THE OPTIMAL LOCATION OF STOCK IN A DISTRIBUTION WAREHOUSE

H. Harrison*

INTRODUCTION AND GENERAL PROBLEM DEFINITION

The increasing sophistication of physical distribution planning systems enables distribution managers to plan their day-to-day and week-to-week routing schedules with near optimum efficiency. Such planning systems[1] permit management to determine many important operational parameters such as warehouse locations, warehouse distribution territories, optimal routing patterns, customer service levels, fleet utilisations and so on. Managements that have used such planning systems have discovered that physical distribution savings of the order of 5% to 20% can be achieved, which are most acceptable results to the manager beset with the burden of ever increasing prices of energy, materials and manpower.

Physical distribution systems, however, are commonly thought to begin *exterior* to the warehouse. What happens *within* the warehouse is, very often, relegated to a position of secondary or tertiary importance. Nothing could be further from the truth.

This paper is concerned with what happens within a distribution warehouse. More narrowly, however, it is directed towards the problem of how to locate stock within the warehouse so that materials handling costs may be minimised. Some of the results that are presented are based on original research carried out by a postgraduate business student group.[2]

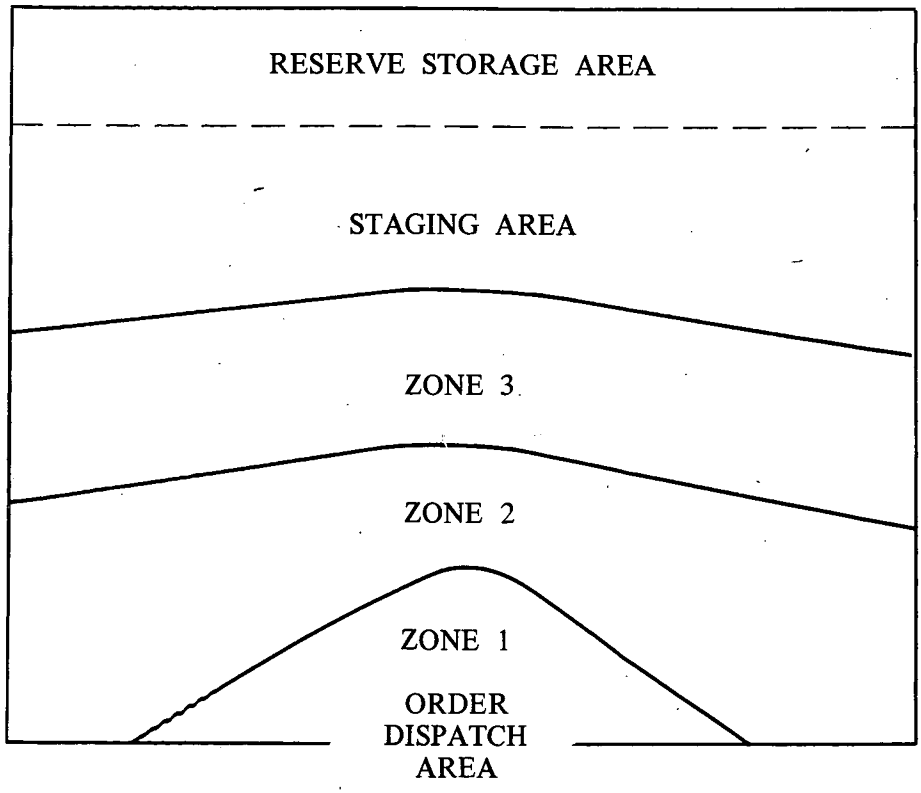
Conceptually, the kind of warehouse layout that is being considered in diagrammed in Figure I.

The warehouse is divided into a reserve storage and staging area. The warehouse stores a variety of items, and customers can place orders for any combination of items. It is assumed that the stock for any given

* The author is Professor of Management Information Systems at University College, Dublin.
He wishes to acknowledge the assistance of Mary Anne Byrne and Conall P. Clancy, who participated in the study.

FIGURE I

CONCEPTUAL DIAGRAM OF A DISTRIBUTION WAREHOUSE



item is divided between locations in the reserve storage and staging areas. When an order for an item is received a warehouse operator travels to the location of that item in the staging area and brings the item to the order dispatch area. At this location, the item is transferred to some outgoing vehicle for delivery to the customer. This process is known as "picking an order", whereas a "pick" relates to one trip of the operator to a location of one of the items on the order.

PICKING DISCIPLINES

Depending on warehouse design, item bulk and item distribution on orders, various picking disciplines can be employed within a warehouse. Those disciplines most usually employed are:

(a) Out and back selection of each item – mainly used where order items are picked in large quantities;

(b) Picker routing – several items are picked on a single trip through the staging area;

(c) Picker stations – each station is manned and served by a fixed or portable conveyor.

The prime variable cost element in any order-picking discipline is the time spent by the operators on their picks – assuming that the average rate of travel is the same for all orders. It is therefore obvious that the labour cost of picking an item is directly proportional to the distance of that item from the order dispatch area if the "out and back" discipline is used within a warehouse. Indeed, even if the picker routing discipline is the one in vogue, it is not difficult to show that the operator's total time on a single trip can be distributed across all of the items picked on that trip. Thus, the time allocated to a given item can be made proportionate to the item's distance from the order dispatch area. Conveyorised systems are extremely varied in design and capacity and are therefore beyond the scope of this analysis. However, the ideas put forward in the paper would apply to picking stations located in the proximity of such systems. In general, therefore, we can say that no matter what the picking discipline, the picking cost per item is correlated to the distance of the item from the order dispatch area.

