

RECENT WORK IN MULTI-ECHELON INVENTORY THEORY

E. P. Durbin
R. D. Wollmer*

The RAND Corporation, Santa Monica, California

PREFACE

The authors were discussants of three recent papers on multi-echelon inventory theory at the Twelfth Annual Meeting of The Western Section of the Operations Research Society of America, June 14, 1966.

Since the material presented at the multi-echelon inventory theory session is considered to be pioneering work in this area, and since the papers are not yet generally available, it appeared worthwhile to make summaries of the three papers and the discussants' comments available to those interested in the field.

The comments included are essentially the same as those delivered at the multi-echelon inventory session.

*Any views expressed in this paper are those of the authors. They should not be interpreted as reflecting the views of the The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

Summary of Network Analysis of Multi-Product Production and Inventory Models, by W. I. Zangwill, University of California at Berkeley.

This presentation described how a wide variety of multi-product, multi-echelon, multi-period production and inventory problems could be modeled by means of networks. The central problem Zangwill deals with is the minimization of a concave function defined on a convex set. He has developed an efficient algorithm for determining the minimum cost flow in a single-source, single-destination, single-commodity network when the cost of shipment across any arc is a concave function of the amount shipped. Zangwill indicates how this type of network models the Wagner-Whitin production smoothing problem under concave production and inventory holding cost.

Zangwill notes that optimal flows are extreme flows and further characterizes extreme flows in certain types of acyclic single-source multiple-destination networks as arborescences. An arborescence flow has the property that each node may have only one input. Using the properties of arborescence flows, Zangwill presents a dynamic programming algorithm for determining the optimal flow in acyclic networks when shipment costs are concave. Given these general algorithmic results, Zangwill indicates how multi-echelon, multi-period problems have a natural network formulation.

COMMENTS ON ZANGWILL'S PAPER

I thoroughly enjoyed the opportunity of reading a preliminary version of portions of Dr. Zangwill's talk. In particular, I think that the ease and natural manner with which the Wagner-Whitin Planning Horizon Theorem emerges from Dr. Zangwill's preliminary network definitions and results is that quality which is termed "mathematical elegance."

The algorithm for determining an optimal solution to the single source, single destination, single commodity, concave cost minimization problem in very general networks extends the class of problems which can now be treated in a computationally efficient manner to include a realistic class of practical problems. I note that in explaining this algorithm, Dr. Zangwill shows that the existence of a solution to the concave problem implies the existence of a solution to the linear problem, and that the non-existence of a solution to the linear problem implies that a solution will not exist in the concave cost problem. However, due to the concavity of the cost function there may be a solution to the linear problem and yet no solution to the concave cost minimization problem. Does Dr. Zangwill have a method for determining whether or not a solution will exist, and is there an easily demonstrated case in which a solution will not exist?

I will briefly review the main steps in the development of the single source, multiple destination, single product, concave cost minimization problem since this has independent value as an algorithm for minimizing a concave function on a useful type of convex set. Dr. Zangwill first defines source to destination chains, next defines arborescences, and then shows that extreme flows in acyclic multiple destination networks are arborescences since arcs are non-capacitated. It is this acyclic property which is sufficient to allow the use of a dynamic programming algorithm, and the arborescences which make dynamic programming computationally feasible. Next the flow at a node in the network is shown to be the sum of the amounts demanded at those destinations which are served by chains passing through that node. This illustrates the computational difficulty for networks with large numbers of destinations, for the number of flow values (or routings) that must be considered at each node is $2^d - 1$, where d is

the number of destinations. Are there any (easily stated) situations which result in a simpler network structure and hence a computationally simpler algorithm?

The use of a multiple destination, single source network to portray a multi-echelon, multi-product, multi-period production and inventory problem permits a variety of very general situations to be treated. The consideration of separate products arising from common sub-products or sub-assemblies whenever possible, avoids the assumption that cost is a function of the sum of the production, and still allows the cost functions to be concave. (Note that in general, however, Zangwill's multi-commodity case still treats cost as a function of the sum of production of all types.)

Dr. Zangwill has treated an important class of production and inventory problems, and has created a very powerful, very general set of efficient algorithms for dealing with the minimization of concave functions defined on convex sets. That the computational problems become somewhat unwieldy is an indication of the complexity of the multi-echelon and multi-period decision problem itself.

Summary of Optimal Policy for a Dynamic Multi-Echelon Inventory Model, by Stuart A. Bessler and Arthur F. Veinott, Jr., Stanford University

A general multi-period, multi-echelon supply system consisting of n facilities each stocking a single product is studied. At the beginning of a period each facility may order stock from an exogenous source with no delivery lag and proportional ordering costs. During the period the (random) demands at the facilities are satisfied according to a given supply policy that determines to what extent stock may be redistributed from facilities with excess stock to those experiencing shortages. There are storage, shortage, and transportation costs. An ordering policy that minimizes expected costs is sought. If the initial stock is sufficiently small and certain other conditions are fulfilled, it is optimal to order up to a certain base stock level at each facility.

