either of the two boundaries given by "no/all" stock at the MWH.

5.8.1   Further details

The simulation values plotted on figs. 5.5 and 5.6 were obtained for an inventory system comprising 1 MWH and 5 satellites with equal distributed demand. Other

characteristics are mentioned in the printout of fig. 5.7.

```
* SIMULATION ON A 2-LEVEL MULTIFACILITY STOCK *

SETTING PARAMETERS:
      SYSTEM DAILY DEMAND, D1N=              50.
      EXTERNAL ORDER OVERLAPPING, D/Q=       0.80
      EXTERNAL REORDER LEVEL RATIO, RO/D=    1.30
      EXTERNAL LEAD TIME, LE=                25.
      FINAL LUMP DELIVERY, YFO=             100.
      No.OF SATELLITES, NWH=                 5.
      SATELLITE DAILY DEMAND, D1=            10.
      SATELLITE DAILY DEMAND MODULUS =       0.50
      INTERNAL REORDER RATIO =               2.50
      INTERNAL LEAD TIME, LI=                5.
      MAX/FINAL ALLOCATION =                 1.5
      LOSS FACTOR, ALFA=                     0.00

      SIMULATION LENGTH =            5000.
      RUN-IN TIME =                  200.
      RANDOM GENERATOR =             531.

JOINT ALLOCATION TO SATELLITES


PERFOMANCE   EXPECTATIONS:      LIMIT CASES:
FINAL STOCKOUT RATE= 0.074    0.005 TO 0.123
FINAL SHORTAGE RATE= 0.006    0.000 TO 0.013


* PERFOMANCE PARAMETERS *
Ave.No.INT. CYCLES = 649.4
No.EXTERNAL CYCLES = 129.
INTER STOCKOUT   RATE = 0.010
FINAL STOCKOUT   RATE = 0.075
INTER SHORTAGE   RATE = 0.003
FINAL SHORTAGE   RATE = 0.007
AVERAGE STOCKS  (no.of weeks):
      SYSTEM NOTIONAL =     11.45
          ''   ON HAND =     6.43
      SATELLITES ON HAND =   2.72
      ON INTERNAL TRANSIT =  1.00
      MAIN WH ON HAND =      2.71

ACTUAL DEMAND CHARACTERISTICS:
      SAT.     RATE     ST.DV.     MOD.
       1       9.64     13.68      0.50
       2      10.04     14.66      0.47
       3      10.09     14.21      0.50
       4      10.20     14.19      0.52
       5      10.15     14.12      0.52
      ALL     10.02     14.18      0.50
*ET=5:54.2 ET=52.4 IO=0.3
```

Fig. 5.7 - Simulation printout

The daily demand was set at 10 units a day for each satellite and assumed to be Gamma distributed. Two cases were considered for the satellite daily demand modulus: 'g=0.5' and 'g=1.0'. The system daily demand is then Gamma distributed with 'mean=50' and 'modulus=5*g'.

There is no overlapping. The ratio 'D/Q' was set as 0.8,

-5.8.1-

where 'D' is the lead time mean demand and 'Q' the reorder quantity.

The external reorder level ratio 'RO/D' was set at 1.1, 1.2, and 1.3. As the orders do not overlap, the notional and physical reorder levels, 'RO' and 'R' respectively, coincide. Lower values for 'RO/D' have not been considered for they are outside of the range of interest and, furthermore, the Normal approximation becomes very poor, there.

The lead times were fixed in 25 and 5 days for the external supplies and for the internal replenishments, respectively. The system external lead is then their sum: 'L=30'.

The final lump delivery was set as 'YOf=100', i.e., 2 days of system demand. The clearing level is taken as twice that quantity.

The internal reorder ratio was set at 2.5 times the standard deviation of the satellite demand during the internal lead time.

Capacity constraints have been considered for the satellites. Each delivery 'Q' for the satellite cannot exceed 1.5 times the final allocation. The final allocation is the internal reorder level plus 1/5 of 'YOf'.

The loss factor was set at 'a=0'. This means that demand

-5.8.1-

during the stockouts is entirely backlogged.

The simulation runs lasted 5000 days. An initial period of 200 days was allowed to stabilise the system, before recording the perfomance variables. Three seeds, 13, 531 and 5531 were used.

Service expectations were computed based on the depletion time approach formulas in (5.79). The limit cases mentioned relate to the expectations for the "no stock at MWH" and "all stock at MWH" situations.

The perfomance parameters recorded along the simulation give the number of intermediate and external (final) cycles for each satellite; and the intermediate and final cycles stockouts and shortages. Intermediate and final cycles and services have been defined in section 5.2, fig 5.1. The average stocks are also computed.

The stream seed mentioned above originates a set of seeds, one for each satellite, from which demand streams are generated. A check is made on the characteristics of the demand streams to see how they conform with values as set for the daily demand mean and modulus.

The difference from the set to the actual demand characteristics causes a distortion between the expected values and the simulation results insofar as the former are

derived from the set demand and the latter from the actual stream. In order to eliminate such a distortion, the results of simulation were plotted on figs. 5.5 and 5.6 against the 'equivalent R/D'.

The 'equivalent R/D' is the value which, with the demand parameters as set, would give the same expected services as the 'set R/D' with the actual demand parameters. Denote by 'ro' the 'R/D' as set and by 'r' the equivalent ratio. Denote by 'Do' and 'Go' the mean and the modulus for the lead time demand as set; and by 'D' and 'G' the corresponding values out of the simulation stream. If we accept the Normal approximation for the demand distributions, an equivalent service would be obtained in both cases if the standard Normal variable had the same value. The standard Normal value for 'ro' with the actual demand parameters is

$$x = (ro*Do-D)/s \qquad\qquad (5.175a)$$
$$= (ro*Do/D-1)*\sqrt{G} \qquad (5.175b)$$

The standard Normal value for the equivalent 'r' with the demand parameters as set is

$$x = (r*Do-Do)/so \qquad\qquad (5.177a)$$
$$= (r-1)*\sqrt{Go} \qquad\qquad (5.177b)$$

Hence

$$r = (ro*Do/D-1)*\sqrt{G/Go} +1 \qquad (5.179)$$

5.9 Conclusion

The goal set up for this chapter has been the extension of the Johnston approach to inventory situations with two hierarchical levels. Earlier work on this area is not only difficult to implement but also it is based on assumptions which do not fit the rationale followed in the preceding chapters. These facts prevented the J-model from being extended to the 2-level systems by using results already established.

The complexity of the multilevel systems makes their analytical treatment extremely heavy unless one works on a model tailored for a specific situation. This has been the way followed here: the system was modelled having in mind the Organisation which had already provided the framework for the analysis carried out along the previous chapters.

In the case referred to, system replenishments come to the warehouse and external demand appears at the satellites. The Johnston model is used for the main warehouse whose control is independent and ignores the actual stocks in the remaining system: an aggregate reorder level is fixed for the main warehouse based on the system demand and aiming at a high level of service. Each satellite, in turn, is controlled independently, actually on a (r,R,T) control, i.e., the cyclical reorder level.

