CESifo *Working Paper Series*

HOW EFFICIENT IS A CONTESTABLE NATURAL MONOPOLY?

Alfons J. Weichenrieder*

Working Paper No. 186

May 1999

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1344
Fax: +49 (89) 9224-2344
http://www.lrz-muenchen.de/~ces/c02.htm

* I thank Geir Asheim and Marcel Thum for valuable discussion. All remaining errors are mine.

How Efficient is a Contestable Natural Monopoly?

by

Alfons J. Weichenrieder*

University of Munich

June 1999

Abstract

This paper considers the efficiency of a contestable natural monopoly if consumers are heterogeneous and the monopolist can differentiate prices imperfectly. With restricted entry, the standard result in this case is that the monopoly offers a menu of price-quantity combinations which leads to the well-known 'no–distortion–at–the–top' pricing. Low demand consumers are induced to consume less than their first-best quantity, while high demand consumers buy a quantity where their marginal willingness to pay equals marginal cost. The paper shows that this type of inefficiency may also appear in a contestable market. Depending on cost and demand structures, first best efficiency can also be a sustainable equilibrium. However, due to the existence of a continuum of equilibria, first best efficiency is never guaranteed. Most notably, even a stable 'distortion–at–the–top' result is possible.

Keywords: natural monopoly, contestability, contract theory
JEL classification: D42, D82

Address of the author
Alfons J. Weichenrieder
Center for Economic Studies
University of Munich
Schackstr. 4
80539 Munich
Germany

a.weichenrieder@LRZ.de

*I thank Geir Asheim and Marcel Thum for valuable discussion. All remaining errors are mine.

Introduction

In markets where resale by consumers is difficult or impossible, price differentiation between consumers is a frequent empirical phenomenon. Examples of prices which depend on the quantity purchased (non-linear prices) can be found in the electricity and airline industries, in telephone services, in insurance and in many other areas. A great deal of theoretical analysis has been devoted to such pricing strategies. The focus of most existing studies is the problem of a profit maximising monopolist who is informed about the overall distribution of consumer types but who cannot tell any single buyer's type. For important contributions see Spence (1977), Mussa and Rosen (1978), Maskin and Riley (1984), Spulber (1993) and, for the case of insurance, Stiglitz (1977).

The present paper is concerned with price discrimination between consumer groups in the case of a contestable natural monopoly. A natural monopoly is defined as a market where, because of decreasing average cost, cost efficiency requires only one active firm. Contestability occurs if the market allows free and costless entry and exit of competitors [Baumol, Panzar and Willig (1982)]. This implies, among other things, that there are no sunk costs of production.

Price discrimination in a contestable natural monopoly has received little attention in the literature.[1] This is surprising since the "paradigmatic contestable market" which is often presented as an example of a contestable natural monopoly, namely airline service to a small town, typically does show price differentiation between buyers which depends on the quality of service.

Modeling price discrimination between consumer groups, the present paper addresses the question of how efficient a contestable natural monopoly is compared to a non-contestable monopoly on the one hand and to perfect competition on the other hand. It turns out that its

[1] A few exceptions should be mentioned. Shaffer (1987) discusses two-part tariffs but does not consider the case of different consumer types as the present paper does. Heyword and Pal (1993) discuss two-part pricing with heterogenous consumers but do not consider the case of a natural monopoly with decreasing average cost. Allard, Cresta and Rochet (1997) consider price discrimination and fixed cost in an insurance market with adverse selection. Unlike the present paper, which considers firm specific fixed cost, these authors assume contract specific fixed cost, which, in insurance markets, may lead to pooling equilibria.

market equilibrium may be first best. However, stable equilibria with 'distortion-at-the-top' and 'distortion-at-the-bottom' results are also possible.