STOCK PLACEMENT WITHIN THE WAREHOUSE

From the aforegoing discussion it is evident that a warehouse's order selection costs are very closely related to stock location. So how should stock be located? Some of the most popular ways of locating stock are:

(i) popularity, i.e. demand;
(ii) cubic footage, i.e. size;
(iii) alphabetically;
(iv) supplier orientated.

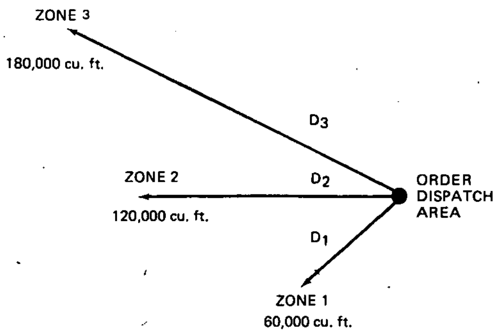
Very few warehouses in Ireland locate stock on the basis of the Cube Per Order Index (C.P.O.) method, first proposed by J. L. Heskett[3] in 1963. There is no doubt, based on empirical research carried out at this university, that the C.P.O. method offers significant efficiencies over all other methods, in the location of stock within a warehouse. The following simple example will initially illustrate the Cube-Per-Order Index philosophy.

AN INITIAL COMPARISON OF STOCK LOCATION PROCEDURES

Let us consider the following (simple) situation:

ITEM NO.	ITEM VOL.	AVERAGE ORDER SIZE	NO. OF ORDERS PER DAY	NO. OF DAYS ON HAND	REQD. CUBIC FOOTAGE
1	4 cu. ft.	5 units	100	3	6,000
2	20 cu. ft.	3 units	300	2	36,000
3	32 cu. ft.	1 unit	450	1	14,400
4	2 cu. ft.	2 units	350	3	4,200
5	32 cu. ft.	2 units	900	1	57,600
6	8 cu. ft.	10 units	300	3	72,000

Further let us consider that the warehouse is laid out as follows:



where D_1, D_2 and D_3 represent the distances to the various stock areas.

Location Rule A – Popularity Arrangement of Stock					
ITEM NO.	NO. OF ORDERS/ DAY	REQD. CU. FT.	ZONE	Order Select. COST/ ORDER	Order Select. COST/ ORDER
5	900	57,000	1	£0.40	£360
3	450	14,000	1 & 2	£0.74y	£333
4	350	4,200	2	£0.80	£280
2	300	36,000	2	£0.80	£240
6	300	72,000	2 & 3	£0.82y	£246
1	100	6,000	3	£1.20	£120
	2,400	190,200			£1,579

Location Rule B – Total Cubic Footage Arrangement of Stock					
ITEM NO.	CU. FTGE.	ZONE	Order Select. COST/ ORDER	NO. OF ORDERS/ DAY	Order Select. COST/ DAY
4	4,200	1	£0.40	350	£140
1	6,000	1	£0.40	100	£ 40
3	14,000	1	£0.40	450	£180
2	36,000	1 & 2	£0.50[y]	300	£150
5	57,000	2	£0.80	900	£720
6	72,000	2 & 3	£0.86[y]	300	£258
	190,000			2,400	£1,488

Thus on comparison of location rules A and B, location by Cubic Footage is somewhat cheaper than by popularity.

Let us now look at the same situation again, but this time locate the stock by the Cube-Per-Order Index method where:

$$\text{C.P.O. Index} = \frac{\text{Required cubic footage of item}}{\text{Number of items ordered/day}}$$

Location Rule C – C.P.O. Index and Stock Location						
(1) Item No.	(2) Reqd. Cu. Ftge.	(3) Orders /Day	(4) C.P.O. (2÷3)	Zone	Order Select. Cost/ Order	Order Select. Cost/ Day
4	4,200	350	12	1	£0.40	£140
3	14,400	450	32	1	£0.40	£180
1	6,000	100	60	1	£0.40	£ 40
5	57,000	900	64	1 & 2	£0.48[y]	£432
2	36,000	300	120	2	£0.80	£240
6	72,000	300	240	2 & 3	£1.00[y]	£300
	190,000	2,400				£1,332

x Based on measurements of D_1, D_2 and D_3

y Order selection cost adjusted to reflect stocking in 2 zones.