The system for analysis was restricted to the inventory domain only, freezing deliberately all other decisions concerning physical distribution. The inventory system was decoupled from the wider distribution system, primarily, for sake of simplicity; but also on the grounds that inventory control, essentially, involves decisions for short periods while the management of the other areas of the distribution system, as number and location of depots, transportation facilities, etc. has a more distant horizon.

The system modelled comprises one main warehouse and its satellites, all controlled centrally. The flows of materials are from the external source to the main warehouse, from here to the satellites and, finally, to the external customers. No transhipments or returns are considered. Holding cost rates are assumed to be the same for the main warehouse and for the satellites. Therefore, the justification for retaining stock in the main warehouse is, first, to provide a better global service to customers for the same investment in stock; and, second, to store product that exceeds the capacity in satellites. Other assumptions and more details were given at the end of section 5.2.

Major decisions in the control of multilevel systems relate to the allocation quantities. That is, if stock exists in the main warehouse and a shipment is required for a satellite, how much should be sent and how much should be retained. The allocation should take into account that other satellites are also likely to need more supplies and that

keeping stock at the main warehouse, for later redistribution, may provide a better service.

The allocation rules discussed in sections 5.3 and 5.4 can be stated as follows (assuming that services are equally important for all satellites):

"Each time an allocation is made to a satellite,

it should be made to all of them, so that they

should have the same probability of depletion by

the time the next external replenishment is available.

Furthermore, that probability of depletion should be

the same for the successive allocations in a cycle".

This rule has led to the formulas (5.49) to compute allocation quantities.

The probability of depletion required for the allocation can be computed from the quantities which, in prospect, would be involved in the last allocation of each cycle. By applying the rule of constant probability, the allocations decrease in quantity towards the end of the cycle. There is, normally, a practical lower limit for a shipment. This limit together with the reorder levels of the satellites fix the minimum for the final allocation in a cycle. The probability of stockout can be calculated for that allocation; and that probability can be used as the depletion probability for the successive allocations (which should be the same for all).

The ability to estimate analitically the stockout and

shortage rates is fundamental to the Johnston approach. In the case of multilevel systems, the evaluation of those rates from the demand distribution, as it is done for the single level inventories, does not work: so, an alternative way was used based on the depletion time distribution, i.e., the distribution of the time to deplete a quantity 'R' of an item. Referring to (5.71)

$$P= \int_0^L q(t)*dt \qquad\qquad (5.183a)$$

$$Zt= \int_0^L (L-t)*q(t)*dt \qquad\qquad (5.183b)$$

where 'P' and 'Zt' are the probability of stockout and the expected time out of stock in a cycle, respectively, 'q(t)' the density for the depletion time distribution and 'L' is the lead time. The shortage rate is then 'V=Zt/T'.

Aditionally, reasons were found to support the Normal distribution as a reasonable approximation for the depletion time, under the circumstances. Furthermore, expressions for the mean and the variance of the depletion time were proposed, so that those parameters can be approximated from the mean and the variance of the demand. Referring to (5.15)

$$tm= R/d \qquad\qquad (5.185a)$$
$$tv= tm/g \qquad\qquad (5.185b)$$

where 'tm' and 'tv' are the mean and the variance of the depletion time distribution, 'R' the reorder level and

'g=(1/dc)**2' with 'dc' standing for the coefficient of variation of the daily demand ('g' is the modulus if demand is Gamma distributed). Details are given in section 5.5.

The setting of the internal reorder levels is discussed in section 5.7. The internal reorder level relates to the probability of stockout in intermediate cycles, 'PI', which could be calculated from (5.171b). It is suggested that the 'PI's are calculated periodically for a sample of items and then related to the final or system probability of stockout 'P' such as in (5.135):

$$PI = c*P \qquad (5.187)$$

This latter relationship could then be used in the routines for the decision procedure with the advantage of alleviating the calculations. Another problem discussed in the same section is the use of two internal reorder levels: a higher level when there is plenty of stock in the system and a lower level at the end, when the stock becomes scarce.

The approximations introduced in the course of these derivations were submitted to a simulation check. For this purpose, values for the stockout and shortage rates obtained by simulation were compared with the analytical predictions. The accuracy seems to be reasonable for the range of high services which have been implied. Details of these checks are given in section 5.8.

The extension of the Johnston approach to multilevel systems is discussed in section 5.6. In the case of 'twin' satellites, i.e., when the satellites are balanced in respect to commercial importance and demand characteristics, the expressions for the service rates can be simplified considerably. In such circumstances, the control would be almost as simple as for the single level case. Furthermore, all the results achieved in the previous chapters apply. The non-twin case raises difficult analytical problems, and because of that, it was not contemplated in this thesis.

Chapter SIX

CONCLUSION

6.1   Summary

The entire analysis presented in this thesis stems from the Johnston approach to multi-item inventory control. Johnston devised a decision aid system which does not require, necessarily, individual decisions for each item and awkward estimates for holding and shortage penalties. The choice of the control parameters is based upon a vector of performance variables which reflect a trade-off between service to customers and capital invested in stock.

The distribution of the capital invested between the items is made according to the criterion of maximum profit explained in section 2.4. This allocation is embedded in the decision aid system and therefore does not require the attention from the user, except perhaps for some key products. Thus, the decision maker can concentrate mainly on the broad figures for families of items and for the whole inventory.

These figures are generated by setting values for the so called "stock factor" which is used as a tuning knob. Section 2.5 extends the meaning of the stock factor and stresses that it should be compared with the current borrowing interest rate. There is another 'factor' implied in the Johnston model which eventually could constitute a second knob: the relation of the shortage cost to the profit of the item. The implications of such a relationship is

dealt with in section 2.6

The Johnston model throws a new light on to the decision making process in inventory control. Nevertheless, some important shortcomings exist, several of which are discussed in this work.

The first relates to the interaction or feed-back effect of lost sales on the perfomance variables. The perfomance measures are estimated according to formulas which have been derived as if demand were captive. But, on the other hand, the criterion to distribute the investment by the items is strongly dependent on the loss of sales (i.e., on demand being non-captive) during shortages. That loss of sales, actually, affect the perfomance variables, an aspect which was ignored by Johnston.

The sales loss effect had been recognised previosly but had not been quantified. In chapter 3 approximate expressions to relate the expected shortage quantities under captive and non-captive demand were derived. Further, the new concept of notional reorder level was introduced so that the original Johnston approach could be extended to non-captive situations. Values obtained by simulation, as those shown in figs.3.10 to 3.12, are close to the predictions, especially for the high service levels. Therefore, the approximations introduced seem to be acceptable, and the corrections are recommended.

The effect of the lost sales on services and on average stock was shown to exist only if the decision periods overlapped. Under the reorder level control, that period coincides with the procurement lead time, therefore, lost sales effects appear only when more than one order is outstanding at the same time. Under a periodical review control, as the relevant decision period is the lead time plus the review period, the lost sales always influence the services.

