An Equilibrium Theory of Rationing

By Richard J. Gilbert and Paul Klemperer

January 1997

Abstract

Setting a price that results in rationing may be optimal for a seller whose customers must make a specific investment to be able to use its product. Rationing results in ex-post inefficiency, but the resulting distribution of ex-post surplus can compensate consumers for their transaction-specific investments at a lower cost to the seller's profits than would market-clearing prices. Similarly, it may be optimal for a purchaser to procure some of its requirements from a high-cost "second source" rather than purchase only from the lowest-cost supplier. (91 words.)

JEL classification: D45, L10, L14

Keywords: rationing, sunk costs

Richard J. Gilbert
Department of Economics
University of California
Berkeley
California 94720
USA

Paul Klemperer
Nuffield College
University of Oxford
Oxford, OX1 3UJ
England

1. Introduction

Economists praise the virtues of price as a mechanism to equate supply and demand, but markets often clear by non-price means. For example, Intel has sometimes rationed supply of its microprocessors rather than raise its price to clear the market (Burke, 1990). There is also evidence of rationing of several metals, electronic parts, metal fasteners, gypsum board, personal computers, semiconductors, compact disks, titanium dioxide, polypropylene, and petrochemicals (Carlton (1991), Ghemawat (1986), Haddock and McChesney (1992), McKinnon and Olewiler (1980), and Slade (1991)). Restaurants and movie theaters usually turn away late arrivals at peak hours rather than invite them to bid for scarce seats. A mild form of rationing that is very commonly observed is that some customers receive a good without delay while others must wait longer for all or part of their orders to be fulfilled without receiving any compensating price discount (Rotemberg and Summers (1990)). Finally, and conversely to suppliers rationing purchasers rather than raising price, purchasers sometimes "second-source" some of their requirements, rather than cutting costs by procuring only from the lowest-cost supplier.

This paper explains why a rational seller may prefer to set prices at which there is excess demand, and ration output rather than set market-clearing prices. The key feature is that consumers must incur sunk costs that are specific to the seller before they can use the seller's product. For example, a computer manufacturer must invest in product design, and, perhaps, software and other complementary products to be able to use a new microprocessor chip. In such a situation prices must be set to compensate consumers for the sunk costs they incur, or consumers will not make the investments necessary to use the seller's product. While it is feasible in these markets to set prices that do not result in rationing, we show that this may not be optimal for the seller. Although rationing is ex-post inefficient, it changes the distribution of surplus among consumers relative to market-clearing prices. As a consequence, rationing in some states of demand may allow higher prices in other states, and thus higher profits.

Section II describes the structure of a simple model and Section III develops the basic argument. Setting prices that avoid rationing requires high prices that clear the market in high-demand states, together with prices that are low enough in low-demand states to give consumers sufficient ex-post surplus to cover their investment costs. But this pattern of prices favors consumers who are more likely to be in the market when demand is low, and such consumers will typically be the inframarginal investors. If the marginal investor (i.e. potential consumer) whom the seller wishes to induce to invest is relatively unlikely to wish to consume when demand is low, it is costly to the seller to attract this potential investor by offering a low price in the state of low demand. Instead, charging lower than market-clearing prices in high-demand states may attract this potential consumer at a lower cost to the seller, even though the ensuing rationing is ex-post inefficient.

We show that rationing is most likely to be profitable when the marginal consumer gains more from lower prices in high-demand than in low-demand states (relative to other consumers), when consumers' valuations are not too dissimilar (so rationing is not too ex-post inefficient), and when consumers' sunk costs are neither insignificant nor so large that it is not worth attracting more consumers to invest than the firm has capacity. Our model also makes two assumptions that are important for rationing to arise. First, capacity must not be too large relative to potential demand. This is most likely in new, growing, or booming markets, and is consistent with, for example, the fact that microprocessors are very commonly rationed when a new product is first introduced, and less commonly rationed thereafter. Second, the seller must be able to pre-commit to a price (or to a demand-dependent price schedule), perhaps by developing a reputation to charge a "fair" price, before consumers invest.[1]

Another way of viewing our result is in the context of *auction theory with endogenous entry*: when buyers have costs of entering an auction (i.e. sunk investment costs) the seller may wish to precommit to running an inefficient auction (i.e. rationing) in order to

---

1 Modelling how the seller makes commitments is beyond the scope of our paper. However discussions with industry executives suggest that Intel, for example, believes that its reputation for charging "fair" prices is very important to it and also provides an effective commitment (supported by implicit threats by customers to punish) even though the market for computer microprocessors is not an especially stable market. In the example we develop in the next paragraph (second-sourcing), the government makes explicit promises. Our discussion is also consistent with evidence of stable prices in, eg, Carlton (1986). Note also that absent our commitment assumption, the need for consumers to invest can prevent any sales in equilibrium. If, after investments are sunk, the firm can set a profit-maximizing price that gives the marginal consumer no surplus, that consumer will do better not to invest, so there will be no marginal consumer who invests, and therefore no investment at all. Diamond (1971) emphasized the importance of commitment when consumers have transaction-specific costs.

encourage the entry of buyers who are likely to have lower values and who it would otherwise be more costly to induce to enter.[2,3]

Reversing the roles of consumers and the seller in our model shows that rationing can emerge in a procurement market where potential suppliers must incur sunk costs. This corresponds to a "second-sourcing" contract, in which the buyer commits to purchase at least a fraction of its needs from an (ex-post) inferior source of supply in order to provide that supplier with incentives to participate in the market. For example, the U.S. government has for defense procurement (e.g., missiles) on occasion asked two suppliers to invest in developing prototypes and manufacturing technology and to make bids, with the promise that even the high bidder will receive a fraction of the order (e.g., 30%) if the high bid is within "the competitive range".[4]

More broadly our model provides a rationale for offering "second prizes". While the second prize may provide little incentive (or a disincentive) for investment by a strong competitor, offering the second prize may be a very effective way to persuade a weaker competitor (who thinks he has relatively little chance of actually winning first prize) to enter the

---

2 Thus our assumption that the distributions from which buyers' values are drawn are not ex-ante symmetric, distinguishes our model from most earlier work on auctions with entry costs e.g. McAfee and McMillan (1987), Harstad (1990), Englebrecht-Wiggans (1987, 1993), and Levin and Smith (1994). Also by contrast with this literature, our primary focus is on comparing simple market institutions rather than on developing optimal auctions. However section V shows that a form of rationing is the optimal solution to a general mechanism design problem. King, Welling and McAfee (1992) is in a similar spirit to our work (although their buyers are ex-ante symmetric) arguing that if buyers must make unobservable investments prior to bidding, then the seller may prefer a first-price auction to a second-price auction even where this is not socially efficient.

3 In our main model, rationing does not affect the number of customers who enter (i.e. invest) but induces these customers to enter at a lower cost to the seller. We also analyze an extension in which rationing results in additional customers entering (investing).

4 "The competitive range" does not seem well-defined. Industry executives suggest that a 10% higher per-unit price would be within the competitive range, while a higher bid might lead to a re-negotiation process to lower the bid. A too-high bid (e.g. after failure to achieve a low-cost technology) would receive no order at all. (See Pyatt, 1989.)

Consistent with our model that emphasizes an asymmetry between the stronger and weaker (lower-and higher-cost) supplier, Burnett and Kovacic (1989) suggest that guaranteeing a minimum share of production (subject to a "reasonable" bid) is particularly important where the Department of Defense wishes to induce a firm to bid against an established producer. In addition, "DOD ... may need to offer minimum production guarantees even when the entire production run will be dual-sourced .... Without the guarantees, suppliers ..... may decline to invest their own funds in the development and prototyping efforts that would precede the award of production contracts".

game.[5]

Procurement practices that are similar to those of the U.S. Department of Defense are also observed in Japan's telecommunication industry.[6] Another example of a customer rationing suppliers is that in the period when General Foods was the dominant buyer of coffee, it largely engineered the 1962 International Coffee Agreement that precommitted it to pricing policies that involved rationing suppliers. There are both substantial sunk costs and uncertainties in coffee production (see McLaren (1995)).

Similarly, our model offers a possible explanation for efficiency wages in monopsonistic labor markets when workers have to invest in employer-specific human capital.

Section IV begins by noting that because rationing is ex-post inefficient, buyers' surpluses in the rationed state can be raised if resale is permitted. The seller can then extract this extra surplus and increase his own profits by raising the general level of prices. Thus rationing with resale permitted is, in principle, even more attractive than rationing and not allowing resale. A problem is that permitting resale may encourage third parties to enter the market purely for purposes of arbitrage, and this may explain why the resale of rationed goods in "gray markets" is often discouraged. (For example, restaurants may not permit the sale of places in their queues because to do so would invite non-serious customers to queue purely to obtain a seat that can be resold.) However, we show how practices such as airlines' buying back of tickets when they are overbooked can be interpreted as mechanisms that implement rationing-plus-resale while making it hard for third parties to enter.