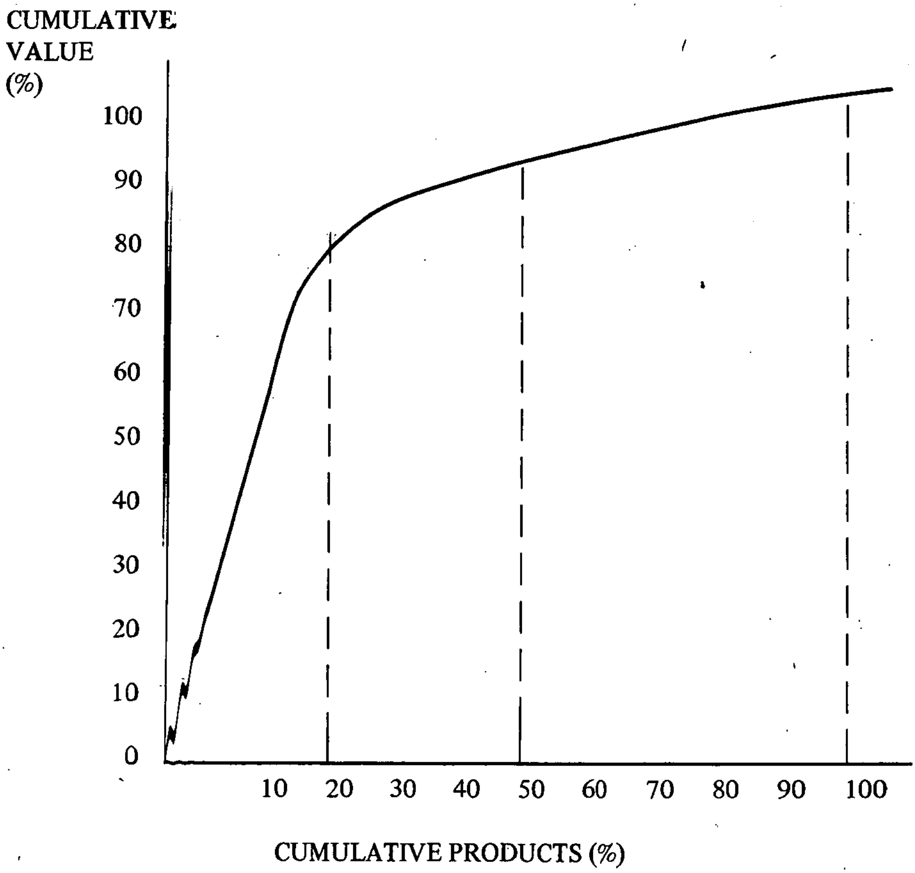
Thus for this simple, hand-computable, exercise the Cube-Per-Order Index method of stock location is 11.7% cheaper than the Total Cubic Footage discipline and 18.5% cheaper than the Popularity Arrangement of Stock procedure.

EMPIRICAL ANALYSIS FOR A DISTRIBUTION WAREHOUSE

Although these examples demonstrate the superiority of the Cube-Per-Order Index approach such "hand-cranked" procedures could not be applied to a full-scale distribution warehouse. Many distribution warehouses have thousands of different items in stock at any one time. Clearly, therefore, a sophisticated approach is needed.

FIGURE II

THE PARETO DISTRIBUTION AND STOCK CLASSIFICATION



As with most product ranges, the principles of the Pareto distribution can be used to good effect in the stock location problem. Put at its simplest, the Pareto distribution shows us that a large amount of value of stock is accounted for by a relatively small number of items. Figure II explains the situation.

Generally speaking the Pareto shows us that 20% of the stock items account for 80% of the *value* of the items contained within a distribution warehouse. This then, is our starting point – namely, ascertain those items contained within the 20/80 rule of the Pareto distribution.

In many warehouses the number of items contained within the 20/80 rule could be several hundred, or, exceptionally, several thousand. When Heskett first introduced (in 1963) the C.P.O. Index method he intended it to be a manual approach to solving the stock location problem. However, the manual approach has two major disadvantages:

(i) Even after a Pareto analysis, the number of items remaining can be several hundred, or more; by hand, therefore, very time consuming.

(ii) The manual method can only with difficulty take into consideration the *relative* location of merchandise throughout the warehouse.

These disadvantages can be simultaneously overcome if one uses the well-known management science technique, linear programming, and a standard solution procedure such as IBM's MPS/360 – or the equivalent systems of ICL, Digital, and Honeywell.

The informational needs of this approach are as follows:

(i) The number of different products considered → N.

(ii) The cubic measurement of each of these (i) products → V_i.

(iii) The cost of moving a product from the (j^{th}) sector to the assembly despatch area → C_j (and linear).

(iv) The total number of storage sectors → T.

(v) The number of storage sectors in the assembly/dispatch area → L.

(vi) The amount of product (i) to be stored in the assembly/dispatch area → A_i.

(vii) The amount of product (i) to be stored in the warehouse → W_i.

(viii) The cubic capacity of the (j^{th}) sector → S_j.

If we now define the amount of product (i) to be stored in storage (j) as X_{ij} and have as our objective the minimisation of stock handling costs, we can formalise the problem as:

$$\text{Minimise Cost} = \sum_{i=1}^{N} \sum_{j=1}^{T} C_j X_{ij}$$

Subject to:

$$\sum_{j=1}^{L} X_{ij} = A_i \qquad \text{for } i = 1, 2, \ldots N$$

$$\sum_{j=1}^{T} X_{ij} \leqslant W_i \qquad \text{for } 1, 2, \ldots N$$

$$\sum_{i=1}^{N} v_i X_{ij} \leqslant V_i \qquad \text{for } j = 1, 2, \ldots L$$

$$\sum_{i=1}^{N} v_i X_{ij} \leqslant V_i \qquad \text{for } j = L, \ldots T$$

$$X_{ij} \geqslant 0$$

This is a linear programming formulation, that uses easily accessible information, and can be analysed using IBM's MPS/360 software or its equivalent. The empirical work of Byrne and Clancy, based on a distribution warehouse in Ireland, has shown that the computer-based approach is superior to existing known systems of stock location.[4] The work of Kallina and Lynn conclusively proved that the C.P.O. Index approach yields the optimum method of stock location in a distribution warehouse.[5]