Another shortcoming in the Johnston model relates to the reorder frequency. This is taken as a exogenous variable to be set up on an empirical basis. As indicated in chapter 4, the choice of a frequency consistent with the objectives assumed by Johnston is not straight forward. Increasing the reorder frequency decreases the average stock 'S' for the same stockout rate 'P'; but it increases simultaneously the number of times in a year that a stockout may occur, and hence, the shortage rate 'V'. The interaction is such that to keep 'V' fixed while increasing the reordering frequency, more stock may be required, as shown in fig 4.1a. In such a case, reordering more often brings no benefit either for the investment or for the shortages, and of course, it increases the procurement costs.

The situation described above, in which a poorer performance vector (S,V) is achieved by reordering more often, relates to demands with large variances or to cases of already high

reordering frequency. Usually, an increase in the frequency can lead, after a convenient re-adjustment of the reorder level, to lower stocks, better shortage rates or some trade-off between those two. This is shown in fig. 4.1b. But in any case, the marginal benefits upon 'S' and 'V' of reordering more often decrease steeply. Therefore, high overlapping degrees, with many orders outstanding at the same time, is unlikely to be a good practice from the point of view being discussed.

It is not satisfactory to leave the choice of a reorder frequency to good sense alone, due to the ambiguities just mentioned. Section 4.4 gives the expressions for obtaining the coherent values for that frequency. Basically, those expressions result from considering the number of orders per year as an independent variable in the maximisation of the profit as expressed by Johnston. A limit is imposed either on the global number of orders for the whole system or on the marginal cost of ordering. Then, the frequency for each 'buying family' (i.e., the items with a common supplier) can be determined to meet those limits.

The treatment of the reorder frequency includes a new tuning knob and a new variable in the perfomance vector. Deciding on the most preferred vector may not be easy, for the preference relationships are not structured and often are dependent upon time and place. Decisions could be helped if the decision maker were provided with the means for cross-checking the consistency of his own preferences.

The sketch for a possible cross-checking system was described in section 4.5. It is based on the fact that each performance variable can be associated with some cost rate: stock investment relates to a holding cost; service to a shortage cost; and order frequency to a procurement cost. Therefore, a cost or, say, a 'shadow price' could be easily calculated for each performance vector. The more preferred the vector the higher its shadow price should be. The shadow prices may, not only provide a test for the preference relationships, but also help the quantification in money units of the preference differences.

The Johnston model is directed towards single level inventories, however, it is being used to control a system comprising a main warehouse that supplies a number of satellites. This fact motivated the analysis in chapter 5 which extends that model to a 2-level system shaped on the real situation.

Criteria borrowed from the Johnston approach and applied to that specific system have produced the rules and the mathematical expressions to govern the successive allocations from the main warehouse to the satellites and amongst the latter. The basic rule is discussed in sections 5.3 and 5.4 and states that allocations amongst satellites at one time as well as successive allocations at different time moments should have the same probability of depletion.

In section 5.5, formulas to compute the stockout and the shortage rates have been derived on the assumption of those rules being in force. These results have been achieved through the depletion time distribution, an approach not entirely new but rarely used for that purpose.

Some approximations were introduced in the course of the derivations in order to achieve implementable expressions. The magnitude of the errors so introduced in the calculation of service rates was checked against results obtained by simulation. Three sets of results are depicted in figs. 5.5 and 5.6, section 5.8. The predictions look reasonably good within the range of services usually adopted.

The ability to estimate the stockout and the shortage rates paved the way to extend the J-model and the results in previous chapters to the 2-level situations. In section 5.6, formulas have been derived for the case of twin satellites, i.e. satellites having comparable demand distributions and commercial importance. The distortions which might be incurred by using these same formulas for unbalanced configurations is a matter that would require further investigation.

6.2   Achievements and shortcomings

The major achievements reported in this thesis relate to the extensions made to the model proposed by Johnston. Such extensions are depicted in fig 6.1
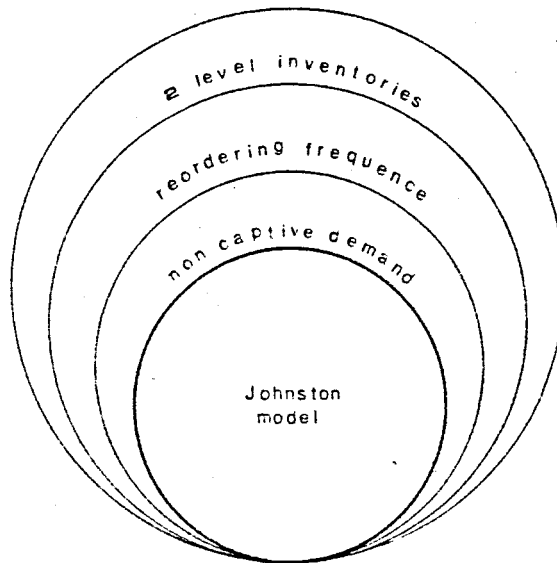
Fig. 6.1 - Extensions to the Johnston model

The first extension, discussed in chapter 3, contemplates situations in which more than one replenishment may be outstanding and demand during shortages may represent a loss in sales. The analysis led to the concept of the notional reorder level which enabled the use of established formulas to relate control parameters to performance variables, and hence, the use of the Johnston approach. An approximate expression for the shortage rate was also derived.

The second extension, covered in chapter 4, gives an analytical process to calculate, for each group of items,

the reorder frequency consistent with the services. The best combination of reorder levels and reorder frequencies can now be looked for, analytically.

Finally, chapter 5 extends the model to centrally controlled inventory systems comprising a main warehouse and its satellites. The analysis involved the establishment of allocation criteria coherent with the general approach, and the derivation of expressions enabling the estimate of service variables from the control parameters. The potentialities of the depletion time distribution have been explored for this purpose.

Other results, though less important, may deserve a mention. They relate to the attempt in providing analytical tools to help the decision maker to be consistent in his choices. Namely, in section 2.5, by exploring the economic meaning of the Lagrangean parameters; in section 2.6, by discussing possible relationships between shortage costs and profits; and, in section 4.5, by introducing the shadow prices for the performance variables.

There are, nevertheless, weak points in the analysis which should be referred to. First, in relation to the assumptions. Most of them are commonly accepted, e.g. the ability to know the demand distribution, and the independence of demands in successive periods and on different outlets. Some others, however, are less common,

such as, the restrictions made to fast moving items with Gamma demands and to inventory policies with high service levels.

Assumptions on operating policies limit the replenishments to the reorder level control. Overshooting problems due to lumpy demands and to stock reviews being discrete, were not examined. Assumptions on the structure of the 2-level system and on the internal flows of goods restrict the applicability of the results to relatively simple multilevel situations.

