Economic Analysis of Non - Price Rationing With Particular Reference to Petrol

Occasional Paper

A thesis that no longer provokes comment is that domestic fuel supplies are likely to be curtailed suddenly and without warning. An idea less well established, but one that has widespread intuitive appeal, is that non-price rationing can claim to efficiently allocate resources. This idea arises out of the notion that during unexpected fuel shortages, the ability of individuals and firms to adjust might not be related to their contribution to the community's well being.

Economic Analysis of Non-Price Rationing

with Particular Reference to Petrol

Ross Parish
Monash University

AUSTRALIAN GOVERNMENT PUBLISHING SERVICE, CANBERRA 1981

ISBN 0 642 06794 5

Printed by Canberra Reprographic Printers, 119 Wollongong St, Fyshwick, A.C.T. 2609

FOREWORD

A thesis that no longer provokes comment is that domestic fuel supplies are likely to be curtailed suddenly and without warning. This is so well established that various governments in Australia have introduced legislation or revised administrative procedures for the central management of fuel supplies when events threaten their orderly distribution.

An idea less well established, but one that has widespread intuitive appeal, is that non-price rationing can claim to efficiently allocate resources. This idea arises out of the notion that during unexpected fuel shortages, the ability of individuals and firms to adjust might not be related to their contribution to the community's well being.

This latter notion accords with the prevailing belief in the community that governments should ration fuel during times of scarcity, if only to bring about an equitable distribution of available stocks.

Given that fuel may be rationed in some way, it is useful to examine the various forms of non-price rationing so as to discover their merits and flaws, particularly their costs, relative to price rationing.

This is the intent of Professor Parish's paper, wherein he examines in some detail both the efficiency and equity of proposed schemes. Although it is clear that much work remains to be done, the BTE believes that this report will make a useful contribution to the needed analysis of energy policies now being formulated.

R.H. HEACOCK
Acting Assistant Director
Planning and Technology Branch

Bureau of Transport Economics
Canberra
May 1981

CONTENTS

FIGURES

CHAPTER 1 - INTRODUCTION

Non-price rationing is very prevalent in our society. Many goods, services, and valuable privileges are allocated among demanders by criteria other than willingness to pay, or by a mixture of price and non-price rationing. Access to the courts is limited by long waiting times as well as by legal costs; medical services are rationed largely according to criteria determined by doctors; licences to enter certain closed industries (eg taxi ownership, commercial fishing) are allocated partly on the basis of the recipient meeting certain minimum criteria (eg being of good character), and partly on the basis of queuing (long service as an employee in the industry); quantitative restrictions are frequently imposed on markets, and quotas (eg import quotas, grain marketing quotas) are allocated according to historical market shares; access to some good schools is determined partly on a hereditary basis (preference given to children of 'old boys' or 'old girls'); access to some campsites in national parks is deliberately made difficult (eg by the closing of roads) in order to screen out those insufficiently dedicated to the rigours of outdoor life.

While most of the above examples are drawn from the public sector, private firms also not infrequently make use of non-price rationing. Price lists are often adhered to even though demand exceeds supply, and supplies are rationed partly by waiting time, or by giving preference to regular customers. For example there have been times when some new cars have been able to be resold at a substantial profit immediately after purchase; in periods of sudden disruption of petrol supplies, service stations often voluntarily impose transaction limits and/or give preference to regular customers.

Non-price rationing of goods normally rationed by price is also commonly imposed by governments in times of emergency, for example during wartime or in response to acute shortages (or threatened shortages) of some 'essential' good. It is with non-price rationing in this context that this paper is largely concerned. In particular it was motivated by the fact that, in response to the large rise in oil prices, and the disruptions of supplies from a major exporter, petrol rationing has been resorted to in varying degrees in many countries, and that Australia, or some parts of it, may experience oil supply disruptions from external or internal causes (eg strikes, production difficulties). However, much of the analysis is of a very general nature, and not restricted to non-price rationing in any particular context.

MOTIVES FOR NON-PRICE RATIONING

The reasons why firms, institutions, and governments adopt non-price rationing are various. Rather than to seek to find the market clearing price, firms may find it less costly to adhere to price lists and ration available supplies - at least for short periods, and within limits. They may perceive advantages in binding customers to them by 'looking after regular customers' in times of scarcity. Some trades are much more used to price fluctuations and coping with them than others. For example, potatoes and onions are often in short supply, but their scarcity does not become a shortage (that is, an inability by the consumer to obtain supplies). Rather, price adjusts, and the reason why this adjustment occurs relatively smoothly is presumably that the greengrocery trade, at all levels, is used to a great deal of price volatility. By contrast, petrol distributors are accustomed to a relatively stable price - only recently, with the growth of discounting, have retailers had to think about their pricing policies - and may find it difficult to achieve a price rationing solution to a sudden shortage. Their apparent readiness voluntarily to ration supplies in a shortage may simply reflect their unfamiliarity with widely fluctuating prices - though it must be admitted that it may be because they know from experience that governments often do impose non-price rationing and denounce as profiteers those who raise their prices.

Economists generally stress the efficiency benefits of price rationing. Allowing the price of a scarce commodity to rise to a market-clearing level (or levels, in spatially-separated markets) ensures that it goes to its highest-valued uses (as measured by consumers' willingness to pay for it) and that it will be moved to an appropriate extent, given transport costs, from areas of relative abundance to those of relative scarcity. It also provides an incentive for producers to increase their output, and for stockholders to run down their stocks before the scarcity is alleviated by new supplies.

Sometimes governments implement non-price rationing so as to replace the market-determined allocation by one intended to further some national purpose, over-riding individual preferences. Thus, during war, individual preferences are made subservient to the aim of national survival, and many market activities are replaced by direct controls. It would be pointless in these circumstances to evaluate the resulting allocation by reference to that achievable by the undirected price mechanism, since the validity of that norm requires the acceptance of individualistic values that have been deliberately rejected by the government. (This is not to say that economic criticism of non-price allocations is irrelevant. For example, it might be arguable that a market system extensively directed by taxes, subsidies and similar interventions would enable the aim of winning the war to be prosecuted more efficiently than is possible by direct controls.)

Usually, however, the government's aim in opting for non-price rationing is a less compelling one than national survival. The following are some of the more usual motives:

- to shift the burden of adjustment to the scarcity of some good or goods from some groups to other groups - for example, from the business sector to the household sector;

- to avoid or mitigate the income-distributional consequences of sudden and sharp rises in the price of a good;

- to suppress the symptoms of inflation; and

- to respond to a perceived demand from the electorate that the government does something positive in a crisis situation.

These motives do not involve the rejection of the efficiency arguments for price rationing. Rather, the pursuit of efficiency is seen to be in conflict with some other value or values, and the government is willing to compromise the former in order to pursue the latter. Thus it would be legitimate for the economist to argue, for example, that a method of rationing that achieved the income-distributional goal at small cost in efficiency should be preferred to a method that involved a large loss of efficiency.

This paper is concerned both to analyse how various methods of rationing work, and to evaluate them in terms of economic efficiency criteria. It thus involves both positive and normative analysis. For example, positive analysis of queue rationing involves a consideration of the factors determining the aggregate time spent queuing and its cost in money terms; also, whether, and in what circumstances and in what senses, queue rationing can be said to favour the poor. Normative analysis is conducted in terms of the criterion of aggregate economic surplus: that is, the various rationing schemes are ranked in accordance with the sum of consumers' and producers' surplus that each generates.

CHAPTER 2 - METHODS OF NON-PRICE RATIONING

The term 'non-price rationing' is intended to include all types of rationing where price is held below its market-clearing level, and not merely instances where the money price is zero. The price may be held down by price-fixing laws or by moral suasion. Black markets may come into being, particularly in the longer run, but for the most part this possibility will be ignored. Distribution, it is assumed, will take place through normal commercial channels, in accordance with instructions issued by governmental authorities, and that the distribution so achieved will not be modified subsequently by unofficial transactions between consumers.

There are two basic types of non-price rationing, although actual practice often involves some mixture of both. These are queue rationing, and allocation schemes. Queue rationing allocates the good on the basis of 'first-come, first-served', together, usually, with some constraint on the quantity transferred per transaction. It is essentially a market form of allocation, analogous to price rationing, except that 'willingness to queue' rather than 'willingness to pay' is the characteristic that is rewarded by the competitive process. Under allocation schemes, however, market competition is suppressed, and supplies (or the right to buy them at the fixed price) are allocated among individuals on the basis of some non-economic or quasi-economic principle, such as 'need', or 'equity', or 'essentiality'.

The feasibility and costs of different rationing schemes are affected by the nature of the rationed item and its demand. For example, some goods and services are ordinarily consumed only in unit quantity per consumer; hernia operations, seeing a particular movie, particular titles of books, or gramophone records are cases in point. If the demand for such a commodity exceeds the supply at the fixed price, there is no possibility of rationing it on a 'fair shares all round' basis. It is inevitable that some consumers will have their demand completely satisfied, while others will be completely excluded. On the other hand, if each demander ordinarily consumes many units of the commodity, it can be rationed in such a way that each consumer is given access to some of it. The storability of a commodity will affect the extent to which it might be hoarded or black marketed. Again, some goods have a natural transaction size - or upper limit to transaction size - which may obviate the need for, or otherwise affect, any constraint on transaction size imposed in connection with queue rationing. For example, the upper limit to the petrol transaction for most motorists is the car tankful - and for many this is also their maximum storage capacity. The latter can be raised, of course, by the purchase of drums, pumps, syphons, etc, but this involves some cost and inconvenience.

As compared with price rationing, non-price rationing involves three main types of cost, these are:

- administrative and enforcement costs;
- misallocation of the good; and
- real transaction costs incurred by consumers (especially queuing costs).

The costs of misallocation and of queuing are affected by the nature of the individual demand curves underlying the market demand curve. Differences in the valuations placed on different units of the good - and hence the slope of the market demand curve - arise from both differences among consuming units in their tastes and opportunities for consuming the good, and changes in each consumer's marginal valuation of a unit of the good as the quantity he consumes varies - or, in brief, from both interpersonal and intrapersonal differences in valuation. It will be convenient to consider two pure types, where the downward slope of the market demand curve derives solely, in Type 1, from interpersonal differences, and, in Type 2, from intrapersonal differences.

Type 1 is the one already mentioned above, where each potential consumer consumes only one unit of the good, but where different consumers value the unit differently. The market demand curve is then simply an array, in descending order, of all potential consumers' maximum willingness to pay for a unit of the good.

Type 2 requires the assumption that all consumers have identical, downward-sloping demand curves for the good. The market demand curve is then the individual demand curve with the quantity axis multiplied by the number of consumers.

Allocation schemes and queue rationing are now analysed in the context of those different basic demand structures.

ALLOCATION SCHEMES

It is not possible to consider allocation schemes exhaustively, since the criteria for and methods of allocation are so various. Some schemes attempt to mimic, in some degree, the allocation that would be achieved by price rationing. For example if productive inputs are allocated on the basis of priority for essential services, then, presumably, the result would bear some similarity to that of price rationing, since willingness to pay would be correlated with 'essentiality' (that is, provided the activities so designated really were essential). Other schemes are based on quite different criteria, and, indeed, are prompted precisely because they produce an allocation different from the price-rationing outcome. Examples include the allocation of medical services by 'need', as judged by doctors: rationing of essential foodstuffs on the basis of equal quantities per person - or per adult equivalent allocation of soldier settlement blocks, or taxi licences or the obligation to undertake National Service, by ballot.

In order to keep the analysis within manageable limits, only two simple cases will be considered, namely *random* allocation of the good among consumers willing to pay the fixed price, and *equal* allocation to all of these consumers.

Type 1 demand

Random allocation:
Suppose that the supply available is sufficient to meet only three-quarters of the quantity demanded at the fixed price. One fourth of consumers will have to be excluded from consuming the good. If these are selected randomly, they are equally likely to come from any point on the demand curve (above the fixed price) and, on average, they would be distributed uniformly along the curve. The average surplus accruing to those receiving the good, and lost by those excluded from consuming it, would be equal to the average surplus received per consumer in the pre-rationing situation, ie 1/2DP (see Figure 2.1). Since one fourth of consumers would be excluded, a quarter of

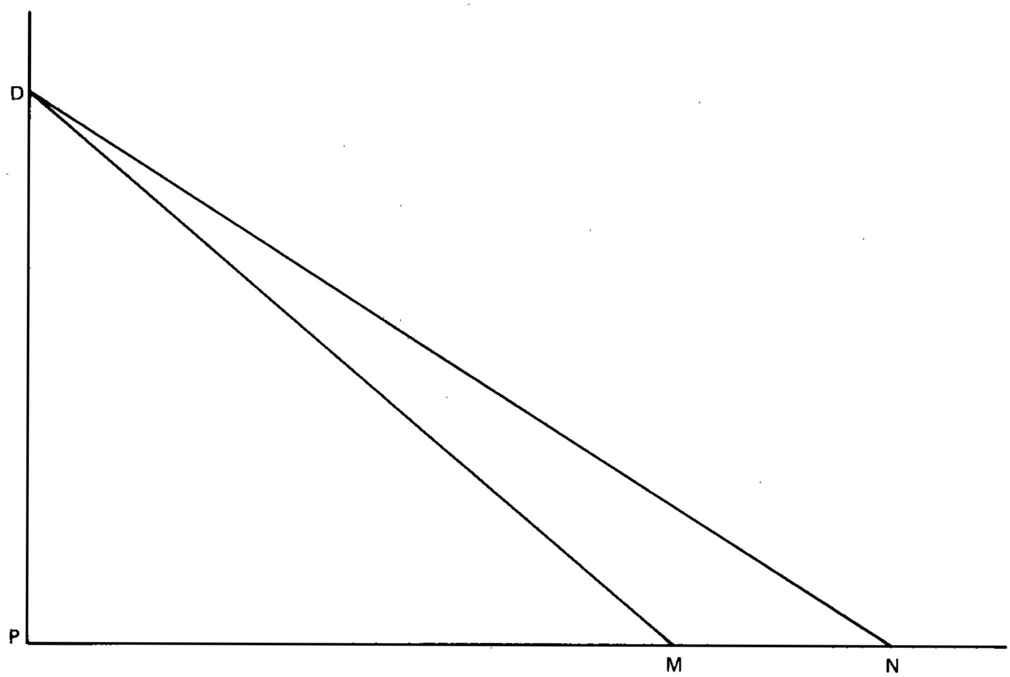

Figure 2.1 Effect of random allocation of supplies

the total surplus would be lost. This is shown by the triangle DNM in Figure 2.1, the area of which is 1/2DP.MN: since MN = 1/4PN, the area of DNM is 1/8DP.PN, which is a quarter of the whole pre-rationing surplus, 1/2DP.PN. In general, random allocation reduces the total pre-existing surplus in the same proportion as supply is reduced.

Equal allocation:

Despite what was said above, to the effect that equal allocation was inapplicable to a Type 1 demand structure, equal allocation would be possible if each consumer demanded one unit of the good, not once, but, say, once a week. Then an equal cut-back of one-fourth of his consumption could be achieved by requiring each consumer to go without his unit one week in every four. The loss of surplus would then be identical to the average loss under random allocation: one-fourth of the total surplus would be lost.