Our analysis in Sections III and IV shows that committing to a single fixed price

---

5 In the context of our basic model of one supplier with two customers, a "first prize" is a low price to a customer when only his investment is successful (the low demand state) while a "second prize" is a low price, but a rationed quantity, to a customer when both customers' investments are successful (the high demand state).

Anton and Yao (1989, Section 5) also suggest that "split-award auctions" (auctions with multiple winners) can provide desirable incentives for investment, but they assume the purchaser always minimizes its total procurement costs, given suppliers' bids (and this also minimizes total production costs) so they do not explain trade with an ex-post inferior supplier or customer.

On the standard theory of prize-based incentive schemes, see, e.g. Nalebuff and Stiglitz (1983).

6 "Competition between the suppliers .... to NTT .... is not of the 'winner takes all' variety. Rather, it involves *controlled* competition in so far as, contingent on reasonable performance as judged and monitored by NTT, each supplier can expect to receive a sizeable share of NTT's order .... NTT rewards co-operative behaviour and good performance, and punishes poor performance, by somewhat increasing or decreasing a supplier's share of its orders." (Fransman 1995 pp. 22-3, author's italics.)

independent of demand, and rationing when necessary, can dominate the best (typically demand-dependent) market-clearing prices. In Section V we analyze our model as a more general mechanism-design problem: we allow the firm in each demand state to force consumers to choose from a menu of contracts, each contract specifying a price that the consumer pays and a quantity of output that the consumer receives. Since our focus is on the effects of sunk costs when consumers cannot be contracted with ex ante, we assume consumers select contracts only after demands are reported and so insist that the firm respect consumers' ex-post individual rationality and ex-post incentive compatibility. We show that for a wide range of parameters the solution to this more general problem is equivalent to simply rationing in the high demand state, as in Sections III and IV, but generalized so that the price at which the good is rationed in the high demand state no longer equals the price in the low demand state. Although the simple fixed-price rationing described in Sections III and IV is less profitable than if the seller could compensate consumers directly for their investment costs, the more general rationing scheme of Section V is as profitable as if the seller could pay all consumers' sunk costs directly. (Of course, direct compensation may be hard because of the difficulties of verifying either whether investment has taken place or whether it has been successful.)

Our explanation of rationing differs from the existing literature. One argument is that rationing can be an optimal policy if low prices improve the "quality" of a firm's demand. Thus, Stiglitz and Weiss (1981) maintain that a lender may hold borrowing costs at a level that generates excess demand in order to attract more desirable borrowers. A similar argument is Basu's (1987) and Becker's (1991) suggestion that a seller may choose a price that generates excess demand if consumers' valuation of the seller's product increases with the nominal (unrationed) demand (e.g., a restauranteur might not raise prices to balance supply and demand if customers prefer restaurants with queues).[7] In Allen and Faulhaber's (1991), Slade's (1991) and Haddock and McChesney's (1992) models of rationing, future demand depends on the current price, so a seller may choose a non market-clearing price to increase his subsequent demand. Kenney and Klein (1983) contend that deBeers rations consumer choice of individual

[7] See also Karni and Levin (1994).

diamonds in favor of bundled sales as a means to economize on information costs, and so increase aggregate consumers' willingness-to-pay.[8]

Carlton (1991) shows that rationing can be efficient if a seller has perfect information about the relative demands of his buyers.[9] However, in Carlton's paper, rationing does not improve upon market-clearing prices unless it has lower administrative costs. In our model, rationing emerges as an alternative for the seller that strictly dominates market-clearing prices when there are no administrative costs involved in using either mechanism.

The most common argument in favor of rationing is that it is "fair". Many people feel that all loyal customers deserve a chance of purchasing a scarce commodity. The concept of fairness has received little respect in most economic models.[10] However, our paper suggests that an appeal to fairness as a justification for rationing can be understood as the need to guarantee positive surplus ex-post to all customers whom the seller wishes to induce to make seller-specific investments ex-ante. The seller rations inefficiently (acts fairly ex-post), in order to send the right signals for investment ex-ante.

II. The Model

A monopolist has one unit of a good for sale, and no costs. Each of two risk-neutral potential customers $j = H,L$ independently chooses whether or not to make a sunk investment, s, that succeeds with probability r_j. Customer j has valuation v_j for the seller's product if his investment succeeds and zero otherwise. The investment has no value in alternative transactions. Conditional on both consumers choosing to invest, we let q_{LH} be the probability

[8] Rationing can also arise in models in which a monopolist's marginal revenue is not downward sloping, see for example Bulow and Roberts (1989). Bohm et al. (1983) show how rationing can arise in multi-good contexts in general equilibrium. DeGraba (1995) argues that rationing can induce consumers to purchase before they learn the good's value. See also DeGraba and Mohammed (1996). Denicolo and Garella (1996) obtain rationing in a durable goods monopoly model (a la Coase (1972), Bulow (1982)). Png (1991) discusses rationing in a setting which assumes the firm must set price before demand is known.

[9] Carlton also models fixed set-up costs, but for the seller rather than the buyers, and the set-up costs are unimportant for his rationing result provided the seller has a capacity constraint.

[10] Exceptions include Weitzman (1977) and Kahneman et al. (1986), but these papers do not provide an economic explanation for a preference for fairness.

that both investments are successful, and q_j be the probability that only j is successful, j = H,L. For simplicity we assume consumers' investment successes are uncorrelated, so $q_j = r_j(1 - r_k)$, $j \neq k$ and $q_{LH} = r_L r_H$, but none of our propositions require this assumption.

All agents know the values of the parameters s, v_H, v_L, r_H, r_L, q_{LH}, q_H, q_L, but the seller cannot identify which consumer is which.

The order of events is: (1) the seller commits to, and makes public, his pricing policy; (2) each consumer independently chooses whether or not to invest; (3) each consumer finds out whether or not his investment is successful, and any consumer who wishes to do so publicly announces that his investment is successful; (4) the seller announces the price, and consumers who reported a successful investment choose whether or not to purchase.[11]

Note that a consumer's stage-3 announcement does not commit him to purchase. This assumption is consistent with a market environment in which consumers cannot easily be individually contracted with. Because consumers do not make an ex-ante commitment to purchase, this forces the firm to respect consumers' ex-post individual rationality in each demand state and not merely guarantee non-negative surplus averaged across all demand states.[12]

We assume the seller must stick to his stage-1 pricing commitment in stage 4, and set his price on the assumption that consumers report honestly in stage 3.[13] We will distinguish cases according to whether the seller can or cannot prevent purchases by unsuccessful investors (or non-investors) who might hope to make a profit by resale. (Whether or not such "arbitrage" can be prevented will depend on what the seller and consumers can observe about other

---

11 Our model is equivalent to one in which each of two potential consumers j = H, L has demand v_j with probability r_j and with probability $(1-r_j)$ has no demand, and after learning his own demand and the seller's pricing policy, the consumer can choose to incur a sunk investment, $s_j = (s/r_j)$, which always succeeds and allows the option of making a purchase. This models, for example, a retail market in which consumers' sunk costs are their seller-specific costs of visiting the store. It is not hard to extend our model to allow $s_H \neq s_L$.

12 The results would be unaffected if consumers were required to make binding commitments to purchase at stage 3, if we also assume that successful consumers can observe the number of other successful consumers prior to making their commitments.

13 We assume (as is conventional) that indifferent consumers report honestly. (As the model is stated, unsuccessful consumers have no reason not to always report success, but adding a small cost of reporting success would remove their indifference in all our equilibria.)

consumers' investment successes.[14)15]

We assume that $v_H > v_L$, and $r_H > r_L$. Since $r_j = q_j + q_{LH}$, it follows that $q_H > q_L$. This assumption that r_j is positively correlated with v_j is not required for our model to yield rationing,[16] but it simplifies the analysis by assuring that the expected surplus of the high-value consumer is at least as large as the expected surplus of the low-value consumer. We also assume that the seller's maximum profit at prices that clear the market, but induce both consumers to invest, is higher than the seller's profit at prices for which only the high-value consumer would invest.[17] It follows that the seller always will induce both consumers to invest and also, therefore, that the low-value consumer is always the marginal consumer who must be induced to invest. This assumption, also, is not necessary to our results,[18] but it too reduces the number of different cases to consider and so simplifies the analysis.

In the following sections we contrast different pricing and rationing policies that the seller might commit to in stage 1.

III. "Sticky prices"-plus-rationing dominates market-clearing.

This section shows that committing to a single price that applies to all demand states, and rationing when there is excess demand, can be more profitable than the best market-clearing price schedule. This is true even though rationing is inefficient ex-post (and consumers cannot resell between themselves) and there are no administrative costs to charging different prices in different states.

---

14 Section III forbids all resale so it does not matter what agents can directly observe about others' successes. Sections IV and V distinguish different cases.

15 We assume that the seller cannot directly compensate consumers for their investment costs, or directly pay consumers whose investments are successful, but we note that our most general rationing scheme will yield seller profits that are as high as if he could do either or both of these things. See Section V.

16 It is required for rationing that the customer j with the lower r_j is the "marginal consumer" who expects the lower surplus from investing at market clearing prices, either because he has the lower v_j or because his r_j is sufficiently much lower to outweigh a higher v_j.