Another weakness relates to the need for approximations in the course of the analysis. To start with, a few expressions have been simplified on the assumption that some terms involving stockouts and shortage rates could be neglected because services were high. Further, the expected value of a variable, instead of the variable itself, has been used in a few instances to undo analytical ties. Furthermore, maldistributions of stock when replenishment orders are raised have been disregarded. However, the magnitude of the errors so introduced are not large if services are high, as simulation results have shown. The same appears to be true in relation to the errors caused by the Normal approximation for the depletion time distribution, and by the simplified formulas for the mean and the variance of the depletion time, when the depletion time approach is used to estimate services in a 2-level inventory system.

The analysis of the specific 2-level inventory led to usable results in situations of twin satellites, only. This will be the general case, for systems with satellites of quite different order of magnitude are rare. In this latter case, how satisfactory the results would be is unknown.

Finally, the analysis isolated the inventory from the whole distribution system. It is well known that tactical decisions which ignore the strategy are liable to be sub-optimal. Nevertheless, integration has not been considered, even for the internal transportation procedures and costs.

In brief, the extensive analysis presented in this thesis has reached results which are neither exact nor general. The approximations, however, seem to be reasonable, compared with the uncertainty about some of the exogenous variables and the intangibility of most of the decision parameters. The benefit of dealing with a specific system rather than a general one was to have achieved results almost ready for implementation. There may be improvements to make and further areas to cover; but the analysis worked through has enlarged considerably the potentialities of the initial model.

The implementation, which is outside the scope of this work, will prove, hopefully, the merits of those achievements, and will suggest, surely, many other refinements that could be made.

6.3   Leads for further research

The analysis in earlier chapters had the implementation in mind, but results may have to be reformulated to fit the requirements of the specific inventory system, of the specific users and of the specific information system. The implementation stage and the practice later on will give, surely, many hints for further analysis. There are, however, two points which can already be admitted: the effects on the real systems of the mal-distribution of stocks and of the twin satellite approximation.

Mal-distributions, in the present context, refer to the imbalances in stocks due to the differences from actual to expected values. In section 4.2, a comment was made about the fact that, when an order for a buying family is raised, the family is at its aggregate reorder level but the stock for each individual item is unlikely to be at the respective reorder level. Such a fact brings a decrease in control effectiveness which leads to a poorer performance: for instance, for a given service, the required global investment in stock is higher than if items were reordered individually.

That increment in stock represents a cost which should be off-set by the savings in procurement costs derived from family reordering. Each extra item added to the family represents, potentially, a marginal cost in effectiveness

and a marginal saving in procurement. If the latter is always higher than the former, enlarging the buying family would be always favourable, from the present perspective; but if not, the point where both margins equate should constitute a limit for the number of items to be included in the family. Currently, this number is fixed on a practical basis.

Mal-distribution effects may become amplified in multilevel inventories. Here, imbalances exist not only among the items of a family, but also among the satellites, for a same item. In the specific model considered, mal-distribution effects appear also at the level of the last partition, as mentioned in section 5.8. Nevertheless, for the situations simulated, central control gave considerable better performance than individual control of the satellites. This matter, however, would deserve further consideration.

There is another source of mal-distributions resulting from bad demand forecasts. Quoting Lampkin (1967, p64): "(...) the oversestimate of demand cause overstocking which persists for a long time, while underestimates cause understocking which only persists for a short time. The result is that the average stock holding is considerable greater than one would expect from the average demand rate. Correspondingly ordering rates are, in practice, lower than estimates suggest". That author reports also that having average stocks 50% higher for some goods is not uncommon. That being so, the overal effect on the inventory may be

quite important and should be investigated.

The analysis on 2-level inventories used the twin satellite simplification to get through analytical difficulties and reach tractable results. The use of those same formulas for situations where that assumption does not hold would bring errors on the estimates of the performance variables and lead to sub-optimal use of the resources. The magnitude of these effects should be investigated, eventually, on live situations and data.

There are logical extensions to the proposed model whose interest only implementation will reveal. For instance, the control of the slow movers. Often these products are associated with high services: essential spare parts, prestige products, etc. The profit criterion used earlier, may not be the most appropriate for those items. That being so, the basic approach would require to be reviewed, but some of the approximations might still stand. Low service items, on the other side, do not require a sophisticated control.

Other extensions might be towards more complex multilevel structures or towards further integration in a wider distribution system. In order that such attempts have success in reaching usable results, mathematical models may have to be shaped on particular situations, otherwise, they are likely to clash with analytical difficulties.

Internal transportation and picking costs were not quantified in the analysis of the 2-level inventory system. It was shown in section 5.3.1 how the number of internal replenishments 'n' could be related to the allocation parameter 'p', if there were no capacity restrictions for the satellites. Delivery costs are likely to increase with the number of internal orders, but the strength of such a dependence can vary widely from one situation to another. At one extreme, if a delivery trip were made for each order, delivery costs would be roughly proportional to the number of orders. At the other extreme, one can imagine that with systems involving thousands of items and several satellites, full advantage of vehicle capacity can be obtained by rationalising cargoes and trips. Then, delivery costs would depend essentially on the global throughput and not so much on how that throughput is actually ordered; thus, the number of internal orders would not be a preponderant factor for those costs.

Restrictions on satellite capacities are likely to exist and to be binding for the distribution shipment immediately following external replenishment and this might increase substantially the number of internal deliveries. It is a guess that in most of real situations capacity restrictions will be dominant in fixing the number of internal shipments. The analysis of the internal delivery costs might then require that satellite capacities were also considered.

The cost of internal delivery and its relation to the average number of deliveries per cycle, to the allocation parameter, to capacity constraints in satellites, and to the other logistic facilities might be relevant for the decision process. However, the analysis of this point to go beyond the theoretical generality and to reach practical results should be geared to the particular situation.

The estimation of the demand characteristics is an ever present problem in inventory control models. In this thesis, the knowledge of the demand has been taken for granted. Furthermore and in the wake of Burgin and Johnston, demand was assumed to follow a Gamma distribution.

There is a large amount of literature on forecasting demands for inventory control and on appropriate probabilistic distributions for those demands. Section 5.5 discussed the use of the depletion time approach for control purposes, namely, to estimate the services for each reorder level. The shape of this distribution and the legitimacy of the assumption made should be investigated further. The analysis of the virtualities of this approach as a usable alternative for demand distribution might be rewarding.

Hypothetically, there are many other areas which could be thought of. Enumerating them would be as much fastidious as irrelevant. Real issues will arise after the implementation of the results arrived at.

END

I need no gravestone, but

If you need one for me

I would like it to bear these words:

He made suggestions. We carried them out.

Such an inscription would

Honour us all.

(From Bertolt Brecht Poems, p.218

Ed. John Willet and Ralph Manheim

London, Eyre Methuen Ltd. 1976)

REFERENCES AND BIBLIOGRPHY

ABRAMOWWITZ, M.; STEGUN, I.A. (1964): Handbook of
Mathematical Function. National Bureau of Standards

ADELMAN, L. (1982): Involving Users in the Development of
Decision Analytic Aids: The Principal Factor in Successful
Implementation. J.Opl.Res.Soc., vol.33,4 pp333-342.

ADELSON, R.M. (1966): Compound Poisson Distribution.
Opl.Res.Quart., vol.17,1 pp73-75.