I. A SIMPLE TWO-CONSUMER TYPE FRAMEWORK

The objective of the paper is to evaluate the efficiency properties of a contestable natural monopoly when non-linear pricing is feasible. There are two possible reference points which are useful in this evaluation. On the one hand, if fixed cost were absent, then perfect competition would lead to marginal cost pricing without any price differentiation between consumers.[2] Another point of reference is the market outcome in a natural monopoly if new entry of firms is excluded.[3] The following section will reproduce the no-distortion-at-the-top result in this case. Section III will then introduce contestability. Section IV adds some remarks on insurance before Section V concludes.

I assume that there are two consumer groups, labelled H and L. The absolute number of consumers in the respective groups is α_H and α_L. The utility of an individual from group i = H, L depends on the quantity x consumed and on a total payment t made to the supplying firm: $U_i(x_i, t_i) = u_i(x_i) - t_i$, where $u_i' > 0, u_i'' < 0$. Note that x may also be interpreted as a quality index in the case of a good of which every consumer demands only one unit. In the airline service example, business class would imply a higher x than coach class. H's marginal willingness to pay is supposed to exceed that of L for any level of x: $u_H'(\bar{x}) > u_L'(\bar{x}),\ \forall \bar{x}$. This ensures that in (t,x)-space indifference curves of the two types cross only once (single crossing condition). Reservation utility of consumer type i is given by $\overline{U_i} = u_i(0) - 0$. The cost function C of a monopolist is assumed to consist of a fixed cost F (which is not sunk in the case of the contestable monopoly) and constant marginal cost, c:

$$C(\alpha_H x_H + \alpha_L x_L) = F + c \cdot (\alpha_H x_H + \alpha_L x_L),\ \forall (\alpha_H x_H + \alpha_L x_L) > 0$$

[2] Compare Oren, Smith and Wilson (1983).
[3] This is assumed in most of the existing literature.

II. Restricted Entry

Without potential competitors, the monopolist maximises profits by offering pairs of quantities and payments (x_i, t_i) for H type and L-type consumers $(i = H, L)$ such that each consumer type voluntarily chooses its designated contract. Formally, the firm's problem is:

$$\max_{\{t_H, t_L, x_H, x_L\}} \pi = \alpha_H t_H + t_L \alpha_L - C(\alpha_H x_H + \alpha_L x_L) \tag{1}$$

s.t.

$$u_L(x_L) - t_L - u_L(0) \geq 0 \tag{1a}$$

$$u_H(x_H) - t_H - u_H(0) \geq 0 \tag{1b}$$

$$[u_H(x_H) - t_H] - [u_H(x_L) - t_L] \geq 0 \tag{1c}$$

$$[u_L(x_L) - t_L] - [u_L(x_H) - t_H] \geq 0 \tag{1d}$$

First note that, from (1a), (1c) and the single crossing condition, (1b) is redundant and can be eliminated. Then in an optimum (1a) must be binding. If this were not the case, the firm could increase profits by marginally increasing t_L and t_H while keeping quantities constant and still satisfying all other constraints. Further, (1c) must be binding. Otherwise, it would be possible to increase the payment t_H. Under the assumption that both groups are served in equilibrium, the first order conditions are

$$u_H' = c \tag{2}$$

$$u_L' - c = (\alpha_H / \alpha_L)[u_H'(x_L) - u_L'(x_L)] > 0 \tag{3}$$

which imply underprovision for low demand consumers and the 'no–distortion–at–the–top' result.

III. The Contestable Monopoly

Now consider free entry and exit of firms.[4] In many natural monopolies, in particular those where significant sunk costs have to be incurred, the assumption of free entry and exit will be much too optimistic. However, there clearly are markets with fixed costs but no sunk costs. In the example of airline service to a small city, the wage bill for the pilot and the landing fee are

[4] Potential entrants can use the same technology as an incumbent and can serve the same demand.

largely independent of the number of passengers and therefore lead to fixed costs. However, these costs are certainly not sunk. Another example of a market with economies of scale but no sunk costs is mailing services. While it is economically more efficient if a certain area is served by only one postman instead of two competing ones, the postman's wage is not a sunk cost.