17 We show in Section III, equation (A), that this assumption corresponds to $s \leq q_L\left[q_L v_L/q_H - \left[v_H - v_L\right]\right]$, so we are assuming v_L and q_L not too small and s not too large.

18 We discuss the converse assumption at the end of Section III.

With market-clearing prices, the firm sets a price P_1 at which the unit will be sold if one customer demands the good, and a price P_2 at which the unit will be sold if both customers demand the good. Recall that the firm cannot distinguish the customers, and so cannot make P_1 contingent on which customer demands the good. Consequently, to guarantee market-clearing, the prices must satisfy $P_1 \leq v_L$ and $v_L \leq P_2 \leq v_H$.

The alternative we consider is to set a single price P^r for both states of demand and ration if there is excess demand. The two alternatives are set out below.

	Low Demand	High Demand
Market-clearing Prices	P_1	P_2
Fixed Price and Rationing	P^r	P^r and ration

Table 1. Prices in low and high demand states

Because there is no resale, no unsuccessful investor can gain by misreporting in order to acquire the good and profit from resale. Consequently, it also does not matter whether or not any agent can or cannot directly observe any other agent's success.

Market-clearing prices

With market-clearing prices, the low-value consumer, L, is able to buy only in the low-demand state, when his investment, and not H's investment, is successful. The probability of this is q_L, so he obtains expected ex-ante surplus $q_L(v_L - P_1) - s$. Setting this equal to zero yields

$$P_1 = v_L - \frac{s}{q_L} .$$

(Our assumption that the seller's profit using market-clearing prices with both consumers investing exceeds the profit with only consumer H investing guarantees $P_1 > 0$. See equation (A) below.)

Clearly the seller maximizes profits by setting the high-demand price equal to the highest price the high-value consumer will pay, that is,

$$P_2 = v_H \quad .$$

The seller's maximized profit from market-clearing prices is therefore

$$\Pi^{mc} = (q_L + q_H)(v_L - \frac{s}{q_L}) + q_{LH}v_H \, . \tag{1}$$

Setting market-clearing prices that induce both consumers to invest, rather than only consumer H, is preferred by the seller as assumed if

$$\Pi^{mc} \geq r_H v_H - s = (q_H + q_{LH})v_H - s \, ,$$

that is, if

$$s \leq q_L \left[\frac{q_L}{q_H} v_L - (v_H - v_L) \right] . \tag{A}$$

Fixed Price with Rationing

Suppose the seller sets a fixed price P^r and rations when there is excess demand. With rationing, the low-value consumer receives half a unit on average in the high-demand state, and so obtains total expected ex-ante surplus

$$q_L(v_L - P^r) + q_{LH}(\tfrac{1}{2}\left[v_L - P^r\right]) - s \, .$$

Setting this equal to zero yields

$$P^r = v_L - \frac{s}{q_L + \frac{1}{2}q_{LH}} \, .$$

(Assumption (A) ensures $P^r > 0$.)

So the seller's maximized profit from setting a fixed price and rationing when there is excess demand is

$$\Pi^r = (q_L + q_H + q_{LH})(v_L - \frac{s}{q_L + \frac{1}{2}q_{LH}}) \, . \tag{2}$$

Proposition 1: *The seller prefers setting a fixed price and rationing (with resale between consumers prohibited) to market-clearing prices, if and only if*

$$q_{LH}(v_H - v_L) \leq \left(\frac{q_H}{q_L} - \frac{q_H + \frac{1}{2}q_{LH}}{q_L + \frac{1}{2}q_{LH}}\right)s\,. \tag{3}$$

Proof: follows directly from comparison of (1) and (2). ■

Given $r_H > r_L > 0$ and $s > 0$, the right-hand-side of (3) is strictly positive so rationing is always preferred for sufficiently small $(v_H - v_L)$.

Figure 1 illustrates the costs and benefits to the seller of the rationing strategy. The diagonal lines labelled π_1, π_2, π_3 are iso-profit lines, assuming that prices are such that both consumers participate in the market and at least one consumer is willing to buy in each state. The thick line labelled "L" is the locus of state-contingent prices for which the ex-ante expected surplus of the low-value (marginal) consumer is zero. In any optimal strategy, the seller chooses a price pair on this locus. The optimal market-clearing strategy is the price pair labelled "MC" in Figure 1. (Only the vertical segment of "L" is market-clearing, and for $P_2 > v_H$ the seller would make no sales in state 2.) A rationing strategy corresponds to $P_2 \leq v_L$, and a single-price rationing strategy is a price on the diagonal OA. The optimal single-price rationing strategy is the price pair labelled "R".

The cost to the seller of choosing a rationing strategy can be measured by the movement from "MC" to "B", where the marginal consumer is just willing to buy in state 2 so rationing occurs. This movement reduces the seller's profit by $q_{LH}(v_H - v_L)$, the left-hand side of equation (3), with no effect on the marginal consumer. The benefit to the seller of rationing is measured by the movement from point "B" to "R", which increases the seller's profit by the right-hand-side of equation (3) while leaving the marginal consumer no worse off. The reason is that the seller's indifference curves have a slope $(q_L + q_H)/q_{LH}$ (the relative likelihood of the two demand states), which is steeper than the slope of the marginal consumer's indifference curve, $q_L/(\frac{1}{2}q_{LH})$, when P_1, $P_2 < v_L$ and the seller rations in the high-demand state.

The relative slopes of the indifference curves reflect the lower cost to the seller of subsidizing L in the high-demand than in the low-demand state. A \$1 reduction in the low-demand state price provides an expected benefit of $\$q_H$ to H and only $\$q_L$ to L, while a \$1 reduction in the price in both the (rationed) high-demand state and the low-demand state

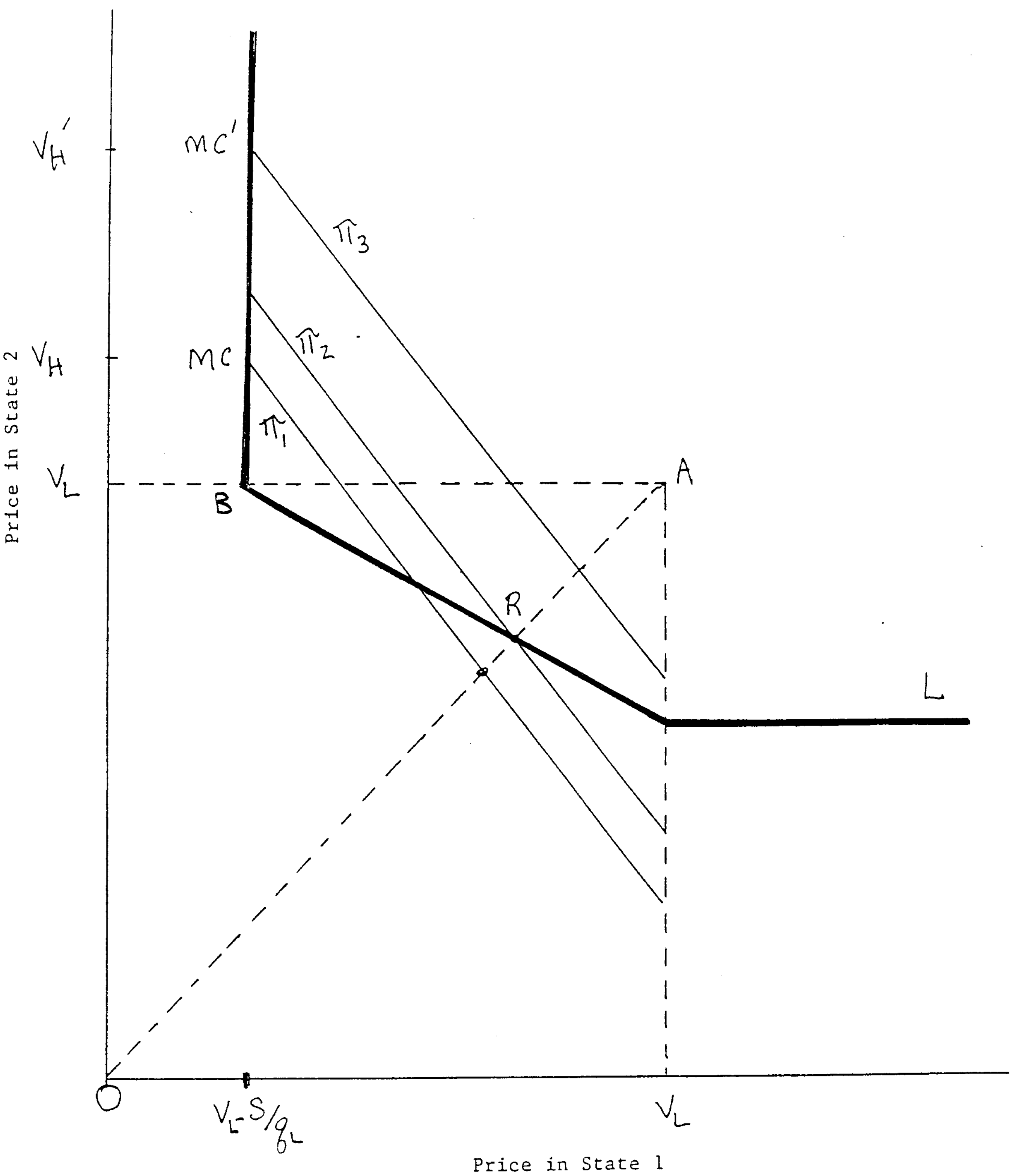

Figure 1. Level curves for seller's profit and marginal consumer's surplus (without resale)

yields expected benefits of $\$(q_H + \frac{1}{2}q_{LH})$ to H and $\$(q_L + \frac{1}{2}q_{LH})$ to L, and so gives out less surplus to H relative to the surplus given to L. The difference in the amounts of surplus that are given to H while giving s to L is the right-hand-side of (3).