Type 2 demand

Random allocation:

With Type 2 demand, the meaning of random allocation is ambiguous. One possibility is that randomly-selected consumers are each given the opportunity of purchasing as much of the good as they wish. They would each then purchase their normal pre-shortage quantity, while the unlucky remainder would go without. The loss of surplus in this case would be the same as in the Type 1 demand situation discussed above, ie it would equal the area DNM in Figure 2.1, and, given that MN = 1/4PN, would equal one-fourth of the total pre-shortage surplus.

The other possibility is that, for each unit that is demanded, the probability that it will be satisfied is reduced from 1 to 0.75. Suppose, for simplicity, that each consumer wishes to buy four units at the fixed price He is obliged, however, to participate in a lottery four times, each time with probability of 0.75 that he will be permitted to buy one unit of the good. Consumers will thus be permitted to buy from 0 to 4 units, the fraction of the population in each category being as follows:

0	$\frac{1}{256}$
1	$\frac{12}{256}$
2	$\frac{54}{256}$
3	$\frac{108}{256}$
4	$\frac{81}{256}$

Suppose that the maximum amount, in addition to the fixed price, that each consumer was willing to pay for each unit as follows:

	$
First unit	4
Second unit	3
Third unit	2
Fourth unit	1(1)

Then the average loss of surplus, per consumer, under this rationing scheme, can be calculated to be $1.375, as follows:

Quantity consumed per consumer	Loss of surplus per consumer	Proportion of consumers	Average Loss
	$		$
0	10	.0039	.039
1	6	.0469	.281
2	3	.2109	.633
3	1	.4219	.422
4	0	.3164	0
			1.375

(1) The willingness of consumers to pay more than zero in addition to the fixed price for the fourth unit is not inconsistent with the postulated initial equilibrium, provided the good is more finely divisible than into the 'units' under discussion here.

By comparison, the first type of random rationing scheme mentioned above whereby the consumption of one-fourth of the consumers is reduced to zero, and that of three-fourths is unaffected, results in an average loss of surplus per consumer of $10 x 0.25 = $2.50.

The reason why the second scheme reduces the loss of surplus by almost a half is, of course, that when units of consumption (rather than consumers' whole entitlements) are withdrawn at random, the impact on the value of consumption is not random: the units eliminated tend to be the lower-valued ones, and thus the impact is softened.

Equal allocation:
If all consumers have identical demand curves, under price rationing each would consume the same quantity of the good. If price was allowed to ration supplies during the shortage, each consumer would cut back his consumption in the same proportion. Exactly the same result would be achieved by a rationing scheme that allocated the available supplies equally among all consumers. In this case there would therefore be no additional loss of surplus occasioned by the use of non-price rationing, as compared with the inescapable - and minimum - loss resulting from price rationing.

QUEUE RATIONING

Under queue rationing, consumption is discouraged both by the fixed money price and by the cost of having to wait in line(1). The price thus consists of two parts, the money price and the 'queue price'. It is usual to identify the queue price with the opportunity cost of the buyer's time spent in the queue, but it should be noted that queuing involves other costs as well:

---

(1) Rationing by queuing is distinguishable from rationing by waiting, where one has to join a waiting list, rather than wait in a line, in order to obtain supplies of the good or service. A third method, which might be called 'rationing by ordeal', requires the would-be consumer to perform some time-consuming and/or disagreeable task to become eligible to obtain supplies: making access to campsites difficult is a case in point, while another is limiting the use of water by banning the watering of gardens by sprinklers, but allowing the use of hand-held hoses. While the following analysis is confined to queue rationing proper, much of it applies to these related methods of rationing.

waiting in line may be an irksome way of spending one's time and may involve the risk of becoming engaged in altercations and even of being physically assaulted; also, queuing might involve direct money costs - as, for example, when the consumer has to take his automobile with him when queuing for petrol, and thus use petrol and incur other running costs. The consumer can be assumed to take these various cost elements into account in deciding how long he is willing to wait in line in order to buy a quantity of the good. His demand will thus be expressed through his willingness to queue as well as his willingness to pay the fixed price.

With supplies fixed, and an exogenously-fixed price, queuing time is the only variable available to equilibriate the market. It is possible to visualise a demand curve for the right to buy the good at the fixed money price, with the queue 'price' being measured in waiting time per unit bought. The equilibrium queuing time would then be established at the intersection of this demand curve with the vertical supply curve. If every individual's valuation of time (and the other elements of the cost of queuing) were the same, the time demand curve would be a direct translation of the ordinary demand curve to the left of its intersection with the supply curve. However, if people differ in their evaluation of time spent queuing, their relative willingness to pay in time will differ from their relative money evaluations of the good, so that the time demand curve will not be related in any simple way to the ordinary demand curve.

It is useful to distinguish two elements of queuing time - waiting time and transaction time. A reduction in the latter - say, by the use of faster petrol pumps or more attendants - would not, on the basis of the theory outlined above, result in a reduction in total queuing time; the reduction would be taken up by an increase in waiting time. This is because - and it is an important point - queuing time is a demand-determined market equilibrium price. Its magnitude is determined by the tastes and opportunities of consumers, and, in particular, by those of 'the marginal queuer'. This is not to say that the equilibrium queuing time cannot be influenced by regulations affecting transactions. For example, if transactions were limited to times when the opportunity cost of time was high, the equilibrium queuing time would be expected to be shorter (although the cost to consumers would be the same). Also, as will be shown below, constraining the size of transactions can also affect the time spent queuing. However simply making transactions more or less efficient should have no effect on the time costs incurred by consumers, since it does not affect their willingness to pay.

Queue rationing is customarily associated with some constraint on the size of each transaction - usually an upper limit on the amount that can be purchased. The purpose of such a limit might be to make it more difficult for potential scalpers to obtain supplies for resale on the black market; or it might be to help ensure that all consumers have some chance of obtaining at least a minimal amount of the good at the fixed price. Sometimes, however, a minimum transaction size is imposed. In this case the purpose might be to reduce sellers' transaction costs (sell out quickly and close the shop until the next batch of supplies is delivered); or, in the case of petrol, to discourage motorists from frequently topping up their tanks - a practice that increases the aggregate in-car inventory at a time of scarcity when hoarding should be discouraged. Another motive might be to reduce the amount of queuing: however, from our previous discussion it would seem unlikely that it would have this effect.

Type 1 demand

If each consumer demands only one unit of the good per unit time (and storage is impossible) neither maximum nor minimum limits on transaction size are necessary or feasible.

If the number of consumers has to be reduced to three-quarters of those willing to pay the money price, the equilibrium queuing time is the time that the marginal consumer is willing to queue. If the willingness to queue of all consumers is arrayed in descending order, thus to form the 'time' demand curve, the marginal consumer is the one located at the bottom of the third quartile. All other consumers will enjoy some consumer's surplus, in that their willingness to spend time queuing exceeds the amount of time they are obliged to spend. The result is analogous to price rationing, except that:

- the ordering of consumers along the demand curve does not necessarily correspond to the ordering on the ordinary demand curve, and the money value of the surplus, individually and in aggregate, may be different; and

- the time price paid represents a real cost, and not simply a transfer from buyers to sellers.

Type 2 demand

With queue rationing the possibility exists - as Barzel (1975) has pointed out - of the consumer being forced to operate on his all-or-nothing demand curve, rather than on his ordinary demand curve. This possibility arises because *the queue price is a price per transaction:* from the point of view of the individual queuer, it remains the same irrespective of the number of units he buys. Having paid the transaction 'time price', no rational consumer would buy less than the maximum amount he is prepared to buy at the fixed money price, or the maximum permitted amount, if the latter is smaller than the former. The queuing process confronts the consumer with the same choice as he has to make when a discriminating monopolist offers him 'x units for y dollars', or 'as many as you want for y dollars'. In responding to such offers the consumer compares the money price, or time-and-money price of the package with his total evaluation of it. The maximum price per unit that he is willing to pay is his average valuation per unit of the quantity offered, ie a point on his all-or-nothing demand curve. Hence it is possible that consumers will bid up the queue price to the point where all of the surplus associated with the transaction is dissipated in queuing time.

If a maximum transaction limit of one unit of the good is imposed - where a 'unit' (which may or may not be a 'natural' unit) is a small quantity relative to the consumer's equilibrium level of consumption - the consumer will be operating on (or approximately on) his ordinary demand curve. Each consumer will then be queuing more frequently than, say, once a week for his weekly supplies, and his decision to queue or not to queue (for a given queuing time) or his evaluation of the benefit to be gained or lost - and hence the time worth spending or saving - by queuing one additional or one fewer time per week will be based on the value of the surplus attaching to a marginal change in consumption. Since the surplus per marginal unit is less than the average surplus per unit, the willingness to queue, and hence the equilibrium queuing time, will be less with a low transaction limit than with a limit so large that each consumer would queue only once a week. *Low transaction limits, though they make for more frequent queuing, may mean less total time spent queuing per week, or per unit consumed.*

However, although a relatively large transaction limit is a necessary condition, it is not a sufficient condition for consumers to be forced into making all-or-nothing choices. For if the commodity is storable, consumers can regulate their rate of consumption as finely as they please, irrespective of the transaction size, by changing the frequency of transacting. For example, suppose that the consumer's demand function is stable on a weekly basis (ie his consumption follows a weekly rhythm, and repeats itself week after week), and the transaction limit is four units. For a rate of consumption of four units per week he will queue once a week, 4.33 times per month, or 12 times per quarter. A planned decision to queue one fewer time per quarter would not mean the loss of a whole week's consumption and the associated surplus, but a reduction in the weekly rate of consumption from four to 3.67 units, and a loss of surplus of 12 times the surplus associated with the last third of the fourth weekly unit (or, if the good can only be consumed in discrete units, four times the surplus associated with the fourth unit).

A similar argument would apply if, although the good were perishable, the demand for it were, so to speak, storable. If the uses to which the good was put could be postponed or brought forward without loss of utility, and if they yielded surplus at varying rates per unit of the good, a failure to queue would result in the loss of the least valuable uses.

An element of all-or-nothing choice will be present whenever transaction size influences the willingness to pay per unit for the units comprising a marginal transaction, ie whenever it is not possible to reallocate these units - or the demand for them - over time so freely as to ensure that they constitute marginal units. If the demand curve is stable from week to week, the divisibility in consumption of the batch of goods comprising one transaction depends on the number of weeks over which consumption of a batch can be averaged out. Call this the planning period. With a transaction size of four units, and a four-week planning period, consumption can be adjusted as finely as by one unit more or less per week; with a two-week planning period, by two units; and with a one-week period, by four units. The consequences for demand are illustrated in Figure 2.2, which shows an individual's weekly 'time' demand for a good, AB, and the four-weekly demand curve, AC. (Both of these curves are drawn on the assumption that the good may be divided in

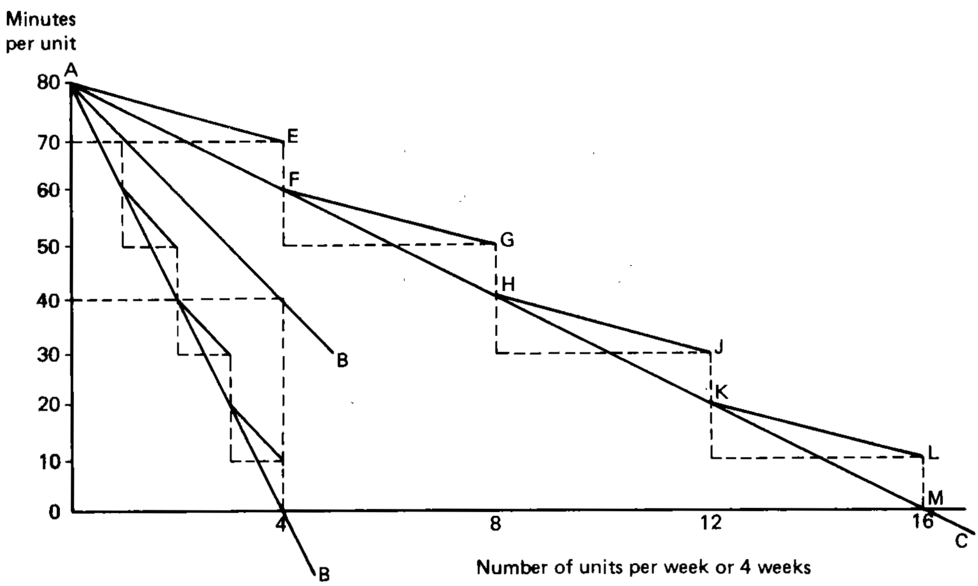

Figure 2.2 Effect of planning period on demand

consumption more finely than into the 'units' on the horizontal axis. These may be interpreted as some arbitrary or customary unit of purchase). With a transaction size of four units, and a four-weekly planning period, the four-weekly curve AC is transformed into the dotted step-function. This translates into the analogous step-function lying along BA, for the weekly demand curve. The lines AE, FG, etc, and their analogues originating along AB, are segments of all-or-nothing demand curves drawn on the assumption that the consumer initially has none (AE), or four units per month (FG), or eight units per month (HJ), etc - and, along AB, none, one, two, etc, units per week.

AB is the all-or-nothing weekly demand curve with a one-week planning period. If offered the choice between four units and nothing, the consumer would be willing to queue for up to 40 minutes per unit so as to be able to buy four units. By contrast, with the four-week planning period, he would be willing to queue for a maximum of only 10 minutes per unit to obtain his fourth unit per week. Since equilibrium queuing time would be determined by the willingness to queue for the marginal unit, total queuing time would be up to four times as great with the shorter than with the longer planning period. A reduction in transaction size from four units to one unit would have the same effect as quadrupling the planning period.

The length of the planning period is determined by technical, institutional, and psychological factors. For example, if a perishable good with a life of one week were distributed by queuing every Monday, the planning period would be restricted to one week, by virtue of the perishability of the good and the timing of distribution, which together ensure that none of the good could be carried over from one week to the next. The width of the steps in the demand function would be equal to the transaction limit. The same would be true of once-only distributions, such as queues for tickets to sporting events, concerts, etc. A shortage might be expected to be alleviated within a given time - eg the time elapsing until the new harvest came in - which would provide an upper limit to the planning period. If the shortage were of uncertain severity and duration, and with supplies becoming available only intermittently, consumers would tend to regard each distribution as, in effect, a once-only event. With no assurance as to the availability of future supplies, each consumer would have to decide how much time to spend obtaining supplies now for use in a future of uncertain duration. Expectations and risk preferences, together with his consumption preferences and wealth, would determine his demand functions for supplies *now*.

In general, then, in some situations the maximum permitted transaction size will affect the willingness to queue, and hence the equilibrium queuing time of consumers comprising a Type 2 demand structure, while in other situations the transaction limit will be irrelevant. In the latter case, the outcome is similar to that obtaining with a Type 1 demand structure - the equilibrium queuing time per unit will be the time consumers are willing to spend in order to buy the marginal unit, and the queue price is determined by the intersection of the time demand curve with the supply curve. In the former case, the larger the transaction size relative to equilibrium consumption over the planning period, the more exposed is the consumer to all-or-nothing choice situations, the greater his willingness to queue per unit obtained, and the greater the loss of surplus. In the extreme case, the whole of the area under the demand curve could be disipated in time spent queuing.