These examples indicate that the case of a contestable natural monopoly does indeed deserve attention. Moreover, the framework of a contestable market is intellectually interesting as it serves as a benchmark case which usually gives the most optimistic prediction about the working of private markets.

For further analysis, it is helpful to first restate the equilibrium concept in a contestable market (Baumol, Panzar and Willig 1982).[5]

Definition 1: In a *sustainable industry configuration* the following holds. (1) Aggregate demand equals firms' aggregate output. (2) Each active firm's revenues are no less than its production costs. (3) There must be no opportunities for entry that appear profitable to potential entrants who regard the price-quantity offers made by incumbent firms as given.

Note that a situation where each consumer group is served by a different firm cannot be a sustainable industry configuration. The fixed costs imply that a single firm jointly supplying both consumer groups can always outperform two separate firms.

In the current setting, a sustainable configuration according to Definition 1 can only be attained if the utility of one group is maximised given the utility of the other group and the non-negativity constraint on profit. Either the utility level for the low demanders or the level for the high demand consumer may be taken as given. In the following, assume a certain utility level U_0 for type L consumers. Then a sustainable industry configuration must be a solution to the following problem:

$$\max\ u_H(x_H) - t_H \tag{4}$$

s.t.

[5]This is a slight generalisation of the original definition to the case of non-linear pricing. In Baumol, Panzar, Willig (1982), potential entrants regard incumbents' prices (instead of contracts) as fixed.

$$u_L(x_L) - t_L - U_L^0 \geq 0 \tag{4a}$$

$$u_H(x_H) - t_H - u_H(0) \geq 0 \tag{4b}$$

$$[u_H(x_H) - t_H] - [u_H(x_L) - t_L] \geq 0 \tag{4c}$$

$$[u_L(x_L) - t_L] - [u_L(x_H) - t_H] \geq 0 \tag{4d}$$

$$\alpha_H t_H + \alpha_L t_L - C(\alpha_H x_H + \alpha_L x_L) \geq 0 \tag{4e}$$

The maximisation problem then leads to the Lagrangian

$$\begin{aligned} L = {} & u_H(x_H) - t_H + \Theta[u_L(x_L) - t_L - U_0] + \Gamma[u_H(x_H) - t_H - u_H(0)] \\ & + \Lambda[u_H(x_H) - t_H - u_H(x_L) + t_L] + \Omega[u_L(x_L) - t_L - u_L(x_H) + t_H] + \\ & + \Phi[\alpha_H t_H + \alpha_L t_L - C(\alpha_H x_H + \alpha_H x_L)] \end{aligned}$$

with the following first order conditions:

$$(1+\Gamma)u'_H(x_H) + \Lambda u'_H(x_H) - \Omega u'_L(x_H) - \Phi \alpha_H c = 0 \tag{5}$$

$$\Theta u'_L(x_L) - \Lambda u'_H(x_L) + \Omega u'_L(x_L) - \Phi \alpha_L c = 0 \tag{6}$$

$$-(1+\Gamma) - \Lambda + \Omega + \Phi \alpha_H = 0 \tag{7}$$

$$-\Theta + \Lambda - \Omega + \Phi \alpha_L = 0 \tag{8}$$

Note that not every solution of (4) is necessarily a stable industry configuration. An entrant could try to enter by increasing the utility level of H types. Although he could not attract L types and therefore fixed costs are spread among a smaller number of consumers, H types could still benefit from not subsidising L types. To rule out this case, the incumbent must not subsidise L types 'too much'. More specifically, he must comply with

Restriction 1: A solution to the maximisation problem (4) is only a sustainable industry configuration if $t_L > x_L c$, i.e. if low demanders bear at least the variable cost necessary to produce the units they consume.

