1982

On the Use of Distributional Waits

Sam Bucovetsky

Follow this and additional works at: http://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:
RESEARCH REPORT 8202

ON THE USE OF DISTRIBUTIONAL WAITS

by

Sam Bucovetsky

Abstract

Commodities in excess demand are often allocated by queues. The effective prices of such commodities to individuals are thus increasing functions of the individuals' value of time. In simple models of optimal taxation, this value equals the net wage. In this paper, the hypothesis that such rationing by waiting is an efficient method of redistribution to the less productive is examined. It is shown that when the other government instruments are proportional commodity taxes, and lump-sum grants, the case for such rationing is quite strong. Conditions for optimality of rationing by waiting are derived, and several examples are presented.

Sam Bucovetsky
Department of Economics
University of Western Ontario
London, Ontario N6A 5C2

January 1982

Comments from Chin Lim, John McMillan, Mike Veall, Antal Deutsch and participants at the Queen's Public Finance Workshop and the U.W.O-U of Michigan Theory Workshop were very helpful. The usual disclaimer applies.
On the Use of Distributional Waits

It is noteworthy how many commodities are allocated by waiting. To the non-economist, these waits are due to "shortages". Economists tend to prefer to ascribe them to excess demand at the posted prices. Such waits are a commonplace in many centrally planned economies. Presumably there are political reasons for not raising prices there, yet central planners are unable or unwilling to provide market-clearing quantities at the official prices. The recent (Autumn 1981) Polish experience illustrates these problems, and Deutsch (1981) offers an interesting explanation of its development. But waiting in line is not a habit confined to Warsaw Pact countries. Metered parking spaces (during peak hours), popular concerts and football games, and squash courts in university gyms are examples of goods whose low price and limited supply give rise to long lineups.

Perhaps the most obvious recent experience of such queues were the intermittent periods in which long lines formed at gas stations in the United States. Clearly such lines could have been eliminated by raising gasoline prices. But even given the gasoline price, there appear to be less wasteful mechanisms for allocating excess demand than "first come, first serve". An appropriate number of ration coupons would presumably end these lineups, whether or not resale was permitted. And the "odd-even" license number rules eventually imposed also served this purpose. Although actual queues may not be observed, other underpriced commodities give rise to similar phenomena. It has often been noted that rent controls, by reducing mobility, give rise to longer journeys to work than would otherwise obtain. In addition people may devote more time to hunting for accommodation. Subsidized public housing leads people to devote time and
energy towards being among the select few who receive the subsidized units. In medical care as well, low-priced public provision has been observed to cause a substitution of time costs for money costs.

In all the above examples, pricing below the market-clearing level led to time being "wasted". Quite apart from the efficiency loss due to the "incorrect" price, there is an additional loss due to the fact that property rights to the subsidized commodity have not been assigned. People wait in line to acquire the property right. This phenomenon has been well analyzed by Barzel (1974). He notes in such situations the waiting is not due to the speed at which the sellers of the commodity serve the buyers. Speadier gas pumps will not reduce the lineups at gas stations. The length of the line is determined by how long the marginal consumer is willing to wait. As Cheung (1974) notes in a related paper, "The 'rent' thus dissipated constitutes a waste in the sense that valuable resources are allocated to the waiting, which produces nothing of specifiable value." (Cheung, 1974, p. 71.) My purpose in this paper is to present the case in favour of such "wasteful" waiting.

Certainly such a case involves "second-best" type arguments. Given costless redistribution of endowments, any outcome involving "wasteful" waiting can be dominated by one in which all markets clear. Thus the range of government instruments will be restricted in my analysis. Several authors have demonstrated that it may not be optimal for markets to clear when government instruments are somehow limited. Weitzman (1977) showed how (non-transferable) ration coupons may be superior to the price system in allocating some good. Guesnerie (1981) showed that redistribution in kind may be preferable to cash transfers, provided that recipients of the
redistribution could not resell their transfers. Southey (1979) has
shown that it may be optimal for the government to set prices and quantities
so that markets do not clear, with the short side of markets rationed
by some known rule.

In the above work there is no queuing. The property rights to sub-
sidized goods have been costlessly assigned. By contrast, Porter (1977)
shows the potential superiority of queuing to (non-transferable) coupons,
for given price and quantity. His point is that queuing will allocate the
unit to those who value it most highly. Donaldson and Eaton (1981) show
that a monopolist may wish to sell some of his output so cheaply that a
queue forms. Despite the fact that this policy implies some of consumers'
expenditure is in the form of time, which does not accrue to the seller,
this loss may be outweighed by the potential advantage of price discrimination.

My analysis is in a sense "dual" to that of Donaldson and Eaton.
It rests on the differences in the valuation of time between different agents.
Whereas the monopolist uses this difference to price discriminate, the
welfare-maximizing government uses it as a redistributive tool. The argument
is a simple one, using the familiar model of optimal taxation. Suppose
individuals differ only in ability, measured by the amount of output they
can produce per hour. If such ability were observable, utilitarianism would
dictate redistribution of income between ability groups so as to equalize
the marginal utility of income. But if ability cannot be observed this
policy cannot be implemented. However income, the product of ability and
hours worked, may be observed. Income may be used as a signal of ability,
so that the optimal income tax is a form of self-selection device, enabling
the government to separate income classes. But this device is not costless;
it may distort individuals' labour-leisure choices. Therefore, any additional signals of ability may be useful tools of tax policy. By reducing dependence on the income tax, they may reduce its distortions.