Thus, in the best case, the costs of queue rationing are equal to the familiar 'welfare triangle' plus the 'expenditure rectangle' associated with price rationing. At worst, they are equal to the whole of the area under the demand curve, above the fixed price. Type 1 demand structures belong to the best case, since, if each consumer demands only one unit, the all-or-nothing and ordinary demand curves coincide.

COMPARISON OF RATIONING METHODS WITH RESPECT TO LOSS OF ECONOMIC SURPLUS

The results of the preceding discussion are summarised schematically in Figure 2.3 in which the hatched areas represent the loss of surplus for each method of rationing and each demand structure. Note that the demand curves for queue rationing relate to the time price of the good, and are directly comparable with the other curves only on the assumption that all consumers value time equally.

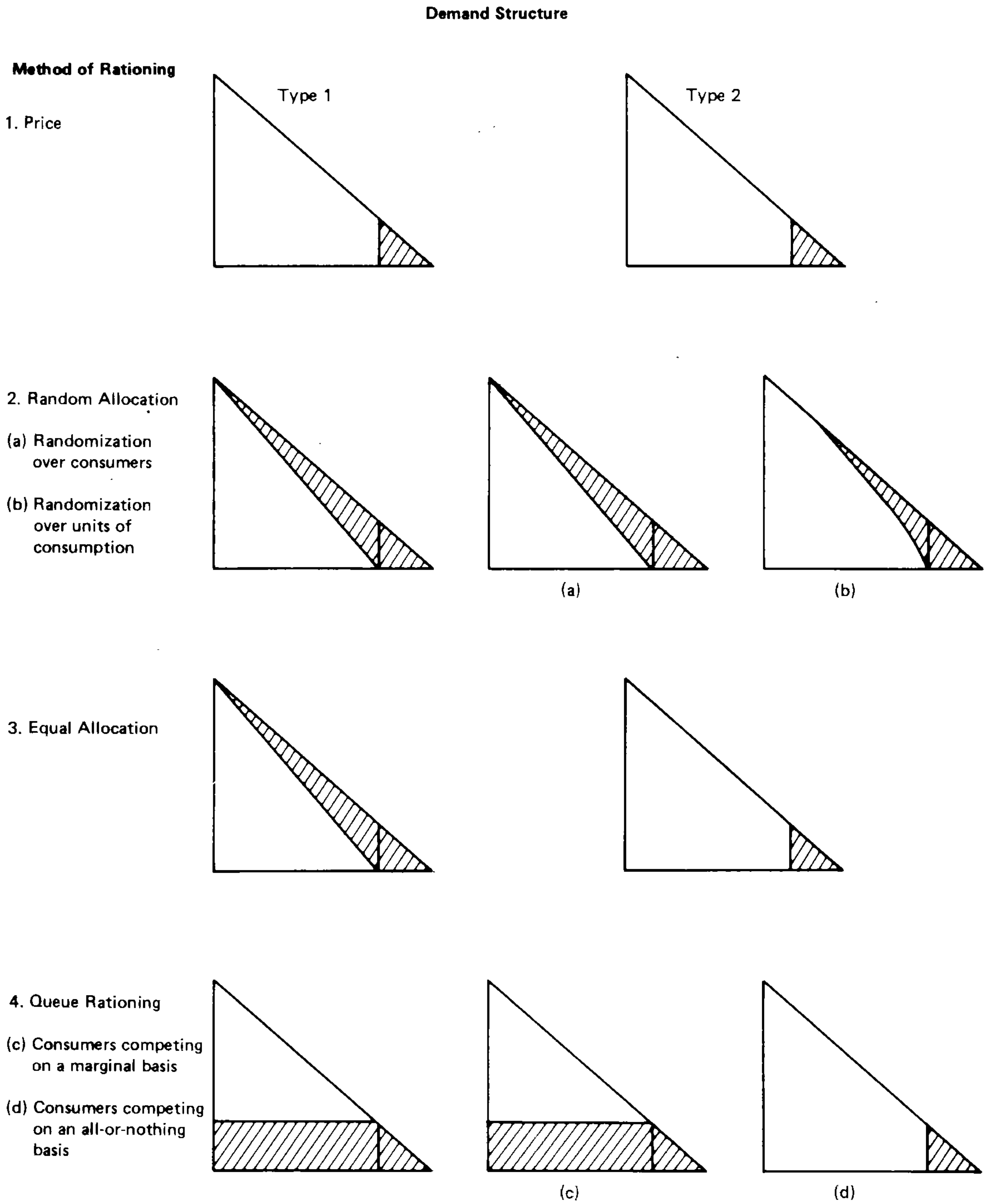

Figure 2.3 Loss of surplus under different rationing methods

The simpler demand structure, Type 1, gives the more clear-cut results, the principal one being that, subject to the just-mentioned caveat about the time demand curve, and for the case of a linear demand curve, queue rationing gives rise to precisely twice as much 'excess burden' as does random allocation, where excess burden refers to the loss of surplus over and above the inescapable minimum loss which is associated with price rationing(1). This result is reinforced by considering the Type 2 demand structure, in which the queue rationing burden can be greater, and the random allocation burden less, than for the Type 1 structure. It is therefore perhaps worthwhile to consider it in a little detail.

Linear case

For a linear demand curve, if the fraction whereby consumption has to be reduced (as compared with the quantity demanded at the pre-existing price) is k, then the fractions of the surplus lost under the different rationing schemes are as follows:

$$P/S = k^2$$
$$Q/S = 2k - k^2$$
$$R/S = k$$

where P is the loss of surplus under price rationing,
Q is the loss of surplus under queue rationing,
R is the loss of surplus under random allocation, and
S is the total initial consumer surplus(2).

The losses of surplus under the non-price rationing schemes, expressed as a proportion of the loss with price rationing, are

(1) The excess burden under queue rationing is the 'expenditure' rectangle, while that under random allocation is a triangle whose base is one side of, and height is equal to the other side of, that rectangle.
(2) Under price rationing the loss of *consumer surplus* is the welfare triangle, k^2, plus the expenditure rectangle $2(k-k^2)$. However, the latter is a corresponding gain in producer surplus, so that the loss of *total surplus* is only k^2.

$$R/P = \frac{1}{k}$$

$$Q/P = \frac{2}{k} - 1 \quad .$$

It is evident that as the shortage becomes more severe, the relative disadvantage of the non-price rationing schemes falls. As k approaches 1, the loss of surplus approaches 100 per cent, under all three types of rationing.

The relative 'excess burdens' of the non-price rationing schemes, that is, their extra losses of surplus as compared with price rationing, expressed as a proportion of the total initial surplus, are

$$\frac{R-P}{S} = k - k^2$$

$$\frac{Q-P}{S} = 2(k - k^2) \quad .$$

These quantities reach a maximum when k = 0.5. Thus the *additional* losses associated with non-price rationing increase, in absolute size, as k increases to 0.5, and diminish thereafter. (This is obvious from the fact that the expenditure rectangle (= the excess burden of queue rationing) is greatest at the mid-point of the demand curve.)

Hence, the non-price rationing methods are at their greatest disadvantage, relative to price rationing, when the reduction in supply is small; however, the losses of surplus for small reductions in supply are also relatively small, for all types of rationing. Conversely, with a severe shortage, the losses of surplus are great for all types of rationing, and the non-price methods are then relatively least disadvantageous compared with price rationing. The non-price methods impose their greatest additional costs for supply reductions of 50 per cent.

Non-linear cases

The above results hold only for linear demand curves. However, it is obvious that with any demand curve the loss of surplus under price rationing increases, at an increasing rate, from zero to 100 per cent as k increases from 0 to 1. The excess burden of non-price rationing must therefore eventually diminish and approach zero as k approaches 1. Also, the proportional loss of surplus with random allocation is always equal to k, irrespective of the shape of the demand curve. For curves that are concave to the origin, the excess burden of queue rationing and random allocation will reach their peaks at lower values of k than 0.5, and for convex curves, at higher values(1).

One category of convex curves, namely the constant elasticity type, is of particular interest, on account of its familiar and convenient properties, and also because a constant proportional demand response is generally regarded as being a more plausible assumption than a linear response. For demand functions of the form $q = ap^{-\alpha}$, where α is the elasticity of demand, the loss of surplus associated with the different types of rationing are as follows(2):

$$P/S = 1 - \alpha(1-k)^{1-\frac{1}{\alpha}} + (\alpha-1)(1-k)$$

$$Q/S = 1 - (1-k)^{1-\frac{1}{\alpha}}$$

$$R/S = k$$

The excess burdens of the non-price rationing methods as compared with price rationing, are

---

(1) Q-P reaches a maximum where the elasticity of the demand curve is -1. For a convex curve, this point corresponds to a quantity closer to the origin than to the quantity axis intercept, and for a concave curve, to a quantity closer to the intercept. This can be seen by drawing a line, intercepting both axes, and tangent to the curve, such that the point of tangency is midway between the line's two intercepts. (The author is indebeted to M. Burns for clarification of this point.)

(2) Proofs are given in Appendix I.

$$(Q-P)/S = (\alpha-1)\left[(1-k)^{1-\frac{1}{\alpha}} - (1-k)\right]$$

$$(R-P)/S = \alpha\left[(1-k)^{1-\frac{1}{\alpha}} - (1-k)\right] .$$

Since the surplus associated with unit-elastic or inelastic curves is infinite, these results apply only to cases where $\alpha > 1$. The relative excess burdens are in marked contrast to those arising in the linear case: there, queue rationing gave rise to twice as much excess burden as random allocation; here, random allocation has the larger excess burden, and the difference is greater, the lower the elasticity (in absolute value). For an elasticity of -2, the excess burden of random allocation is twice as great as that of queue rationing - precisely the reverse of the relationship existing in the linear case. As the demand elasticity increases, the difference between the excess burdens diminishes, and approaches zero as α approaches infinity.

The loss of surplus, as a function of k, is plotted in Figures 2.4 and 2.5, for the linear and log-linear cases respectively. For price and queue rationing, the loss of surplus, for all values of k less than 1 is lower for log-linear than for linear demand curves. One way of explaining this result is to note that, with convex curves, a much greater part of the surplus is accounted for by small quantities of the good valued at very high prices, and much of this remains intact, with queue rationing (and all remains intact with price rationing) unless the shortage is extremely severe. By contrast, random allocation makes inroads on the surplus equally throughout the length of the demand curve. Hence random allocation performs very badly if demand curves are of the constant elasticity type.

To summarise our findings to this point: random allocation has been compared with queue rationing on the assumption that all consumers value time equally. Queue rationing may, or may not, involve what is, in effect, price discrimination. If it does not, its excess costs (as compared with price rationing) are twice as great as those associated with random allocation, if the demand curve is linear. However if the demand is of constant elasticity (and the elasticity greater than unity) queue rationing involves a smaller excess burden than random allocation. If the demand elasticity is -2 or less (absolutely), the excess burden of random allocation is twice (or more than twice) as great as that of queue rationing.

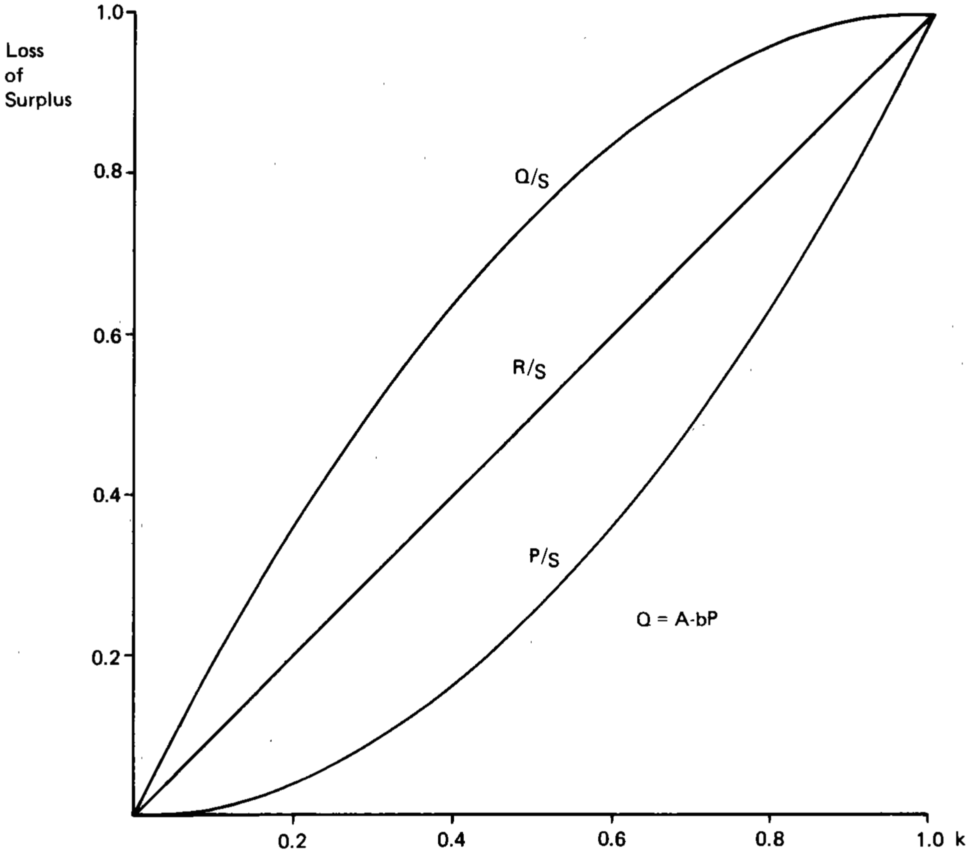

Figure 2.4 Proportion of initial surplus lost as initial supply reduced by proportion k, under price rationing (P/S), queue rationing (Q/S), and random allocation (R/S), with a linear demand curve.

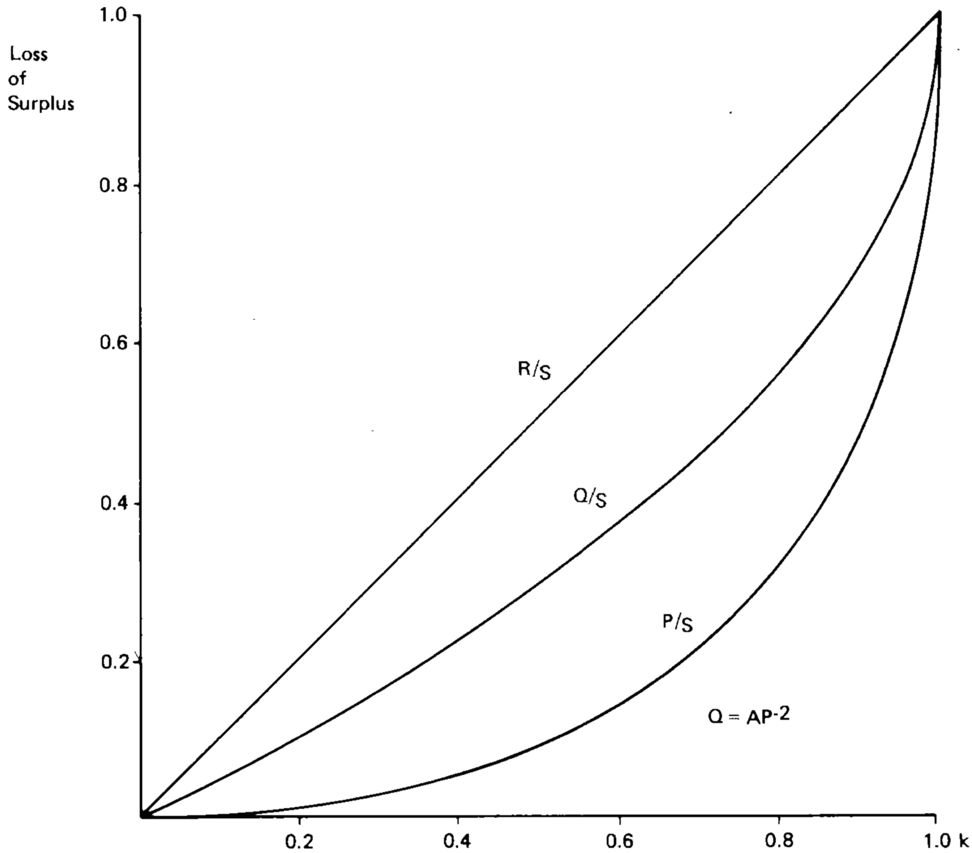

Figure 2.5 **Proportion of initial surplus lost as initial supply reduced by proportion k, under price rationing (P/S), queue rationing (Q/S), and random allocation (R/S), with a demand curve of elasticity −2.**

If queue rationing involves 'price discrimination', its excess burden may be much greater. In the extreme case, virtually all of the surplus may be lost, irrespective of the degree of supply reduction, or of the shape of the demand curve.