Finally, a solution to (4) must also rule out L-type consumers being better off if they are served by a separate firm. This introduces

Restriction 2: A sustainable industry configuration requires $U_L^0 \geq u(x_L^*) - c \cdot x_L^* - F/\alpha_L$, where x_L^* is the first best quantity for L-type consumers.

It is clear from the contestability of the market that (4e) must be binding with equality. Moreover, positive sales to H types require that type H's participation constraint (4b) is complied with ($\Gamma = 0$). Next, note that (4c) and (4d) cannot both be binding with equality. If this were the case, $t_H > t_L$ would follow from (4c) and (4d). However, given the single crossing condition, it is not possible to find two contracts with different t's which are preferred equally by both types. This contradiction shows that the two self-selection constraints (4c) and (4d) cannot be binding simultaneously. Therefore, only three cases need to be distinguished. First, neither (4c) nor (4d) is binding. In this case, there is a first best outcome. Second, we may have solutions in which only the self-selection constraint for the H type (4c) is binding. Third, only the self-selection constraint for the L type (4d) may be binding. Consider the optimality conditions in the latter two cases.

The self-selection constraint for the high demand consumers is binding ($\Lambda > 0, \Omega = 0$).

From the first order condition, we have

$$u'_H(x_H) = c, \tag{9}$$

$$u'_L(x_L) = c + \frac{\Phi\alpha_H - 1}{\Phi\alpha_H}\left[u'_H(x_L) - u'_L(x_L)\right]. \tag{10}$$

Since for $L > 0$, $\Gamma = \Omega = 0$ eq. (7) shows that $\Phi\alpha_H$ is larger than unity, the first order conditions (9) and (10) imply a less than first best efficient consumption for the L type and an efficient consumption for the H type.

The self-selection constraint for the low demand consumers is binding ($\Lambda = 0, \Omega > 0$).

The first order conditions yield

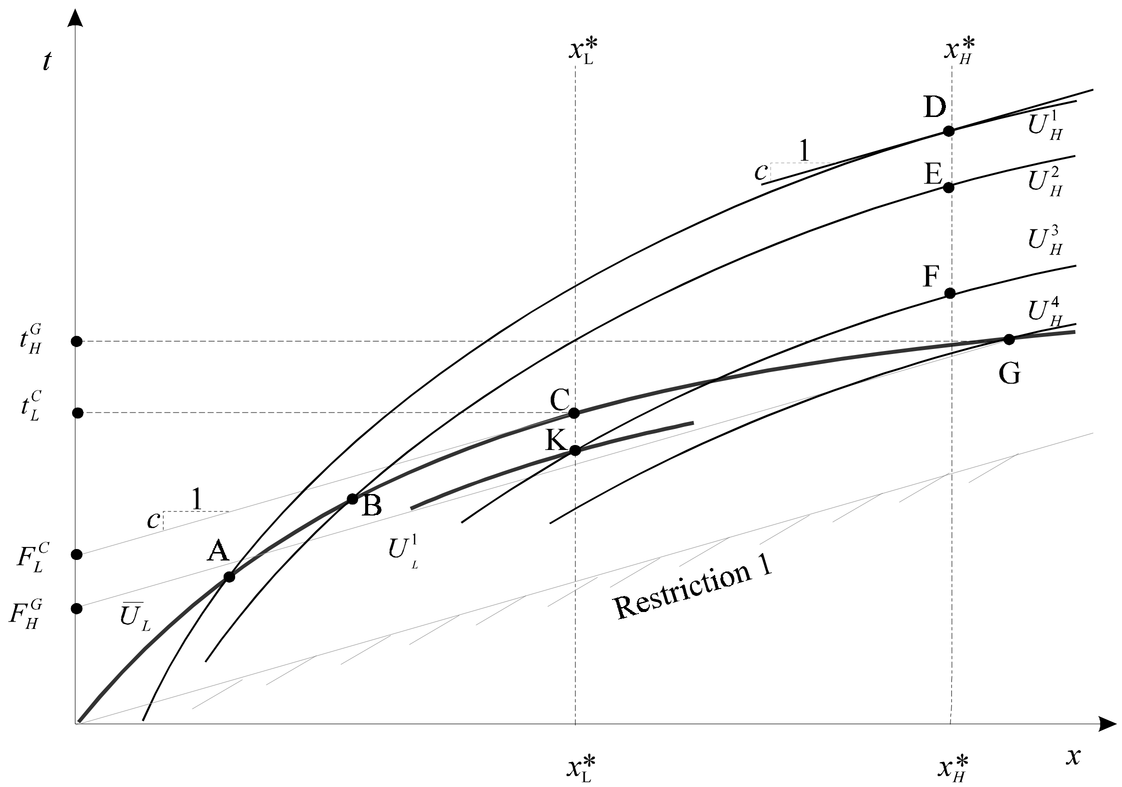
$$u'_H(x_H) = c + \Omega\left[u'_L(x_H) - u'_L(x_L)\right] \tag{11}$$