Rationing by waiting is such a tool. The price of waiting is higher to the able than to the less able, since they can earn more in their alternative uses of time. This is clearly not a costless form of redistribution, since time waiting in line is a total waste. But as second-best theorists from Ramsey on have noted, such costs do not rule out some use of this tool by a suitably constrained government.

My model of the mechanics of queuing is similar to Barzel's. There is an endogenous unit wait $m_i$ for commodity $i$. Purchase of $x_i$ units entails a unit of $m_i x_i$ hours (and a payment of $p_i x_i$ dollars). Implicit here is that there are constant returns to scale in waiting. This assumption is primarily for analytic tractability. Note however that increasing returns to scale (waiting in line entitles you to all you want to buy) would relatively benefit the higher-ability people, at least for normal goods. It also creates problems if resale is allowed. The constant-returns-to-waiting assumption can certainly be imposed, by assuming a rule such as "one to a customer" is in force. The endogenous wait clears the market; in aggregate people are willing to buy $\bar{x}_i$ units if they have to wait $m_i$ hours per unit and pay $p_i$ dollars per unit, where $\bar{x}_i$ is the total supply. Of course, $m_i$ cannot be negative. If supply exceeds demand, even with no queue, then $m_i = 0$.

Throughout the paper it is assumed that there are only 2 ability classes. Each class has the same utility function, and differs only in the amount of output $w_i$ that people of class $i$ can produce per hour. If the government could set an arbitrary income tax function, and knew the relevant
parameters, then it is very plausible that a first-best Pareto optimum could be achieved. Instead, I will allow only linear taxes. In addition, a lump-sum grant $g$ will be allowed. This grant may be negative, but it must be the same for all individuals.

The question posed in this paper is the following. Suppose the government has chosen its tax instruments (the grant and the linear taxes) optimally. Will it ever be the case that welfare could be further increased by making available some quantity of a commodity at a price below the prevailing market price? It is assumed that a queue will form for such a subsidized item, its length endogenously determined as described above. Notice that the question implies the commodity may also be available at the (higher) market price with no waiting. It will be shown that the continued existence of such a private market is always optimal. Therefore, for example, the coexistence of private entrepreneurs with cheaper state markets in which one must wait is supported by the analysis.

In Section 2 below the model just presented will be analyzed formally, and conditions derived under which the answer to my question is "yes". These conditions seem surprisingly plausible. But prior to this, a very contrived example is presented, which demonstrates why rationing by waiting may be optimal in some cases.

1. A Contrived Example

Suppose there is only one consumption good. Let utility be of the Stone-Geary form

$u(x, \lambda) = (x-x^*)^\alpha (\lambda-\lambda^*)^{1-\alpha}$ if $x \geq \bar{x}$ and $\lambda \geq \bar{\lambda}$

$= 0$ otherwise

(1)

where $x$ is consumption of the numeraire commodity, $\lambda$ is leisure consumption, and $\bar{x}, \bar{\lambda}$ are positive "required" levels of consumption. Production technology
is linear. Each type-2 worker produces $w_2$ units of output per hour. But type-1 workers are totally useless at productive labour. Therefore, in the absence of government intervention, such workers will attain zero utility. And whatever the government policy, such workers will never work.

In the absence of rationing by waiting government policy can be described by 2 numbers, the commodity tax rate $t$ and the lump-sum grant $g$. For government policy to raise the utility level of the low-ability above 0, $g > \bar{x}(l+t)$. But the demogrant $g$ must be given to everyone (since ability is unobservable). If leisure has a high income elasticity, the grant will greatly reduce work done by the high-ability group. Let $s_i$ denote the "surplus" obtained from group $i$, that is the excess of the output produced per person over the consumption per person (measured in terms of the numeraire). Thus $s_i = w_ih_i - x_i$, where $h_i$ is hours worked per capita, and $x_i$ the consumption. Utility maximization implies

$$
\lambda_2 = \bar{x} + \frac{(1-\alpha)}{w}[wL + g - \bar{x}(l+t) - w\bar{x}]
$$

(2)

$$
x_2 = \bar{x} + \frac{\alpha}{l+t}[wL + g - \bar{x}(l+t) - w\bar{x}]
$$

where $L$ is hours available for work, and $\lambda_2$ is leisure consumption ($\lambda = L-h$).

Thus if $g > (l+t)\bar{x}$, it then follows from (2) that $s_2 < \omega_2(L-\bar{x}) - \bar{x}$. The consumption of the low-ability group ($\frac{g}{l+t}$) must come from the surplus of the high-ability group. A consumption level of $\bar{x}$ can be financed only if a surplus

$$
\frac{N_2}{N_1}\bar{x}
$$

can be extracted from each high-ability worker, where $N_1$ is the number of people in class 1. Thus if condition (3) holds,

(3)

$$
\frac{N_1+N_2}{N_1}\bar{x} > \omega_2(L-\bar{x})
$$
there is no way that the low-ability group can be given a "subsistence" level $\bar{x}$ of consumption by a linear income tax system.

