International Journal of Industrial Organization 5 (1987) 435–446. North-Holland

MONOPOLY, QUALITY UNCERTAINTY AND 'STATUS' GOODS

Kaushik BASU*

Delhi School of Economics, Delhi 110 007, India

Final version received April 1987

This paper provides an explanation of price and wage rigidities in certain industrial structures. It considers products for which quality is important or the possession of which enhances a person's 'status'. It is argued that an individual's desire for such a good or service may be positively related to the amount of aggregate excess demand for it. It is then shown that this gives rise to a natural discontinuity in demand which may result in price rigidities. This explains, for instance, why it may be profit-maximising for a doctor facing a long waiting-line of patients not to raise his fees.

1. Introduction

To buy certain products where quality is important, for example Jaguar cars, the newly released shares of a company, meals at a popular Chinese restaurant, or consultation with a 'good' doctor, one has often to join a waiting list.[1] In other words, these 'firms' price their products so as to maintain an excess demand. At first sight this is paradoxical. Why does the doctor not raise his fees and eliminate or shorten the waiting line of patients? While this *may* be attributed to the benevolent spirit of the medical profession, the present paper provides an explanation of such behaviour in terms of selfishness and, in particular, profit maximisation. It presents a theory of why and under what circumstances it pays a firm to follow a strategy of maintaining excess demand.

There is a long tradition in economics – dating back to at least Scitovsky's paper of 1944 – of modelling markets where consumers judge quality by price. In a situation of price rigidities, a natural analogue of this is that consumers will judge quality by also observing the aggregate excess demand for the good. At first sight, this seems to provide an explanation of why a

*This problem was suggested to me by Avinash Dixit. I am also grateful to two anonymous referees for very helpful comments. A first version of this paper was written at the Institute for Advanced Study, Princeton.

[1]It has, for instance, been common practice for companies liquidating their foreign holdings and selling shares to Indian citizens in response to the Foreign Exchange Regulation Act to price them so as to have a large excess demand. The same has been observed for debenture issues of private companies.

firm may wish to maintain a price at which there is an excess demand: because by this it manages to increase the desirability of its product to its individual consumers.

This argument, however, turns out to be inadequate if demand functions are continuous. Consider a monopoly. In the presence of excess demand its sales need not be affected by small changes in price. In other words, it is possible for the firm to raise price a little without having to sell less, thereby increasing its profit. Hence, an excess-demand situation cannot be a profit-maximising one.

Note that the above reasoning makes use of the continuity of demand. Hence, *if it were the case* that people judge quality by excess demand *and* the demand function is discontinuous, only then may it be possible to explain excess demand in equilibrium. What is interesting, and this is the central point of this paper, is that if people actually judge quality by excess demand, then the demand function automatically turns out to be discontinuous and under certain conditions this discontinuity is precisely of the kind that results in price rigidities and excess demand. It is important to emphasise that the discontinuity of demand which plays a crucial role in this paper is *not* an assumption but arises naturally in this framework. It arises even though the utility function or the n-function (defined below), i.e., all the basic ingredients of the model, are continuous.

While the model in this paper is motivated by considering products where quality is judged by excess demand, there are several other situations where the formal model would be applicable. First, it can be applied to the supply side. For instance, by thinking of jobs as work environments or certificates of prestige which people buy, my argument can be inverted to explain wage rigidities in the face of excess supply of labor. Secondly, it could apply to markets where the quality of the product is not in question but where people buy a good or service to establish their own social status or worth. The membership of a club could often be a symbol of status. Calcutta's old colonial Calcutta Club provides an example of this. One reason why an individual covets the membership of this club is because there are others who covet the same but nevertheless fail to become members. Aggregate excess demand for membership enhances an individual's value of the membership. These alternative motivations for the same formal model are discussed in section 4. The next two sections develop the model and its properties. In the present paper the problem is analysed within one specific industrial structure, namely, a simple monopoly.