The principal conclusion to be drawn is the importance of having knowledge of the demand curve throughout its length if practical applications of these findings are to be made. This is in contrast to the normal situation in applied economics, where it is sufficient to know how demand behaves over the range of normal fluctuations in supply. As a result of this circumstance, little is known about the shapes of demand curves in their higher reaches, thus limiting the usefulness of these findings.

All of the foregoing is based on a model of queue rationing that abstracts from an important aspect of reality, namely variation in consumers' valuation of time. Also, no account has been taken of administration and enforcement costs of the rationing schemes. These matters will be discussed in the next two sections.

QUEUE RATIONING WITH NON-UNIFORM VALUATION OF TIME

Variation among consumers in their evaluation of time may or may not be related systematically to variation in their valuation of the rationed good. Economic theory suggests that willingness to pay for the good and the value of time would generally be positively correlated, since both would tend to increase with income (inferior goods being the obvious exception). However, there are reasons (see Chapters 3 and 4) to believe that the correlation between the value of time and income may be rather weak. Similarly, willingness to pay for a good is affected by numerous factors, usually subsumed under the rubric of differences in taste, in addition to capacity to pay, or income. Hence while it would be expected that, except for inferior goods, the value of time would be positively correlated with willingness to pay for the good, it would also be expected that there would be a good deal of variation in the value of time that was uncorrelated with willingness to pay. The effects on the costs of queueing of the systematic and random components of variation in the value of time will be discussed separately.

Systematic variation

It is clear that insofar as the value of time is positively related to willingness to pay for the good, the loss of surplus under queue rationing will be greater than if all consumers valued time uniformly. Suppose, for example, that as one moves up the Type 1 demand curve, each successive consumer's valuation of time increases, but less than proportionately to the successive increases in the value of the good. The ordering of consumers by willingness to queue will be the same as their ordering by willingness to pay, so that the same consumers will be successful in obtaining the good under queue as under price rationing, and all will pay the same time price, measured in hours. However when the time price is measured in dollars, successive intramarginal consumers will be paying successively higher prices for the good, so that the loss of surplus will be higher than if all consumers valued time equally.

If the valuation of time rose faster than the valuation of the good, the ordering of consumers by their willingness to queue for a discrete quantity of the good would be the reverse of their ordering by willingness to pay. Those who, under price rationing (or queue rationing with a uniform valuation of time) would enjoy the greatest surplus, would now be excluded from consumption. Since the largest surplus would accrue to the consumer with the lowest willingness to pay, this, and the total surplus, must necessarily be very small.

The essential point is that if, among consumers, the valuation of time is positively related to the valuation of the good, so is the dollar equivalent of the time price paid. Those who would otherwise enjoy a high consumer surplus will have it reduced, or, if they refuse to pay the price, eliminated entirely.

Random variation

Variation in the value of time that is random with respect to willingness to pay for the good has two effects on the consumer surplus. First, those with a relatively low value of time will tend to queue, and this will tend to reduce the cost of queuing. Second, since a low valuation of time can compensate for a low valuation of the good in determining willingness to queue, some who queue will value the good less than some who are excluded from consumption, so that some misallocation of the good will occur. The task is to determine which of these effects predominates.

If it is assumed that the time valuations observed among a group of consumers take on a number of discrete values, the effect is that of dividing the group into a number of sub-groups, one for each time value. Whatever equilibrium queuing time is established, each sub-group will face a different money-equivalent time price for the good, this being the queuing time multiplied by the sub-group's valuation of time. If the value of time varies randomly with respect to willingness to pay for the good, the expectation is that each sub-group's demand curve for the good will be a horizontally-contracted replica of the group demand curve, the degree of contraction depending on the proportion of the whole group belonging to the particular sub-group. To take the simplest case: assume the group consists of two sub-groups, equal in size, one having twice as high a valuation of time as the other. The group demand curve would be expected, on average, to divide into two identical demand curves, as shown in Figure 2.6.

The equilibrium queuing time would translate into a money-price equivalent that was twice as high for one sub-group as for the other. Thus rearranging the two demand curves back-to-back the result is as indicated in Figure 2.7.

Now suppose a uniform time valuation, equal to the existing average valuation, was established for the whole group. Both sub-groups would now be identical in all respects and each would consume one half of the fixed quantity of the good at an equilibrium money-price equivalent of $P_u = 1/2\ (P_h + P_\ell)$.

Comparing the consumers' surplus in the two situations, it is evident that it is greater when the sub-groups value time differently than when they have the same valuation. As compared with the latter situation, the gain of surplus by the low time value sub-group - the strip marked with plus signs - exceeds the loss of surplus by the high time value sub-group - the strip delineated with minus signs.

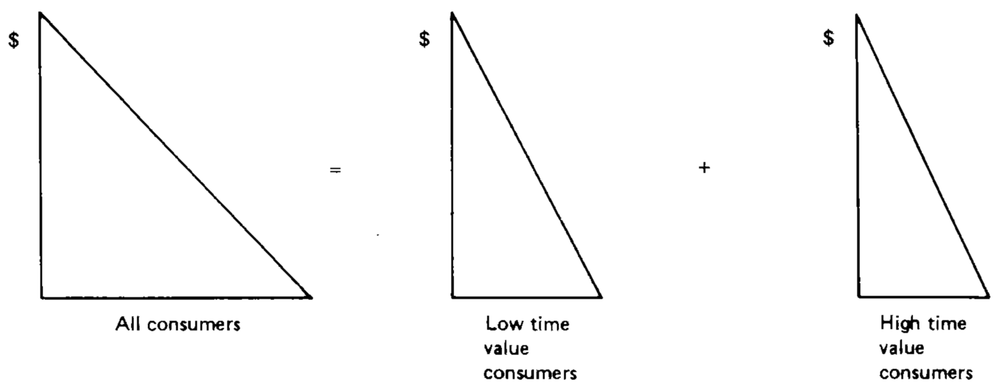

Figure 2.6 Division of demand into sub-groups

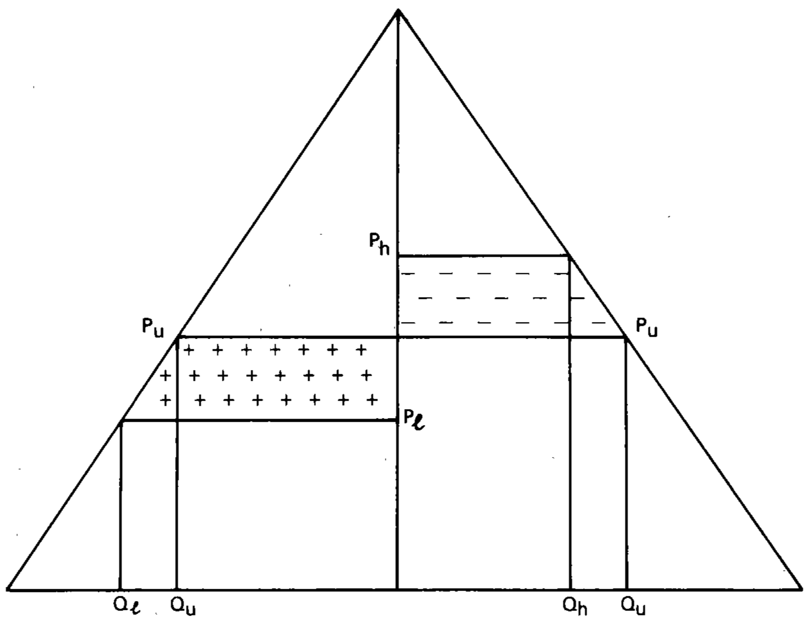

Figure 2.7 Effect of time valuation on surplus

This result is essentially the same as that demonstrated by Waugh many years ago. Waugh (1944) showed that the stabilisation of a fluctuating price at its arithmetic mean made consumers worse off. His analysis was similar to the above, except that the two sub-groups were the same group at different times.

The reason that these results hold, in Waugh's case, is that the fluctuations in price give consumers opportunities to concentrate their purchases in periods when price is low, opportunities that are denied when price is stabilised. In the present case, variation in the effective prices facing different consumers allows greater specialisation in consumption by those consumers for whom the price is low.

The foregoing argument is based on a linear demand curve and only two value-of-time sub-groups, of equal size. Increasing the number of sub-groups, and allowing them to vary in size, do not alter this conclusion. However, restrictions do have to be placed on the shape of the demand curve for Waugh's result - and the analogous finding - to hold.

The means whereby price stabilisation is achieved is the stabilisation of quantity. Complete stabilisation of a fluctuating quantity by stock management implies stabilisation at its arithmetic mean(1). With a linear demand curve, price is also stabilised at its arithmetic mean; with a concave curve, the stabilised price is higher than the arithmetic mean, and, with a convex curve lower. It is clear that, in the concave case, Waugh's proposition holds *a fortiori*: the strip of surplus which consumers gain when price falls is both wider and longer than that which they lose when price rises. However it is also clear that if the demand curve is sufficiently convex, Waugh's result would be negated or reversed. It transpires that <u>any</u> constant elasticity demand curve is sufficiently convex to reverse Waugh's finding, ie, stabilisation of quantities sold on a constant elasticity demand curve <u>increases</u> consumers' surplus (Turnovsky 1976). Analogously,

---

(1) Assuming that stocks are not accumulated indefinitely, and that exogenous stocks are not available.

the loss of surplus with queue rationing will be less if the value of time is uniform across consumers than if it varies randomly about a mean equal to the uniform value.

Conclusions

Making allowance for variability of consumers' valuation of time greatly complicates the analysis of queue rationing, and, generally speaking, shows its welfare costs to be greater than indicated by the earlier analysis.

To the extent that variations in the value of time are positively related to variations in consumers' valuations of the good, the costs of queue rationing are unambiguously increased.

The sources of the additional costs are twofold.

Firstly, if the increases in the value of time associated with increases in the valuation of the good are not great enough to alter the allocation of the good among consumers - as compared with price rationing - consumers suffer as a result of what is, in effect, discrimination with respect to the queue price (this type of 'price discrimination' is quite distinct from that associated with non-marginal transaction quantities, analysed earlier).

Secondly, if the increases in the value of time associated with increases in the value of the good are large enough to cause the ordering of consumers by their willingness to queue to be the reverse of their ordering by willingness to pay, then an exceptionally perverse misallocation of the good occurs. The misallocation is also accompanied by the type of price discriminating just described.

Random variation in the value of time has different effects on the costs of queue rationing, depending on the shape of the demand curve, lowering them for linear, concave, and slightly convex curves, but raising them for strongly convex curves, including all constant-elasticity curves. Since convex curves are generally thought to be more likely to occur than linear or concave curves, this finding is adverse for queue rationing.

These results are subject to an important proviso: they assume that the consumer does his own queuing. If the act of queuing can be contracted out to others, it would be expected that persons with a high value of time would pay those with a low value of time to queue for them. Alternatively, members of the latter group might take over the queuing process and resell the product to other consumers. In either case the results of the earlier analysis, based on the assumption of a uniform valuation of time, would be reinstated. Indeed, the loss of surplus could be even less. If the supply of professional queuers was less than perfectly elastic, then part of the 'expenditure rectangle' would be composed of producers' surplus.

ADMINISTRATION AND ENFORCEMENT COSTS

Administration and enforcement costs, it would seem, are likely to be greater with allocation schemes than with queue rationing. Allocation schemes involve identifying consumers, allocating supplies (or the right to buy supplies) among them, and enforcement of these allocations. With queue rationing, on the other hand, consumers readily identify themselves by their willingness to queue, allocations (in the form of a transaction limit) can be easily enforced, while orderly queuing (in countries with British traditions, at any rate) is usually enforced by the queuers themselves.

Allocation schemes have substantial setting-up costs. They are therefore more appropriate for extended periods of rationing, such as in war time, than for shortages that are expected to be of limited duration.

The costs of administering an allocation scheme will depend on how ambitious it is. For example, a rough-and-ready allocation component of a mixed scheme of rationing would be to give 'essential users' of the good priviliged access to the available supplies. 'Non-essential' users could be allowed to compete for the remaining supplies by queuing. This involves the designation of certain supply points to serve essential users only, and the issuing to the latter of some means of identification. (The improvisation of such schemes is a common reaction by the authorities to emergency shortages, and this method of distribution is a permanent feature of the Soviet Union, where high officials are allowed to purchase goods in special shops that are closed to ordinary citizens.) In its simplest form, essential users would be allowed to purchase supplies *ad lib*, but if it was desired to limit their purchases, these would have to be recorded, perhaps on their identity cards, or on specially printed ration books.

Such a rough-and-ready scheme would be much less costly to set up and operate than fully-fledged coupon rationing, applied to all, or virtually all, consumers. The latter involves, *inter alia*, the printing and issuing of ration books; the determination and enforcement of eligibility criteria; discretionary decision making regarding special cases, making appropriate allowances for wastage of the good between different points in the distribution chain; the provision of a 'working capital' of coupons to retailers, wholesalers, etc; and setting and adjusting the value of coupons in relation to the anticipated quantity of available supplies. (For an interesting account of these and other problems encountered in Australia during the Second World War, the reader is referred to Butlin and Schedvin, 1977, Chapter 11).

CHAPTER 3 - MANAGEMENT OF STOCKS

If it is known that output of a product will be reduced for a period, it is desirable that consumption be reduced now and stocks accumulated until the onset of the production cut-back, and that stocks be run down during the period of reduced production. Such a pattern of response will minimise the disruption accompanying the (assumed unavoidable) reduction in supply. The earlier the shortage is anticipated, the longer the period over which the reduction in consumption is spread, and hence the smaller the cut-back in the rate of consumption in any period. Similarly, the higher the level of initial stocks, and the more elastic the supply of storage, the easier it is to smooth consumption in the face of fluctuating production.