However, if $g = 0$, a larger surplus may be raised from the high-ability group. From equations (2)

$$s_2 = \omega_2 (L - \tilde{L}) \left( \frac{t}{1+t} \right) + (1-\alpha) t \bar{x}$$

if $g = 0$.

Consider a tax rate $t = \frac{N_1}{N_2}$. If condition (5)

$$w_2 (L - \tilde{L}) > \left( \frac{N_1 + N_2}{N_2} \right) \bar{x}$$

holds, then this tax can raise a surplus of at least $\left( \frac{N_2}{N_1} \right) \bar{x}$ per high-ability worker, as can be confirmed from (4) above. Since (4) also shows $s_2$ to be increasing in $t$, any tax rate greater than $\left( \frac{N_1}{N_2} \right)$ can raise enough of a surplus to put the low-ability group above subsistence. Clearly conditions (3) and (5) are not inconsistent (and are most likely to be satisfied when (i) there are relatively many high-ability people (ii) the high-ability people are productive and (iii) leisure has a high income elasticity).

The problem remains of getting the surplus from the high-ability group to the low-ability group. But this (finally) is where rationing by waiting enters. Suppose that a quantity $N_1 x^*$ of the consumption good is made available free of charge ($x^* > \bar{x}$). The good remains available on private markets at a price of $1+t$. If only the less able were to wait in line for this largesse, then the length of the wait in line $m$ would be determined so as to equate the value of an hour's leisure $u_\ell$ with the value of an hour's waiting $\frac{u_\ell}{m}$. From the definition of $u$ above (equation (1)),

$$m = \alpha(L - \tilde{L})/(x^* - (1-\alpha) \bar{x})$$

Since $m$ decreases with $x$,

$$m < \frac{L - \tilde{L}}{\bar{x}}$$

$$\lim_{x^* \to \bar{x}} m = \frac{L - \tilde{L}}{\bar{x}}$$
What remains to be shown is that the queue functions are a self-selection mechanism. To the able, the return to working is \( \frac{w_2}{1+t} \) in terms of the numeraire. The return to waiting is \( \frac{1}{m} \). But if \( t = \frac{N_1}{N_2} \), and \( x^* = \bar{x} \), the conditions (5), (6) imply \( \frac{w_2}{1+t} = \frac{N_2}{(N_1+N_2)} \frac{\bar{x}}{L-\bar{x}} = \frac{1}{m} \). By continuity, then, there is some tax rate \( t > \frac{N_1}{N_2} \) and some \( x^* > \bar{x} \) such that the high-ability still prefer to work (and generate enough surplus to provide each of the less able with \( x^* \) units of consumption).

Therefore this example shows that it may be optimal to ration by waiting. The unique Pareto optimum when only a linear tax system could be used was the no-tax initial situation, if condition (3) holds. But when the consumption good is provided free to those willing to queue, the utility of the less able can be increased above the no-tax situation (if condition (5) holds). Thus the set of Pareto optima has been expanded by the use of rationing by waiting. The policy described involved a grant of \( g = 0 \); there is no reason why this need be optimal. But if decision makers' values weight the welfare of the low-ability group sufficiently greatly, the optimal policy should involve rationing by waiting.

2. The Formal Model

The general model is of a 2-class, J-commodity world, in which the government's instruments are \( J \) commodity tax rates, a demogrant available to each individual, and \( J \) quantities of the commodities to be sold at \( J \) subsidized prices. Some notation:

- \( q \): vector of \( J \) producer prices \((q_1 = 1)\)
- \( t \): vector of \( J \) tax rates
- \( g \): level of lump-sum grant
- \( x_i \): vector of commodity consumption of class \( i \)
\( l_i \): leisure consumption of class \( i \)

\( h_i \): hours worked by class \( i \)

\( w_i \): output (of good \( i \)) produced per hour by class \( i \)

\( X \): vector of quantities provided on subsidized markets

\( p \): vector of prices charged on subsidized markets

\( m \): vector of waiting time per unit purchased on subsidized markets

\[ s_i = w_i h_i - q_i x^i \]: surplus extracted from class \( i \)

\( N_i \): number of people of class \( i \)

The government seeks to maximize some welfare measure

\[ W[u(x^1, l_1), u(x^2, l_2)] \] subject to a revenue constraint \( \Sigma N_i s_i \geq \bar{R} \), where \( \bar{R} \) is some revenue requirement. Without loss of generality, the subsidized prices can be restricted to be less than or equal to market prices: \( p \equiv q + t \).