2. The model

Let $H=\{1,\dots,h\}$ be the (finite) set of individuals. It will be assumed that there is only one producer of a certain good, for example, cars. Each person i

in H can buy at most one car, and the maximum amount of money, v_i, that he is willing to pay for the car depends on the excess demand (i.e., the difference between aggregate demand and aggregate supply), z, that exists: Hence, for all i in H,

$$v_i = v_i(z), \qquad v_i' \geqq 0. \tag{1}$$

If p is the price of a car and z the excess demand for it, a person i will want to buy the car if and only if $v_i(z) \geqq p$, and the total demand for cars, n (which is equal to the total number of people who want to buy), is given by[2]

$$n = n(z, p) = \#\{i \in H \mid v_i(z) \geqq p\}. \tag{2}$$

Given (1), it follows that as z increases or p falls, n will not fall. Since each individual can buy at most one car, n can take values between 0 and h. Finally, the finiteness of H ensures that for a sufficiently large value of p, n is zero. Within these restrictions, the function $n(z,p)$ can take any form. Hence, we may treat the function $n(z,p)$ as a 'primitive' as long as we keep the above restrictions in mind. In this paper, we do precisely that and, in addition (this is a harmless and mathematically convenient assumption), ignore the fact that n and z can take only integer values.

Hence, from now on we shall treat n as a mapping (R being used to denote the set of all real numbers and R_+ all non-negative ones),

$$n: R \times R_+ \to [0, h],$$

which is continuous, monotonically non-decreasing in the first variable, non-increasing in the second variable, and for a sufficiently large value of the second variable n takes the value of zero. I shall refer to this as an n-function.

Given a price of $\hat{p}$ and a supply of $\hat{x}$, what are the equilibrium amounts that may be demanded? Before giving a formal answer, it is useful to have a diagrammatic representation.

Fig. 1 shows the function $n(z, \hat{p})$ for different values of z. The amount supplied is measured from the origin in the left-ward direction and excess demand is measured in the right-ward direction. In this case, the amount supplied is marked by $\hat{x}$. Given any level of demand d, to find out the amount of excess demand, we simply have to mark a distance of d from $\hat{x}$ (to the right of $\hat{x}$). Thus, if d_2 is demanded, the excess demand is equal to $0A$.

[2]It is worth noting that a more sophisticated theory would require v_i to be a function of z and p, i.e., $v_i = v_i(z, p)$, instead of (1). But this complication is inessential, given the problem addressed in this paper. It will be clear as we go along that as long as aggregate excess demand figures in the individual demand function – no matter in what form – the model developed in this paper will be relevant. Hence there is no harm in using the simple idea captured in (1).