The special supply policy in which each facility except facility 1 passes its shortages on to a given lower numbered facility called its direct supplier is examined in some detail. Bounds on the base stock levels are obtained. It is also shown that if the demand distribution at facility j is stochastically smaller ("spread" less) than that at another facility k having the same direct supplier and if certain other conditions are fulfilled, then the optimal base stock level ("virtual" stock out probability) at j is less than (greater than) or equal to that at a facility k .

COMMENTS ON BESSLER'S AND VEINOTT'S PAPER

I was very impressed with Dr. Bessler's and Dr. Veinott's approach to the multi-echelon inventory theory. The approach is classical and the assumptions are realistic when speaking of steady state inventory systems. It is

quite possible that this paper will serve as a take-off point for future research in this field.

The problem that Drs. Bessler and Veinott consider assumes each of n facilities has its own demands. In addition it assumes that all but facility one has a unique direct supplier of lower number to which it passes its shortages in the form of demands. Thus the demand at a facility depends not only on its own demands but also on the demands and stock levels at all facilities it supplies, either directly or indirectly. It would be interesting to know what changes would be necessary in order to allow the possibility of a facility having several direct suppliers of lower number. Assumptions are made such that the optimal ordering policy will be a base stock policy of ordering up to a stock level $\bar{Y}_i(H_i) \geq 0$ in period i where H_i is the past history. This essentially means that no set-up costs are involved. Furthermore, assumptions are made such that the stock will always be below the base stock level at the start of a period. In other words, we always order. In many finite period problems this is unrealistic as the base stock level often decreases in later periods due to the increased danger of having an item left over at the end. However, this assumption is quite valid in the infinite period or steady state case.

The authors show an upper bound for the optimal stock level at facility k may be found by restricting the stock level at all facilities supplying k and supplied by k to equal numbers known to be lower bounds on their optimal stock levels and optimizing subject to these restrictions. This occurs because under-stocking at facilities supplied by k increases the demand at k and under-stocking at facilities supplying k inflates the cost of being short at facility k . Both factors tend to raise the desired stock level at k . Similarly lower bounds on optimal stock level at facility k can be

calculated if one knows upper bounds at all facilities supplying and supplied by k . Since zero is a lower bound at all facilities we can get upper and lower bounds and refine them by an iterative process. In general, this process will not converge. Are there any special structures that do converge or are there any simple counter-examples that do not converge?

Probably the most interesting results are those that give conditions under which one can establish relationships between the optimal stock levels at different facilities. Basically, if j and k have the same direct supplier and the demand at j is less than that at k , then the optimal stock level at j is less than that at k provided the holding cost is not less or the shortage cost greater at j than at k . In the event that the facilities supplied by these two are symmetric and their holding and shortage costs equal, the two have identical optimal stock levels. Thus one has the computational advantage that he does not need to solve for optimum stock levels at all facilities.

The general characterization of demand distributions on the Bessler-Veinott paper will undoubtedly underlie future work in multi-echelon inventory models.

Summary of A Multi-Echelon, Multi-Item Inventory Model for Recoverable Items, by C. C. Sherbrooke, The RAND Corporation.

A base depot system is modeled where demand is compound Poisson. When a unit fails at base level there is a probability r that it can be repaired at the base according to an arbitrary probability distribution of repair time and a probability $1-r$ that it must be returned to depot for repair according to some other arbitrary distribution. Bayesian procedures are employed in demand estimation. For high-cost, low-demand items the appropriate policy is $(s-1, s)$ which means that items are not batched for repair. In this case there is an analytic solution. An extremely efficient computer program has been designed to show the cost-effectiveness trade-off for a large group of recoverable items.

COMMENTS ON SHERBROOKE'S PAPER

The Sherbrooke paper is of special significance in that it may be the first multi-echelon inventory model to be implemented successfully. By making assumptions which are quite valid in practice, he has been able to devise a model which is not only analytically tractable but also very efficient computationally.

The model deals with a job shop type situation in which equipment is of high cost and high reliability. The actual operations are carried on at the level of the lower echelon, or as referred to in the Air Force application, at the bases. When pieces of equipment break down they are repaired at the base if possible. Otherwise they are repaired at the depot or higher echelon. The fraction of items that are base reparable is given for each base. When an item is demanded (repaired) at a base or depot, another item from that particular facility's stock replaces

it. In the event stock is zero, the demand is satisfied by the first item to finish repair after all demands ahead of it are satisfied.

Item demands or failures are given by a Compound Poisson distribution. This means item demands come in batches, the occurrence of batches being characterized by a Poisson distribution. The number of items in the batch may have an arbitrary distribution. However, in this report it is assumed that the compounding distribution is logarithmic due to its computational desirability (gives negative binomial distribution) and since it also turns out to be a good approximation in the base stockage application. It would be interesting to know if there are any other compounding distributions that have desirable computational properties, what properties are necessary for computational feasibility, and which additional ones are sufficient. The mean of the base demands are estimated by a Bayesian procedure and the variance to mean ratio, which is assumed to be the same for all bases, is estimated by a maximum likelihood test.