It also can be assumed that a queue forms if and only if the subsidized price is strictly below the market price. For if \( m_j = 0 \) and \( p_j < q_j + t_j \) no one would choose to shop on the private market. In such a case (if \( m_j = 0 \)) demand would be less than the quantity available on the subsidized market. Hence nothing would be changed if the tax \( t_j \) were replaced by \( t'_j = p_j - q_j \). Then we would have \( m_j = 0 \), \( p_j = q_j + t'_j \) (and still no transactions on the private market, since there is excess supply on the subsidized market). Without loss of generality, then,

\begin{equation}
(7) \quad p \equiv q + t; \quad m_j > 0 \text{ if and only if } p_j < q_j + t_j
\end{equation}

One can define i's "full price" \( p^i_j \) on a subsidized market as the cash price plus the money equivalent of the waiting price.\(^5\)

\begin{equation}
(8) \quad p^i_j = p + w_i m
\end{equation}

This formulation is not correct if class \( i \) chooses to do no remunerative work. Nothing of substance is changed in this case, and so it will not be
discussed further. By construction, class $i$ will strictly prefer the private (subsidized) market when $q_j + t_j < p_j^i (q_j + t_j > p_j^i)$. Without loss of generality, I will assume that the less able class transacts all its purchases on the subsidized market when $m_j = 0$. If $q_j + t_j = p_j$, it does not matter on which market a transaction takes place.

It must be true that

$$ p^1 \leq q + t \leq p^2 $$

(9)

If anyone waits in line in the subsidized market, it must be because they get at least as good a deal there as on private markets, which (along with (7) above) implies $p^1 \leq q + t$. If we had $p_j^2 < q_j + t_j$, then $t_j$ could be replaced by $t'_j = p_j^2 - q_j$. Such a replacement would not affect anyone's behavior, except, perhaps by inducing high-wage people to shop on private markets. A switch of this type would be all to the good, since government revenues would have increased and no one's utility changed. In fact, I will assume that if $p_j^2 = q_j + t_j$, the high-wage group will all choose to purchase on private markets. This sort of assumption is common in self-selection models, such as Rothschild and Stiglitz (1976). Under that assumption it can be shown that it is always optimal to have $q + t = p^2$.

The reason for this optimality is that it is only the full prices which matter to class $1$. Their consumption vector $(x^1, k_1)$ can be written as a function of $(p^1, w_1, s)$. Thus if the price-waiting time combination $(p_j, m_j)$ is replaced by $(p_j + \epsilon, m_j - \frac{\epsilon}{w_1})$, neither the consumption vector nor the utility of group $1$ will be changed. Suppose then that $p_j^2 > q_j + t_j$. If $p_j$ is replaced by $p_j + \epsilon$ and $m_j$ by $m_j - \frac{\epsilon}{w_1}$, it will still be true that group $2$ strictly prefers the private market, if $\epsilon$ is small enough. Therefore the consumption vector of neither group has been affected by the replacement.
However, group 1's time spent waiting has decreased by $\frac{e}{w_1} X_j$. Since leisure time is unchanged, they must be working more. Therefore their surplus $s_1$ must have increased. As long as the revenue constraint ($\sum s_1 \geq R$) is binding, this means there is room for further welfare increases. As long as some $p^2_j > q_j + t_j$, such improvements can continue. Thus only when $p^2 = q + t$, so that the high-wage group is on the margin between markets can we be at an optimum. As a corollary, the low-wage group must strictly prefer the subsidized market if it charges a lower price. Their waiting earns them a rent.

Given that $p^2 = q + t$, the government's policy choice on subsidized markets can be reduced to the selection of a "full price" vector $p^1$. For then the quantities $X$ are determined from the low-wage group's demand. And the waiting time $m_j$ and the cash price $p_j$ are the solutions to the two equations $p^1_j = p_j + w_1 m_j; p_j + w_2 m_j = q_j + t_j$. These yield solutions

$$\begin{align*}
\text{(10)} & \quad p_j = \frac{1}{w_2 - w_1} \{p^1_j w_2 - w_1 (q_j + t_j)\} \\
\text{(11)} & \quad m_j = \frac{1}{w_2 - w_1} \{q_j + t_j - p^1_j\}.
\end{align*}$$

Note that $p^1_j = q_j + t_j$ implies $m_j = 0$ and $p_j = q_j + t_j$.

If $v$ denotes the indirect utility function, then the government's problem can be stated as that of maximizing

$$\begin{align*}
\text{(12)} & \quad w*(t, g, p^1) = w[N_1 v(p^1_j, w_1, w_1 L^g), N_2 v(q + t, w_2, w_2 L^g)] + \lambda \{N_1 s_1 \{p^1_j, w_1, g\} + N_2 s_2 (q + t, w_2, g) - R\}
\end{align*}$$

Consider the optimal choice of tax variables $(t, g)$ first. If we constrain $p^1$ to equal $q + t$ this is just the "two-class Ramsey problem". Therefore the solution to this problem derived by Mirrlees (1975) can be used. Using my notation, with the addition of
\( \alpha_i \): social marginal utility of income of class \( i \equiv \frac{1}{N_i} w_i v_3 \)

\( x_g \): vector of income derivatives of demand

\( S^1 \): Slutsky matrix of class \( i \) (\( J \times J \) sub-matrix - leisure is excluded)

Mirrlees' equation (7) can be written

\[
(13) \quad \left( \frac{1}{\lambda} - 1 + t \cdot x_g \right) N_1 (x^1 - x^2) = (N_1 S^1 + N_2 S^2) \cdot t
\]

The case for rationing can now be examined. This case requires \( \frac{\partial W*}{\partial p_j} \) to be negative at \( p^1 = q + t \), for some \( j \), when the taxes and lump-sum grant have been optimally chosen. For this is precisely the condition that welfare rise with a small decrease in some subsidized price, when everything else is optimized out. From Roy's Identity, and the definition (12)

\[
(14) \quad \frac{\partial W*}{\partial p_j} = - \alpha_1 N_1 x_j^1 + \lambda N_1 \frac{\partial s_1}{\partial p_j}
\]