Byrne and Clancy ranked the computer-based C.P.O. Index method against five other frequently used procedures and obtained the following results:[6]

Although the manually operated "Total Cubic Requirements" procedure incurred costs only 4% greater than the computer based C.P.O. Index System, post optimal sensitivity analysis is only possible using the computer based approach. The main conclusions emanating from the post-optimal analysis were:

(i) A greater cost advantage could be gained using the C.P.O. Index procedure if management adopted a more flexible approach with regard to the stock levels, on an item-by-item basis, contained within the staging area of a distribution warehouse.

(ii) With respect to the stock items contained within the 80/20 Pareto classification, (ideally) no such items should be located in the reserve storage area of a warehouse.

ALLOCATION PROCEDURE	RANK ORDER BASED ON TOTAL W'HOUSE COSTS	% INCREASE IN W'HOUSE COSTS – C.P.O. BASE
MANUALLY BASED		
Unit Size	5	33.0%
Total Cubic Requirements	2	4.0
Demand	6	68.0
COMPUTER BASED		
Demand	3	10.0
No. of Times Picked	4	17.0
Cube-Per-Order Index	1	–

(iii) An important system variable is the "average quantity picked per order". Small orders are definitely expensive. For popular products, therefore, additional promotional effort could be significant.

(iv) As is already well known, stocks in any warehouse should be kept as low as is practically feasible. The analysis that has been described reinforces this view by showing that an additional saving of 17% (in handling costs) could be achieved if the Reserve Storage Area were to be eliminated. In many business areas this, of course, would not be practical. However, it is further evidence for the need for an efficient management information system for the warehousing/distribution function.

CONCLUSIONS

There is no doubt that the C.P.O. Rule represents a major breakthrough in materials handling methodology. It must become the *only* procedure for the location of stock within a warehouse. The implementation of the C.P.O. rule may represent some difficulties initially; however, once stock items are located on this basis operators will quickly appreciate its advantages. From the managerial point of view C.P.O. implementation will result in a more careful approach to short term demand forecasting and a more ordered approach to data collection, storage and retrieval. In addition management and operators alike will be faced with a radically different approach to stock location and order-picking procedures.

Finally, considerable thought will have to be given by organisations as to the relative merits of hand computation of the Index versus the more all-embracing and powerful linear programming and computer analysis approach. No doubt, initially, many companies will opt for the more straightforward hand computation procedures. However, in the author's opinion, the rapid (and ever increasing) advance in computer-based information systems will, within a decade, bring even the smaller distribution warehouses within the sphere of influence of the mindless efficiency of the computer.

REFERENCES

1. H. Harrison: "Corporate Distribution Planning", IBAR, Vol. 1, April, 1979.

2. Mary Anne Byrne and Conall P. Clancy: "A Computer-based Linear Programming Model for the Optimal Product Layout of a Warehouse". Unpublished Thesis for the Degree of M.Mangt.Science, 1978.

3. H. L. Heskett: "Cube-Per-Order Index – A Key to Warehouse Stock Location", *Transportation and Distribution Management*, Vol. 3, April, 1963.

4. M. A. Byrne and C. P. Clancy, *op. cit.*

5. C. Kallina and J. Lynn: "Application of the Cube-Per-Order Index Rule for Stock Location in a Distribution Warehouse", *Interfaces*, Vol. 7, No. 1, November, 1976.

6. M. A. Byrne and C. P. Clancy, *op. cit.*

FUEL PRICING — THE COSTS OF QUEUEING

Moore McDowell*

The purpose of this article is to examine some of the consequences of the method of fuel rationing which was permitted to develop in the spring of 1979. The refusal of the Government to allow the price of fuel oils to be increased resulted in an overall supply short-fall at the regulated price, and in considerable uncertainty as to supply availability at any price to the individual consumer.

It will be remembered that the allocation system which emerged was one whereby supply was made available to customers on the basis of a composite price – the money element of which was the controlled price, the remainder being a time-price paid in queueing. In addition, retailers imposed maximum volume sales per transaction, (£2 to £3), and later, with Government encouragement and regulation, introduced minimum or even single quantity sales (typically £5 min./max. per customer). Furthermore, there was widespread uncertainty as to whether particular garages would have supplies, or be open, at any given time.

It will be shown that the allocation system adopted was costly and inefficient (in the sense that any rationing system is costly but this one was unnecessarily so, and, temporarily at least, exacerbated the shortages). It will also be shown that the system adopted had some redistributive implications which were hardly desired by those who allowed the system to emerge. This paper examines the demand consequences of supply uncertainty and lists and explains some of the now standard results of rationing by queueing as developed by Barzel.[1] The consequences of supply constraints for market clearing are also examined.