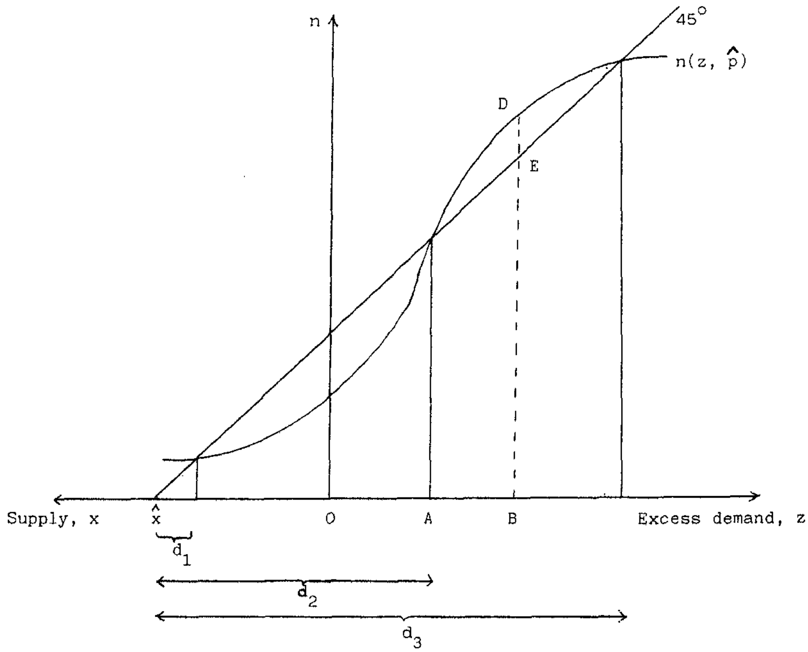

Fig. 1

Now draw a 45° line through $\hat{x}$, as shown. Let us check whether a particular level of demand is compatible with $\hat{x}$ being supplied at price $\hat{p}$. Consider the demand given by the distance $\hat{x}B$ in the diagram. This implies an excess demand of $0B$, which in turn implies that aggregate demand will be BD. But BD is greater than BE, which equals $\hat{x}B$. Hence, demand equilibria are given by the points of intersection between the n-function and the 45° line. Therefore, demand levels of d_1, d_2 and d_3 are the only demands compatible with $(\hat{x}, \hat{p})$. This is an essential part of a rational expectations equilibrium. According to this an individual's valuation of the good depends not on excess demand, as in (1), but his *expectation of* excess demand. An equilibrium occurs when the expected value matches the actual one.[3]

Now for the formal derivation. The set of demands compatible with (x,p) is denoted by $D(x,p)$ and defined as follows:

$$D(x,p)=\{r\in R_+ \mid n(r-x,p)=r\}. \tag{3}$$

Let us define $d(x,p)$ as the largest element in the set $D(x,p)$.

It is assumed that given (x,p), the producer knows that only the demand levels in the set $D(x,p)$ can persist and he can choose any of these. As will be

[3]A similar idea is used and elaborated upon in Basu (1986).

obvious as we go along, from his point of view it is always best to set the demand at $d(x,p)$. What we have to determine next is which (x,p) is the best from the monopolist's point of view.

Let $c(x)$ be the cost function, which is continuous and monotonically increasing in x. Hence, the producer's profit function is given by

$$Y(x,p)=p\min\{x,d(x,p)\}-c(x). \tag{4}$$

It is assumed that the producer maximises Y by choosing (x,p). Throughout the remaining pages I use (x^*,p^*) to denote an output level and a price which maximises profit.[4] Thus

$$(x^*,p^*)=\operatorname{argmax} Y(x,p) \tag{5}$$

(x^*,p^*) is at times referred to as the equilibrium. The trivial case of zero production is ruled out by assuming throughout that $x^*>0$.

To aid intuition, let us draw a picture of $D(x,p)$ for a fixed value of x, say at $\hat{x}$. This is done in fig. 2. As p is raised, the n-function shifts down. By plotting the points of intersection between the 45° line and each n-function corresponding to a price, in the lower diagram, we get a picture of the correspondence $D(\hat{x},p)$ for different values of p. Thus, for example, $D(\hat{x},p')=\{d'_1,d'_2,d'_3\}$. It is easy to see that $d(\hat{x},p)$ plotted against p is the thick discontinuous line *abce*, which at p'' takes the value of c instead of b.[5]

[4]The existence of a maximum is guaranteed by the following argument. Let P be a price such that for all prices above this Y is non-positive. The existence of such a P follows from the properties of the n-function. Clearly there is no advantage in raising x to above h. Hence x may be taken as varying between 0 and h. Hence the domain of the function, $Y(\cdot)$, is $[0,h]\times[0,P]$, which is compact. It is easy to check that $Y(\cdot)$ is upper semi-continuous, thereby ensuring that it achieves a maximum somewhere in the domain [Berge (1963, p. 74)].

[5]It is interesting to note that this demand curve, $d(x,p)$, is downward sloping in p. However, the proof is very different from the one used to establish the downward slope of the textbook demand curve. For a proof suppose $p^0<p^1$. We have to show that $d(x,p^1)<d(x,p^0)$.

Define $z^0\equiv d(x,p^0)-x$. Since $n(\cdot)$ is non-decreasing in p,

$$n(z^0,p^1)\leqq n(z^0,p^0).$$

Next, it is proved that

$$\text{for all}\quad z>z^0,\quad n(z,p^0)<x+z. \tag{i}$$

Let $z'>z^0$ and suppose

$$n(z',p^0)\geqq x+z'. \tag{ii}$$

Since $n(\cdot)$ is bounded, there must exist $\hat{z}>z'$ such that

$$n(\hat{z},p^0)<x+\hat{z}. \tag{iii}$$

(ii), (iii) and the continuity of $n(\cdot)$ imply that there exists $\ddot{z}\in[z',\hat{z}]$ such that $n(\ddot{z},p^0)=x+\ddot{z}$. This implies that

$$\begin{aligned} d(x,p^0) &\geqq x+\ddot{z} \\ &> x+z^0=d(x,p^0). \end{aligned}$$

This is a contradiction. Hence, (i) must be true, which implies that, for all $z>z^0$, $n(z,p^1)<x+z$. Hence, $d(x,p^1)\leqq x+z^0=d(x,p^0)$.