From the low-ability group's budget constraint \( s_1 = (p-q) \cdot x_g \). The definition (10) of \( p_j \) implies \( \frac{\partial p_j}{\partial p^1_j} = \frac{w_2}{w_2 - w_1} \), so that

\[
(15) \quad \frac{\partial s_1}{\partial p_j} = \left( \frac{w_2}{w_2 - w_1} \right) x_j^1 + \Sigma (p_k - q_k) \frac{\partial x_j^1}{\partial p_j}
\]

If there is no rationing by waiting, \( (p-q) = t \). Therefore substituting (15) into (14),

\[
(16) \quad \left. \frac{\partial W*}{\partial p_j} \right|_{p=q+t} = \left( \frac{w_2}{w_2 - w_1} \right) - \frac{\alpha_1}{\lambda} - t \cdot x_g^1 + \frac{\Sigma w - w_k}{w_1} ^1 \frac{\partial x_j^1}{\partial p_j} \lambda x_j^1 N_1
\]

The first term in parentheses on the right-hand side of (16) is \( \frac{w_2}{w_2 - w_1} \), which is larger the smaller is the difference in ability between groups. Not surprisingly, a small difference in ability weakens the case for rationing by waiting. The
other two terms \( \frac{\alpha_1}{\lambda} + t \cdot \frac{x^1}{g} \) in the parentheses are what have been referred to (e.g., Atkinson and Stiglitz [1980, p. 387]) as the net social marginal valuation of income. A result of optimal commodity taxation is that these net social marginal valuations average to 1, over all the population (e.g., Mirrlees [1975, eq. (14)]). Hence a necessary condition for the term in parentheses to be negative is that the low-ability group have a larger net social marginal valuation of income.

Equation (13) has been used by Mirrlees to derive a sufficient condition for the low-ability group to have the higher marginal social evaluation of income. If equation (13) is pre-multiplied by the vector \( t \), one gets

\[
(17) \quad \left( \frac{\alpha}{\lambda} - 1 + t \cdot \frac{x^1}{g} \right) N_1 = \frac{N_1(t' S^1 t) + N_2(t' S^2 t)}{t \cdot \left( \frac{x^1}{x^2} \right)}
\]

The numerator on the right-hand side of (17) must be negative, since the \( S^i \)'s are Slutsky sub-matrices. Hence \( \frac{\alpha}{\lambda} - t + t \cdot \frac{x^1}{g} \) will be positive if and only if the high-quality group pays more taxes \( t \cdot x^2 > t \cdot x^1 \). Therefore if \( t \cdot x^2 > t \cdot x^1 \), the expression in parentheses in (16) will be negative for sufficiently small \( \frac{w_1}{w_2 - w_1} \). Further, if all tax rates are positive, the final term in (16) must be negative for some \( k \). Hence, strong conditions which together would guarantee the optimality of some rationing by waiting are:

(i) \( w_1 = 0 \); (ii) \( t \equiv 0 \); (iii) \( t \cdot x^2 > t \cdot x^1 \). This precise condition is not terribly useful; the wage rate in (16) can never be zero if the marginal utility of leisure is positive. (If low-ability people choose not to work, the marginal utility of leisure divided by the marginal utility of income is the appropriate \( w_1 \) to use in equation (16).)
An alternative formula can be obtained by substitution of (16) in (13), yielding

\[
(18) \quad \left| \frac{\partial W^*}{\partial p_j} \right|_{p=q+t} = \lambda N_1 x_j^1 \left( \frac{w_1}{w_2 - w_1} \right) + \frac{N_2}{N_1} \frac{\Sigma k \frac{S^k}{1}}{x_j^2 \frac{1}{x_j^1}} + \frac{x_j^2}{x_j^1} \frac{\Sigma \frac{k}{1}}{x_j^2 \frac{1}{x_j^1}}
\]

In the special case where the optimal tax system is proportional, (18) can be further simplified. Then

\[
(19) \quad \left| \frac{\partial W^*}{\partial p_j} \right|_{p=q+t} = \lambda N_1 x_j^1 \left( \frac{w_1}{w_2 - w_1} \right) - \frac{N_2}{N_1} \left( \frac{t}{1 + t} \right) \left( \frac{\partial L_1}{\partial p_j} \right) \left( \frac{1}{x_j^2 \frac{1}{x_j^1}} \right) - \frac{x_j^2}{x_j^1} \left( \frac{t}{1 + t} \right) \left( \frac{\partial L_2}{\partial p_j} \right) \left( \frac{1}{x_j^2 \frac{1}{x_j^1}} \right)
\]

where

\[
\left( \frac{\partial L}{\partial p_j} \right)_c \text{: compensated derivative of leisure demand with respect to price of } j^{th} \text{ good.}
\]

This formula indicates that strong substitutability of goods for leisure leads to optimality of rationing by waiting (given that the optimal commodity tax structure is proportional). It also is used to examine the plausibility of rationing by waiting in a Cobb-Douglas example in Section 3 below.