$$u'_L(x_L) = c. \tag{12}$$

When the L-type consumer is indifferent between the contracts offered (i.e. $\Omega > 0$), and the H-type consumer prefers his own contract, $t_H > t_L, x_H > x_L$ is required. Together with condition (11), this implies that the high demand consumer consumes 'too much'. Intuitively, if (4d) is binding, then it is better to increase high demand consumers' welfare by offering an 'excessive' amount of x_H than by reducing t_H. The reason is that the first measure is better suited to deter low demand consumers from buying the contract designed for H types.

Figure 1 illustrates why all three cases may appear. It depicts the sets of indifference curves for the L and H types, respectively. The slope of the indifference curves measures the marginal willingness to pay for *x*. Since income effects are assumed to be absent, the indifference curves' slopes are only dependent on the amount of *x* and the consumer's type. The single crossing property guarantees that, for any *x*, H's indifference curves are steeper than L's curves. For simplicity the figure assumes $\alpha_L = \alpha_H = 1$.

Figure 1. Contestable self-selection contracts



Consider first a monopoly without entry. Here the monopolist will provide a pair of contracts like A and D. The low demand consumer is on her reservation utility $\overline{U}_L$ and she consumes less than the first best efficient quantity x^*_L, i.e. at A the slope of the indifference curve $\overline{U}_L$ is larger than *c*. The self-selection restriction for type H is binding and his contract D yields the same utility as contract A would. Now consider free entry. Because of contestability, the monopolist will earn zero profit. Moreover, it must not be possible for an entrant to offer a contract which is strictly preferred by one group and which breaks even. To reduce profits the monopolist may either increase H's welfare, or L's welfare, or both. Figure 1 assumes that L keeps her reservation utility. Then, which constraint is binding depends on the exact cost structure. If, for given marginal cost *c*, fixed cost is high and therefore profits were low under the non-contestable monopoly, then the new utility level of H (due to a decrease of t_H) could be at U_H^2 and the resulting contract for L would shift from A to B. x_L would increase but would still be inefficiently low. If, because of low fixed cost, monopoly profits in the no-entry case were higher, the reduction of t_H may lead to a utility level like U_H^3. In this case the relevant indifference curves intersect between the lines $x_L^* - x_L^*$ and $x_H^* - x_H^*$. Neither restriction (4c) nor (4d) is binding and L's contract C implies a first best efficient quantity of x_L^*. Finally, disposal of very high monopoly profits may lead to the contract pair (C,G), for which (4d) is binding but (4c) is not. In this case a distortion at the top emerges: $u'(x_H) < c$.