THE DEMAND CONSEQUENCES OF SUPPLY UNCERTAINTIES

There is a well known formula to determine optimum inventories under conditions of market certainty. Inventory holdings impose costs on the

*The author is Lecturer in Economics in the Department of Economics at University College, Dublin. He wishes to acknowledge the valuable discussions of the contents with Professor J. W. MacManus (Carleton University) and the comments of Dr. W. K. O'Riordan (University College, Dublin). An earlier version of this article was given to the Economic Policy Conference of the Dublin Economics Workshop in Killarney, October 1979.

holders – the opportunity cost of the capital involved. On the other hand, there are transaction costs involved in purchasing. Transaction costs per unit of output are minimised by maximising the time between purchases; but this involves holding larger average inventories. Total cost minimisation involves a trade-off. The optimal inventory purchase under these circumstances is given by the familiar square root formula:

$$Q^* = \sqrt{\frac{2aS}{r}}$$

where Q* is the optimal maximum value of inventories, a is the (fixed) transaction cost per purchase, r is the opportunity cost of funds and S is the use rate per unit of time of the commodity. If we assume that the use rate is constant, and inventory usage also, then T*, the optimal average inventory is ½Q*, and actual inventories will decline linearly from Q =Q* to zero over the inventory period. In practice the inventory level will never actually reach zero.

In the case of the motorist and petrol, the tank (ideally) never reaches zero – there is some residual uncertainty as to the fuel gauge accuracy, proximity to supply and consumption. But we can model his behaviour under normal conditions as under total certainty without undue strain on credibility. In one respect, however, his behaviour may be constrained significantly: he may be unable to hold his optimal stock due to the limits on size of purchase imposed by car tank size. This means that while he cannot reach his global minimum cost level of purchase, his constrained minimum cost is obtained if he fills his tank and runs it down to the level he would run it down to were he unconstrained.

Hence, for some motorists the optimal purchasing pattern under certainty is to fill up each time; for the remainder the optimal purchase is a quantity less than a tankful. In passing, it is interesting to note that since inventory costs reflect the opportunity cost of time (in transaction costs) and of funds, for any given tank size, the lower one's income the more likely is one to have an average petrol stock of less than half a tankful.[2] Also, for those whose optimal Q* exceeds tank size, small changes in transactions or opportunity costs of holding inventories will have no effect on their behaviour.

A priori then, on average we may expect that a stock of petrol held by the motoring public in car-tanks is less than 50% of the tank-capacity of the car-stock. (A survey quoted in the newspapers during the petrol shortages suggested that the average was of the order of 25%.) Hence, without incurring *any* significant increase in transactions costs it is feasible for the public to increase the stock held in car tanks. This will happen when those individuals whose cars' tank capacity exceeds Q* experience an increase in Q*. *A fortiori,* if the motoring public is prepared to accept a higher level of transactions costs or if the opportunity cost of inventories falls, or if both occur, it is possible to increase the stock of petrol held in tanks by reducing the time period between purchases, for any given rate of consumption.

It is a standard result of inventory theory that uncertainty as to requirements over lead-time – the consequence of a stochastic demand function for the firm's output – implies the holding of buffer-stocks. Utility maximisation in an uncertain environment, where the firm is risk averse, requires that the firm hold a level of stocks such that the expected minimum stock size is greater than zero. The greater the expected minimum stock size (the desired buffer stock) the smaller is the risk of "stock-out" with its attendant costs.[3] Since the transactions costs of purchasing are usually, and fairly realistically, held to be constant, the optimal buffer stock is a function of expected sales, the variance of expected sales and the opportunity cost of holding inventories. Formally, this is:

$$B^* = B^*(S, \hat{\sigma}^2, r); B_1^* > 0; B_2^* > 0; B_3^* < 0;$$

In inventory studies the result is usually derived from a model where supply is certain, but demand is stochastic. The same result is obtained if the reverse holds true – viz. demand is certain, but the timing of supply is uncertain. It is further re-inforced if both the timing and the quantity available are not known with certainty.

It is possible, without over-simplifying, to model the position of the average motorist in the Spring of 1979 as being one in which his flow demand of petrol was known, but the timing and quantity availability of supply was subject to risk. Consequently, independently of any problems of over-all excess demand due to price control, the method of allocation "chosen" by the authorities resulted in a temporary increase in demand for petrol as utility-maximising and risk-averse motorists attempted to increase the average stock of petrol held in their tanks. This was socially inefficient since it obliged people to accept higher transactions costs per mile and higher total opportunity costs, all because of the needless uncertainty which resulted from the rationing mechanisms adopted by retailers and wholesalers. These in turn stemmed from the unwillingness of the authorities to free the market price of petrol, which not only caused overall supply difficulties but also introduced incentives for both retailers and wholesalers to hold on to stocks in anticipation of price rises which everyone knew were inevitable.

This clearly implies that the stocking-up behaviour of motorists, castigated at the time as perverse and irrational, was a predictable, rational response to the price/supply regime which resulted from Government action (or inaction?).

CONSEQUENCES OF RATIONING BY QUEUEING

It is obvious that if the price of a commodity is held by decree below the market clearing level, the available supply will be allocated by some mechanism other than pricing. The standard arguments concerning the inefficiencies which are the likely consequence of departure from allocation by price are well known. In this section we examine the implications of adopting one particular method of allocation, rationing by time, which was adopted in 1979. These come under two headings – efficiency and distribution.