AGGARWAL, S.C.; DHAVALE, D.G. (1975): An Empirical Study
of an Inventory Distribution System. Omega, vol.3,2
pp203-211.

AITCHINSON, J.; BROWN, J.A.C. (1957): The Lognormal
Distribution (Monograph no5). Cambridge University Press.

ALLEN, S.G. (1958): Redistribution of Total Stock over
Several User Locations. Naval Res. Logistics Quart., vol.5,4
pp337-345.

ARCHIBALD, B.C.; SILVER, E.A. (1978): (s,S) Policies under
Continuous Review and Discrete Compound Poisson Demand.
Manag.Science, vol.24,9 pp899-909.

ARROW, K.J.; KARLIN, S.; SCARF. H. (1962) Editors: Studies
in Applied Probability and Management Science. Standford
University Press, California.

ARROW, J.K. (1963): Historical Background. In Arrow et al
(1963a), Studies in the Mathematical Theory of Inventory and
Production, Ch.1 pp3.15.

ARROW, K.J.; KARLIN, S.; SCARF, H. (1963): The Nature and

structure of Inventory Problems. In Arrow et al (1963a),
Studies in the Mathematical Theory of Inventory and
Production, Ch.2 pp16-36.

. . . .(1963a): Studies in the
Mathematical Theory of Inventory and Production. Standford
University Press, Standford, Calofornia.

BARBOSA, L.C.; FRIEDMAN, M. (1978): Deterministic
Inventory Lot Size Models - a General Root Law.
Manag.Science, vol.24,8 pp819-826.

BARRETT, D.A. (1969): Automatic Inventory Control
Techniques. Business Books Ltd, London, 1969.

BARTAKKE, M.N. (1981): A Method of Spare Parts Inventory
Planning. Omega, vol.9,1 pp51-58.

BECKMAN, M.J. (1964): Dynamic Programming and Inventory
Control. Oper. Res. Quart., vol.15,4 pp389-400.

BELL, P.C. (1980): A Decoupling Inventory Problem with
Storage Capacity Constraints. Opers. Res., vol.28,3, part 1
pp476-488.

. . . . . .(1981): Adaptive Sales Forecasting with Many
Stockouts. J.Opl.Res.Soc., vol.32,10 pp865-873.

BESSLER, S.A.; VEINOTT, A.F. (1966): Optimal Policy for a
Dynamic Multi-echelon Inventory Model. Naval Res. Logistics
Quart., vol.13, pp355-389.

BOTHROYD, H.; TOMLINSON, R.C. (1963): The Stock Control of
Enginneering Spares - A Case Study. Opl. Res. Quart.,
vol.14,3 pp317-332.

BRAMSON, M.J. (1962): The Variable Lead Time Problem in
Inventory Control, Part 1. Opl.Res.Quart., vol.13,1 pp41-53.

BROWN, R.G. (1959): Statistical Forecasting for Inventory

Control. McGraw Hill.

. . . . . . (1962): Smoothing, Forecasting and Prediction
of Discrete Time Series. Prentice Hall.

. . . . . . (1963): Estimate Aggregate Inventory
Standards. Naval Res. Logistics Quart., vol.10,1 pp55-71.

. . . . . . (1967): Decision Rules for Inventory
Management. Holt, Rinehart and Winston. New York, London.

. . . . . . (1977): Materials Management Systems: A
Modular Library. Wiley. New York.

BRYAN, J.G.; WADSWORTH, G.P.; WHITIN, T.M. (1955): A
Multi-stage Inventory Problem. Naval Res. Logistics Quart.,
vol.2,1-2 pp25-37.

BUCHAN, J.; KOENIGSBERG, E. (1963): Scientific Inventory
Management. Prentice Hall.

BURGIN, T.A.; WILD, A.R. (1967): Stock Control -
Experience and Usable Theory. Opl.Res.Quart., vol.18,1
pp35-52.

BURGIN, T.A. (1969): Time to Sell a Fixed Quantity of
Stock Depleted by Distributed Demand. Opl.Res.Quart.,
vol.20,4 pp421-428. See Comments (1970).

. . . .(1970): Back Ordering in Inventory Control.
Opl.Res. Quart., vol.21,4 pp453-461.

. . . . . . (1972): Inventory Control with Normal Demand
and Gamma Lead Times. Opl.Res.Quart., vol.23,1 pp73-80.

. . . . . . (1975): The Gamma Distribution and Inventory
Control. Opl.Res.Quart., vol.26,3 pp507-525.

BURGIN, T.A.; NORMAN, J.M. (1976): A Table for Determining
the Probability of a Stock Out and Potential Lost Sales for
a Gamma Distributed Demand. Opl.Res.Quart., vol.27,3

pp621-631.

BUZACOTT, J.A. (1975): Economic Order Quantities with Inflation. Opl.Res.Quart., vol.26,3 pp553-558.

CAWDERY, M.N. (1976): The Lead Time and Inventory Control Problem. Opl.Res.Quart., vol27 pp971-989.

CHAKRAVARTY, A.K. (1981): Multi-item Inventory Aggregation into Groups. J.Opl.Res.Soc., vol.32,1 pp19-26.

CHAMBERS, M.L.; JARRATT, P. (1965): Controlling Rota Scheduled Production by Adjusting Stock Requirements. Opl.Res.Quart. vol.16,3 pp367-376.

CHENTNIK, C.G. (1969?): Physical Distribution Management and its Environment. Intern.J.Phys.Distrib. vol.5,1 pp62-73.

CLARK, A.J.; SCARF, H. (1960): Optimal Policies for Multi-Echelon Inventory Problem. Manag.Science, vol.6,4 pp475-490.

. . . .(1962): Approximate Solution to a Simple Multi-Echelon Inventory Problem. In Arrow et al (1962), Studies in Applied Probability and Management Science, Ch.5.

CLARK, A.J. (1972): An Informal Survey of Multi-Echelon Inventory Theory. Naval Res. Logistics Quart., vol.19,4 pp621-650.

CLARK, C.E. (1957): Mathematical Analysis of an Inventory Case. Ops.Res., vol.5,5 pp627-643.

CLEROUX, R.; McCONALOGUE, D.J. (1976): A Numerical Algorithm for Recursively Defined Convolution Integrals Involving Distribution Function. Manag.Science, vol.22,10 pp1138-1146.

CLINTON, J. (1969): Inventory Control in a Multi-Warehouse

System. Unpublished Ph.D. thesis submitted to the University of London.

COMMENTS (1970) to Burgin (1969). Opl.Res.Quart., vol.27,1 pp385-386.

COX, D.R.; LEWIS, P.A. (1966): The Statistical Analysis of Series of Events. Methuen and Co. London, 1966.

CRAN, J.A. (1966): A Two Level Inventory Control in a Distribution System. Proceedings of 4th Int.Conf. in O.R., Boston 1966, pp864-873. Edited by D.B.Hertz and J.Malese. Published by J.Wiley-Interscience.