The pattern of price changes consistent with the above responses is as follows:

- an immediate rise in the price of the good once it is known that the supply disruption will occur;
- a continuing rise in the price during the shortage; and
- an immediate fall in the price back to normal levels once normal production is resumed.

The first ensures that conservation of the good for use during the supply cut-back commences as soon as possible; the second is necessary to provide an incentive for stocks to be held until the end of the shortage - the required rate of price increase being that which is sufficient to cover the marginal costs of storage; and the third is needed to ensure that only minimal stocks are carried over from the period of shortage into the succeeding period of normal supplies.

A well-functioning market mechanism will bring about such a pattern of price change. Inter-temporal price spreads will reflect the marginal costs of storage, so long as it is expected that stocks will be carried continuously. The expectation of excess demand during the shortage period will thus be expressed in the market as soon as the expectation is formed, thus causing the price to rise. Expectation of the resumption of normal supplies will ensure that the forward or futures price for dates after the expected resumption will be at normal levels.

One disadvantage of non-price rationing is that the price signals that would otherwise facilitate rational stockholding decisions are suppressed. Producers and distributors have no direct price (profit) incentive to build up stocks in anticipation of the shortage or to run them down in anticipation of the resumption of normal supply. However, they would presumably have some indirect incentive to do so, such as a desire to 'look after' - and be seen to be looking after - their customers.

Consumers have an incentive to accumulate stocks in anticipation of a shortage. During the shortage they also have an incentive to use up stocks, since retaining them would be at the expense of relatively high-priority consumption uses. However, a consumer fortunate enough to be able to obtain adequate supplies would suffer no penalty from keeping large stocks on hand or from carrying them over from the period of shortage.

The principal differences between non-price and price rationing, with respect to stock management are, then, the incentives to accumulate and to rundown stocks remain with consumers, and are not communicated, via price changes, to middle-men and merchants; as a result, the stock management function will be shifted at least in part from producers and merchants to consumers, and specialist expertise will not be utilised fully; the lack of price changes will mean that decisions regarding stocks will be largely unco-ordinated: each economic agent will be acting more or less in isolation, having regard only to his own situation and expectations. Since, especially with allocation schemes, the shortage will impinge with differential severity on different users, the marginal productivity of stocks will vary substantially among users. Thus in addition to the misallocation of consumption, the costs of non-price rationing include the mismanagement of stocks.

A tendency for the de-specialisation and fragmentation of the stock management function would seem to be an inevitable concomitant of the suppression of the price movements that would occur naturally in a shortage situation. Also the tendency is inevitably deplored. However, in its defence, it can be argued, that the accumulation of stocks by consumers in correct anticipation of a shortage is not undesirable; that dispersed and unco-ordinated stock management is better than no stock management at all and that the dispersion of stocks in small quantities among final consumers is no barrier to rational stock management if price is allowed to adjust. For example, if the price is allowed to rise and is expected to fall when the shortage is over, consumers will tend to run down their stocks and not carry them into the period of relative abundance.

In the case of petrol, the principal storage capacity outside of the normal production and distribution system is that of commercial users of petroleum products, eg road transport firms, taxi depots, farmers, etc. It seems likely that at least some of these users are at present keeping their stocks high and installing additional storage capacity. If shortages become more probable or more imminent private motorists might be expected to invest in out-of-car storage facilities. At the onset of a shortage, and so long as uncertainty exists about future supplies, motorists also tend to keep their tanks topped up.

Topping up is a natural response to a situation of uncertainty about the availability of future supplies. Under price rationing topping up would be expected to occur if the price were expected to rise. With non-price rationing there is a greater fear that supplies will become unavailable (infinite price) or very hard to get, and hence a greater incentive to top up. Motorists resorting to frequent topping up would increase their average in-car petrol stock by probably no more than a quarter of a tankful. (This assumes a change in the timing of refilling from when the tank is a quarter full to when it is three-quarters full). This is a once-and-for-all effect. In a shortage such that private motorists curtailed their mileage by about one half, a quarter of a tank of petrol would represent little more than half a week's consumption(1). It seems likely that notice of the impending resumption of normal supplies would be sufficient to allow this excess stock to be run down. Complaints about topping up would seem to be exaggerated.

---

(1) Assuming that the 'average' motorist travels about 320 kilometres per week, and that his car's range is about 400 kilometres.

An externality argument can be mounted against the storage of petrol on domestic premises and in loose cans in cars, on the grounds that it represents a fire risk. The most effective disincentive to private storage would be price rationing, where both the high price of petrol and expectation of its continued availability would discourage stock holding. With non-price rationing, there would be less incentive to hold out-of-car stocks if topping up were not discouraged, and if restrictions were not imposed on the times when petrol could be sold.

CHAPTER 4 - RATIONING AND INCOME DISTRIBUTION

The most common reason for price controls and other interventions in response to substantial disruptions of supply is the desire to alter the distribution of the resulting gains and losses. In the case of a consumer good, consumers suffer because they consume less of it and pay a higher price for the supplies they do obtain; sellers lose because of the reduced volume of sales, but gain from the higher price they are able to charge. If demand is inelastic, sellers are net gainers, and the more inelastic the demand, the greater the transfer from consumers to sellers. This transfer is frequently deemed to be undesirable, and price control is imposed in order to prevent it. Condemnation of 'profiteering' by sellers in these circumstances may stem from envy, or from a variety of moral considerations, including the notion of a 'just price', the idea that it is immoral for anyone to benefit as a result of a general misfortune, and the view that such profits are 'unearned'.

Price control is also sometimes rationalised on the more pragmatic grounds that allowing prices to rise in periods when supply is fixed serves no useful purpose. This view reflects an inadequate understanding of the function of the price mechanism, in that it acknowledges the role of price in directing production, but ignores its functions of allocating supplies to the highest-valued uses and of promoting efficient management of stocks. Whatever its psychological origin, or however it is rationalised, opposition to producers and sellers profiting from scarcity is a deeply-rooted attitude which often provides the prime motivation for the imposition of price control and the consequent need for non-price rationing.

If demand for the good is inelastic with respect to income, the income effects of a rise in its price are greater, as a proportion of initial real income, for low-income than for high-income consumers. The poor are thus burdened disproportionately by the price rise. They may also bear a disproportionate share of the cost of adjustment by cutting back more severely on their consumption of the good. This regressive incidence of the costs of scarcity is another frequent reason for intervention in the market, by way of price control together with rationing via either an egalitarian allocation scheme, or queue rationing - the latter being generally considered to favour the poor, since their valuation of time is likely to be low.

A shortage of a producer good results in reduced activity in industries that use the good directly and perhaps in those that use it indirectly. Demand for complementary inputs falls, and demand for substitute inputs rises. Some employers may lay off workers, or put them on short time. Producers with more inelastic demands for the good tend to bid the available supplies away from those with more elastic demands. There is a presumption that market allocation tends to direct supplies to the highest-valued uses and hence minimises the costs of the supply disruption, but as with the case of a consumer good, the government may prefer a different distribution of the costs of disruption from that provided by the market. In particular, it may prefer to minimise the unemployment rather than the total costs resulting from the supply disruption, and it may seek to achieve this result by some scheme of direct allocation among industries. Or it may deem some activities to be more essential than others, and allocate supplies preferentially to the former.

All of the aforementioned considerations are relevant in the case of a good, like petrol, that is both an intermediate good and a consumer good. In addition the question arises of its allocation between business users and private users. Preference is almost invariably given to business users. This may reflect the preservation of employment motive, and/or the notion that production is more essential than consumption. It may also be justified on the grounds that households are more flexible and adaptable production units than business firms, or that they have more available substitutes for the scarce good.

If the primary aim of government intervention is to prevent or modify the income effects and transfers accompanying simple price rationing, rather than to modify the allocation of the good *per se*, then it is possible to achieve this aim while continuing to rely on the market to allocate supplies. For example, if the aim is to prevent or reduce the 'profiteering' of producers, price can be allowed to rise and perform the role of allocating supplies among consumers, while an 'excess-profits' tax is levied on producers. This is the path that has been taken with respect to normal petrol pricing in Australia and the United States. Similarly the aim of shielding poor consumers from the adverse effects of a sharp rise in the price of an 'essential' good can be accomplished by subsidising the poor's consumption of the good, but allowing price to be determined by market forces. A convenient method of doing this is to issue 'stamps' to eligible consumers which entitle them to purchase a quantity of the good at a concessional price, food stamps are a case in point, and suggestions have been made in the United States that an 'energy stamp' program be implemented.

COUPON RATIONING

The more ambitious aim of preventing or reducing the consumer-producer income transfer can be achieved by price control together with a transferable-coupon rationing scheme. Coupons confer on consumers the right to buy a certain quantity of the good at the fixed price. If a black market in coupons is tolerated, or better still if a 'white' market is encouraged then the good will tend to reach those who value it most highly, irrespective of the initial distribution of coupons. Market forces will determine or at least influence the allocation of the good, but will operate via the price of coupons rather than the price of the good. Income transfers take place, but these are within the set of users of the good - from those with an excess demand for coupons to those with an excess supply at the market price of coupons.

Rationing by transferable coupons can also accommodate the aim of achieving an equitable distribution of the burden of scarcity among consumers. Indeed, the question of what constitutes an equitable allocation of coupons arises, irrespective of whether inter-consumer equity was on the government's original agenda. Coupons can be allocated in any manner deemed to be equitable. To be sure, the fear that the poor will take a disproportionate cut-back in their consumption of the good cannot be assuaged completely, since they may sell all or some of their coupons. But if they do, it will be in the pursuit of a perceived gain as compared with the option (always open to them) of consuming their (by assumption, equitably-determined) ration bought at the fixed price. It would be possible to object to this exercise of free choice only on paternalistic grounds. For example, if the good were a basic food, that those selling their coupons were thereby risking malnutrition.

Thus, by redefining property rights in the scarce commodity, a transferable-coupon rationing scheme would seem to allow certain income distributional desiderata to be met without abandoning the advantages of allocation of the good by market forces. This is why such schemes are much favoured by economists. Some reservations are in order, however.

Firstly, exchange of coupons involves transaction costs, even if the market is a white one, so that some misallocation costs will remain.

Secondly, the establishment of a white market is highly unlikely, its rationalisation requiring a degree of economic sophistication rarely found among either politicians or the public at large. Hence probably the best that one could hope for would be a tolerated black market.

Thirdly, real costs would be incurred by consumers, individually and as members of groups, in lobbying the government in an attempt to increase their allocations of coupons. Part of the surplus intended for consumers would thus be dissipated.

Fourthly, when it comes to the point, finding an equitable basis for allocating coupons will prove to be troublesome; any scheme of allocation will leave many people dissatisfied, and not merely because of how they fare personally. The fact is that there is no generally accepted notion of what does constitute a fair allocation of a scarce good - especially a good like petrol which is used both in 'production' and 'consumption' and may be income elastic. (The case is rather different for staple food, where equal allocation per person, or per adult equivalent, would command a good deal of assent.)

It must be stressed that the use of transferable coupons is an appropriate means only of suppressing or modifying the income redistributional effects of the scarcity of a good. It does not affect the final allocation of the good, except insofar as this is influenced by the modified distribution of income, and by the transaction costs associated with the use of coupons in addition to money. If it is desired to alter the allocation of the good more purposefully for the reasons mentioned above, (ie to minimise the unemployment resulting from the scarcity to favour the business sector at the expense of the household sector, or to maintain essential services where essentiality is defined in some way other than having relatively inelastic demand for the good as an input) then effort has to be expended in order to make the desired allocation 'stick'; that is, black marketing of the good or of coupons has to be discouraged.

QUEUE RATIONING

If the aim of price control is to prevent income transfers from buyers to sellers, then the use of queue rationing to allocate the good is only partially consistent with this aim. Queue rationing does not prevent the real price paid by consumers from rising, since it allows the time cost to rise from being an insignificant to being a substantial component of the total price, and relies on the higher time cost to discourage consumption. It does, however, prevent sellers and producers from benefitting from the scarcity, and to this extent is consistent with the presumed aim of price control. It may therefore be an appropriate method of rationing if the objection to buyer-seller transfers is based more on envy, or other moral objections to 'profiteering', than on compassion for consumers in general. However, insofar as consumers differ in their valuation of time and inconveniences costs of queuing, it imposes different real price rises on different consumers. There is a widespread belief that the pattern of price rises generated by queue rationing is positively correlated with income, so that it discriminates against the well-off and is more favourable to the poor than is price rationing. Hence if the principal concern is with the effect of a price rise on the welfare of poor consumers, rather than consumers in general, queue rationing might not be inconsistent with the purpose of price control.

The notion that an individual's valuation of time is closely related to his income is subject to a number of caveats.

Firstly, the relationship is asserted only insofar as income is derived from working. A person whose income is derived solely from property does not have to make an income-leisure choice. The subjective cost of queuing to the idle rich might be quite low.

Secondly, for workers, the marginal value of an hour's leisure may differ from the marginal hourly rate of pay for two reasons.

- Institutional constraints, such as standard hours, the non-availability of, or compulsory nature of overtime, may prevent the worker from making marginal adjustments. His marginal evaluation of leisure may therefore be greater or less than his net marginal evaluation of working.

- The non-pecuniary aspects of work may be valued positively or negatively at the margin. If negatively - ie if there is marginal disutility of effort - and if the worker is free to make marginal labour-leisure choices, then the marginal evaluation of leisure will be less than the wage rate. The latter will have to compensate him not only for the leisure foregone at the margin, but also for the marginal disutility of work. Conversely, if work yields positive utility at the margin, his marginal evaluation of leisure will exceed the wage rate. (Johnson 1966, Oort 1969)

These considerations presumably make for a good deal of random variation in the income, valuation-of-time relation, thus weakening the correlation between the two. They may also bias it in either direction; that is, make the evaluation of leisure rise faster, or more slowly, than income. On one view of the functioning of labour markets, highly-qualified workers, or those with skills in high demand secure both high salaries and non-pecuniary advantages. If, for such people, work yielded positive utility at the margin, the value to them of an extra hour's leisure would exceed the wage rate. The value of leisure would then rise faster than the wage rate. On the other view, wage differences tend to equalise the total advantage of different jobs, by being negatively correlated with non-pecuniary advantages. Again assuming that non-pecuniary aspects are operative at the margin, there would be a tendency, in this case, for the value of leisure to rise more slowly than the wage rate. It is not, I think, possible to say which of these tendencies predominate in labour markets and in any case, the assumption that 'good jobs' yield utility at the margin is a strong one, so that speculation along these lines must remain inconclusive. However, it can fairly be concluded that the income, valuation-of-time relation is subject to a good deal of 'noise'.

Thirdly, constraints on the timing of activities prevent a full equalisation of the marginal value of time in different uses. Introduction of a new time-consuming activity, queuing, itself subject to timing constraints (eg hours of opening of sellers) into a time budget is likely to involve higher costs in terms of activities foregone if the existing activities are themselves subject to numerous timing constraints, than if they can readily be shifted. In particular, the cost of queuing is likely to be higher, *cet par*, for persons working during prescribed periods than for those whose hours of work are flexible. Hours of work tend to be more flexible in higher status (professional and managerial) occupations. In this respect queue rationing tends to be the opposite of egalitarian.