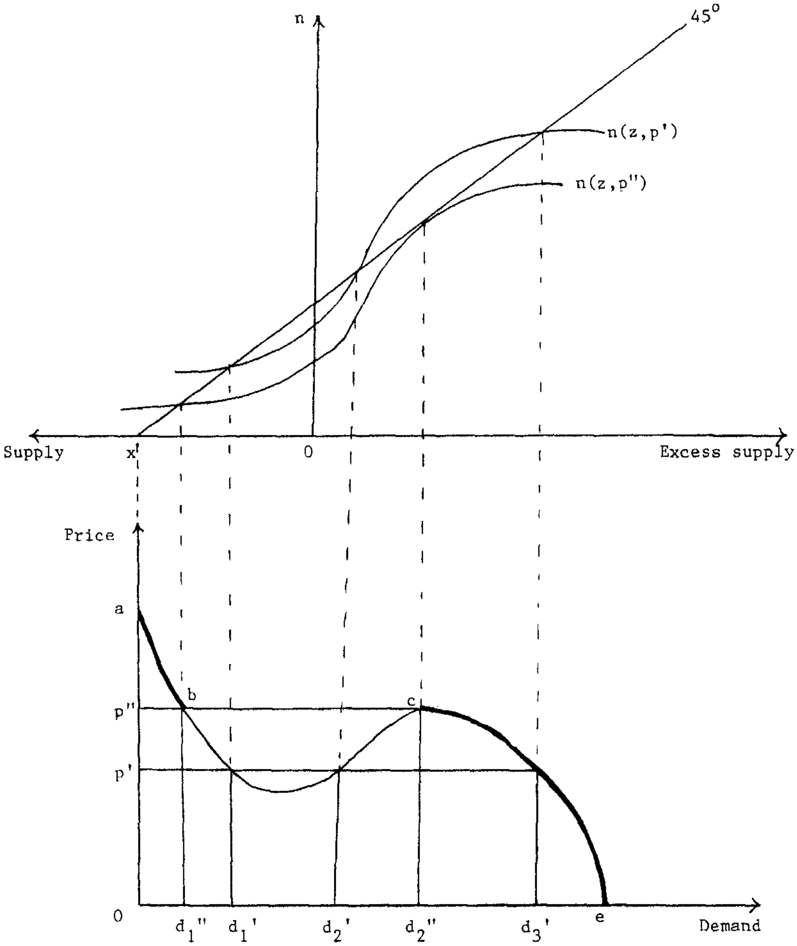

Fig. 2

The discontinuity is now clear. If, starting from a price of p'', the producer increases price a little, the maximum sustainable demand for his product drops from d''_2 to below d''_1. If, holding p'' constant, he raises supply to slightly above $\hat{x}$ (which will shift the 45° line up in fig. 2), a similar sharp deline in demand occurs. It is this kind of discontinuity which can explain why it may be profit-maximising for a producer to maintain an excess demand for his product.

3. Excess-demand equilibria

There are different ways of illustrating excess-demand equilibria. One is to construct an example. I shall, however, first use the more general – and hopefully more instructive – approach of establishing a sufficient condition for equilibria to exhibit excess demand. Later, an example is constructed to illustrate this.

For this we need a definition. Let us use $p(x)$ to denote the price (wherever such a price exists) which ensures $n(0,p(x))=x$. If the n-function has the property that for all $x>0$ there exists $z>0$ such that $n(z,p(x))>x+z$, then we shall say that *individual valuation of the good is strongly responsive to excess demand.*[6] This property essentially requires the n-function (see fig. 1) to rise sufficiently sharply with z.

Theorem. If individual valuation of the good is strongly responsive to excess demand, then all equilibria exhibit excess demand, that is, $d(x^*,p^*)>x^*$.

Proof. Assume individual valuation is strongly responsive to excess demand and $d(x^*,p^*)\leqq x^*$.