CROSTON, J.D. (1972): Forecasting and Stock Control for Intermittent Demands. Opl.Res.Quart., vol.23,3 pp289-304.

. . . . (1974): Stock Levels for Slow Moving Items. Opl.Res.Quart., vol.25,1 pp123-130.

CROUCH, R.B.; OGLESBY, S. (1978): Optimization of a few Lot Sizes to Cover a Range of Requirements. J.Opl.Res.Soc. vol.29,9 pp897-904.

CROWSTON, W.B.; WAGNER, M.; WILLIAMS, J.F. (1973): Economic Lot Size Determination in Multi-Stage Assembly Systems. Manag.Science vol.19,5 pp517-527.

CROWSTON, W.B.; WAGNER, M.H. (1973): Dynamic Lot Size Models for Multi-StageAssembly Models. Manag.Science, vol.20,1 pp14-21.

CURLEY, M.J. (1978): Service Level and Averagg Stockholding in a Re-Order Level System. J.Opl.Res.Soc., vol.29,8 pp803-805.

DAS, C. (1976): Approximate Solution to the (Q,r) Inventory Model for Gamma Lead Time Demand. Manag.Science, vol.22,9 pp1043-1047.

DAVE, V. (1979): On a Discrete-in-Time Order Level
Inventory Model for Deteriorating Items. J.Opl.Res.Soc.,
vol.30,4 pp349-354.

DONALDSON, W.A. (1974): The Allocation of Inventory Items
to Lot Size / Reorder Level (Q,r) and Periodic Review (T,Z)
Control Systems. Opl.Res.Quart., vol.25,3 pp481-485.

. . . . (1975): Inventory Replenishment Policy for
a Linear Trend in Demand - An Analytical Solution.
Opl.Res.Quart., vol.28,3 pp663-670.

. . . . (1981): Grouping of Inventory Items by
Review Period. J.Opl.Res.Soc., vol.32,12 pp1075-1076.

EILON, S. (1962): Elements of Production Planning and
Control. McMillan Co., New York.

. . . . . (1964): Dragons in Pursuit of the EBQ.
Opl.Res.Quart., vol.15.4 pp347-354.

EILON, S.; ELMALEH, J. (1968): An Evaluation of
Alternative Inventory Control Policies. Int.J. Production
Res., vol.7,1 pp3-14.

. . . .(1970): Adaptive Limits in
Inventory Control. Manag.Science, vol.16B,8 pp533-548.

EILON, S.; WATSON-GANDY, C.D.T.; CHRISTOFIDES, N. (1971).
Distribution Management: Mathematical Modelling and
Practical Analysis. Charles Griffin and Co, London.

EILON, S. (1972): Goals and Constraints in Decision
Making. Opl.Res.Quart., vol.23,1 pp3-15.

. . . . . (1979): Management Control. Pergamon Press, U.k.

FENSKE, R.W. (1968): Non-Stocking Criterion.
Manag.Science, vol.14B,12 pp705-714.

FILDES, R. (1979): Quantitative Forecasting - The State of

the Art: Extrapolative Models. J.Opl.Res.Soc., vol.30,8 pp691-710.

FORTUIN, L. (1980): Five Popular Probability Density Functions: a Comparison in the Field of Stock Control Models. J.Opl.Res.Soc., vol.31,10 pp937-942.

. . . . . . (1981): A Comparison of SIC and MRP, Two Methods for Material Procurement in Industry.

GALLIHER, H.P.; MORSE, P.M.; SIMOND, M. (1959): Dynamics of Two Classes of Continuous Review Inventory Systems. Opers. Res., vol.7, pp362-384.

GELDERS, L.F.; LOOY, P.M. (1978): An Inventory Policy for Slow and Fast Movers in a Petrochemical Plant: A Case Study. J.Opl.Res.Soc., vol. 29,9 pp868-874.

GLUSS, B. (1960): Costs of Incorrect Data in Optimal Inventory Computation. Manag.Science, vol.6,4 pp491-497.

GOYAL, S.K. (1974): Optimum Ordering Policy for a Multi-Item Single Supplier System. Opl.Res.Quart., vol.25,2 pp293-298.

. . . . . . (1977): An Integrated Inventory Model for a Single product System. Opl.Res.Quart., vol.28,3 pp539-545.

. . . . . . (1978): A note on "Multi-Product Inventory Situations with One Restriction". J.Opl.Res.Soc., vol.29,3 pp269-271.

GRIFFITHS, B. (1961): A Production Planning and Stock Control Problem in the Textile Industry. Opl.Res.Quart., vol.12,4 pp244-260.

GROSS, D. (1963): Centralized Inventory Control in Multilocation Supply Systems. In Scarf et al (1963), Multi-stage Inventory Models and Techniques, Ch.3.

HADLEY, G.; WITHIN, T.M. (1961): A Model for Procurement Allocation and Redistribution for Low Demand Items. Naval Res. Logistics Quart., vol.8 pp395-414.

. . . .(1963): Analysis of Inventory Systems. Prentice Hall, Englewood Cliffs, N.J.

. . . . (1963a): An Inventory Transportation Model with N-Locations. In Scarf et al (1963), Multistage Inventory Models and Techniques, Ch.5 pp116-142.

HANSSMANN, F. (1962): Operations Research in Production and Inventory Control. John Wiley, New York.

HENERY, R.J. (1979): Inventory Replenishment Policy for Increasing Demand. J.Opl.Res.Soc., vol.30,7 pp611-617.

HOADLEY, B.; HEYMAN, D.P. (1977): A Two Echelon Inventory Model with Purchases, Dispositions, Shipments, Returns and Transhipments. Naval Res.Logistics Quart., vol.24,1 pp1-19.

HOCHSTAEDTER, D. (1970): An Approximation of the Cost Function for Multi-Echelon Inventory Model. Manag.Science, vol.16 pp716-727.

HOLLIER, R.H.; VRAT, P. (1976): A Review Of Multi-Echelon Inventory Control Research and Applications. Technical Report, Dept. Production Engineering, Univ.Birmingham.

HURTER, A.P.; KAMINSKY, F.C. (1968): Inventory Control with a Randomly Available Discount Purchase Price. Opl.Res.Quart., vol.19,4 pp433-444.

IGLEHART, D.L.; MOREY, R.C. (1971): Optimal Policies for a Multi-Echelon System with Demand Forecasts. Naval Res. Logistics Quart., vol.18 pp115-118.

IGNALL, E.; VEINOTT, A.F. (1969): Optimality of Myopic

Policies for Several Substitute Products. Manag.Science, vol.15A. pp284-304.

JOHNSTON, F.R.; MINLE, D.J. (1972): A Comparison of Two Stock Control Models. Opl.Res Quart., vol.23,4 pp573-577.

JOHNSTON, F.R. (1974): The Calculation of Order Quantities and Protection Levels in a Re-Order Level Stock Control System. Opl.Res.Quart., vol.25,1 pp169-171.

. . . .(1975): Forecasting in a Real Time Stock Control System. Working Paper no.24. SIBS, University of Warwick.