If, despite the foregoing considerations, queue rationing does nevertheless have the effect of imposing a higher price for the rationed good on the rich than on the poor, it still does not necessarily follow that the good will be distributed in a more egalitarian fashion than with price rationing. If the rich value the good sufficiently more highly than the poor, they may secure the bulk of the available supplies, despite the higher price. Consider again a Type 1 demand structure and suppose that demand for the good (willingness to pay for a unit of it) is income elastic (ie $\eta>1$) and that the value of time is directly proportional to income. The ordering of consumers by willingness to pay (in money) will correspond exactly to their ordering by income and by valuation of time. If the money-price demand curve is now converted into a time-price demand curve, it will be flatter than the original - the higher demand prices will be deflated by higher dollar prices per minute than the lower ones - but the ordering of consumers will be unchanged and so will the identity of those receiving the available supplies. This is because the valuation of the good increases faster than the valuation of time, as income rises(1).

The foregoing argument, while based on a Type 1 demand structure, can nevertheless be extended, with modifications, to more realistic cases. Although, under price rationing, a rich man's valuation of a marginal unit of the good would be equal to a poor man's, his valuation of successive intramarginal units might increase much more rapidly, and his willingness to queue might exceed the poor man's, once each has given up some marginal units.

---

(1) Some notes on the analysis of this situation by Barzel (1975) are given in Appendix II.

A final point bearing on the distributional effects of queue rationing is the possibility that consumers with a high valuation of time would employ persons with a lower valuation to queue for them. Ignoring transaction costs, the result would be that each consumer's willingness to pay in money would be converted into a willingness to queue, by proxy, at a common wage rate for unskilled labour. The distribution of the good would be as under price rationing, and not all of the payments made for queuing would be matched by real costs incurred by specialist queuers: the intramarginal ones would earn some surplus.

In summary, conventional wisdom that queue rationing has egalitarian distributional consequences may not be so wise, after all. This is because:

- the association between income and the value of time may be weak;
- the poor tend to be more subject to inflexible work schedules, which increase the inconvenience of queuing;
- if demand for the good is more elastic, with respect to income, than is the valuation of time, queue rationing will not deter the wealthy from consuming the good; and
- the discrimination against those with a high valuation of time could be thwarted by the contracting out of queuing to specialists.

The third point above might be particularly apposite in the case of petrol. United States data suggest that the income elasticity of demand for petrol is slightly greater than unity, except at the extremes of the income distribution (Freeman 1977). Also, for many purposes, the use of a car saves time. Therefore one would expect a positive relationship between the valuation of time and the demand for petrol.

CHAPTER 5 - RESTRICTIONS ON QUEUING AND/OR USE OF CARS

In recent years, in response to short-term supply disruptions, or to anticipated increasing scarcity of oil products, governments have introduced a number of measures intended to regulate queue rationing, or to reduce petrol consumption via non-price restrictions. These include limitations on the times when cars may be used, constraints on the timing of petrol purchases, and constraints on the size of petrol transactions. These will be discussed in turn.

LIMITATIONS ON THE TIMES WHEN CARS MAY BE USED

Limitations that have been imposed include bans on the use of cars on particular days, eg at weekends, or on Sundays; and the institution of 'carless' days, whereby each car owner nominates one day of the week (or working week) when he will not use his car. Enforcement of the latter requires that each car display a sticker stating the day it will not be used.

Banning private motoring at weekends or on Sundays should be effective in greatly reducing the use of cars for recreational excursions (including church attendance), since, for most families, work commitments of the breadwinner(s) substantially rules out major excursions during the working week. However, during summer, at least, one could expect some substitution of after-work trips to the beach, for example, in place of longer Sunday beach visits.

Similarly, banning the use of one's car on one working day per week should also be effective in securing a reduction in the use of cars for commuting purposes. However, those whose work attendance is flexible, or who work less than a five-day week, or can work at home, may be able to adjust to the restriction with little inconvenience or reduction in petrol use. The same is true of two-car families.

These rationing devices fit the model of random allocation fairly well. Some motorists would be quite unaffected by them, while, among those that are, the impact would be quite variable: in some cases high-priority activities and in others marginal activities would be curtailed. The proportion of surplus lost could therefore be expected to be roughly equal to the proportion by which petrol consumption was reduced; or rather, given that through rearrangement of activities some amelioration of the impact on high-priority uses could be achieved, the loss of surplus would be somewhat less than this.

CONSTRAINTS ON THE TIMING OF PETROL PURCHASES

Measures in this category include the closing of petrol stations in the evenings and at weekends, and the staggering of sales to different groups of motorists by such means as requiring motorists with even-numbered license plates to purchase petrol only on even-numbered days of the month, and those with odd-numbered plates, only on odd-numbered days.

These restrictions make motoring more inconvenient and costly. The consumer has to take greater care to anticipate his petrol requirements, for he may buy it only every second day (with odd-and-even-days) or only on five days per fortnight (with odd-and-even-days plus week-end closing). Mistakes will be made, and cars immobilised for periods. Also, his rate of consumption is limited, during dry periods (of up to three days) to the car's tank capacity divided by the number of dry days, unless he makes use of supplementary storage. It is possible that most motorists would invest in some storage containers, but, even so, private motorists would be deterred to some extent from making long trips, particularly at weekends - which, of course, is the time when most people are ordinarily best able to make long trips. In some cases it would be physically impossible to carry sufficient supplementary supplies to make the length of trip desired. There would therefore be some reduction in car use and in the demand for petrol.

The rationing effect of restricted trading hours has two sources:

- the added inconvenience of having to plan and execute petrol purchases within restricted times; and

- limitations on the mileage that can be covered during dry periods, and/or the costs of avoiding this limitation by the use of supplementary storage.

In the absence of queuing, the demand for petrol would be reduced by reason of both sources.

If these restrictions were imposed under a regime of queue rationing, the added inconvenience attaching to petrol purchases would be much greater, since the transaction time is much longer, and evening and week-end closing of petrol stations would limit transactions to periods when much of most people's time is already committed to work activities. Assuming the costs of queuing per unit time were raised for the marginal queuer, the time he was willing to queue would be reduced. However, since the costs he is willing to incur by queuing are determined by his evaluation of the surplus to be obtained from a marginal petrol transaction, his total costs of queuing would not be reduced, except in so far as his valuation of the surplus was also reduced for the reasons listed in the second point in the preceding paragraph. If his valuation of petrol was unaffected by these restrictions, his willingness to pay the queue price would also be unaffected. The price would be the same, but it would be incurred through a reduced expenditure of more valuable time.

This example illustrates the general proposition that attempts to reduce the *time* spent in queues by making it more difficult to queue are ineffective, by and large, as a means of reducing the *costs* of queuing. This is because the queue price is a demand-determined price. However, such attempts to regulate the queuing process are likely to have distributive consequences which may have second-order effects on the social costs of queuing. For example, intra-marginal queuers might gain or lose as a result of a shortening of the equilibrium queuing time, depending on whether their own time costs rose less or more than those of the marginal queuer. Or the identity of the marginal queuers might change.

Constraints on the timing of petrol purchases may act as a petrol rationing device in their own right, and choke off some demand. Insofar as the demand for petrol is reduced, the queue price is also likely to be reduced, and the social costs of queue rationing reduced. However in that one form of rationing is being partially substituted for another, so are one set of social costs being partially substituted for another.

The social costs of rationing by restricting trading hours are partly those of misallocation (motorists with limited storage capacity will be unable to equate the marginal utility of petrol during dry periods with its marginal utility at other times), and partly those incurred in order to reduce misallocation (as when motorists acquire supplementary storage capacity and experience costs partly of money, partly of inconvenience, and partly of the danger associated with storing and handling petrol on domestic premises and carrying it in loose containers in cars). The latter are akin to the costs of queuing, in that real resources are expended in order to improve the allocation of the good.

There is an important qualification to the argument that restricted trading hours will not affect the cost of queuing directly, but only indirectly. This is discussed in the next section, dealing with transaction limits.

In a situation of no excess demand (and no queuing) sellers as a group could be expected to support the idea of evening and weekend closing, since it would reduce one form of competition and lower their costs. (Sellers specialising in evening and weekend sales would presumably protest, however.) However with queue rationing in force, it is doubtful whether legal restrictions on trading hours would reduce retailing costs - and thus provide a partial offset to the costs imposed on buyers - since presumably most stations which are able to sell their weekly allocation in a fraction of normal operating hours would restrict their trading hours in any case. Indeed, regulations forcing them to close at particular times would harm those sellers who otherwise would have chosen a different pattern of trading hours.

CONSTRAINTS ON THE SIZE OF PETROL TRANSACTIONS

Maximum limits

The simplest limitation on the amount of petrol bought per transaction is a prohibition on the filling of cans, drums, etc from bowsers. This limits each transaction to a maximum equal to the capacity of the car's tank and in practice to something a bit less than this. Tank capacities vary from car to car, but are positively correlated with the rate of consumption, so that such a limit is less variable when expressed in terms of distance travelled per transaction. Motorists wishing to accumulate stocks in excess of their car's tank capacity would have to siphon from the tank into other containers - an inconvenient and in some cases difficult process. The prohibition would therefore tend to discourage hoarding of petrol, whether in anticipation of increasing scarcity, or as a means of accumulating sufficient supplies for a specific purpose (eg for a long trip).

Transaction size can be further limited by the imposing of a maximum dollar limit, eg $5.00, or by a sliding dollar limit based on the size of the car (in practice this would probably mean number of cylinders).

It was argued in Chapter 2 that in some circumtances a relatively low transaction limit can lower significantly the real costs of queue rationing. Are the relevant circumstances likely to exist in the case of petrol rationing?

In the ensuing discussion it shall be assumed that <u>some</u> maximum transaction limit is imposed; say a tankful *as a high maximum*. The question then is whether some smaller limit might be preferable. (The interesting question of what the transaction size would be if no exogenous limits were imposed on it will not be considered.)

Waiting in queues is a transaction cost which the consumer will attempt to minimise by queuing infrequently and making a large purchase each time. The size of the transaction may be limited by the exogenously-imposed maximum limit, or by the consumer's own circumstances. He might be inhibited from purchasing up to the maximum limit by an inability to consume the maximum amount in the time available if the good is perishable, or by a lack of storage capacity or the high cost of storage in the case of a storable good. If the costs of storage are negligible up to a capacity limit, if the good is available for purchase continuously and if the consumer's storage capacity is greater than or equal to the maximum transaction limit, then the consumer can always purchase up to the transaction limit. Furthermore, he can adjust his rate of consumption however he desires by varying his frequency of purchase. However if his storage capacity is not or little greater than the transaction limit, the timing of purchase would become quite critical.

Petrol is storable and each motorist has the primary storage capacity of his car's tank. If the transaction limit were equal to or less than the capacity of his car's tank and if supplies were continuously available, each motorist could adjust his required purchases, each of maximum size, to his desired consumption. Variations in the transaction limit provided it did not exceed one tankful would not affect his willingness to queue, per unit purchased. A lower limit would mean he would queue more frequently, but would be willing to wait a short time on each occasion. If an equilibrium queuing time became established as market participants gained experience, it would be unaffected by the size of the transaction limit.

If supplies were not available continuously but were regularly interrupted, as in the week-end closing of petrol stations and odd-and-even-day scheme mentioned above, motorists would experience difficulties in matching their purchases to their desired consumption activities. Having a half-full tank immediately preceding a dry period in which he desired to use more than half a tank would present the motorist with the choice of curtailing his activities in the dry period, or queuing for half a tankful now. If the maximum transaction limit was one tankful, the latter option would involve him in paying almost double the queue price per litre that he would pay if supplies were continuously available and he could wait until his tank was nearly empty before queuing. His willingness to pay this higher price would depend on his evaluation of petrol's productivity in the activities he would otherwise have to curtail during the dry period. He thus has to make an all-or-nothing choice with respect to these activities: he either buys petrol now or foregoes them entirely. If they were high-priority activities involving a lot of surplus, he would be willing to pay the higher price and forego some or all of the surplus(1). Alternatively, if the price was too high he would forego the activities and the entire surplus. However if the transaction limit was half a tankful, buying now would not involve any increase in the normal queue price (although it might involve some additional cost in the

---

(1) If the motorist's desired consumption during the dry period exceeded his storage capacity, he would have to eliminate the less valuable uses, so that those remaining would be non-marginal ones.

form of inconvenience) so that although the choice would still remain an all-or-nothing one, he would not be forced to pay more than the normal queue price.

Uncertainty about future supplies

When sudden supply disruptions occur, there is often considerable uncertainty about the severity and likely duration of the shortage and about how long supplies will continue to be available for sale, or whether some rationing scheme will be introduced. There is then a scramble for supplies while they last. Consumers endeavour to meet the demands of future days and weeks by buying supplies now. The demand curve swings upwards, pivotting about its price-axis intercept. (Refer back to Figure 2.2. If it were expected that new supplies would not be available for four weeks, the four-week demand curve AC would come into operation immediately.) Consumers cut back on their rate of consumption and attempt to accumulate stocks. To the extent that they are successful, they move down AC. How might the willingness to queue to accumulate stocks be affected by different transaction limits? Firstly assume that motorist's desired stock holdings are not constrained by a lack of storage capacity.

On the basis of the earlier discussion of transaction limits and all-or-nothing choices it might seem that a low transaction limit would reduce the time spent in queues while stocks were being transferred from service stations into motorists' tanks, cans and drums. However that discussion was concerned with the determination of equilibrium queuing times; here the concern is with a disequilibrium situation and one, moreover, involving a good deal of uncertainty about how long supplies will last and hence about how many opportunities consumers will have to queue before supplies run out. It is not implausible to assume that consumers would bid up the queue price to higher levels in the earlier stages of stock accumulation than in the later stages. This pattern would emerge if consumers, fearing that each transaction might be their last, heavily discounted the possibility of making another transaction and based their willingness to pay mainly on their evaluation of this single increment to their supplies, given the stocks they already held.

As their stocks increased they would move down their demand curves and their willingness to pay would decline. In the process they could lose a large proportion of their consumer surplus. Moreover, the smaller the steps in their demand functions, the greater the potential loss of surplus. In this case then, the smaller the transaction limit, the greater the time spent queuing for a given aggregate supply.

Of course the foregoing scenario is just one of many possibilities: events could unfold quite differently. Therefore perhaps not too much weight should be attached to it. It is of interest however as an example of a low transaction limit having the opposite effect in a dynamic model from what it has in a comparative static framework of analysis.

Now, assume that motorists' desire to accumulate stocks of petrol is constrained by limited storage capacity. For simplicity assume that their only capacity is that of their cars' petrol tank. Each consumer will wish to fill his tank at the onset of the shortage and, so long as supplies remain available, to keep it full. This requires frequent topping up by small purchases of petrol. Whatever the queuing time, the smaller the transaction, the higher the time-price per unit purchased. If the transaction limit is relatively large - say a tankful -different consumers' different valuations of incremental petrol stocks can express themselves in their paying different time prices, by spending the same time queuing for different quantities. Hence those with high demand, and/or pessimistic expectations about the availability of supplies in the future, will top up with smaller purchases than others. The queuing process, in conjunction with a rigid individual storage limit, thus allows price discrimination to occur with resulting additional losses of surplus by intra-marginal consumers. This problem can be mitigated by relatively low transaction limits which reduce, if they do not entirely eliminate, the differences in queue prices per unit paid by different consumers. (This case is analytically the same as that of known interruptions to supply, discussed earlier.)