Step 1. Let us first suppose $d(x^*,p^*)<x^*$. There are two possibilities: (i) $n(0,p^*)\geqq x^*$, or (ii) $n(0,p^*)<x^*$. If (i), then since n can take a maximum value of $\#H$, and the n-function is continuous, there must exist $\hat{z}\geqq 0$ such that $n(\hat{z},p^*)=\hat{z}+x^*$. Hence, $\hat{z}+x^*\in D(x^*,p^*)$. Since $\hat{z}+x^*>d(x^*,p^*)$, this is a contradiction. Now suppose (ii) is true. Define $\hat{x}=n(0,p^*)$. It follows from definition (3) that $\hat{x}\in D(\hat{x},p^*)$. Hence,

$$d(\hat{x},p^*)\geqq\hat{x}. \tag{6}$$

Since n is non-decreasing in z, we have $n(0,p^*)\geqq n(d(x^*,p^*)-x^*,p^*)$. Hence,

$$\hat{x}\geqq d(x^*,p^*), \tag{7}$$

$$\begin{aligned} Y(\hat{x},p^*) &= p^*\hat{x}-c(\hat{x}) \quad \text{by (6)},\\ &\geqq p^*d(x^*,p^*)-c(\hat{x}) \quad \text{by (7)},\\ &> p^*d(x^*,p^*)-c(x^*) \quad \text{since} \quad \hat{x}<x^*,\\ &= Y(x^*,p^*) \quad \text{since} \quad d(x^*,p^*)<x^*. \end{aligned}$$

[6]The choice of terminology is clear once it is appreciated that this property is equivalent to requiring that for a 'large' number of people v_i rises sharply in response to an increase in excess demand.

This contradicts the fact that (x^*, p^*) maximises profit, and thereby establishes that $d(x^*, p^*)$ cannot be less than x^*.

Step 2. Suppose now that $d(x^*, p^*) = x^*$. Hence, $p(x^*) = p^*$. It follows from the fact that individual valuation of the good is strongly responsive to excess demand, that there exists $z > 0$, say z', such that $n(z', p^*) > x^* + z'$. By the continuity of n and the fact that n goes to zero for a sufficiently large p, it follows that there exists $p' > p^*$ such that

$$n(z', p') = x^* + z'.$$

This implies $x^* + z' \in D(x^*, p')$, by (3). Hence,

$$d(x^*, p') \geqq x^* + z'.$$

Therefore,

$$\begin{aligned} Y(x^*, p') &= p'x^* - c(x^*) \quad \text{since} \quad z' > 0, \\ &> p^*x^* - c(x^*), \\ &= Y(x^*, p^*) \quad \text{since} \quad x^* = d(x^*, p^*). \end{aligned}$$

This contradiction establishes that $d(x^*, p^*)$ is not equal to x^*.

Together with Step 1, this establishes a contradiction of our initial assumption that $d(x^*, p^*) \leqq x^*$. □

The remainder of this section is used to construct a numerical example to illustrate a monopoly equilibrium which exhibits excess demand, that is, the kind of situation described in the above theorem.

Suppose the n-function is given as follows:

$$\begin{aligned} n(z, p) &= 100 - 10p \quad \text{for all} \quad z \geqq 20, \\ &= \max\{0, 60 - 10p + 2z\} \quad \text{for all} \quad z < 20. \end{aligned} \tag{8}$$

Thus if price is 5 and excess demand is 25, the demand for this good will be 50. Assume, further, that this good can be produced at zero cost by the monopolist. That is, $c(x) = 0$, for all x.

It will now be shown that, at equilibrium, the price of the good will be 4, supply 40 and demand 60; thereby implying an excess demand of 20.

Let us first derive the function $d(x, p)$ from eq. (8).

$$\begin{aligned} d(x, p) &= 100 - 10p \quad \text{for all} \quad p \leqq 8 - (x/10), \\ &= 0 \quad \text{for all} \quad p > 8 - (x/10). \end{aligned} \tag{9}$$

Fig. 3 depicts the derivation of (9). Recall that the points of intersection between an $n(z,p)$ curve and the 45° line through x depict equilibrium levels of demand for a given x and p, and the largest of these demands is $d(x,p)$.