The goal is to minimize expected back-orders at the bases. It would be interesting to know how this criteria compares with others such as fill rate, service rate, ready rate, and operational rate. Note that the system is conservative; i.e., no items enter or leave the system. Of course, some condemnations might be able to be handled if one considered condemning an item and purchasing a new one as one method of repair. However, for the Air Force application for which this model was designed, condemnations are not replaced on a one for one basis and cannot be handled in this model. Given the fraction base repairable at each base, the mean and variance to mean ratio of the distributions, the mean repair time and turn-around time at the bases and depots, and the stock at each base and depot, one can calculate the expected number of

back-orders at a random point in time. For a given depot stock level and total stock level one can allocate the base stock so as to minimize total back-orders at the bases by starting with zero stock and then allocating the stock to the bases one at a time by means of an incremental approach. One may solve this for several stock levels and plot a curve of back-orders versus total stock.

An alternative problem is to put upper bounds on the expected back-orders at each base and then find the minimum total stock necessary to achieve these bounds. This is equivalent to minimizing a weighted sum of total stock and expected back-orders at the bases, these weighted sums being Lagrange multipliers. Presently there are no methods for determining these multipliers that are known to be efficient. It would be interesting, though, to know which methods, if any, look promising.

COMMENTS ON THE THREE PAPERS IN GENERAL

I think that at this early stage in the development of multi-echelon inventory theory we might expect pioneering papers to be characterized by a common recognition of problems peculiar to the topic and diverse attempts to deal with these problems. It appears that this is the case here today.

The central difficulty of multi-echelon, multi-period inventory problems arises from the combinatorial interdependence caused by the variety of ways of meeting or dealing with demands at the several facilities in the several periods. In addition, demands at lower echelons affect the demand at higher echelons, and in some cases stock levels at lower echelons may affect stock levels at higher echelons.

An element that underlies the special models developed in these papers is the selection of conditions and decision rules which avoid some of the combinatorial complexity. This occurs in the supply and redistribution policies which indicate how a facility is to react to a demand, given the stock available at the facility. This is also seen in other assumptions. For instance, the simple version of the Zangwill model assumes that demand is deterministic and must be met at the facility and in the period in which it occurs. This assumption allows consideration of the full N-period problem. The Bessler-Veinott model requires that each facility (except the highest level) have a unique supplier. This requirement creates a network which is an arborescence, and which, as you have seen, makes computation possible. The Bessler-Veinott paper further assumes the conditions necessary to insure the optimality of a "base stock" policy, which allows the consideration of the N-period problem, one period at a time. In the Sherbrooke model the depot-base interaction or multi-echelon interaction

is made tractable by the assumption that demands of a certain type go to base and are serviced or backlogged there, while demands of other types pass immediately from base to depot and are serviced or backlogged there. Thus the demand at the depot or higher echelon is not a function of stock at the base or lower echelon.

I will briefly review the models with respect to the demand process, the cost structure, the criterion, the optimization method, and the major assumptions to facilitate more detailed discussion.

The Zangwill model is a very general multi-echelon, multi-period, multi-product production and inventory model. In it, demand is known precisely. The supply policy states that demand must be satisfied from production or inventory when and where that demand occurs. Production and inventory holding costs may be concave. The criterion is the minimization of total cost, and the optimal policy is determined exactly by means of a dynamic programming algorithm.

The Sherbrooke model is a repair shop or job shop model, and assumes a particular demand distribution which provides analytic tractability and yet is still sufficiently general. Purchase cost of stock is linear. There is an implied linear stockout penalty cost. The criterion is the minimization of expected back-orders at the bases (lowest echelon facilities) of the system, and this minimization is accomplished by marginal analysis in the case where stockout costs are equal at all bases. The supply policy or decision rule in the Sherbrooke case essentially determines which items shall be considered at base level and which shall be filtered back to the depot.

The Bessler-Veinott model is a warehousing and inventory problem in a more classical sense. It permits a perfectly general demand process to be postulated so long as the stock remaining at a facility after the

occurrence of any demand possible under the distribution is not greater than the level of stock which is optimal for the next period and that excess stock can issue at the next period purchase cost. Costs considered are purchase, holding, penalty, and a unit transportation cost incurred by the facility which finally supplies a demand. All costs are linear. The criterion used is minimization of total expected cost, and the optimization approach is by determining upper and lower bounds and iteratively sharpening them. In addition, the Bessler-Veinott paper orders the optimal policies in terms of general orderings of the demand distributions, and also seeks symmetry conditions in the costs and demands which can be used to simplify the computations.

Thus it appears that the models discussed here are characterized more by their common focus than by similarities of approach.

REFERENCES

1. Zangwill, W. I., "A Deterministic Multi-Product, Multi-Facility Production and Inventory Model," Operations Research 14, 486-507, (1966).

2. Bessler, S. A. and Veinott, A. F., Optimal Policy in a Dynamic Multi-Echelon Inventory Model, Decision Studies Group, Palo Alto, California, and Department of Industrial Engineering, Stanford University, Stanford, California, February 23, 1966.

3. Sherbrooke, C. C., METRIC: A Multi-Echelon Technique for Recoverable Item-Control, The RAND Corporation, Draft RM-5078-PR, to be published.