First of all, it must be remembered that all methods of allocation are costly. Allocation by market pricing involves information, marketing and distribution costs. The question is: do allocation methods other than market pricing impose excess costs? For a variety of reasons, the answer to this question is in general, yes. In general, allocation by point rationing or by time interferes with exchange efficiency.[4] This can in principle be eliminated if re-trading in the rationed goods is permitted, or if a market in points (or time) is established. Of course, if this is the case, then the rationing system results in a market pricing allocation subject to some redistribution of income. In this case, efficiency could have been improved by a simple prior income redistribution, with the market allocation mechanism being left untouched. If income redistribution is the objective, optimal intervention would require a direct income transfer; to achieve the same objective by rationing plus re-trading (where possible) involves incurring unnecessary costs. It is, therefore, inefficient.[5]

This proposition is in line with the conclusions of the literature on optimal intervention in the presence of distortions. An interesting, if restrictive, rationalisation of direct commodity rationing as a superior means of redistribution has been made by Weitzman.[6] This, however, depends, as Weitzman explicitly admits, on a "one dimensional" equity criterion – matching the supply of a good to an arbitrarily defined and measured "need". Even then he concludes that the preferability of rationing varies inversely with heterogeneity of tastes and directly with the distribution of income. The more heterogeneous are tastes, and the more even is the distribution of income, the better is pricing plus redistribution of income.

At the time of the supply problems induced by price controls the authorities refused either to allocate supply petrol directly to final customers (with some exceptions), or to allow the available supply to be allocated by a point rationing system, of which many variants were possible. Instead, the supply was rationed by queueing, coupled to a variety of quantitative and other constraints.

Queueing is a highly effective way of rationing. It is also arguably the most costly. The costs arise from two sources: the opportunity cost of the time spent in the queue and the loss of consumers' surplus. The first of these needs little explanation: time has an opportunity cost, either in terms of leisure foregone or income and output lost. Further, these are lost, wasted, not simply transferred. Under a price allocation system, a higher price of a product means a transfer of resources from buyer to seller. When queues are used, however, the seller does not appropriate the increase cost to the purchaser, which is a nett loss incurred by the latter as the cost of acquiring a property right in the rationed commodity. Time spent in the queue is an exclusion device, similar to money prices, but it is an exclusion device which is socially costly. From the point of view of the consumer, of course, it ought to be, in principle, irrelevant whether he pays in money or in time equivalent. But even the most casual observation suggests that those waiting in line were far from content at having to do so.

Why do individuals object to queueing? The basic reason is because usually queueing involves a loss of consumers' surplus. Under ordinary circumstances, for all individuals possessing downward sloping demand curves, the total utility derived from consumption (in money terms approximated by the area under the demand curve above the price per unit) exceeds the total cost of purchase. The difference, the welfare gain, is consumers' surplus. It might be thought that since queueing (by substituting payment in time partly or wholly for payment in money) changes nothing. Unfortunately, this is not the case. When goods are rationed by queueing, individuals are not usually allowed to decide whether they will or will not buy an incremental unit of the good at a price of n minutes per unit. Instead, they are faced with a decision to queue or not to queue, in return for which they will obtain or will not obtain a fixed allotment. Marginal time-spending on consumption is not feasible. It is an all-or-nothing decision. (If garages gave more petrol to those who came earlier, this would not be the case.)

Consequently, for the marginal individual – the man who decides it is only just worth joining the queue – the value of the petrol allotment is all but exhausted by the time plus money cost of obtaining his allotment. In other words, he derives no consumers' surplus. This "all-or-nothing" decision reduces consumers' surplus for at least some intramarginal consumers for whom some consumers' surplus survives.[7] It is this all-or-nothing allotment aspect of queuing, rather than the fact of paying a time price, which is the basic cause for the exasperation caused by the queues. Of course, tempers were not helped by the uncertainty of obtaining a supply.

Despite all this, queues there were – which leads on to the questions: what determines the size of queue, and what effects have government sales restrictions on them? Before considering the problems involved in analysing queueing equilibrium as a rationing device, certain verbal ambiguities must be cleared up. For the most part these reflect confusion about the "length" of a queue. Is 100 cars, each queueing for 20 minutes, to obtain 3 gallons each, a longer queue than 10 cars, each queueing for an hour, to obtain 5 gallons? To clarify this, let us define the length of the queue as the number of people in the queue, irrespective of the time they have to wait. The other relevant magnitudes are queueing time per car and queueing time per gallon (TPC and TPU) which are basically independent of the length of the queue.

Remember first that the decision to queue reflects a willingness to pay with time for the right to obtain a commodity; secondly that TPC and TPU perform the function of *excluding* others from consuming.

We simplify the model by assuming that:–

(i) price is zero;
(ii) supply per customer is instantaneous (no time per delivery);
(iii) information concerning supply availability and required queueing time is perfectly disseminated.

(i) is purely a simplification, and clearly does not affect the argument, since TPU is greater than zero only if money price/unit is less than the market clearing money price; (ii) may seem less innocuous – but it can be shown that TPC and TPU are basically independent of the mechanics of distribution – sales time will be taken into account in deciding when or if to join a queue;[8] (iii) is to eliminate uncertainty and searching behaviour considerations.