In summary, the intuition is that to get the marginal consumer L to participate, the seller must provide L an expected ex-post surplus, s. Handing out this surplus in the low-demand state is a relatively ineffective way to reward this consumer, who is relatively less likely than H to be present in this state. Handing out a subsidy in the high demand state means that half of it reaches L (each consumer is rationed with probability one-half) and so is preferred by the seller if the resulting inefficiency due to rationing is not too great.

Figure 1 shows that rationing may be desirable for the seller in some situations, but not in others. As drawn, rationing with a single price at "R" is better than using market-clearing prices at "MC". However, if the high-value consumer had a higher reservation price -- say v_H' in Figure 1 -- the cost of rationing would exceed its benefits, and the seller would be better off with market-clearing prices (at "MC′"). Figure 1 also shows that rationing at a single price is not the seller's optimal rationing strategy; we discuss more general rationing strategies in Section V.

When does rationing arise?

Examining (3) indicates we should anticipate rationing under the following conditions:

First, the inefficiency caused by rationing should not be too great. (The welfare cost of rationing relative to market clearing prices is $\frac{1}{2}q_{LH}(v_H - v_L)$.[19])

Second, the marginal consumer's sunk cost should be large, so that the seller's prices must provide significant ex-post compensation to this consumer. This means that it is worth the seller's while to accept the inefficient rent-extraction in the high-demand state that rationing

[19] Under our assumption (A), both customers always invest so that the only welfare effect of rationing is that L rather than H receives the unit (which has social cost v_H-v_L) half the time when both customers' investments are successful (probability q_{LH}). However if we drop assumption (A), rationing can bring about investment that would not otherwise occur (see below), and in this case rationing always improves welfare in our model.

causes, in order to be able to pay more compensation in the state in which the marginal consumer is relatively more likely to be present. (The inefficient rent-extraction is a fixed cost, $q_{LH}(v_H - v_L)$, independent of the rationing price and hence independent of the marginal consumer's sunk cost, s.) [20]

Third, it should be significantly more desirable to pay this compensation to the marginal consumer in the high-demand state. That is, the marginal consumer should gain more, relative to other consumers, from a lower price in the high demand state than from a lower price in the low demand state. In our model this implies q_H should be large relative to q_L. More generally, any other factor that means that the marginal consumer accounts for a higher fraction of total demand in states when total demand is high (or supply is low), than in states when demand is low (or supply is high) will suffice to favor rationing when demand is high (or supply is low). (See the discussion of general supply and demands in our working paper, Gilbert and Klemperer (1995).)

Other important conditions for rationing are implicit in the assumptions of the model. Capacity must be constrained -- if the seller had two units of capacity in our model, all consumers who successfully invested would always be served. (Thus rationing is most likely in new, growing, or booming markets.) The seller must not be able to discriminate directly between different types of consumers (perhaps because he cannot distinguish between the different types, or perhaps for legal reasons), and the seller must not be able to pay consumers' sunk costs directly (perhaps because of difficulty in verifying the expenditures). Finally, as we

---

20 We also need s small enough and $\frac{q_H}{q_L}$ not too large, so that it is worth inducing the marginal consumer to invest. See assumption (A). Note that for any $s < (q_L^2 v_L / q_H)$ both (A) and (3) are satisfied for all sufficiently small $(v_H - v_L)$.

emphasized in the Introduction, the seller must be able to commit to future prices.[21]

Second-sourcing

By reversing the roles of the seller and buyers, our model can be applied directly to a procurement setting: Assume each of two suppliers, $j = H, L$, can make a sunk investment s which, with probability r_j, would allow the supplier to deliver up to one unit of a good at a marginal cost c_j, with $c_H > c_L$. If only one supplier is successful, the buyer (who has demand for up to one unit) would purchase one unit from that firm at a unit price at least as large as the seller's marginal cost. If both sellers are successful, the buyer can choose between a market-clearing price P with $c_L \leq P \leq c_H$, or a price above c_H and rationing. Exactly as before, rationing may be optimal; rationing allows the buyer to compensate the marginal seller's sunk cost, so permitting a lower price in the state where only one seller is successful. Thus the buyer rations his lowest cost supplier, and purchases some of his requirements from a second source. For example, the U.S. Department of Defense has on occasion promised to use multiple suppliers even when it would be cost minimizing to use a single supplier, in order to encourage suppliers to invest in improved technology (Burnett and Kovacic (1989), Pyatt (1989); see Section I); second-sourcing is also common in a number of other industries e.g., the automotive, computing, and telecommunication industries.

Another possible application is to a monopsonist employer which wants to provide incentives for workers to make investments that are specific to this labor market: our analysis

---

21 As noted in the Introduction (note 1) price commitments by the seller may be <u>required</u> if consumers are to make any specific investments. The need to commit may be a further reason for rationing at a fixed price, because it is then easier for consumers to monitor the seller's behaviour. (Market-clearing prices that compensate consumers for their sunk costs would have to be higher in high-demand states, so the seller may be tempted to misrepresent demand.) An argument with this flavor can be found in Okun (1982), p.154. "The antagonism toward price increases based on rising demand as "gouging" or "exploitation" is understandable ... Once a firm draws a clientele with attractive implicit contracts, any deviation unfavorable to customers is seen as a violation of these contracts ... If the firm raises its prices... it is bound to create suspicions that it is starting to depreciate its investment (or "goodwill") in customer loyalty".

implies the monopsonist may choose to allocate available jobs randomly in states with excess labor supply, or reduce all workers' hours below their preferred levels, rather than clear the market by cutting wages. This provides an interpretation of "efficiency wages" as wages that are purposely held above market-clearing levels in order to compensate workers for investment in human capital.

Rationing to Increase the Size of the Customer Base

Our assumption (A) implies that both consumers invest under either rationing or market clearing. If, instead, (A) were not satisfied, the optimal market clearing prices would induce just one consumer to invest; if also Π^r (as defined in (2)) exceeds $r_H v_H - s$, then setting a fixed price, P^r, and rationing would induce both consumers to invest and also increase the seller's profits.[22] In this case rationing increases the seller's profit by increasing the low-value consumer's surplus from investing, hence stimulating additional investment and so increasing the size of the customer base.

IV. Rationing with resale, and "buy-backs"

Since rationing results in ex-post inefficiencies, it is natural to ask how permitting resale affects the results. We show in this section that resale can make rationing more attractive to the seller. As in Section III, the seller commits to a fixed price and rations if there is excess demand.

Of course, allowing resale is attractive only if the seller can prevent arbitrageurs, including firms whose investments were unsuccessful, from misreporting their demands and purchasing the good purely in order to profit from resale. We therefore assume in this Section that arbitrage can be prevented.

---

22 When the best market-clearing prices induce just one consumer to invest, the conditions for rationing and inducing both consumers to invest to be optimal can be written $\left[\frac{q_H}{q_L} - \frac{q_H + \frac{1}{2} q_{LH}}{q_L + \frac{1}{2} q_{LH}}\right] s - q_{LH}(v_H - v_L) \geq \left[\frac{q_H}{q_L}\right] s + q_H(v_H - v_L) - q_L v_L \geq 0$. So when (A) does not hold, (3) is necessary but not sufficient for rationing to be optimal.

The simplest assumptions that rule out arbitrage are that the seller can directly observe whether or not a consumer who reports success has indeed been successful,[23] and that the seller can exclude non-genuine customers.

An alternative set of assumptions that rules out arbitrage in this Section are that consumers who have not invested successfully cannot observe who else has announced a successful investment (or who has actually invested successfully) before deciding (in stage 4) whether or not to purchase (note that with a fixed price and rationing a buyer cannot infer demand from the price) and that "third-parties" (those who are neither of the two potential customers in the model) can be excluded from the market. The reason is that unsuccessful investors will then not take the risk of buying and being unable to resell.[24] (With these assumptions, it does not matter what the seller can observe, or what successful investors can observe.)

<u>Proposition 2:</u> *Assume rationing is followed by resale between consumers at price* v_H, *and there is no arbitrage by unsuccessful investors or third-parties. The seller prefers a fixed price and rationing to setting market-clearing prices if and only if* $q_{LH}(v_H - v_L) \leq (\frac{q_H}{q_L} + 1)s$.