. . . .(1980): An Interactive Stock Control System with a Strategic Management Role. J.Opl.Res.Soc., vol 31,12 pp1069-1084.

KEEN, P.G.W.; MORTON, M.S.S. (1978): Decision Support Systems: An Organizational Perspective. Addison-Wesley Publishing Co.

KEENEY, R.L.; RAIFFA, H. (1976): Decisions with Multiple Objectives: Preferences and Value Trade-offs. John Wiley and Sons.

KHUMAWALA, B.M.; NEEBE, A.W.(1978): A Note on Warszawski's Multi-Commodity Location Problem. J.Opl.Res.Soc., vol.29 pp171-172.

KICKS, P.; DONALDSON, W.A. (1980): Irregular Demand: Assessing a Rough and Ready Lot Size Formula. J.Opl.Res.Soc., vol.31,8 pp725-732.

KOENIGSBERG, E. (1961): A Multiple Re-Order Point Inventory Policy. Opl.Res.Quart., vol.12 pp27-40.

KOTTAS, J.F.; HON-SHIANG LAU (1980): The Use of Versatile Distribution Families in Some Stochastic Inventory

Calculations. J.Opl.Res.Soc., vol.31,5 pp393-403.

KRUSE, W.K. (1980): Waiting Time in a (S-1,S) Inventory System with Arbitrarily Distributed Lead Times. Opers.Res., vol.28,2 pp348-352.

KUMAR, U.; VRAT, P. (1979): Optimization of Single-Item Multi-Stage Stochastic Production-Inventory Systems. J.Opl.Res.Soc., vol.30,6 pp547-553.

KUMARASWAMY, S.; SANKARASUBRAMANIAN, E. (1981): A Continuous Review of (S-s) Inventory Systems in which Depletion Is Due to Demand and Failure of Units.

LAMPKIN, W.; FLOWERDEW, A.D.J. (1961): On a Multiple Reorder Point Inventory Policy. Opl.Res.Quart., vol.12,3 pp187-190.

. . . . (1963): Computation of Optimum Reorder Levels and Quantities for a Reorder Level Stock Level Stock Control System. Opl.Res.Quart., vol.14,3, pp263-278.

LAMPKIN, W. (1966): The Expected Size of an Order in a (s,S,T) Policy with Poisson Demand. Opl.Res.Quart., vol.17,1 pp91-93.

. . . . . . (1967): A Review of Inventory Control Theory. Production Eng., vol.46,2 pp57-66.

LAWRENCE, J.R.; STEPHENSON, G.G.; LAMPKIN, W. (1961): A Stock Control Policy for Important Spares in a Two Level Stores System. Opl.Res.Quart., vol.12,4 pp261-271.

LEV, B.; SOYSTER, A.L. (1979): An Inventory Model with Finite Horizon and Price Changes. J.Opl.Res.Soc., vol.30,1 pp43-53.

LEWIS, C.D. (1970): Scientific Inventory Control.

Butterworth and Co.(Publishers), London.

. . . . . . (1970a): Industrial Forecasting Techniques.
The Machinery Publishing Co. Ltd, U.K.

. . . . . . (1981): Forecasting. In Lewis (1981b),
Operations Management in Practice, Ch.14 pp241-269.

. . . . . . (1981a): Inventory Control. In Lewis (1981b),
Operations Management in Practice.

. . . . . . (1981b), Editor: Operations Management in
Practice. Philip Allan Publishers.

LOCKYEAR, K. (1972): Stock Control: A Practical Approach.
Cassel, London.

LOVE, R.F. (1967): A Two Station Stochastic Inventory
Model with Exact Methods of Computing Optimal Policies.
Naval Res. Logistics Quart., vol.14 pp185-217.

LOW, R.A.; WADDINGTON, J.F. (1967): The Determination of
the Optimum Joint Replenishment Policy for a Group of
Discount Connected Stock Lines. Opl.Res.Quart., vol.18,4
pp443-462.

MAGGEE, J.F.; BOODMAN, D.M. (1967): Production Planning
and Inventory Control. McGraw Hill.

MERCER, A. (1970): Strategic Planning of Physical
Distribution Systems. Intern.J. Physical Distr., Oct.1970
pp20-25.

MISRA, R.B. (1979): A Note on Optimal Inventory Management
under Inflation. Naval Res.Logistics Quart., vol.26,1
pp161-165.

MITCHEL, G.H. (1962): Problems of Controlling Slow-Moving
Engineering Spares. Opl.Res.Quart., vol.13,1 pp23-39.

MOLE, R.H. (1979): A Survey of Local Delivery Vehicle

Routing Methodology. J.Opl.Res.Soc., vol.30,3 pp245-252.

MUCKSTADT, J.A.; THOMAS, L.J. (1980): Are Multi-Echelon Inventory Methods Worth Implementing in Items with Low Demand Rates? Manag.Science, vol.26,5 pp483-494.

MUHELEMANN, A.P.; VALTIS-SPANOPOULOS, N.P. (1980): A Variable Holding Cost Rate EOQ Model. European J.Opl.Res., vol.4 pp132-135.

MUMFORD, R.J. (1977): The Numerical Generation of Lead Time Demand Distribution for Inventory Models. Opl.Res.Quart., vol.28 pp79-85.

NADDOR, E. (1956): Some Inventory Models and an Application. Manag.Science, vol.2 pp299-312.

. . . . . .(1966): Inventory Systems. John Wiley and Sons.

. . . . . .(1975): Optimal and Heuristic Decision in Single and Multi-item Inventory Systems. Manag.Science, vol.21,11 pp1234-1249.

NAHMIAS, S. (1975): Optimal Ordering Policies for Perishable Inventory II. Opers.Res., vol.23 pp735-759.

. . . . . . (1976): Myopic Approximations for the Perishable Inventory Problem. Manag.Science, vol.22 pp1002-1008.

. . . . . .(1977): A Comparison between two Dynamic Perishable Inventory Models. Opers.Res., vol.25 pp169-172.

. . . . . . (1978): The Fixed Charge Perishable Inventory Problem. Opers.Res., vol.26 pp464-481.

ORAL, M.; SALVADOR, M.S.; REISMAN, A.; DEAN, B.V. (1972): On the Evaluation of the Shortage Costs for Inventory Control of Finished Goods. Manag.Science, vol.18B,6 pp344-351.

PAGE, E.; PAUL, R.J. (1976): Multiproduct Inventory Situation with One Restriction. Opl.Res.Quart., vol.27,4 pp815-834.

PHELPS, R.I. (1980): Optimal Inventory Rule for Linear Trend in Demand with a Constant Replenishment Period. J.Opl.Res.Soc., vol.31,5 pp439-442.

PRASTACOS, G.P. (1978): Optimal Myopic Allocation of a Product with Fixed Lifetime. J.Opl.Res.Soc., vol.29,9 pp905-913.

PSOINOS, D. (1974): On the Joint Calculation of Safety Stocks and Replenishment Order Quantities. Opl.Res.Quart., vol.25,1 pp173-177.