The customer's time price demand curve for petrol is negatively sloped – the larger the allotment he is offered, the lower the time price per gallon he will pay. His total time price (queueing time, or TPC) depends on the position and slope of this curve. The length of the queue at opening time, and the amount of time each has to wait is then simultaneously determined. The numbers in the queue, and the number of hours before opening they arrive, will be such that for the marginal individual the entire value of an allotment is equal to his evaluation of the time cost of purchasing the allotment.

If the number of allotments is increased, the number of individuals (the queue length) increases – but TPU must fall, and TPC must fall, too. That is, more people queue – but arrive a shorter time before opening time. Why? Because if the market is to clear, and the price is measured in time, then increased supplies will only be taken off the market at a lower price, therefore, TPU must fall. Since we restrict quantity per customer, we must increase the number of customers, i.e. include those for whom TPC had previously been too high. All must queue for the same amount of time – so TPC must fall and TPU must fall.

If the number of allotments is held constant, but the size of each allotment is increased, then obviously the number of satisfied customers cannot increase. That is, the equilibrium length of the queue remains unchanged. Further, the TPU must fall – but TPC must rise. Why? Because if TPU falls, more individuals will want to join the queue for a given TPC. To *exclude* them, TPC must rise until, once again, for the marginal individual total time cost equals the value of the petrol purchased.

During the spring of 1979 it was often said that the correct response of the authorities would have been to introduce a minimum sale of £5 – or increase the allotment over than the commonly found £2 or £3 maximum. The foregoing analysis suggests that *cet. par.* this may not have been correct. Given the total supply, increasing the allotment size will reduce the number of cars in the queue – and the number of "satisfied" customers – but will cause the length of time spent in the queues to increase.[9] Conversely, to reduce the allotment – a £3 maximum on sales – would increase the number of cars queueing – but reduce the length of time they would spend in the queue. TPU, however, would increase.[10]

Despite this, it seems in retrospect to have been the case that the increase in size of allotment coincided with shorter queues. How can this be explained? The most obvious answer is that the increased allotments and queue reductions took place *after* the Government had allowed the price of petrol to rise. The inevitable rise in supply, even if it took some time, coupled to a demand response, even if slight, would reduce the queues, in terms of length, TPC and TPU. It is possible, of course, that the increased allotments just coincided with the achievement by motorists of the new, higher desired inventories described earlier.

The only other effect of increasing the size of the allotment on equilibrium queue length, on TPC and on TPU, would be a temporary one. If the increase in the allotment size by a £5 minimum regulation prevented some motorists from holding their desired average inventory of petrol, then while average tank content was being lowered to the newly constrained optimum level demand would fall, *ceteris paribus*. This, of course, is precisely the opposite effect to that described in section I above.

But once the new quasi equilibrium stock level is reached, the effects of the increase in allotment size would be as already described – a reduction in the length of the queue, a rise in TPC and a fall in TPU.

Finally, there is the question of the redistributive impact of rationing petrol by queueing. Since no effective market in queue rights was established, we do not have to concern ourselves with the possibility that the incidence of queueing on income distribution was materially different from its impact.

There is a widespread view that queueing is a means of progressive redistribution of income. The basis for this is the (reasonable) assumption that those on lower incomes have a lower opportunity cost of time. Hence, *ceteris paribus,* they will be more willing to queue than those on higher incomes, and they will obtain a greater quantity of the rationed good. However, as Barzel has shown, the conclusion is unwarranted, as it ignores the demand side – those on higher incomes do pay (in money equivalent) a higher price when asked to queue. But they may reasonably be expected to continue to queue because petrol has a very high income elasticity of demand. In general, the lower is price elasticity of demand for a commodity and the higher is income elasticity of demand, the more likely is queuing as an allocations device to result in the "rich" obtaining the commodity. And petrol clearly has a low price elasticity and a high income elasticity of demand.

Queuing, then, would re-allocate the available supply to the better-off – although they would pay, in opportunity cost terms, a higher price for the petrol. As impressionistic evidence that this in fact took place, consider the widespread reports of housewives queueing during daytime at suburban petrol stations. It seems reasonable to assume that a fair proportion of this represents two car families.[11]

THE CONSEQUENCES OF SUPPLY CONSTRAINTS FOR MARKET CLEARING

It is well known, both in theoretical and applied economics, that price elasticities of demand are greater in the long run than in the short run. Similarly, changes in relative factor prices may be expected to have relatively little effect on factor mixes in the short run. Basically, the reason for the smaller short-run response is because in the short run there are constraints which prevent the full effect of the price change from operating.

These examples illustrate a general principle in economics, described by Samuelson[12] in terms of the generalised Le Chatelier Principle. This principle may be reduced to the statement that the introduction and/or addition of constraints to an optimisation problem result in the reduction of the responses of changes in the equilibrium values of dependent variables in response to changes in the parameters of the system; the greater the number of constraints, the greater the reduction. Formally, the modulus of values of own-substitution effects, which are negative, will increase as the number of constraints is reduced.

The policy implication of this result in the present context is that if there are quantity constraints on consumers' behaviour, then the price elasticity of demand for petrol will be lower, *cet. par.*, than if those constraints did not exist. The greater the number of constraints, the lower the price elasticity of demand. Price in this case may be taken to refer either to money price, time price or a combination of the two.