Since $p \leqq 8-(x/10)$ implies $\min\{x, d(x,p)\}=x$, and since $c(x)=0$, we get

$$Y(x,p)=0, \quad \text{for all} \quad p>8-(x/10),$$
$$=px \quad \text{for all} \quad p \leqq 8-(x/10).$$

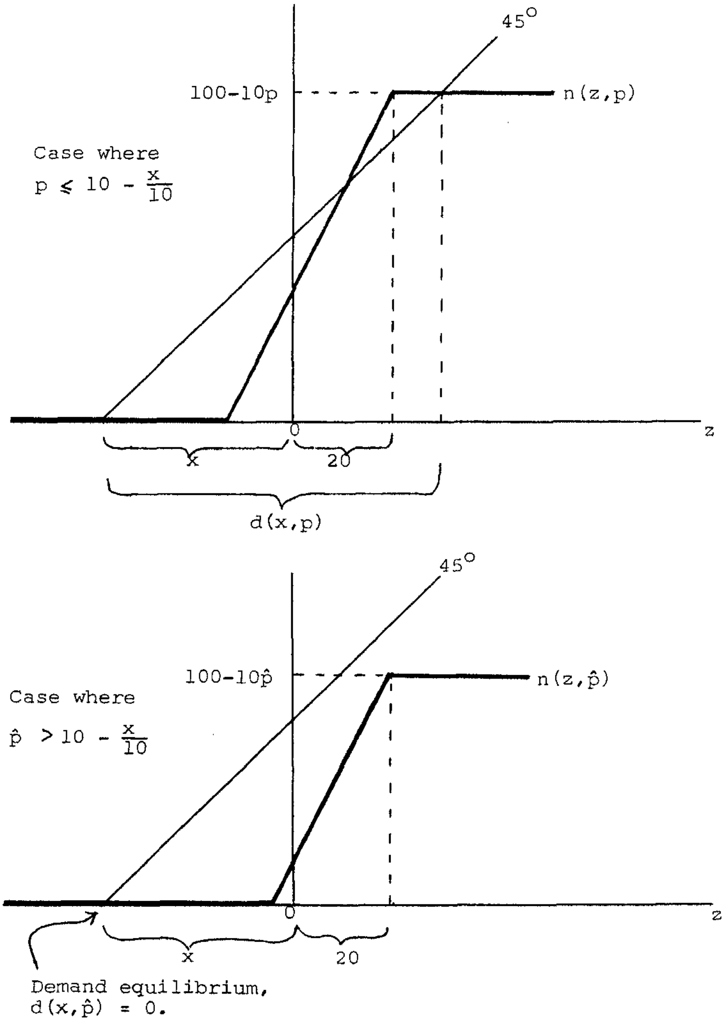

Fig. 3

For each value of x, $Y(x,p)$ is maximised by setting

$$p=8-(x/10). \tag{10}$$

It follows that to maximise $Y(x,p)$, we have to maximise $(8-(x/10))x$. The first-order condition for achieving this is

$$8-(x/5)=0.$$

Hence, $x=40$. By inserting this in (10), we get $p=4$. (9) implies $d(40,4)=60$. This establishes that at equilibrium there is an excess demand of 20. The monopolist cannot respond to this by raising price because this would cause demand to plummet so much that he would be worse off in the end.

4. Extensions

This section discusses some motivational issues and open questions. This paper was concerned with goods for which each consumer's subjective valuation increases as the aggregate excess demand for the good rises. This feature, captured in (1), may be referred to as the *basic equation*. This has been treated as a 'primitive' in the present paper. An interesting direction to pursue is to explain this basic equation from more fundamental assumptions of consumer theory. Without going into a formal analysis of this here, I want to comment on two alternative and reasonable motivations that can be used.

There is a large body of literature which argues that individual welfare depends, among other things, on what an individual consumes *relative to others* [see, e.g., Frank (1985)]. One interpretation of this is that, if a certain product or service is desired by many but not everybody who desires it gets it, then this is an additional reason for individuals to covet it. In other words, human beings seek *exclusiveness*. Leibenstein's (1950) analysis of 'snob' effect was based on a similar argument. This provides an immediate and direct justification for the basic equation. What is interesting but was overlooked by the earlier writers in this field is the fact that the basic equation can explain price rigidities and excess demand equilibria.