<u>Proof:</u> Under market-clearing prices, no resale takes place so the optimal prices are unchanged and profits, Π^{mc}, are given by (1) as before. Let rationing now be at price $P^{rr} \leq v_L$. The low-value consumer will resell the supply he receives in the high-demand state, so obtains total expected ex-ante surplus $q_L(v_L - P^{rr}) + q_{LH}(\frac{1}{2}\left[v_H - P^{rr}\right]) - s$. Maximizing the seller's profit

---

23 The simple fixed-price rationing-with-resale and "buyback" schemes, considered in this Section could be improved upon if the seller could make payments to consumers that are contingent on his direct observations of investment successes. However the more general rationing scheme we will consider in Section V could not be improved upon in this way.

24 Consider L when he has either not invested or unsuccessfully invested (and assume he must offer to buy a whole unit or nothing). Assume he offers to buy without knowing about H's success. His payoff is the same as the ex-post payoff of a successfully-invested L when H has demand one (with probability r_H) and is v_L lower when H has demand zero (with probability $(1 - r_H)$ -- since the good is now unresaleable). However the ex-post payoff of a successfully-invested L is (s/r_L), since the ex-ante payoff of an investing L is zero in an optimal scheme. So the payoff of the non-investing L equals $(s/r_L) - (1-r_H)v_L = (s-q_Lv_L)/r_L < 0$ by (A). It is easy to check that H, also, would not find arbitrage profitable, since a non-investing H would have a lower chance of being able to resell than a non-investing L, and would have to resell at a lower price.

from rationing, $\Pi^{rr} = (q_L + q_H + q_{LH})P^{rr}$, subject to L's surplus being non-negative and $P^{rr} \leq v_L$ (so that L buys in the low-demand state) yields $P^{rr} = \min\left[v_L, v_L-(s-\frac{1}{2}q_{LH}(v_H-v_L))/(q_L+\frac{1}{2}q_{LH})\right]$. By comparison with (1), $\Pi^{rr} > \Pi^{mc}$ if s is such that $P^{rr} < v_L$. When s is smaller so $P^{rr} = v_L$, $\Pi^{rr} \geq \Pi^{mc}$, if and only if the equation in the statement of the Proposition holds. ■

Resale removes the ex-post inefficiency of rationing and at any given prices results in the low-value consumer gaining additional surplus in the high-demand state. If we had not assumed that the seller is constrained to set a rationing price no higher than v_L, then rationing with resale at v_H would <u>always</u> dominate market-clearing prices. The constraint that $P^{rr} \leq v_L$ is implied by our assumption (see Section II) that consumers can observe how many other investors announce success (if $P^{rr} > v_L$, then L would not buy if he was the only successful investor). However, if a successful L could not observe before buying whether or not he was the only successful investor, the seller might be able to set a rationing price $P^{rr} > v_L$ such that L's losses when he is the only successful investor are compensated by his gains from resale when H also succeeds. (With rationing at a fixed price, L cannot infer the state from the price.) In this case, rationing with resale unambiguously dominates market clearing.[25]

In Section V we will explore other, more general rationing-with-resale strategies which always dominate market-clearing without violating consumer L's ex-post rationality (that is, without charging L more than v_L per unit in any demand state).

<u>Gray markets and buy-backs</u>

In fact firms often object to a "gray market"[26] in which customers resell their products. One reason is that unsuccessful investors or third-party interlopers might arbitrage the

---

25 This can easily be checked by solving for the seller's profit maximizing P^{rr} without the constraint that $P^{rr} \leq v_L$.

26 A gray market is one in which trades that are not authorized by the manufacturer are made between retailers or between a wholesaler and distributors or retailers (see e.g. Inman (1993)).

market and so lower the expected surplus earned by the genuine customers that the firm has targeted for subsidy.[27] Furthermore, even absent arbitrageurs, the profitability of rationing-plus-resale depends on the terms of trade in the resale market and on its efficiency; relative to resale at v_H, a lower resale price, or any inefficiency such as trading costs, reduces consumer L's ex-post surplus from any given rationing price and so reduces the seller's profit. Therefore, as an alternative to allowing uncontrolled resale, the firm might prefer itself to conduct a resale market. One way for the firm to operate such a market is to sell more units than it has available and then buy back the difference. Airlines, for example, do just this. They sell more tickets than seats in some states of demand, but then repurchase the excess when their flights are overbooked. Of course a buy-back scheme, like an uncontrolled resale market, is only effective if the seller can prevent arbitrage. (Airlines can discourage arbitrage by limiting the buy-back offer to passengers who have taken a seat on the plane, so an arbitrageur risks being forced to take a flight.)

Proposition 3: The seller prefers setting a fixed sale price together with a (higher) price at which it will buy back any demand in excess of one unit, to setting market-clearing prices if and only if $q_{LH}(v_H - v_L) \leq (\frac{q_H}{q_L})s$, *assuming there is no arbitrage by unsuccessful investors or third-parties.*

Proof: Let P^b denote the fixed sale price and B the buy-back price. The seller's profit under the buy-back scheme is $\Pi^b = (q_L+q_H)P^b + q_{LH}(2P^b-B)$. Noting that at the optimum the low-value consumer's ex-ante surplus, $q_L(v_L-P^b) + q_{LH}(B-P^b)$ - s, is zero, we obtain $P^b = v_L - (s-q_{LH}(B-v_L))/(q_L+q_{LH})$ and $\Pi^b = (v_L q_L - s) + (q_H+q_{LH})P^b$. Hence Π^b is maximized where P^b is maximized subject to the constraints implied by $P^b \leq v_L$ (so that L buys in the low state) and $v_L \leq B \leq v_H$ (so that just one consumer wishes to sell back in the high-demand state). Thus, if $s > q_{LH}\left[v_H - v_L\right]$, then $B = v_H$, $P^b = v_L - (s-q_{LH}(B-v_L))/(q_L+q_{LH}) < v_L$ and

27 Another possible concern is that the gray market might prevent the firm from controlling the distribution of its goods in order to maintain the quality of the product or generate positive externalities from its use.

comparison with Π^{mc} shows that $\Pi^b > \Pi^{mc}$. However, if $s \le q_{LH}(v_H - v_L)$, then $P^b = v_L$ and comparison with Π^{mc} shows that $\Pi^b \ge \Pi^{mc}$ if and only if the equation in the statement of the Proposition holds. ■

As for rationing-with-resale, if the seller were not constrained to set his sale price below v_L (as before, L cannot infer the state from the price, so it may be reasonable to set price above v_L if L cannot observe other consumers' demands before buying), a fixed sale price with a buyback would always dominate market-clearing prices.

Both the buy-back scheme and rationing with resale allow the seller to subsidize consumer L in the high-demand state while avoiding inefficient rationing. They are not identical when the seller has to charge the same price for the good in both states: when s is large, the optimal buy-back repurchases one unit from L at v_H in the high-demand state and this generates a greater subsidy to L in that state than rationing with resale, which allows L to resell only half a unit at v_H. The seller therefore benefits by being able to charge more in the low-demand state ($P^{rr} < P^b \le v_L$). However, if the seller's price is constrained to be no higher than v_L, as we assumed, and s is small so this constraint is binding, then $P^{rr} = P^b = v_L$. In this case, the buy-back scheme earns lower profits than rationing at P^{rr} and allowing resale, since under the buy-back scheme the seller must buy back a unit in the high-demand state at a cost $B > v_L$. We will show in the next section that these distinctions between the schemes disappear when the seller is allowed to charge different prices in different demand states.

V. Rationing as an Optimal Mechanism Design

We have so far shown that even fixing a single price and then rationing when necessary can dominate market-clearing prices. This section shows that rationing continues to arise when we allow the seller to use more general pricing mechanisms. We show that if the seller cannot set a negative price to any consumer in any state (i.e. if he cannot exclude non-genuine customers so "arbitrage" is a problem), then offering the optimal set of contracts is equivalent to implementing a version of rationing without resale (as in Section III) generalized to allow different prices in different states.[28] However, if the seller can set negative prices ("award gifts") in some states without attracting consumers who have not successfully invested, the optimal contracts are equivalent to a similarly generalized version of rationing with resale (as in Section IV).

The order of events is exactly as laid out in Section II except that the seller begins (Stage 1) by announcing and committing to menus of contracts, one menu for each state of total demand (where total demand is the number of consumers who announce success). Each contract in each menu specifies an amount of the good to be received by the consumer and a total payment to be made by the consumer. In stage 2, as before, each consumer independently chooses whether or not to invest. At stage 3, also as before, each consumer finds out the success of his investment, and any consumer who wishes to do so publicly announces that his investment is successful. At stage 4 the seller offers the menu that is appropriate to the state of total demand that corresponds to consumers' stage 3 announcements (as before he is constrained to stick to his stage-1 commitment), and consumers who announced success select contracts from that menu.[29,30] Since there are just two types of consumer, there is no loss of generality in restricting the seller to just two contracts for each state, one of which, in equilibrium, will be selected by each type of consumer. Let the contract that the seller intends that consumer

---

28 Peak load prices and interruptible contracts are examples of prices that depend on the state of demand. In the travel industry, the price and availability of a tour may depend on the number of customers that sign up. Klemperer and Meyer (1989) analyse prices that depend on demand, and discuss how these prices may be implemented in an oligopoly market.