The formulae presented here give little insight as to which commodities should be rationed by waiting. Although there is a lower welfare loss (for a given subsidy) in subsidizing price-inelastic goods, these considerations have already been used in the tax system. Since equations (16) and (18) involve endogenous parameters (such as tax rates), they tell little about which goods to subsidize. However, something can be said about the linear expenditure system. Here optimal taxes are proportional, so that equation (19) applies. Since \( \left( \frac{\partial L}{\partial p_j} \right)_c \) is proportional to \( x_j^2 \cdot x_j^1 \) (in the L.E.S.) the only term in brace brackets in (19) which varies among commodities is \( \frac{x_j^2}{x_j^1} \). This will be higher the higher is the ratio of supernumerary consumption to "essential"
consumption. Thus the strongest case for rationing by waiting is for luxuries, not necessities, at least in the linear expenditure system.

3. A Cobb-Douglas Example

To get some idea of the plausibility of the above results, some very simple computations were done for a 1-good Cobb-Douglas example. The fact that rationing by waiting can be justified even in a one-commodity world shows that the particular characteristics of the rationed good are not vital to the case for rationing. The results confirm that the case for rationing is strengthened by

- high welfare weights on the low-ability group
- large numbers of the high-ability group
- large differences in ability between groups
- a high income elasticity of leisure demand

as was suggested as well by the L.E.S. example in Section 1 above.

If \( u = x^{\beta} L^{1-\beta} \), then equation (19) can be written

\[
\frac{\partial w^*}{\partial p} \bigg|_{p=q+t} = \lambda N_1 \left(\frac{\beta(w_1 L + g)}{1 + t}\right) \left(\frac{w_1}{w_2-w_1}\right) - N_2 \left(\frac{(1-\beta)t}{N_1 L}\right) \left(\frac{w_2 L + g}{1 + t (w_2-w_1)L}\right)
\]

Thus \( \frac{\partial w^*}{\partial p} \) will be negative if and only if

\[
w_1 < \left(\frac{N_1 + N_2}{N_1}\right) \left(\frac{(1-\beta)t}{1 + t}\right) \left(\frac{w_2 + \frac{g}{L}}{w_2-w_1}\right)
\]

which confirms all the tendencies noted above (although of course \( t \) and \( g \) are endogenous). The results below were obtained by plugging in values of \( t \), and solving for the demograph \( g \) from the feasibility condition \( N_1 s_1 + N_2 s_2 = 0 \). This implicitly yields a value for \( \lambda \), the relative (gross) marginal social utility of income transfers to the less able group. For various values of \( \beta \), \( \frac{N_2}{N_1} \) and \( \frac{w_2}{w_1} \) the value of \( t \) was found which set \( w_1 = \left(\frac{N_1 + N_2}{N_1}\right) \left(\frac{(1-\beta)t}{1 + t}\right) \left(\frac{w_2 + \frac{g}{L}}{w_2-w_1}\right) \).
Hence for all larger values of \( t \) (and \( \lambda \)) rationing by waiting is optimal. It may also be noted that in several of the examples, there is a case for rationing by waiting when the social welfare function is "Benthamite", when the cardinal form of the utility function is \( x^\beta \). The results in Table 1 seem to suggest a reasonably strong case for rationing by waiting. In the "base case" of \( \frac{w_2}{w_1} = 3, \frac{N_2}{N_1} = 1, \beta = .4 \), rationing by waiting is justified whenever society values \$1\) to a low-ability person at more than or equal to \$2.26\) to a high-ability person. Such welfare weights would imply an optimal commodity tax rate of 35\% (or an income tax of 26\%).

4. Why Resale Should Not Be Allowed

It has been assumed throughout that commodities bought on subsidized markets cannot be resold. By contrast, Barzel argues that resale should always be allowed. His argument stems from the fact that if resale is allowed, then those with the lowest value of time will wait in line. In the absence of resale this may not be the case. Hence the advantages of resale follow from "comparative advantage" considerations. In the model presented here, however, it is only those with the lowest value of time (group 1) who wait in line, even in the absence of resale. In fact, it turns out that any outcome which can be achieved when resale is permitted can also be achieved without resale, provided that the low-income group does some productive work when resale is allowed.

If customers on low-priced markets can sell their subsidized purchases at the market price \( q_j + t_j \) to those unwilling to wait in line, then the return to queuing will increase. This increased return induces more waiting, so that all added rents due to resale are dissipated by longer queues. In
Table 1

Values for which rationing by waiting is on margin of being optimal (rationing is optimal for higher values of \( \frac{w_2}{w_1} \), \( \frac{N_2}{N_1} \) and lower values of \( \beta \) than those in table).

\[
\frac{w_2}{w_1} = 3, \frac{N_2}{N_1} = 1
\]

\begin{array}{ccc}
\beta & t & \lambda \\
.2 & .26 & 2.12 \\
.4 & .35 & 2.26 \\
.6 & .55 & 2.69 \\
.8 & no solution & \\
\end{array}

\[
\frac{w_2}{w_1} = 3, \beta = .4
\]

\begin{array}{ccc}
\frac{N_2}{N_1} & t & \lambda \\
.5 & .52 & 4.3 \\
1.0 & .35 & 2.26 \\
1.5 & .27 & 1.97 \\
2.0 & .22 & 1.84 \\
2.5 & .18 & 1.75 \\
\end{array}

\[
\frac{N_2}{N_1} = 1, \beta = .4
\]

\begin{array}{ccc}
\frac{w_2}{w_1} & t & \lambda \\
2.0 & .59 & 24.33 \\
3.0 & .35 & 2.26 \\
4.0 & .25 & 1.62 \\
5.0 & .20 & 1.42 \\
\end{array}