By assumption, all three contract pairs imply zero profit. Furthermore, in all three case one group's utility is maximised given the other group's utility. Therefore, profitable entry by a firm which tries to attract both consumer types is ruled out. However, it has been emphasised above that not just any solution of (4) is a sustainable industry configuration. A stable solution must also comply with Restrictions 1 and 2 which are sufficient conditions to rule out that an entrant can profitably attract a single consumer group. Note that (C,G) complies with Restriction 2 as long as H pays slightly more than his variable cost. This is indeed the case in Figure 1. To see this, split the consumers' total payments for the contracts C and G into components which represent participation in fixed cost (F_L^C, F_H^G) and components which reflect attributable variable cost $(t_L^C - F_L^C, t_H^G - F_H^G)$. The slope *c* of the cost function is used for this separation of total payments. Because of the zero profit condition, $F_L^C + F_H^G$ equals *F*. If a firm tried to supply only the L consumer with her optimal amount x_L^*, L would have to bear both total fixed cost *F* and variable cost which equals $t_L^C - F_L^C$. This gives L a lower utility than the contract C as long as F_H^G is positive.

Note that L's marginal willingness to pay is larger than, or equal to, marginal cost in all contract pairs considered. Combined with the fact that L is at her reservation utility, this implies that t_L is larger than the variable cost caused by the quantity consumed by L, i.e.

$t_L > c \cdot x_L$. Therefore, all three pairs of contracts [(B,E); (C,F); (C,G)] comply with Restriction 1 (E, F and G lie above the hatched area) and form a contestable market equilibrium. No entrant can lure away the H-type consumer even though he may be forced to take more than the first best quantity.

In the case of pair (C,G) the monopolist could increase L's utility level somewhat, say to level U_L^1. At the same time, however, t_H must be increased to fulfil the zero profit restriction. This obviously reduces H's utility. Although the equilibrium (K,F) provides first best quantities, the incumbent is not forced by competitive pressure to move from (C,G) to (K,F). This clearly introduces some arbitrariness as the utility levels and quantities are not exactly determined. Similarly, as long as Restriction 1 and the zero profit condition is satisfied, the monopolist could move from (B,E) to a pair of contracts which increases L's utility but lowers her consumption of *x*.

If, in the case of restricted entry, the monopolist can profitably serve both consumer groups (i.e. if it is profitable to offer contract pair (A,D) in Figure 1), then there is a infinite set of equilibria under contestability. The reason is that, within bounds, the incumbent can increase U_L^0 in the maximisation problem (4) beyond the L-type's reservation utility $\overline{U}_L$ without fearing entry. The loss in profits can be compensated by a reduction of H-type's utility. The bounds within which this can be done are set by Restriction 1 and 2 and by the H-type's reservation utility.

The discussion may be summarised in the following way.

Proposition 1: *If in a no-entry monopoly both consumer groups were served profitably, then there will be a continuum of sustainable industry configurations (equilibria) under contestability.*

Proposition 2: *A contestable market equilibrium with self-selection contracts may either provide (a) an inefficiently low quantity to L types and a first best quantity to H types, or (b) an inefficiently high quantity for H types and an efficient quantity to L types, or (c) a first best quantity for both groups.*

Proposition 2 can be proven by constructing a simple numerical case.

Example 1: Consider the example where $u_L = x_L{}^{0.5}$, $u_H = 1.5 \cdot x_H{}^{0.5}$, $\alpha_L = \alpha_H = 1$, $c = 1$. The first-best optimal quantities then are $x_L = .25$; $x_H = .5625$. To ease restriction (4d) as much as possible, set $U_L^0 = 0$ (which implies $t_L > x_L c$ and therefore satisfies Restriction 1). Then, with first best quantities and zero profit, constraint (4c) is violated for $F > .5625$. Hence, in a firm's optimal policy from $\Lambda > 0$, $u_L{}' > c$ holds. Since the social value of the first best production level is positive as long as $F < .8125$, according to (7) and (10) in the range $.5625 < F < .8125$ supply for low demand consumers is inefficiently low. If $U_L^0 = 0$, neither (4c) nor (4d) is violated for first best quantities if $.4375 < F < .5625$. A first best outcome indeed evolves. In the range $.25 < F < .4375$ first best quantities violated (4d), and this indicates 'overconsumption' by the H-type consumer. Restriction 2 is complied with as long as *F* exceeds .25. ■

