Two-part pricing under revenue cap regulation

Björn Lantz

Two-part pricing under revenue cap regulation

Abstract: This paper aims at developing the theoretical understanding of revenue capping as a way of regulating monopolistic firms. It is shown that the fact that a standard monopolist regulated by a fixed revenue cap will raise its price above the unregulated monopoly level is robust to two-part pricing. It is also shown that when regulation of a two-part pricing monopolist is based on a hybrid revenue cap defined as a linear function of quantity, it is the slope of the cap that determines its incentives for efficiencient behaviour while the intercept of the cap only affects the profit level of the firm. This also holds if the cap is defined as a hybrid price-revenue cap. The general conclusion of this is that the slope of the hybrid cap needs to be steeper that the slope of the firm's cost function in order to prevent the incentive to raise price above the unregulated monopoly level.

Handelshögskolan vid Göteborgs universitet
School of Economics and Commercial Law at Göteborg University
Företagsekonomiska institutionen
Department of Business Administration
Box 610, 405 30 Göteborg
Björn Lantz, tel. 031-773 5245, e-mail: <u>bjorn.lantz@handels.gu.se</u>

INTRODUCTION

Revenue cap regulation is often used as an alternative to price cap regulation. It is a common regulation regime especially in the electricity industry (Jamasb & Pollitt, 2001; Viljainen *et al*, 2004) and revenue caps are often assumed by regulators to have similar efficiency properties as price caps. However, there are significant differences between price caps and revenue caps with respect to their theoretical implications. The main difference is of course that price caps by definition cap prices and may thus prevent monopolistic firms from using their monopoly power. Revenue caps do not necessarily cap prices.

In fact, revenue caps may even provide incentives for price increases above the unregulated monopoly level (Crew & Kleindorfer, 1996). Consider the following single-period decision problem of monopolist facing a fixed revenue cap:

$$\text{Maximize} \quad P \cdot Q(P) - C(Q(P)) \qquad\qquad (1)$$
$$\text{Subject to} \quad P \cdot Q(P) \le \overline{R}$$

where P is price (the decision variable), $Q(P)$ is the demand function of the firm, assumed decreasing in P and assumed to have decreasing price elasticity for decreasing prices, $C(Q)$ is the cost function of the firm, assumed increasing in Q, and $\overline{R}$ is the fixed revenue cap. If the cap is not binding, the firm will choose its unrestricted monopoly price P^*. If the cap is binding, the firm must obviously choose a price P_R that is either lower or higher than P^* in order to satisfy the cap. Obviously, the former would increase the demanded quantity and therefore cost while the latter would decrease both. Since revenue will be $\overline{R}$ in both cases, the firm will prefer the higher price solution $P_R > P^*$ yielding a lower cost and thus a higher profit. This is sometimes referred to as "The Crew-Kleindorfer effect" (see e.g. Comnes *et al*, 1995) and serves as one of the main theoretical objections against revenue cap regulation.[1]

There is of course nothing wrong with the above Crew-Kleindorfer analysis. However, many revenue cap regulated firms use two-part pricing rather than linear pricing. For example, the typical electricity distributing firm who faces a revenue cap uses a two-part tariff where the individual customer has to pay a fixed fee to get access to the network and a variable fee that depends on the electricity consumption. Furthermore, the revenue caps used in practice are rarely totally fixed, but rather a function of some variable. This highlights the need to deepen the theoretical Crew-Kleindorfer analysis with respect to non-constant revenue caps and two-part pricing structures. The purpose of this paper is to provide such analyses in order to improve the understanding of revenue capping as regulation regime.

The paper is organized as follows. First, we analyze the case of a monopolistic firm that uses two-part pricing under a fixed revenue cap. Then we extend the two-part pricing to analyses of revenue caps that are functions of quantity and price, respectively, before the paper is concluded.

---

[1] The main "practical" objection may be that the regulator needs a lot of detailed information to be able to determine the cap parameters.

TWO-PART PRICING UNDER A FIXED REVENUE CAP

The decision problem of the monopolistic firm using non-discriminatory two-part pricing is

$$\text{Maximize} \quad N \cdot F + P \cdot Q(P) - C(Q(P)) \qquad (2)$$
$$\text{Subject to} \quad N \cdot F + P \cdot Q(P) \leq \overline{R}$$

where F is the fixed part and P the variable part of a two-part pricing tariff (the decision variables) and N is the number of consumers.