It is worth noting that when stocks are being accumulated, or maintained by topping up, the individual demand curves and the market demand curve are more elastic than normally. (An additional unit will serve a high-priority purpose next week rather than a lower priority use this week.) This means that the differences in the value of surplus dissipated through queuing may not be very great, depending on whether queue rationing is operating in a more or less discriminatory fashion.

To sum up. The effect of the size of the maximum transaction limit on the willingness to queue, in various circumstances, has been considered, with the following results.

- In a continuing queue-rationing situation with supplies being continuously available, transaction limits will have little or no effect on the willingness to queue or the equilibrium time-price per unit.

- If supplies are periodically and predictably unavailable and if motorists' storage capacities are limited, a relatively low transaction limit will tend to reduce the costs of rationing.

- If, because of uncertainty about the availability of future supplies together with limited storage capacity, motorists tend to top up their tanks frequently, then a relatively low transaction limit will tend to reduce aggregate queuing costs.

- While motorists are accumulating stocks, at the onset of a shortage, a relatively low transaction limit could conceivably result in their spending more time queuing than a higher limit would.

The first three points illustrate the more general proposition that a large transaction limit in the presence of other constraints such as limited storage capacities, interruptions to the availability of supplies, the unexpected emergence of new demands, etc can make it difficult for consumers to bring desired consumption and purchases into mutual adjustment, and induce them to make all-or-nothing-type choices in which they may have to pay a higher time price per unit than they otherwise would. Since the discontinuities mentioned, and others, are obvious features of real-world shortages, it would seem that a relatively low transaction limit would reduce the real costs of queue rationing.

Transaction costs

Apart from the rather problematical last point above, there is one other consideration working against low transaction limits and this is that the smaller the transaction size, the greater the costs of transactions. This is because in addition to any variable component, each transaction entails fixed costs -entering and leaving petrol stations, stopping and starting engines, opening and closing filler caps, starting and stopping petrol pumps, and the money transaction. Insofar as these costs are borne by consumers, there would be compensating changes in the time they would be willing to wait, so that total queuing cost would be unaffected. However some of these costs would be borne by sellers and, in a full social reckoning, have to be set against any further reductions in queuing time resulting from the reduced likelihood of consumers having to make all-or-nothing choices, as the transaction limit is made smaller. Also if the transaction limit were so small as to be particularly irksome to sellers, they would be unlikely to adhere to it voluntarily, thus jeopardising the scheme if it were a voluntary one, or increasing the required costs of enforcement if the scheme had the force of law.

Another point to emerge from the foregoing is the considerable advantage the motorist secures, in the face of large transaction limits, supply interruptions, etc, by having supplementary stocks of petrol. One of the benefits of holding stocks is their 'convenience yield'; when a good is rationed by making it inconvenient to buy it, the convenience yield of stocks can be indeed high. They give the motorist greater flexibility in choosing when to queue, help him to avoid having to queue for small quantities and lessen the impact of supply interruptions. If periodic shortages of petrol came to be expected, it seems likely that many motorists would invest in supplementary storage facilities.

Should transaction limit be related to car size?

In the discussion so far references to a transaction limit of a tankful or half a tankful have ignored the problem presented by the fact that the size of tank varies among cars. Imposition of a transaction limit of one tankful, or a sliding dollar limit based on size of car, would mean that different motorists would face different time prices per unit purchased when queuing. Those with large tanks (or large cars) would pay a lower time price per litre than those with small tanks or cars. The earlier analysis of queue rationing which assumed that everyone faced the same queue price would no longer apply. With systematic differences in the time prices paid by different motorists there would clearly be misallocation of petrol supplies and a resulting loss of surplus. Aggregate queuing time would also be different, but it could be either more or less than with a common time price. To see this, consider Figure 5.1 which is based on the simplifying assumption that there are just two types of car, big and small, and that big cars' tanks have twice the capcity of small cars'. CD, related to the origin 0, is the demand curve of owners of small cars, while AB, origin 0', is the demand curve of owners of big cars. Both are 'time' demand curves. 00' is the available supply; BD, the overlap of the two curves, is the amount by which demand would exceed supply at a zero time price. If a common transaction limit no greater than the tank capacity of a small car was in force, each group would face a common time price per unit, which would be bid up to OK, and quantities used in small and big cars would be OE and EO' respectively. If however, the transaction limit was one tankful, owners of big cars would pay a time price only half as great as small-car owners. Equilibrium would be established with a time price of OM for small-car owners, and O'N (=1/2OM) for large-car owners, and with small-car consumption of OF and large-car consumption of FO'.

As compared with the operating of a single time price, the two-price arrangement results in a gain of the area hatched with pluses (reduced queuing costs for owners of large cars), a loss of the area hatched with minuses (increased queuing costs for owners of small cars), and a further loss of the area hatched with dots (costs of misallocation). It is also evident that the change in queuing costs for each group (change in each 'revenue rectangle') will depend on the elasticity of its demand curve. Furthermore, since each experiences an equal and opposite quantity change, the changes in queuing costs will be equal and offsetting if the elasticity over the relevant ranges is the same for each curve. If the big-car demand is more elastic than the small-car demand, there will be a net increase in queuing costs, and if it is less elastic, a net decrease. If the elasticity difference were great enough, and in the right direction, the reduced queuing costs would outweigh the misallocation costs introduced by the two-price arrangement.

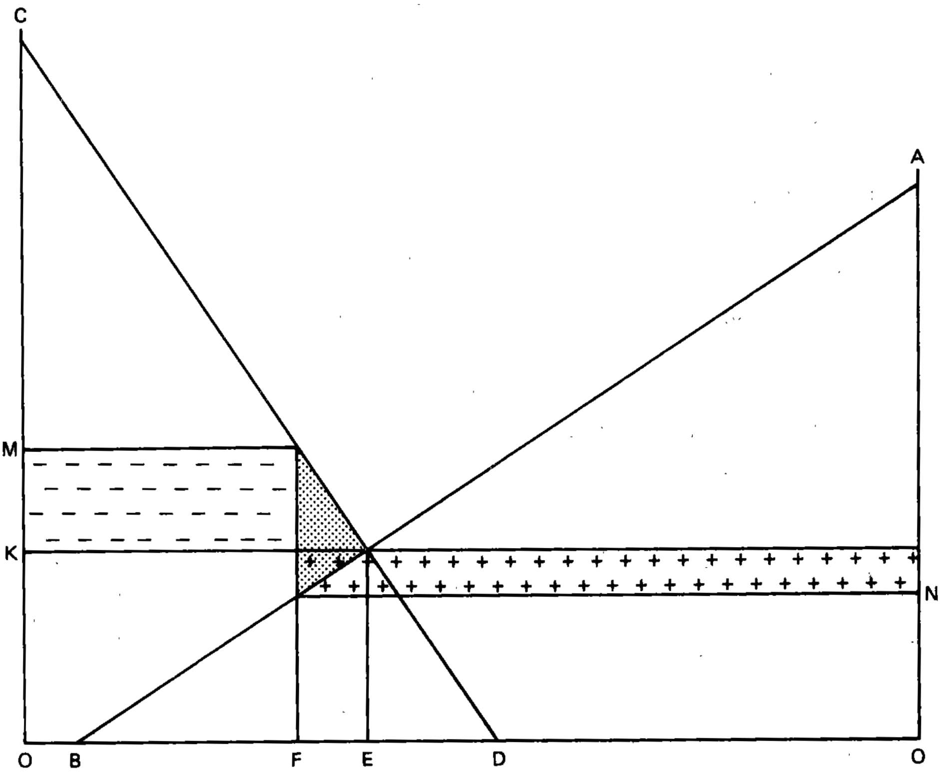

Figure 5.1 Illustration of the effect of car tank size on queuing costs

Minimum limits

A lower limit to the size of petrol transactions may be intended to:

- reduce time spent queuing and petrol used up in the queuing process;
- limit in-car inventories by preventing topping up;
- reduce sellers' transaction costs, and/or
- increase the inconvenience associated with motoring and hence reduce car use.

It will be assumed that, if a minimum transaction limit were imposed, it would be in conjunction with a maximum limit of, at most, one car tankful. Obviously, the minimum limit could not be greater than the maximum limit, but it could be equal to it. It will also be assumed that a motorist with insufficient room in his tank to accommodate the minimum limit could nevertheless purchase a lesser amount, provided he paid for the specified minimum quantity.

On these assumptions, the effect of a minimum limit would be to raise further the costs associated with small transactions: money cost, in addition to the extra time cost, would be imposed. In situations calling for a choice between a small high-cost transaction and foregoing a high-value consumption activity, the latter option would be taken more frequently. There would therefore be smaller queuing costs, but greater costs of misallocation, than if there were no minimum transaction limit. Also there would be a direct rationing effect as the demand for petrol would be reduced somewhat. It could also be anticipated that the greater penalties associated with small transactions would induce more motorists to obtain supplementary storage capacity.

It thus appears that a minimum transaction limit could contribute to all of the aims mentioned above, though probably only in a relatively minor way. Furthermore, the lower the maximum transaction limit imposed, the less effect would the minimum limit have. Hence if the arguments put forward in favour of a relatively low maximum limit are accepted, the question of whether a minimum limit should also be imposed becomes unimportant. In that it would reduce sellers' transaction costs, its imposition, along with a maximum limit, might soften sellers' opposition to the latter.

CHAPTER 6 - CONCLUSIONS

Non-price rationing is widely used in our society, especially in the public and non-profit sectors (eg justice, medical care, education) but also by profit-seeking enterprises (eg waiting lists for new cars). In times of emergency (war, or an acute shortage of an essential good) it is commonly imposed on goods that are normally rationed by price.

Governments' motives for using non-price rationing include the pursuit of some over-riding national objective, the shifting of the burden of adjustment to scarcity of a good among sections of society, mitigating the income-distributional consequences of sharp price rises, being seen by the electorate to be 'doing something' about a problem, and suppressing the symptoms of inflation.

METHODS OF NON-PRICE RATIONING

There are two basic types of non-price rationing, queue rationing, and allocation schemes, although actual practice often involves some mixture of both. Queue rationing is essentially a market form of allocation, analogous to price rationing, except that 'willingness to queue' rather than 'willingness to pay' is rewarded by the competitive process. Under allocation schemes market competition is suppressed, and supplies are allocated among individuals on the basis of some non-economic or quasi-economic principle. For the most part it is assumed here that such allocation is essentially at random with respect to consumers' willingness to pay for the good.

As compared with price rationing, non-price rationing involves three main types of cost:

- administrative and enforcement costs;
- misallocation of the good; and
- real transaction costs incurred by consumers (including queuing costs).

The costs of misallocation and of queuing are affected by the nature of the individual demand curves underlying the market demand curve.

Queue rationing

Under queue rationing, consumption is discouraged both by the fixed money price and by the cost of having to wait in line -the 'queue price'. The major element of the queue price is the opportunity cost of the buyer's time spent in the queue. It is possible to visualise a demand curve for the right to buy the good at the fixed money price, with the queue 'price' being measured in waiting time per unit bought. The equilibrium queuing time would then be established at the intersection of this demand curve with the vertical supply curve. If every individual's valuation of time (and the other elements of the cost of queuing) were the same, the time demand curve would be a direct translation of the ordinary demand curve above the fixed price.

There are two elements of queuing time - waiting time and transaction time. A reduction in the latter would not result in a reduction in total queuing time; the reduction would be taken up by an increase in waiting time. This is because queuing time is a demand-determined market equilibrium price. Its magnitude is determined by the tastes and opportunities of consumers, and, in particular, by those of the 'marginal queuer'.

With queue rationing, what would otherwise be a transfer from consumers to producers becomes a real cost, as surplus is dissipated in time spent and inconvenience suffered in queuing.

In a simple model, queue rationing has the same outcome as price rationing except that the price increase is paid in time. The loss of surplus (in excess of the familiar welfare triangle, common to all methods of rationing) in this case is therefore the 'expenditure rectangle', if price were allowed to rise, minus that part of it paid in money at the fixed price.

In favourable circumstances, the loss of surplus can be less than this, but in most cases will probably be greater. The additional losses arise from the following causes.

- With a 'large' transaction limit and in the presence of various constraints, consumers may be forced to make all-or-nothing choices in their queuing decisions and hence pay time prices that are in effect discriminatory.

- The cost to a consumer of queuing for a given length of time is directly proportional to his valuation of time. If consumers' time valuations are positively correlated with their valuations of the good, so will be the money equivalents of the time prices they pay for the good.

- The pattern of discriminatory time prices may be such as to exclude from consumption those with the highest willingness to pay for the good - a particularly perverse form of misallocation.

- Variations in the value of time that are random with respect to variations in willingness to pay for the good may increase or reduce the costs of queue rationing, depending on whether the demand curve is, or is not, strongly convex to the origin. *All* constant elasticity curves are strongly convex in this sense.

The likelihood of the consumer having to make all-or-nothing choices with queue rationing is less if the good is storable and/or its consumption can easily be postponed or brought forward, and can be reduced by the imposition of a relatively low transaction limit.

The costs of queue rationing are reduced if queuing can be contracted out to persons who place a low value on time.

Random allocation

Random allocation of the good among potential consumers reduces the total surplus by the same proportion as supply is reduced.

Comparison of rationing schemes

In order to assess the likely costs of non-price rationing methods it is usually necessary to know the shape and position of the whole demand curve and not just that part of it lying within the range of expected supply fluctuations.

The loss of surplus with random allocation is affected by the shape of the demand curve only insofar as the shape helps to determine the total initial surplus.

The loss of surplus with queue rationing is quite sensitive to the shape of the demand curve, and, hence, so is its performance relative to random allocation.

Queue rationing and price rationing perform better with constant-elasticity than with linear demand curves.

In a simple model (uniform valuation of time, no 'price discrimination') the excess burden of queue rationing (ie the loss of surplus additional to the loss under price rationing) is twice as great as that of random allocation, with a linear demand curve. However, with a constant-elasticity demand curve, random allocation has the higher excess burden, exceeding that of queue rationing by the factor a/a-1, where a is the absolute value of the elasticity of demand.

The losses of surplus from non-price rationing, relative to the loss from price rationing, are greatest for small cut-backs in supply. Their absolute excess burdens, however, reach a maximum for cut-backs of 50 per cent for linear demand curves and of more than 50 per cent for constant elasticity demand curves. With a severe shortage the loss of surplus is great for all types of rationing, and the differences between them are small.

Allocation schemes will generally involve higher administration and enforcement costs than queue rationing.

It is not possible to give a clear-cut answer to the question of whether random allocation or queue rationing is the more inefficient method of rationing. If constant elasticity demand curves are a better approximation to reality than linear curves, this is a factor favouring queue rationing. However, even so, the costs of queue rationing can greatly exceed the costs of random allocation. If queuing involves all-or-nothing choices, or if the value of time is strongly and positively related to willingness to pay, a high proportion of the surplus may be dissipated, irrespective of the degree of cut-back of supplies.

MANAGEMENT OF STOCKS

The disruption accompanying a reduction in supply of a good will be minimised if consumption is reduced and stocks begun to be accumulated as soon as the shortage is anticipated, and if stocks are run down during the period of reduced production. Market-determined prices will tend to bring about this pattern of stock management.