\( \frac{w_2}{w_1} \): ratio of wage of high-ability group to wage of low-ability group

\( \frac{N_2}{N_1} \): ratio of high-ability population to low-ability population

\( \beta \): exponent of commodities in Cobb-Douglas utility function (1 - \( \beta \) is an exponent of leisure)

\( t \): commodity tax rate

\( \lambda \): social marginal utility of income of low-ability group (relative to high-ability group) implied by \( t \) being optimal.
Barzel's model, heterogeneity of the people in the queue implies that the longer queue only dissipates rents of the marginal waiter; infra-marginal agents gain. But it is not only the difference in the number of income classes which causes the different results concerning resale. In the model presented here, rationing by waiting is one of several government instruments. The case for resale therefore must be considered when all the instruments are adjusted optimally, rather than ceteris paribus.

Formally, suppose resale is suddenly allowed in the model just presented. Low-wage individuals can now sell their goods at market prices. Let \( s \) denote the vector of net sales on the unrationed market. The low-wage agent's optimization problem is

\[
\max u(x - s, \lambda)
\]

subject to
\[
p \cdot x \leq g + w_1 h + (q + t) \cdot s
\]
\[
m \cdot x + \lambda + h \leq L
\]

But note that the return (in dollars) to waiting one hour in the queue for good \( j \) is \( \frac{q_j + t - p_j}{m_j} \). If the individual chooses to work, then the equilibrium wait must equalize the return to waiting with the return to working. Hence

\[ q + t - p = w_1 m \]

Substituting the above equality in the maximization, the problem becomes

\[
\max u(x, \lambda)
\]

subject to
\[
(q + t)z \leq g + w_1 (L - \lambda)
\]

But this is precisely the problem faced by a low-wage worker constrained to shop in the unsubsidized market. Since the queues when no resale is allowed obey \( p + w_1 m \leq q + t \), the low-wage worker is no better off.

Thus whenever the social planners' optimum actually involves rationing by waiting (i.e., \( m_j > 0 \) for some \( j \)), then resale should be banned. The
outcome achieved by resale could be achieved by simply abolishing subsidized markets. In fact, this latter outcome would be superior, since no time would be wasted in line. The above argument required positive hours of work by the low-wage. But if they chose not to work (so that \( q + t - p > w_1^m \)), the "resale" outcome could also be achieved without resale. Let \( m^r \) be the vector of waiting times when resale is allowed, \( z^r \) the consumption vector, and \( w^r \) the value of time spent in line \( (w^r > w_1, w^r = u^s/s/m.u.i.) \). Then if \( z^r \) units are provided at a subsidized price of \( p = q + t - w^r m \), and resale banned, the exact same consumption and work (namely nil) will be optimal for low-wage workers. Thus there is no case in this model for resale.

5. Conclusions

As noted in the introduction, rationing by waiting is a common phenomenon. The purpose of this paper has been to show that such waiting may be an efficient method of redistribution. The conditions under which it is efficient were enumerated in Section 3 below. In particular, they indicate that a society with a relatively small "unemployable" group whose value at productive labour is very low should consider rationing by waiting, if it wishes to increase the welfare of this group. One condition which was not listed above, but which is implicit in the analysis, is that there be no better way of redistribution to the group which gains from rationing.

For instance, it was assumed (as in most optimal tax theory) that there was no exogenous, observable characteristic which could be used to infer an individual's market wage, so that endogenous signals such as income were used. However this may not be the case. In Irkutsk, Siberia, basic foodstuffs were allocated by rationing by waiting, as seems common in the USSR.
But it was observed that only old women were able to purchase under this system. Therefore, rationing by waiting was replaced by a coupon system. Presumably, the beneficiaries of rationing by waiting could be compensated, by some lump-sum transfer to all old women. (The article which described this phenomenon did not indicate whether any such compensation was made.) Here the authorities appear to have discovered the inefficiency of rationing by waiting when some other costless signal is available.

In addition, even if redistribution must be made through an income tax, the form of this tax may be flexible. I have assumed only a linear income tax can be used. If the government is free to select an arbitrary non-linear income tax, the case for rationing by waiting is weakened. If utility is weakly separable in leisure, it will never be optimal to allow rationing by waiting, for any arbitrary distribution of abilities, if a non-linear income tax is allowed. Since all the examples here have involved separable utility functions, the restriction of the income tax to be linear is an important one.