During the petrol supply difficulties the Government encouraged retailers to introduce volume constraints on purchase in addition to the constraint of tank-size – minimum and maximum sales were implemented. In addition, the retail trade introduced constraints to some degree by allocating supplies to "regular customers", card-holders or other arbitrary groups of motorists. This had the effect of tying individuals to specific outlets.

The implication of this is that, knowingly or not, the Government adopted or encouraged measures which had the effect of reducing the price elasticity of demand for petrol at a time when, for reasons already discussed, the demand for it was increasing. Further, since the price which was being exacted was a composite one of money plus time, with the money element being (for the most part) fixed, the constraints must logically have increased the market-clearing time price per gallon during the shortage. Given that time has a positive opportunity cost, we may conclude that the constraints had the effect of lowering aggregate economic welfare.

SUMMARY AND CONCLUSIONS

Rationing is costly; but if some form of non-price allocation is to be adopted, then the cost per unit sold of rationing by queueing is likely to be higher than that of point rationing. In addition, if re-trading is in principle permitted, a rationing system using points or a fixed allocation permits greater distributional efficiency, and, in the case of a commodity like petrol, may be less regressive than rationing by queueing.

Granted all this, the only conclusion possible is that rationing by queueing is undesirable. This, however, can be modified if the rationing is expected to be of short duration. Then it is possible that the social overhead costs of establishing a rationing system which does not involve queueing may not be warranted by the cost-saving of avoiding queues. Under what circumstances is this likely? Short-term excess supply or demand results either from inadequate supply response to demand changes or to regulatory activity by the authorities. If the first is responsible, the onus must be on the authorities to show that the costs of allowing the price mechanism to clear the market exceed those of temporary queues. If it is the second, the effect of regulation induced queues is to lengthen the duration of the excess supply or demand. To do this by imposing queues is to ask the public to bear the cost of delaying a price adjustment which is already conceded as being inevitable.

If, however, in their wisdom the authorities decide to ration by queueing, minimum disruption (i.e., avoiding attempts to increase consumer inventories) requires that certainty of supply at a time plus money price be assured. Garages should, therefore, be obliged to disseminate (truthful) information as to stocks, opening times and supply allotments. This, of course, was not done during the 1979 difficulties.

Given that queueing is already operating, increased supply can reduce the queueing costs per car not by increasing the size of allotments, but by increasing the number of allotments while encouraging, rather than discouraging re-entry into the queues by motorists.[13]

Finally, to reduce the time cost per gallon of allocating a supply at a money price below its market price, the conclusions of the Samuelson analysis of price elasticity of demand are that other quantitive restrictions on consumption should be reduced to a minimum. In each and every one of these respects, the behaviour of the authorities was sub-optimal during the first half of 1979.

REFERENCES

1. Barzel, Y.: *A Theory of Rationing by Waiting:* Journal of Law and Economics, 1975.

2. It is assumed that the income elasticity of demand for leisure is positive: i.e., leisure is not an "inferior" good.

3. cf, for example,: Groff, G.K. and Muth, J.F.,: *Operations Management: Analysis for Decisions:* Homewood, Ill.; Irwin, 1972; Ch. X.

4. Technically, because quantitive constraints on trading force individuals off their contract curves, leaving room for Pareto-optimal re-allocation of resources.

5. Income re-distribution, too, involves costs. But given an existing re-distribution system

(the tax-transfer regime), it would be inefficient to re-distribute income via rationing as this involves (a) a new administrative machinery; (b) individuals having to incur trading costs. Even if a new re-distribution system had to be set up it seems intuitively unlikely that this cost would exceed the administrative and trading costs of re-distribution by rationing.

6. L. Weitzman, *The Price System vs Rationing* (abbrev. title), Bell Journal, 1977.

7. Even for those by whom no consumers' surplus is lost (those for whom the price in time and money per gallon of the allotment is less than the price they would willingly pay per gallon for that quantity if permitted to purchase it on an incremental rather than once-for-all basis), the contraint on purchase lowers welfare since it prevents them from achieving the necessary conditions for pareto optimality.

8. Queueing can take place with instantaneous distribution. Given adequate information concerning distributive time and place in line, the fact that distribution takes some finite time merely gives a queueing individual a choice of times at which to join the queue. It will not affect the length of time between joining it and obtaining the commodity in question.

9. A larger allotment has more value; queueing equilibrium requires that the marginal purchaser "pay" a time price equal to the value of the allotment. Therefore, TPC must rise.

10. From 9 above, TPC must fall; but since the number of gallons per allotment has fallen, the value of a gallon has risen for each customer. Consumers' equilibrium requires that the "price" per gallon rise, too.

11. The implications in terms of the opportunity cost of housewives' time may not be a matter for rejoicing among those liberated members of the sex who have been calling for a "housewife's wage".

12. P. Samuelson: *Foundations of Economic Analysis;* Cambridge, Mass.; Howard U.P., 1947. pp. 36–39.

13. Money-price market clearing is equivalent to an idenfinitely large number of purchasers each queueing for an indefinitely small amount of time. Increasing the number of allotments drives down queueing time. Forbidding re-entry effectively prevents those placing a higher value on a second/further allotment from "bidding" for it with their time in the queue.