29 As before, the assumption that consumers select contracts only after all consumers have revealed their demands is consistent with our focus on the effects of sunk costs when consumers cannot be contracted with ex-ante. However, forcing consumers to accept contracts before they know their demands would affect the solution only for a small range of parameters, assuming ex-post individual rationality is still required --- see discussion in note on page A-3. Even under our current assumption, under the conditions of Proposition 5, the firm is as profitable as if it could directly compensate all sunk costs.

30 In each demand state, a customer's allocation and payment depends only on his own choice of contract. If the seller could design payoffs to each consumer that depend on the choices of other consumers, the seller might be able to perfectly discriminate among consumers and extract their entire surplus. The general utility of such strategies with correlated demands is established in Cremer and McLean (1985), but while these strategies can greatly increase the seller's feasible profits, they seem unrealistic in many market situations.

$j = H,L$ accept in demand state $i = 1,2$ specify an allocation of σ_{ij} of a unit to j and a total payment of t_{ij} from j to the seller. The seller's objective is to maximize expected transfer payments

$\Pi = q_L t_{1L} + q_H t_{1H} + q_{LH}(t_{2L} + t_{2H})$ with respect to $\{\sigma_{ij}, t_{ij}\}$ and subject to

IR_j:	$q_j(\sigma_{1j}v_j - t_{1j}) + q_{LH}(\sigma_{2j}v_j - t_{2j}) \geq s$	$j = H,L$
IR_{ij}:	$\sigma_{ij}v_j - t_{ij} \geq 0$	$i = 1,2 \quad j = H,L$
IC_{ij}:	$\sigma_{ij}v_j - t_{ij} \geq \sigma_{ik}v_j - t_{ik}$	$i = 1,2 \quad j,k = H,L \quad j \neq k$
Resources:	$0 \leq \sigma_{ij} \leq 1, \; 0 \leq \sigma_{2H} + \sigma_{2L} \leq 1$	$i = 1,2 \quad j = H,L$

The ex-ante individual rationality constraints above, (IR_j), are that each consumer j must receive a non-negative expected surplus to be willing to incur the sunk cost s, (since assumption (A) implies it is optimal for the seller to induce the participation of both consumers). The ex-post individual rationality constraints (IR_{ij}) specify that consumer j prefers to accept the contract offered him (a fraction σ_{ij} of the good at a cost t_{ij}) to not participating in the market in each demand state i. The incentive compatibility constraints, IC_{ij}, specify that consumer j always prefers the contract intended for him to the contract intended for the other type of consumer. Finally, the resource constraints require that the seller not distribute more of the good than is available.

We consider two cases corresponding to whether or not it is feasible for a contract to specify a negative payment, that is, a payment from the seller to the buyer. We first consider the case in which this is not feasible, perhaps because the seller cannot exclude unsuccessful investors or other "arbitrageurs" from misreporting their demands in the hope of obtaining a free gift. This case corresponds to imposing the additional constraints:[31]

T_{ij} (no gifts):	$t_{ij} \geq 0$	$i = 1,2 \quad j = H,L$

[31] We assume no resale between consumers after they have received their contract allocations. This assumption increases the firm's power because the firm can achieve the effect of allowing any resale (subject to the constraints T_{ij}, when they are imposed) by announcing the contracts that yield the post-resale allocations of goods and money (as we shall see in Proposition 5).

<u>*Proposition 4:*</u> *If* $2(\frac{v_H}{v_L} - 1) \leq (\frac{q_H}{q_L} - 1)$, (4)

the optimal mechanism subject to the no-gifts constraints, $t_{ij} \geq 0$, *is equivalent to implementing the following rationing strategy with no resale.*

In the low-demand state $P_1 = min\left[v_L, v_L - \frac{s-\frac{1}{2}q_{LH}v_L}{q_L}\right]$

In the high-demand state $P_2 = v_H$ *with probability* $max\left[1 - \frac{2s}{q_{LH}v_L}, 0\right]$

$P_2 = 0$ *with probability* $min\left[\frac{2s}{q_{LH}v_L}, 1\right]$

and, at the zero price, the seller rations the two consumers equally.

If (4) does not hold, the optimal mechanism subject to $t_{ij} \geq 0$ *corresponds to setting market-clearing prices,* $P_1 = v_L - \frac{s}{q_L}$ *and* $P_2 = v_H$.

<u>Proof:</u> See Appendix. ■

The intuition for Proposition 4 parallels that for Proposition 1, but condition (4) is weaker than the condition (3) under which rationing was preferred in Section III. If we multiply both sides of (4) by s, the left-hand side of (4) then represents the cost to the seller of the inefficiency caused by rationing.[32] This is $(\frac{2s}{q_{LH}v_L}) \leq 1$ times as great as the left-hand side of (3), since rationing now occurs only with probability $(\frac{2s}{q_{LH}v_L})$. Furthermore, the right-hand side of (4) (multiplied by s) represents the saving relative to market-clearing prices in the amount of surplus given to H while giving s to L: allowing the seller to choose different prices in the two states allows the entire subsidy to L to be paid in the high-demand state where a $1 reduction in the price when there is rationing benefits H and L equally, so this saving is

32 To check this, note that with probability $\left[\frac{1}{2}q_{LH}\left[\frac{2s}{q_{LH}v_L}\right]\right]$ the seller inefficiently rations the unit to L. This directly causes a (social) inefficiency of $v_H - v_L$ which the seller must bear (since the seller could sell to H for v_H). Furthermore, since this method of giving a subsidy of v_L to L gives L goods that H values at v_H, H must be given an extra $\left[v_H - v_L\right]$ to satisfy H's incentive compatibility constraint. Thus the total (private) cost to the seller of the inefficiency caused by rationing is $2\left[v_H - v_L\right]\left[\frac{1}{2}q_{LH}\left[\frac{2s}{q_{LH}v_L}\right]\right] = 2s\left[\frac{v_H}{v_L} - 1\right]$.

$(\frac{q_H}{q_L} - 1)s$. This exceeds the right-hand side of (3), which is the saving under the conditions of Proposition 1. (The extension of this intuition to the case where $\frac{2s}{q_{LH}v_L} > 1$ is straightforward.)

The cost to the seller of a rationing scheme depends on the probability with which the good is inefficiently allocated. When "gifts" from the seller to buyers are possible without attracting consumers who have not successfully invested (in particular, if the seller can directly observe success or failure and can exclude unsuccessful investors), this probability can be reduced to zero, even while rewarding the low-value consumer, L, in the high-demand state. Since either resale or a buy-back effects a gift (either of these leads to L receiving cash in the high-demand state), it follows that an appropriately constructed rationing scheme with one of these features always dominates market-clearing prices:

Proposition 5: *With no restrictions on t_{ij} (i.e., gifts are possible), the optimal mechanism is equivalent to the following rationing strategy with resale at price v_H:*

In the low-demand state $\qquad P_1 = v_L$

In the high-demand state $\qquad P_2 = v_H - \frac{2s}{q_{LH}}$

with the seller rationing the two consumers equally. It is also equivalent to the buy-back strategy in which:

In the low-demand state $\qquad P_1 = v_L$

In the high-demand state $\qquad P_2 = v_H - \frac{s}{q_{LH}}$

but the seller buys back a unit at v_H in the high-demand state.

Proof: See Appendix. ■

Both the rationing-with-resale and the buy-back schemes of Proposition 5 eliminate the inefficiency of rationing by ensuring that the good is consumed by the high-value consumer when both consumers demand it. Furthermore both schemes pay the entire subsidy to L in the high-demand state (where a \$1 total subsidy generates a \$1/2 subsidy to L and a \$1/2 to H, assuming the resale or buy-back always takes place at v_H). It follows therefore that both schemes always dominate market-clearing prices which pay the subsidy only in the low-demand state (where a price reduction that costs the seller \$1 generates only a $\$q_L/(q_L+q_H)$ subsidy to L while giving $\$q_H/(q_L+q_H)$ to H).

Note that these strategies achieve a profit of $(q_L+q_H)v_L + q_{LH}v_H - 2s$, compared to the maximum profit with market-clearing prices, $(q_L+q_H)v_L + q_{LH}v_H - (\frac{q_H}{q_L} + 1)s$. The former is the same profit that the seller could achieve if it were able to pay all of both consumers' sunk costs directly.

VI. Conclusions

This paper offers a new explanation for rationing. When consumers must incur seller-specific sunk costs to use a product, the seller must commit to prices that compensate the marginal customer for these costs. Thus prices must accomplish the dual job of providing incentives for investment and allocating output. It is not surprising that these requirements may be in conflict, and that a seller may compromise allocative efficiency in order to promote desired investment. That is, rationing may be a profit-maximizing equilibrium phenomenon.