Under non-price rationing, the price signals that would facilitate rational stockholding decisions are suppressed. Unco-oordinated stockholding decisions by final consumers tend to replace price co-ordinated decisions by producers, distributors, and dealers. Stock management is thus less efficient than with price rationing.

The misallocation of supplies resulting from motorists frequently topping up their tanks would appear to be small.

Supplementary storage of petrol by private motorists increases fire risks. This is a reason for avoiding actions tending to encourage it, such as discouragement of topping up and restrictions on the timing of petrol purchases.

INCOME DISTRIBUTION

Distributional motives for controlling the prices of scarce goods include concern for the welfare of consumers generally, and/or of the poor in particular, and objections to 'profiteering' by sellers.

Protecting the poor, and denying profits to sellers, can both be achieved without price control: the former by means of 'food-stamp' type programs, the latter by means of an excess-profits tax on sellers.

The income transfers attendent upon price rationing can be prevented, but much of its allocative advantage retained, by price control combined with rationing by transferable coupons. However, this 'economists' favourite' rationing scheme would have high administrative costs, and a 'white' market in coupons is unlikely to be politically acceptable. There would be disagreement on what is an equitable distribution of coupons, some surplus would be dissipated in activities seeking to influence the allocation, and transaction costs incurred in the transfer of coupons.

Contrary to much received opinion, queue rationing is not a particularly egalitarian method of distribution. There are reasons for believing that the relationship between income and the value of time is weak. High-status workers with flexible working hours are at some advantage in queuing. As income rises, valuation of the good may rise faster than the valuation of time, in which case the rich will be more willing to queue than the poor. Discrimination against those with a high valuation of time may be thwarted by the contracting out of queuing to specialists.

RESTRICTIONS ON QUEUING AND USE OF CARS

Restrictions on the use of cars, eg 'car-less' days and bans on Sunday motoring etc) are akin to a random allocation scheme.

Restrictions on the timing of petrol purchases (eg evening and week-end closing of stations, purchases restricted to odd or even days) make it more difficult for motorists to match purchases with desired consumption, thus:

- choking off some demand;
- putting more motorists in all-or-nothing choice situations;
- causing some misallocation of supplies; and
- inducing motorists to undertake supplementary petrol storage.

The first two points above imply opposite effects on motorists' willingness to queue. Hence the net effect of restrictions on the costs of queuing is not clear.

The last two points above indicate that these restrictions have social costs of their own, apart from any effect they may have on the costs of queuing.

Transaction limits

A relatively low maximum transaction limit, while increasing the frequency of queuing, will not (except for the reasons noted below) affect the equilibrium per unit time price of queuing. If the latter is demand-determined, it will be unchanged by a low transaction limit as the greater frequency of queuing will be offset by a shorter time spent in each queue.

If motorists' petrol storage capacities are limited and there are restrictions on the timing of petrol purchases, a relatively low transaction limit is likely to lower the cost of queuing by reducing the incidence of all-or-nothing choices.

In a situation of uncertainty regarding future petrol supplies, and a scramble by motorists for available supplies, a low transaction limit could increase the cost of queuing (motorists would 'move down' their demand curves in small rather than large increments, and lose more surplus as a result); or, it could have the opposite effect (with limited storage capacities, intensity of demand will express itself in the smallness of the quantity for which a motorist is willing to pay the going queue price. The wider the range of transactions possible, the greater the degree of 'price discrimination' that will occur.)

A low transaction limit will increase transaction costs. Those borne by buyers will be offset by shorter waiting times, but those incurred by sellers will remain a net cost.

A relatively large *minimum* transaction limit will:

- reduce sellers' transaction costs;
- result in some transfer of surplus from consumers to sellers, when buyers pay the minimum transaction price for a smaller quantity, owing to limited tank capacity;

- raise the cost of small transactions, thus reducing queuing costs (less small transactions made), but increasing misallocation costs (more high-value activities foregone); and
- increase the incentive to obtain supplementary storage capacity.

Transaction limits based on size of petrol tanks or size of car result in misallocation of petrol.

CHAPTER 7 - SOME POLICY IMPLICATIONS

It is widely recognised that allowing the price of a good in short supply to rise has the advantage of stimulating increased production and therefore ultimately of alleviating the shortage. The rise in the good's relative price draws resources from less valued to a more highly valued use. Allowing a scarce good to be rationed by price when no supply response can be expected has exactly the same justification - it helps ensure that the good is allocated to its highest-valued uses - yet the advantage of using the price mechanism in this context is much less widely understood, the price rise often being seen as non-functional and distributively disadvantageous. The most general and important contribution of this paper is to spell out in detail the truth that, as compared with other rationing methods, price rationing has substantial efficiency benefits, even when supply is fixed.

The principal advantages of price rationing are that it promotes an efficient allocation of the good, is a form of competition that is not wasteful of real resources, and is self-regulating, requiring little in the way of government intervention and expenditure. Other rationing methods lack one or more of these advantages. Allocation schemes administered by some authority are bound to involve misallocation (in the economic sense) to a greater or lesser degree, and may involve heavy administrative and enforcement costs. Queue rationing may misallocate the good, may require official regulation with respect to transaction limits and orderly queuing, and always wastes real resources in the queuing process.

When judged against the allocative norm of price rationing, no obvious answer emerges as to which non-price rationing method, random allocation or queue rationing, is the more inefficient.

Any non-price method of rationing, in conjunction with price control, serves the distributional purpose of denying scarcity rents to sellers, but, in the case of queue rationing, the rents are not retained by consumers, but dissipated through the queuing process.

Since everyone has the same time endowment, queue rationing is usually considered to be more egalitarian than price rationing, but a number of reasons were put forward in the paper for doubting this proposition.

Despite its disadvantages, queue rationing enjoys widespread community acceptance in emergency situations, and has been used on numerous occasions in Australia in recent years, when petrol supplies have been disrupted temporarily. It is a widely understood method of rationing and requires much less expenditure on administration and enforcement than allocation schemes.

Regulations restricting the timing of petrol purchases and setting minimum transaction limits seem to be intended to reduce the costs of queuing by reducing its frequency. In general this approach would seem to be ineffective, in that the reduced frequency of queuing will tend to be offset by a longer time spent in each queue. To the extent that it is effective it is because it substitutes other forms of rationing for queuing, which is a dubious benefit.

The analysis of queue rationing presented in the paper suggests that queue regulation should aim to reduce the frequency with which motorists confront all-or-nothing choice situations where the surplus associated with intramarginal units of petrol is dissipated in queuing costs. It is argued that this aim is advanced by the imposition of a relatively low maximum transaction limit, and retarded by interruptions to the continuity of supply (week-end closing of stations, odd-and-even day schemes, etc.) and by a high minimum transaction limit. These latter devices also encourage motorists to obtain supplementary petrol storage capacity, with attendant fire hazards, etc.

APPENDIX I - LOSS OF SURPLUS ASSOCIATED WITH DIFFERENT TYPES OF RATIONING

Consumers' Surplus with Isoelastic Demand

Let demand be given by $Q = AP^{-\alpha}$

The consumers' surplus assaociated with some price, P_o, is

$$S_{P_o} = \int_{P_o}^{\infty} AP^{-\alpha} dp$$

$$= \frac{1}{\alpha - 1} AP_o^{1-\alpha}$$

$$= \frac{1}{\alpha - 1} A^{\frac{1}{\alpha}} Q^{\frac{\alpha-1}{\alpha}}$$

Loss of Surplus under Different Rationing Methods

Suppose supply is reduced from $\tilde{Q}$ to $g\tilde{Q}$ ($g<1$)

With *price rationing,* price will rise from P_o to P_1, and the loss of surplus is the initial surplus ($S\tilde{Q}$), minus the new surplus ($Sg\tilde{Q}$), minus that part of the loss of consumer surplus that is transferred to sellers $g\tilde{Q}(P_1 - P_o)$

ie Loss of Surplus, $\Delta S = S\tilde{Q} - Sg\tilde{Q} - g\tilde{Q}(P_1 - P_o)$

$$S\tilde{Q} - Sg\tilde{Q} = \frac{1}{\alpha - 1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} - \frac{1}{\alpha - 1} A^{\frac{1}{\alpha}} (g\tilde{Q})^{\frac{\alpha-1}{a}}$$

$$= \frac{1}{\alpha - 1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} \left(1 - g^{\frac{\alpha-1}{\alpha}}\right) \qquad (1)$$

$$g\tilde{Q}(P_1 - P_o) = g\tilde{Q} \left[A^{\frac{1}{\alpha}} (g\tilde{Q})^{-\frac{1}{\alpha}} - A^{\frac{1}{\alpha}} \tilde{Q}^{-\frac{1}{\alpha}} \right]$$

$$= gA^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} \left(g^{-\frac{1}{\alpha}} - 1\right)$$

$$= A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} (g^{\frac{\alpha-1}{\alpha}} - g) \qquad (2)$$

$$\Delta S = 1 - 2 = A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} \left[\frac{1}{\alpha-1} (1-g^{\frac{\alpha-1}{\alpha}}) - g^{\frac{\alpha-1}{\alpha}} + g \right]$$

Loss of surplus as a proportion of the initial surplus:

$$\frac{\Delta S}{S_{\tilde{Q}}} = \frac{A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} \left[\frac{1}{\alpha-1}(1-g^{\frac{\alpha-1}{\alpha}}) - g^{\frac{\alpha-1}{\alpha}} + g \right]}{\frac{1}{\alpha-1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}}}$$

$$= 1-g^{\frac{\alpha-1}{\alpha}} - (\alpha-1)(g^{\frac{\alpha-1}{\alpha}} - g)$$

$$= 1-\alpha g^{\frac{\alpha-1}{\alpha}} + (\alpha-1)g$$

For consistency with previous notation, for g, write (1-k) and for $\frac{\Delta S}{SQ_0}$, write P/S

$$P/S = (1-\alpha(1-k)^{\frac{\alpha-1}{\alpha}} + (\alpha\ 1)1-k) \qquad (3)$$

Under *queue rationing,* the loss of surplus is given by equation 1 above.

ie $\Delta S = \frac{1}{\alpha-1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} (1-g^{\frac{\alpha-1}{\alpha}})$

Loss of surplus as a proportion of the initial surplus,

$$\frac{\Delta S}{SQ_0} = \frac{\frac{1}{\alpha-1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}} (1-g^{\frac{\alpha-1}{\alpha}})}{\frac{1}{\alpha-1} A^{\frac{1}{\alpha}} \tilde{Q}^{\frac{\alpha-1}{\alpha}}}$$

$$\text{or, } Q/S = 1-(1-k)^{\frac{\alpha-1}{\alpha}} \qquad (4)$$

Under *random allocation* the loss of surplus as a proportion of the initial surplus is

$$R/S = k \qquad (5)$$

The *excess burdens* of queue rationing and random allocation (iethe extra loss of surplus as compared with price rationing, expressed as a proportion of the initial surplus) are as follows:

Queue Rationing

$$\frac{Q-P}{S} = 4 - 3 = 1-(1-k)^{\frac{\alpha-1}{\alpha}} - \left[1-\alpha(1-k)^{\frac{\alpha-1}{\alpha}} + (\alpha-1)\ (1-k)\right]$$

$$= (\alpha\ 1)\ (1-k)^{\frac{\alpha-1}{\alpha}} - (\alpha-1)\ (1-k) = (\alpha-1)\left[(1-k)^{\frac{\alpha-1}{\alpha}} - (1-k)\right] \quad (6)$$

Random allocation

$$\frac{R-P}{S} = 5 - 3 = k - \left[1-\alpha(1-k)^{\frac{\alpha-1}{\alpha}} + (\alpha-1)\ (1-k)\right]$$

$$= k - \left[k-\alpha\ (1-k)^{\frac{\alpha-1}{\alpha}} - (1-k)\right] = \alpha\left[(1-k)^{\frac{\alpha-1}{\alpha}} - (1-k)\right] \quad (7)$$

APPENDIX II - BARZEL ANALYSIS OF REDISTRIBUTION EFFECTS OF QUEUING

Barzel (1974, pp87-88) has argued that in the special case where the demand for a good by every individual is of the form $q=\alpha p^{\beta} y^{\gamma}$ (where q is quantity, p is price, y is income, β is price elasticity and γ is income elasticity) and where individuals differ in their incomes, have time-costs proportionate to their incomes, and may queue for a standard quantity of the good, the queue will be composed of either the richest or the poorest individuals, according as

$-\frac{\gamma}{\beta} > 1$, or $-\frac{\gamma}{\beta} < 1$. He says: 'With $q = \alpha p^{\beta} \underline{y^{\gamma}}$

We can write $p = \bar{\alpha} \, q^{1/\beta} \; {}^{1/\beta} \; {}^{-\gamma/\beta}$
where p may be interpreted as the marginal valuation of the q^{th} unit. To avoid infinite total utility, however, assume that the valuation of the entire first unit is, say, at the same rate as the valuation at q = 1. Given q the relative valuation of the marginal unit by two individuals i and j with incomes y_i and y_j is $(y_i/y_j)^{-\gamma/\beta}$. Since this holds for any value of a q common to the two individuals, it is also true of the total valuation. Given the assumption that across individuals time costs, denoted by w are proportionate to income, the relative cost for the two individuals is $w_i/w_j = y_i/y_j$ and will exceed the relative valuation $(y_i/y_j)^{-\gamma/\beta}$ only if $-\gamma/\beta < 1$. For such values of γ/β, then, time-cost rises faster with income than does the valuation of the good. Consequently, if $-\gamma/\beta < 1$, as the number of units to be distributed increases, the order in which individuals will join the queue is from the poorest to the richest. If $-\alpha/\beta > 1$, the order is reversed. Both Becker (Becker 1965) and Nichols (Nichols 1971) then, err in implying that queuing will necessarily result in a redistribution towards the poor. Finally, if $-\gamma/\beta = 1$, the time-cost rises at the same rate as income. At a low time-cost all individuals are indifferent whether they get the good or not, and for any individual in the queue the waiting cost equals the entire consumer surplus.

REFERENCES

Barzel, Yoram, 1974, A Theory of Rationing by Waiting, *Journal of Law and Economics,* XVII(1), April, pp73-95.

Becker, Gary S, 1965, Theory of the Allocation of Time, *Economic Journal,* p493.

Butlin, S.J. and C.B. Schedvin, *War Economy 1942-45,* Australian War Memorial, Canberra, 1977.

Freeman, Myron A.III, 1977, The Incidence of the Costs of Controlling Automobile Air Pollution, in F. Thomas Juster ed, *Studies in Income and Wealth,* National Bureau of Economic Research.

Johnson, M. Bruce, 1966. Travel Time and the Price of Leisure. *Western Economic Journal,* Spring, pp135-145.

Nichols, D., E. Smolensky and T.N. Tideman, 1971. Disrcimination by Waiting Time in Merit Goods. *American Economic Review,* 61 p312.

Oort, C.J., 1966. The Evaluation of Travelling Time. *Journal of Transport Economics and Policy,* III, pp279-286.

Turnovsky, Stephen J. 1978. The Distribution of Welfare Gains from Price Stabilisation; The Case of Multiplicative Disturbances. *International Economic Review,* 17, pp133-148.

Waugh, Frederick V. (1944). Does the Consumer Benefit from Price Instability. *Quarterly Journal of Economics.* LVIII pp602-614.