RAY, W.D. (1980): The Significance of Correlated Demands and Variable Lead Times for Stock Control Policies. J.Opl.Res.Soc., vol. 31,2 pp187-190.

. . . . . (1981): Computation of Reorder Levels when the Demands Are Correlated and the Lead Time Random. J.Opl.Res.Soc., vol.32,1 pp27-34.

RICHARDS, F.R. (1975): Comments on the Distribution of Inventory Position in a Continuous Review (s,S) Inventory System. Opers.Res., vol.23 pp366-371.

SCARF, H.F. (1963): A Survey of Analytical Techniques in Inventory Theory. In Scarf et al (1963), Multistage Inventory Models and Techniques, Ch.7 pp185-225.

SCARF, H.F.; GILFORD, D.M.; SHELLEY, M.W. (1963): Multistage Inventory Models and Techniques. Standford University Press, Standford, California.

SCHNEEWEISS, C. (1981): Construction and Selection of Quantitative Planning Models - A General Procedure

Illustrated with Models for Stock Control. European
J.Opl.Res., vol.6,4 pp372-374.

SCHNEIDER, H. (1978): Methods for Determining the Reorder
Point of an (s,S) Ordering Policy when a Service Level Is
Specified. J.Opl.Res.Soc., vol.29,12 pp1181-1193.

SCHWARTZ, L.B. (1973): A Simple Continuous Review
Deterministic One-Warehouse N-Retailer Inventory Problem.
Manag.Science, vol.19A,5 pp555-566.

SCHWARTZ, L.B.; SCHRAGE, L. (1975): Optimal and Myopic
Policies for Multi-Echelon Production/Inventory Assembly
Systems. Manag.Science, vol.21,11 pp1285-1294.

SCULLI, D. (1980): Stock Control Using Stock Level
Frequency Distribution. J.Opl.Res.Soc., vol.31,10 pp889-893.

SCULLI, D.; WU, S.Y. (1981): Stock Control with Two
Suppliers and Normal Lead Times. J.Opl.Res.Soc., vol.32,12
pp1003-1009.

SHAKUN, M.F. (1962): An Application of Operations Research
to Multiwarehouse Inventory Control. J.Ind.Eng., vol.13,6
pp470-474.

SILVER, E.A. (1976): A Simple Method of Determining Order
Quantities in Joint Replenishments under Deterministic
Demand. Manag.Science, vol.22,12 pp1351-1361.

SILVER, E.A.; SMITH, S.A. (1977): A Graphical Aid for
Determining Optimal Inventories in a Unit Replenishment
Inventory System. Manag.Science, vol.24,3 pp358-359.

SILVER, E.A. (1979): A Simple Inventory Replenishment
Decision Rule for Linear Trend in Demand. J.Opl.Res.Soc.,
vol.31,1 pp71-75. See corrigenda vol.31,10 p960.

SIMPSON, K.F. (1958): In Process Inventories. Opers.Res.,

vol.6,2 pp863-873.

. . . . (1959): A Theory of Allocation of Stocks to Warehouses. Opers.Res., vol.7,6 pp797-805.

SIVAZLIAN. B.D. (1971): Dimensional and Computational Analysis in Stationary (s,S) Inventory Problems with Gamma Distributed Demand. Manag.Science, vol.17B,6 pp307-311.

. . . . (1974): A Continuous Review (s,S) Inventory System with Arbitrary Interarrival Distribution between Unit Demand. Opers.Res., vol.22 pp65-71.

SMITH, S.A. (1977): Optimal Inventories for an (S-1,S) System with no Backorders. Manag.Science, vol.23,5 pp522-528.

SNYDER, R.D. (1980): The Safety Stock Syndrome. J.Opl.Res.Soc., vol.31,9 pp833-837.

STEVENS, C.F. (1974): On the Variability of Demand for Families of Items. Opl.Res.Quart., vol.25,3 pp411-419.

TADIKAMALLA, P.R. (1978): Applications of the Weibull Distribution in Inventory Control. J.Opl.Res.Soc., vol.29,1 pp77-83.

TATE, T.B. (1964): In Defence of the Economic Batch Quantity. Opl.Res.Quart., vol.15,4 pp329-345.

TAYLOR, C.J. (1961): The Application of the Negative Binomial Distribution to Stock Control Problems. Opl.Res.Quart., vol.12,3 pp81-88.

TAYLOR, P.B.; OKE, K.H. (1976): Tables for Stock Control - Problems of Formulations and Computation. Opl.Res.Quart., vol.27,3 pp747-758.

THOPSON, W.W. (1967): Operations Research Techniques. C.E.Merriel Books Inc.

VEEN, B. (1981): Safety Stocks - An Example of Theory and Practice in O.R. European J.Opl.Res., vol.6 pp367-371.

VEINOTT, A.F. (1965): Optimal Policy for a Multi-Product Dynamic Non-Stationary Inventory Problem. Manag.Science, vol.12A,3 pp206-222.

. . . . (1966): The Status of Mathematical Inventory Theory. Manag.Science, vol.12,11 pp745-777.

WAGNER, H.M. (1962): Statistical Management of Inventory Systems. John Wiley and Sons.

. . . .(1980): Research Portfolio for Inventory Management and Production Planning Systems. Opers.Res., vol.28,3 pp445-475.

WAGNER, H.M.; O'HAGAN, M.; LUNDH, B. (1965): An Empirical Study of Exactly and Approximately Optimal Inventory Policies. Manag.Science, vol.11 pp690-723.

WALLER, A. (1978): Cutting Freight Costs. Professional Administration, Sept.1978 pp26-27.

WARD, J.B. (1978): Determining Reorder Points when Demand is Lumpy. Manag.Science, vol.24,6 pp623-632.

WECKER, W.E. (1978): Predicting Demand from Sales Data in the Presence of Stockouts. Manag.Science, vol.24,10 pp1043-1054.

WEISS, H.L. (1980): Optimal Ordering Policies for Coptinuous Review Perishable Inventory Models. Opers.Res., vol.28,2 pp365-374.

WILSON, R.H. (1934): A Scientific Routine for Stock Control. Harvard Bus.Review, vol.13 pp116-128.

WINTERS, P.R. (1961): Multiple Triggers and Lot Sizes. Opers.Res., vol.9,5 pp621-634.

WITHIN, T.M. (1954): Inventory Control Research. Manag.Science, vol.1,Nov. pp1-40.

WORRAL, B.M. (1981): A Method of Determining Production Load and Size of Inventories when Demand is Variable. J.Opl.Res.Soc., vol.32,7 pp563-575.

ZACKS, S. (1970): A two-Echelon Multi-Station Inventory Model for Navy Applications. Naval Res.Logistics Quart., vol.17 pp79-85.

ZANGWILL, W.I. (1966): A Deterministic Multiproduct, Multifacility Production and Inventory Model. Opers.Res., vol.14,3 pp486-507.

ZOLLER, K. (1977): Deterministic Multi-Item Inventory Systems with Limited Capacity. Manag.Science, vol.24,4 pp451-455.