Our theory addresses related questions such as why sellers might encourage or object to gray markets in rationed goods and why they might organize markets to buy back goods that are in deficit supply. In addition, while we have emphasized that rationing can help a seller extract more surplus from a given customer base, rationing can also be an effective instrument for altering the size and composition of the seller's customer base. Sellers' needs to focus subsidies on key customers may contribute to a range of other phenomena, such as apparently discriminatory refusals to deal.[33]

Our analysis shows that a seller may choose prices that compromise ex-post efficiency in order to promote ex-ante investments. When consumers have to incur seller-specific sunk costs, prices that appear "fair" in the sense that they distribute surplus ex-post to all consumers (and involve rationing), also can be profit-maximizing for the seller.

---

33 See our working paper Gilbert and Klemperer (1995) for some simple examples, in particular an example in which the seller commits to a rationing rule that favors past purchasers.

Appendix. Proofs of Propositions 4 and 5.

Proof of Proposition 4:

The strategy of proof is to use duality theory plus a few short cuts to save algebra. In each regime (according as whether or not (4) is satisfied and as whether or not $s > \frac{1}{2}q_{LH}v_L$) the claimed solution implies a total profit $\widehat{\pi}$ and a vector of shadow values $\widehat{\underline{y}}$ (one for each constraint). If (and only if) the claimed solution is correct, forming the sum of the products of the shadow values and the corresponding constraints will yield an equation that proves that the objective function is no greater than $\widehat{\pi}$ and hence proves that the claimed solution is opimal.[1]

Consider first the program without the constraint IC_{1L}. If in a feasible solution to this program $\sigma_{1L} < 1$, then increasing σ_{1L} by Δ at the same time as increasing t_{1L} by Δv_L and reducing t_{1H} by $(\Delta v_H - \Delta v_L)$ would both be feasible and increase profits for small positive Δ if $q_L v_L - q_H(v_H - v_L) > 0$. But this latter condition is implied by our assumption (A). Furthermore, if either $\sigma_{1H} < 1$ or $\sigma_{2H} < 1 - \sigma_{2L}$, increasing σ_{iH} by Δ at the same time as increasing t_{iH} by Δv_H would be both feasible and profitable for small Δ. So $\sigma_{1L} = \sigma_{1H} = 1$ and $\sigma_{2H} = 1 - \sigma_{2L}$ at the optimum.

Now if (4) holds and $s < \frac{1}{2}q_{LH}v_L$, then substituting these values of σ_{1L}, σ_{1H} and σ_{2H} into the weighted sum of the constraints (in which the weights have been

[1] Write our problem as choose $\underline{x} = \{t_{1L}, t_{1H}, t_{2L}, t_{2H}, \sigma_{1L}, \sigma_{1H}, \sigma_{2L}, \sigma_{2H}\}$ to maximise the objective function $\underline{c}.\underline{x}$ subject to the constraints $\underline{\underline{A}}\ .\underline{x} \leq \underline{b}$, in which the matrix $\underline{\underline{A}}$ and the vectors $\underline{b}$ and $\underline{c}$ are as defined above for our problem. The dual problem is: choose $\underline{y}$ to min $\underline{y}.\underline{b}$ subject to $\underline{y}.\ \underline{\underline{A}} \geq \underline{c}$. The claimed optimal solution $\widehat{\underline{x}}$ implies a total profit $\widehat{\pi}$ and a vector of shadow values $\widehat{\underline{y}}$ (one for each constraint). Now forming the product of the shadow values and the (primal) constraint matrix yields an equation $\widehat{\underline{y}}.\ \underline{\underline{A}}\ .\underline{x} \leq \widehat{\underline{y}}.\underline{b}$. Call this equation (*). But if $\widehat{\underline{x}}$ solves the primal, then $\widehat{\underline{y}}$ solves the dual. In this case $\widehat{\underline{y}}\ .\underline{b} = \underline{c}.\widehat{\underline{x}} = \widehat{\pi}$ (since the primal and dual have the same value) and $\widehat{\underline{y}}.\ \underline{\underline{A}}\ .\underline{x} = (\underline{c} + \underline{v}).\underline{x}$ in which the elements of $\underline{v}$ are all non-negative (since feasibility implies $\widehat{\underline{y}}.\ \underline{\underline{A}} \geq \underline{c}$ implies $\widehat{\underline{y}}.\ \underline{\underline{A}} = \underline{c} + \underline{v}$). So if the claimed solution is correct, then equation (*) will read $(\underline{c} + \underline{v}).\underline{x} \leq \widehat{\pi}$ and so prove that there is no $\underline{x}$ that does better than $\widehat{\underline{x}}$ (since $\underline{c}.\widehat{\underline{x}} = \widehat{\pi}$). We therefore proceed by forming the equation (*), using the $\widehat{\underline{y}}$ that would be the shadow values of the claimed optimal solution. Furthermore, it is clear that in any optimal solution all the constraints corresponding to strictly positive $\widehat{\underline{y}}$ must bind, so it is straightforward to check uniqueness.

chosen using our knowledge of duality),

$$q_H IC_{1H} + q_{LH} IC_{2H} + 2\left(\frac{v_H}{v_L}\right) IR_L + 2q_{LH}\left(\frac{v_H - v_L}{v_L}\right) T_{2L} +$$

$$q_L\left[\left(\frac{q_H - q_L}{q_L}\right) - 2\left(\frac{v_H - v_L}{v_L}\right)\right] IR_{1L},$$

and rearranging, yields$(q_L + q_H)v_L + q_{LH}\left(1 - \frac{2s}{q_{LH}v_L}\right)v_H \geq \Pi$. But this upper bound on profits can be achieved by $t_{1L} = t_{1H} = v_L, t_{2L} = 0, t_{2H} = \left(1 - \frac{2s}{q_{LH}v_L}\right)v_H$, $\sigma_{1L} = \sigma_{1H} = 1, \sigma_{2L} = \left(\frac{s}{q_{LH}v_L}\right), \sigma_{2H} = 1 - \sigma_{2L}$, which parameters satisfy all the constraints, and therefore form a solution to both the program without IC_{1L} and, since the parameters also satisfy IC_{1L}, the full program. Furthermore, this upper bound cannot be achieved unless the constraints $IC_{1H}, IC_{2H}, IR_L, T_{2L}$, and IR_{1L} all bind. Since solving the equations that result from these constraints all binding (together with $\sigma_{1L} = \sigma_{1H} = 1$ and $\sigma_{2L} + \sigma_{2H} = 1$) yields a unique solution (the solution above), it follows that this solution is unique (unless (4) holds with equality).

Now observe IC_{2H} and IC_{2L} together imply $(\sigma_{2H} - \sigma_{2L})v_H \geq t_{2H} - t_{2L} \geq (\sigma_{2H} - \sigma_{2L})v_L$, which implies $\sigma_{2H} \geq \sigma_{2L}$ so $\sigma_{2L} \leq \frac{1}{2}$. Substituting $\sigma_{2L} \leq \frac{1}{2}$and $t_{2L} \geq 0$ into IR_L yields $\sigma_{1L}v_L - t_{1L} \geq (s - \frac{1}{2}q_{LH}v_L)/q_L$. Call this constraint IR'_{1L}. If (4) holds and $s > \frac{1}{2}q_{LH}v_L$, then repeating the argument of the previous paragraph, but with IR'_{1L} substituted for IR_{1L}, shows that the parameters $t_{1L} = t_{1H} = v_L - (s - \frac{1}{2}q_{LH}v_L)/q_L$, $t_{2L} = t_{2H} = 0, \sigma_{1L} = \sigma_{1H} = 1, \sigma_{2L} = \sigma_{2H} = 1/2$ are the (unique) solution to the full program.

If (4) fails, then writing $\sum_{2L}$ for the constraint $\sigma_{2L} \geq 0$ and rearranging the weighted sum

$$q_H IC_{1H} + q_{LH} IC_{2H} + \left(\frac{q_H + q_L}{q_L}\right) IR_L + q_{LH}\left(\frac{q_H - q_L}{q_L}\right) T_{2L}$$

$$+ q_{LH}v_L\left[2\left(\frac{v_H - v_L}{v_L}\right) - \left(\frac{q_H - q_L}{q_L}\right)\right]\Sigma_{2L},$$

yields $(q_L + q_H)(v_L - \frac{S}{q_L}) + q_{LH}v_H \geq \Pi$. But this upper bound can be achieved by the market-clearing parameters $t_{1L} = t_{1H} = (v_L - s/q_L), t_{2L} = 0, t_{2H} =$

$v_H, \sigma_{1L} = \sigma_{1H} = 1, \sigma_{2L} = 0$, $\sigma_{2H} = 1$, which satisfy all the constraints and therefore (uniquely) solve the full program.