Of course, assuming that any income tax schedule can be imposed probably overstates considerably the flexibility of government instruments. Highly non-linear taxes impose high administrative costs (to both authorities and taxpayers), and induce intricate rearrangement of income among family members, and between different years of taxpayers' lives. For such reasons, restriction of the form of the income tax may be quite reasonable. It should also be noted that still more general taxes are available. If non-linear commodity taxes were possible, rationing by waiting could be ruled out in even more general circumstances. But non-linear commodity taxes pose even greater administrative difficulties than income taxes.
Like most results in the theory of optimal taxation, the case for rationing by waiting presented here is not robust. Perhaps the most important (and obvious) message of this paper is that the potential efficiency of some policy depends crucially on the other policies available. Nonetheless, most countries derive most government revenue from proportional taxes. If such taxes are the main redistributive instruments, then the analysis of Section 2 applies. The line-ups observed for underpriced commodities may actually be an efficient form of redistribution to those whose marginal value of time is lowest.
Footnotes

1 See, for instance, Hayek (1975).

2 Stiglitz (1981) presents this model, quite clearly, and demonstrates how it is a problem in self-selection.

3 See Stiglitz (1981), Section 1.

4 Which may be negative.

5 A person of class i chooses \((x, \ell)\) to maximize \(u(x, \ell)\) subject to
(i) \(p \cdot x + wh \leq g + wh\), (ii) \(h + m \cdot x + \ell \leq L\). Substituting for \(h\) from the second constraint (assuming it holds with equality), this can be reduced to

\[
\text{maximize } u(x, \ell) \text{ subject to } (p + w)m \cdot x + \ell \leq wL + g
\]

so that the individual can be viewed as facing prices \(p + wm\) for commodities, \(w\) for leisure and having exogenous income of \(wL + g\).

The above substitution is incorrect if the individual chooses not to work, since then the constraint \(h \geq 0\) becomes binding.

6 The example of Section 1 involved the less able group doing no remunerative work.

7 From footnote 5 above, if \(h > 0\) the low ability group's problem is

\[
\text{maximize } u(x, \ell) \text{ subject to } p^1 \cdot x + w_1 \ell \leq w_1 L + g
\]

which implies their optimal \((x^1, \ell)\) are functions of prices \((p^1, w_1)\) and income \(w_1 L + g\) alone.

8 The assumption here is that production efficiency is desirable at an optimum. This can be demonstrated by using the arguments of Diamond and Mirrlees (1971).

9 \(v(p, w, y) = \max u(x, \ell) \text{ subject to } p \cdot x + w \ell \leq y\).
\[ s_1 = w_1 h_1 - q \cdot x^1 \text{ and } p \cdot x^1 = g + w_1 h_1 \]

11 From the adding-up property of compensated derivatives,

\[ \sum_{p_k} s_{jk}^1 = -w_1 \frac{\partial \lambda}{\partial p_j} \text{ at } p = p^1 = q + t \]

Since taxes are proportional, \( p_k = q_k (1 + t) \) and \( t_k = t q_k = \frac{t}{1 + t} \), yielding equation (18).

12 If \( u = \beta_0 \log(L - \bar{x}) + \sum \beta_j \log(x_j - \bar{x}_j) \), then

\[ x_j = \bar{x}_j + \beta_j [w_1 L + g - w_1 \bar{x}_j - \sum \beta_j \bar{x}_j] . \]

Thus \( x_j^2 = \beta_j (w_2 - w_1) (L - \bar{x}) \), and \( (\partial \lambda / \partial p_j)^c = -\beta_0 \bar{x}_j + \beta_0 x_j = \beta_0 \beta_j (w_1 L + g - w_1 \bar{x}_j - \sum \beta_j \bar{x}_j) . \)

13 From footnote 12

\[ x_j^2 = \frac{1 + w_2 (L - \bar{x})}{1 + w_1 (L - \bar{x})} \frac{\beta_1}{x_1} + \frac{g}{x_1} \frac{\beta_1}{x_1} \frac{\beta_1}{x_1} \frac{\beta_1}{x_1} \]

which is an increasing function of \( \frac{\beta_1}{x_j} \) if \( w_2 > w_1 . \)

14 For the Cobb-Douglas function, the equation \( N_1 s_1 + N_2 s_2 = 0 \) implies

\[ g = \beta t (w_1 + \frac{N_2}{N_1} w_2) L / (1 + \frac{N_2}{N_1}) \]

and \( \frac{dg}{dt} = \frac{g}{t(1 + (1 - \beta) t)} \)

(if the constraint \( h_1 \geq 0 \) is not binding, which was checked). Since \( \frac{dg^*}{dg} = N_1 \lambda_1 + N_2 \), \( \frac{dg^*}{dt} = -N_1 \lambda_1 x_1 + N_2 x_2 \), \( \lambda_1 \) can be solved as \( \lambda_1 = \frac{N_2}{N_1} (x_2 - \frac{dg}{dt} / (\frac{dg}{dt} - x_1)) \).

Note some \((g, t)\) combinations--for high enough \( t \)--are not Pareto optimal, but these would imply \( \lambda_1 < 0 \).

16 This example is from S. Schmemann (1982).

17 This result is derived in an earlier draft of this paper, available from the author. But it is a simple corollary of the theorem that when utility is separable in leisure, income is the only observable signal a government need use to attain a second-best optimum (cf. Stiglitz (1981), or Atkinson and Stiglitz (1980, Chapter 14-3)).

18 Again, see Stiglitz (1981) for a convincing elucidation of this point.
References