Policy Implications

The maximisation problem (4) is very similar to a central planner's problem if the planner and the firm have identical information. This raises the question of whether beneficial policy measures exist. In a Kaldor-Hicks sense, the government could possibly improve the outcome if it levied a uniform tax which is then used to ease a binding constraint. For example, in the equilibrium (B,E) of Figure 1, the government could subsidise the contract which offers more of good *x*. This could shift point E downwards and would enable the firm to increase x_L without violating the self-selection constraint. The problem of such a policy is that the firm might as well use the subsidisation of contract E to increase the L type's utility. While this would increase L's utility, it would lower L's consumption of *x*. Therefore, to guarantee a Kaldor-Hicks improvement, the government must, for example, make the subsidy conditional on a reduction of t_H. A similar argument applies if the constraint (4d) is binding in the market equilibrium.

Another possible reason for government intervention arises from the indeterminacy of the market result. Although the firm's maximisation problem (4) looks like a beneficial government's problem, the government may not share the firm's preferences about the value of the low demand consumers' utility level, U_L^0. In other words, the government may wish to attribute the fixed cost differently from a private firm.

IV. CONCLUSIONS

This paper links the literature on contestable markets with the work on optimal contracts. The choice of optimal self-selection contracts by a monopolist was shown to imply a certain arbitrariness. Cost and demand structures which create high profits in the case of restricted entry enable a firm to supply first-best quantities in a contestable market. However, a distorted equilibrium in which one consumer group's marginal willingness to pay is not equal to marginal cost is also a sustainable outcome. For example, if only small profits are possible in the case of restricted entry, then the monopolist under free entry can potentially increase the inefficiently small quantity low demand consumers want to buy but may not be able to provide first best quantities. Perhaps the most interesting result is that, in a contestable monopoly, high demand consumers may be forced to consume too much of the good, i.e. their marginal willingness to pay falls short of marginal cost. This occurs if the monopolist has a preference for making the high demand consumers well off. Overconsumption then is a device to deter low demand consumers from buying the high demand consumers' contract.

Bibliography

Allard, Marie, Jean-Paul Cresta, and Jean-Charles Rochet (1997), "Pooling and separating equilibria in insurance markets with adverse selection and distribution costs", *Geneva Papers on Risk and Insurance Theory* 22, 103-120.

Baumol, William J., John C. Panzar and Robert D. Willig (1982), *Contestable markets and the theory of industry structure*. San Diego, et al.: Harcourt Brace Janovich.

Heywood, John S. and Debashis Pal (1993), "Contestability and two-part pricing", *Review of Industrial Organization* 8, 557-565.

Maskin E. and J. Riley (1984), "Monopoly with incomplete information", *Rand Journal of Economics* 15, 171-196.

Mussa, M. and S. Rosen (1978), "Monopoly and product quality", *Journal of Economic Theory* 18, 301-317.

Oren, Shmuel S., Stephen A. Smith and Robert B. Wilson (1983), "Competitive nonlinear tariffs", *Journal of Economic Theory* 29, 49-71.

Rothschild, Michael and Joseph E. Stiglitz (1976), "Equilibrium in competitive insurance markets. An essay on the economics of imperfect information", *Quarterly Journal of Economics* 90, 626-650.

Shaffer, Sherill (1987), "Two-part tariffs in a contestable natural monopoly", *Economica* 54, 315-316.

Spence, Michael (1977), "Nonlinear prices and welfare", *Journal of Public Economics* 8, 1-18.

Spulber, Daniel F. (1993), "Monopoly pricing", *Journal of Economic Theory* 59, 222-234.

Stiglitz, Joseph E. (1977), "Monopoly, non-linear pricing and imperfect information: the insurance market", *Review of Economic Studies* 44, 407-430.