The other approach to the basic equation is to claim that human beings are really interested in product quality. But since this cannot be directly observed, they treat excess demand as a *signal* for quality, in the same way that education is a signal for productivity in Spence's (1974) model. This approach, however, leads to an open question which deserves further research: Under what circumstances is there reason for consumers to treat excess demand as an index of quality?

Turning to another problem concerning primitives, note that in this paper we do not begin from utility functions but by directly specifying each person

i's behavioural function (1), which is later aggregated to (2). Can these be derived from standard utility maximisation? The answer to this is yes and this is demonstrated by a utility function which belongs to a class of functions which have been used widely in recent years [see, e.g., Gabszewicz and Thisse (1979), Shaked and Sutton (1983)].

Suppose that a consumer $t \in [0,h]$ has an income of $W(t) = W_1 + W_2 t$, where W_1 and $W_2(\geqq 0)$ are exogenously given parameters. The utility that consumer t gets by consuming a unit of the good, characterised by price p and excess demand z, is $u(z)(W(t)-p)$, where $u(z)$ is a non-decreasing function of z. The utility that he gets if he does not consume the good is given by $W(t)u_0$. It is easy to check that the number of people demanding the good, characterised by (p, z), is[7]

$$n(z,p) = h - u(z)p/(u(z)-u_0) + W_1/W_2.$$

So what was earlier treated as a primitive, to wit, the n-function, is now derived from a standard utility function.[8] Of course, the important question as to how the function $u(z)$ is, in turn, derived is not answered here.

Finally, a comment on labour markets. There has been a large recent literature on 'efficiency' wages [e.g., Shapiro and Stiglitz (1984)], which explains downward wage rigidity in terms of the *employer's* preference for not lowering wages. The model of this paper can be used to construct a new efficiency wage argument.

Consider a person seeking a job and evaluating alternative possibilities. There is a lot of sociological evidence [Jencks, Perman and Rainwater (1985)] that a person will be concerned about, among other things, the social status of a potential job. And one way of judging the status associated with a firm is by the excess supply of employees faced by it. Admittedly this argument would be more applicable to salaried and higher-paid jobs, in contrast to several existing theories which apply mainly to the low-wage sector [e.g., Mirrlees (1975)]. But this complementarity may well be its advantage.

[7]Given that the number of people demanding the good cannot be less than 0 or more than h, a more fastidious specification of the n-function is as follows:

$$n(z,p) = \text{mid}\,\{0, h - u(z)p/(u(z)-u_0) + W_1/W_2, h\},$$

where mid $\{a,b,c\}$ is a number which is second largest (ties being broken arbitrarily) among a, b and c.

[8]Further, $n(z,p)$ is concave in z if $u(\cdot)$ is concave.

References

Basu, K., 1986, The market for land: An analysis of interim transactions, Journal of Development Economics 20.

Berge, C., 1963, Topological spaces (Oliver and Boyd, Edinburgh).
Frank, R.H., 1985, Choosing the right pond: Human behaviour and the quest for status (Oxford University Press, Oxford).
Gabszewicz, J.J. and J.-F. Thisse, 1979, Price competition, quality and income disparities, Journal of Economic Theory 20.
Jencks, C., L. Perman and L. Rainwater, 1985, What is a good job? A new measure of labour market success, American Journal of Sociology, forthcoming.
Leibenstein, H., 1950, Bandwagon, snob and Veblen effects in the theory of consumers' demand, Quarterly Journal of Economics 64.
Mirrlees, J.A., 1975, Pure theory of underdeveloped economies, in: L.G. Reynolds, ed., Agriculture in development theory (Yale University Press, New Haven, CT).
Scitovsky, T., 1944, Some consequences of the habit of judging quality by price, in: T. Scitovsky, Papers on welfare and growth (Stanford University Press, Stanford, CA, 1964).
Shaked, A. and J. Sutton, 1983, Natural oligopolies, Econometrica 51.
Shapiro, C. and J. Stiglitz, 1984, Equilibrium unemployment as a worker discipline device, American Economic Review 74.
Spence, A.M., 1974, Market signalling (Harvard University Press, Cambridge, MA).