Finally, it is easy to check that in each case the optimal t_{ij} and σ_{ij} can be implemented by the strategies specified in the statement of the proposition.[2] $\square$

Proof of Proposition 5:

We use duality theory as in Proposition 4. We use the same argument as in the first paragraph of the proof of Proposition 4, and substitute $\sigma_{1L} = \sigma_{1H} = 1$ and $\sigma_{2H} = 1 - \sigma_{2L}$ into

$$q_H IC_{1H} + q_{LH} IC_{2H} + 2IR_L + (q_H - q_L) IR_{1L} + 2q_{LH}(v_H - v_L)\Sigma_{2L}$$

(where Σ_{2L} is the constraint $\sigma_{2L} \geq 0$ as before), to obtain $(q_L + q_H)v_L + q_{LH}v_H - 2s \geq \Pi$,which upper bound on Π can be achieved by $t_{1L} = t_{1H} = v_L, t_{2L} = -\frac{s}{q_{LH}}, t_{2H} = v_H - \frac{s}{q_{LH}}, \sigma_{1L} = \sigma_{1H} = 1, \sigma_{2L} = 0, \sigma_{2H} = 1$. Since these parameters satisfy all the constraints, they (uniquely) solve the program. It is easy to check that this solution can be implemented using either of the strategies in the statement of the proposition. $\square$.

[2] If consumers had to accept contracts before knowing their demands (i.e. at "stage 2" in the timing), the state by state ex-post incentive compatibility constraints $IC_{ij}, i = 1, 2, j = H, L$, would be replaced by weaker ex-ante incentive compatibility constraints $IC_j \equiv q_j IC_{1j} + q_{LH} IC_{2j}, j = H, L$. This affects the proof only when both (4) and $s > \frac{1}{2} q_{LH} v_L$ hold (assuming that we still require IR_{ij}), and in this case only by permitting the possibility $\sigma_{2L} > 1/2$, that is, the good is rationed even more (ex-post) inefficiently. (Ex-post rationality is a reasonable assumption if the seller cannot observe the success of buyers' investments or for other reasons cannot compel a buyer to purchase the good in each state of demand.) Similarly, it is easy to check that Proposition 5 is unaffected if consumers are forced to accept contracts before observing demand.

References

Allen, Franklin and Gerald R. Faulhaber (1991), "Rational Rationing", *Economica*, 58(230): 189-198, May.

Anton, James J. and Dennis A. Yao (1989), "Split awards, procurement, and innovation", *RAND Journal of Economics*, 20(4): 538-552, Winter.

Basu, Kaushik (1987), "Monopoly, Quality Uncertainty and "Status" Goods", *International Journal of Industrial Organization*, 5: 435-446, December.

Becker, Gary S. (1991), "A Note on Restaurant Pricing and Other Examples of Social Influences on Price", *Journal of Political Economy*, 99(5):1109-1116.

Bohm, Volker, Eric Maskin, Heraklis Polemarchakis, and Andrew Postlewaite, (1983), "Monopolistic Quality Rationing", *Quarterly Journal of Economics*, 98 (Supplement): 189-197.

Bulow, Jeremy I. (1982), "Durable Goods Monopolists", *Journal of Political Economy*, 90(2): 314-332, April.

Bulow, Jeremy, and John Roberts (1989), "The Simple Economics of Optimal Auctions", *Journal of Political Economy*, 97(5): 1060-1090.

Burnett, W. and W. Kovacic (1989), "Reform of United States Weapons Acquisition Policy: Competition, Teaming Agreements, and Dual Sourcing", *Yale Journal on Regulation*, 6: 249-317.

Burke, S. (1990), "Vigorous Chip Sales Boost Intel Earnings, Despite Shortages", *PC Week*, April 6.

Carlton, Dennis W. (1986), "The Rigidity of Prices", *American Economic Review*, 76(4):637-658, September.

Carlton, Dennis W. (1991), "A Theory of Allocation and its Implications for Marketing and Industrial Structure: Why Rationing is Efficient", *Journal of Law and Economics*, October, 34(2): 231-262.

Coase, Ronald H. (1972), "Durability and Monopoly", *Journal of Law and Economics*, 15: 143-49, April.

DeGraba, Patrick (1995), "Buying Frenzies and Seller-Induced Excess Demand", *RAND Journal of Economics*, 26(2):331-42, Summer.

DeGraba, Patrick and Rafi Mohammed (1996), "Intertemporal Mixed Bundles and Buying Frenzies", Working Paper, Federal Communications Commission,1919M Street, Room 538, Washington D.C 20554, March.

Denicolo, Vincenzo and Paulo Garella (1996), "Rationing in a Durable Goods Monopoly", Working Paper, Dipartimento di Scienze Economiche, Strada

Maggiore 45, 40126 Bologna, Italy, August.

Diamond, Peter (1971), "A Model of Price Adjustment", *Journal of Economic Theory*, 3:156-168, June.

Englebrecht-Wiggans, Richard (1987), "On Optimal Reservation Prices in Auctions", *Management Science*, 33(6): 763-70, June.

Englebrecht-Wiggans, Richard (1993), "Optimal Auctions Revisited", *Games and Economic Behavior*, 5(2): 227-239, April.

Farrell, Joseph and Nancy Gallini (1988), "Second Sourcing as a Commitment: Monopoly Incentives to Attract Competition", *Quarterly Journal of Economics*, (4):673-694, November.

Fransman, Martin (1995), *Japan's Computer and Communications Industry*, Oxford University Press.

Ghemawat, Pankaj (1986), "DuPont in Titanium Dioxide", *Harvard Business School Case 9-386- 070*, July.

Gilbert, Richard and Paul Klemperer (1995), "An Equilibrium Theory of Rationing", U.C.Berkeley and Institute of Economics and Statistics, Oxford, Working Paper.

Haddock, David D. and Fred S. McChesney (1992), "Why do Firms Contrive Shortages? The Economics of Intentional Mispricing", Northwestern University working paper.

Harstad, Ronald M. (1990), "Alternative Common-Value Auctions Procedures: Revenue Comparisons with Free Entry", *Journal of Political Economy*, 98(2): 421-29, April.

Inman, James E. (1993), "Gray Marketing of Imported Trademarked Goods: Tariffs and Trademark Issues", *American Business Law Journal*, 31(1): pp. 59-.

Kahneman, Daniel, Jack L. Knetsch and Richard Thaler (1986), "Fairness as a Constraint on Profit Seeking: Entitlements in the Market", *American Economic Review*, 76(4):728-41, September.

Karni, Edi and Dan Levin (1994), "Social Attributes and Strategic Equilibrium: A Restaurant Pricing Game", *Journal of Political Economy*, 102(4): 822-840.

Kenney, Roy and Benjamin Klein (1983), "The Economics of Block Booking", *Journal of Law and Economics*, 26(3): 497-540, October.

King, Ian, Linda Welling, and R. Preston McAfee (1992) "Investment Decisions Under First and Second Price Auctions", *Economics Letters*, 39: 289-

293.

Klemperer, Paul and Margaret A. Meyer (1989), "Supply Function Equilibria in Oligopoly under Uncertainty", *Econometrica*, 57(6):1243-77, November.

Levin, Dan and James L. Smith (1994), "Equilibrium in Auctions with Entry", *American Economic Review*, 84(3): 585-599, June.

McAfee, R. Preston and John McMillan (1987), "Auctions with Entry", *Economics Letters*, 23(1): 343-47.

McKinnon, James G. and Nancy D. Olewiler (1980), "Disequilibrium Estimation of the Demand for Copper", *Bell Journal of Economics*, 11(1):197-211.

McLaren, John (1995), "Why did Big Coffee seek Regulation?" Columbia University Working Paper, September.

Nalebuff, Barry J. and Joseph E. Stiglitz (1983), "Prizes and Incentives: Towards a General Theory of Compensation and Competition", *Bell Journal of Economics*, 14(1): 21-43, Spring.

Okun, Arthur M. (1981), *Prices and Quantities*: A Macroeconomic Analysis, Basil Blackwell, Oxford, United Kingdom.

Png, Ivan P.L. (1991), "Most-Favored-Customer Protection versus Price Discrimination Over Time", *Journal of Political Economy*, 99(5): 1010-1028.

Pyatt, E. (1989), "Procurement Competition at Work: The Navy's Experience", *Yale Journal on Regulation*, 6: 319-331.

Rotemberg, Julio J. and Lawrence H. Summers (1990), "Inflexible Prices and Procyclical Productivity", *Quarterly Journal of Economics*, November.

Slade, Margaret E. (1991), "Strategic Pricing with Customer Rationing: The Case of Primary Metals", *Canadian Journal of Economics*, 24(1):70-100, February.

Stiglitz, Joseph E. and Andrew Weiss (1981), "Credit Rationing in Markets with Imperfect Information", *American Economic Review*, 71(3):393-410.

Stole, Lars A. (1994), "Information Expropriation and Moral Hazard in Optimal Second-Source Auctions", *Journal of Public Economics*, 54: 463-484, July.

Weitzman, Martin L. (1977), "Is the Price System or Rationing More Effective in Getting a Commodity to Those Who Need it Most?" *Bell Journal of Economics*, 8:517-24, Autumn.