Let (F^*, P^*) be the unrestricted monopoly solution, i.e. the solution to the decision problem above when the revenue cap $\overline{R}$ is not binding.[2] Now assume the cap is actually binding, i.e. $N \cdot F^* + P^* \cdot Q(P) > \overline{R}$. This means that the solution (F, P) to the decision problem above must involve a reduction of total revenue so that the restriction $N \cdot F + P \cdot Q(P) = \overline{R}$ is satisfied. Assuming that the number of consumers is exogenously given, there are three ways to accomplish that:[3]

- Setting $F < F^*$, which obviously reduces total revenue
- Setting $P < P^*$, which increases demand but reduces total revenue when a solution in the inelastic portion of demand is obtained
- Setting $P > P^*$, which decreases demand and thereby also total revenue when a solution in the elastic portion of demand is obtained.

Since total revenue will be $\overline{R}$ in all three cases, the monopolistic firm will choose the least-cost solution, which is equivalent to the solution with the highest marginal price and thereby the lowest demanded quantity. This is solution 3 where $P > P^*$. Thus, the profit maximizing solution is to decrease social efficiency.

The monopolistic firm can also combine a higher P with a lower F, or vice versa, in order to reduce total revenue. A solution (F, P) where $F > F^*$ and $P < P^*$ will obviously lead to a higher demand and thereby a higher total cost than a solution where $F < F^*$ and $P > P^*$. The latter alternative will obviously be preferred, since total revenue in both cases are $\overline{R}$. Again, the profit maximizing solution is to raise the marginal price which decreases both demand and social efficiency.

---

[2] Oi (1971) showed that an unrestricted monopolistic firm will maximize profit at a variable price P^* that is lower than the unrestricted linear monopoly price and a fixed fee $F^* > 0$. This means that the possibility for an unrestricted monopolistic firm to use two-part pricing creates a potential for higher social efficiency even though it forces the consumers to use a larger portion of their budget on the monopolist's product or service.

[3] If the number of consumers is an endogenous variable and the consumers are assumed to be sufficiently sensitive to changes in F, there may be a fourth way to reduce total revenue: Setting $F > F^*$ which would reduce the number of customers that buys the product or service at the variable price P^*. In this case, Q must be defined as a function of both P and F, which would make formal reasoning far more complex. However, the basic logic used in this section works even if this assumption is relaxed. The profit maximizing two-part tariff (F, P) under a fixed and binding revenue cap will always be the tariff where total revenue equals $\overline{R}$ at the lowest possible quantity, which implies that $P > P^*$ regardless of whether consumers are sensitive to changes in F or not. In fact, sensitivity to changes in F may actually reinforce the effect described here. Under a binding revenue cap, the monopolist will set $F < F^*$ and $P > P^*$ as we see here. But if consumers are assumed to be sensitive to changes in F, the number of customers with access will be increased, which means that the monopolist has to raise P even more to avoid violating the revenue cap.

So, what solution will the monopolistic firm prefer: A solution (F_1, P_1) where $F_1 = F^*$ and $P_1 > P^*$ or a solution (F_2, P_2) where $F_2 < F^*$ and $P_2 > P_1 > P^*$? Again, the firm will choose solution based on cost since total revenue in both cases are $\overline{R}$. The profit maximizing solution must be to lower F as much as possible in order to be able to raise P as much as possible in order to reach total revenue of $\overline{R}$ at lowest possible demand and thus lowest possible cost. Thus, a monopolistic firm using two-part pricing under a fixed and binding revenue cap will raise marginal price and thus reduce social efficiency compared to a situation where no binding revenue cap is applied.

The fact that the monopolistic firm will seek a pricing solution that raises marginal price can easily be proved. The decision problem of the monopolistic firm was

$$\text{Maximize} \quad N \cdot F + P \cdot Q(P) - C(Q(P)) \qquad (3)$$
$$\text{Subject to} \quad N \cdot F + P \cdot Q(P) \leq \overline{R}$$

which can be rearranged to

$$\text{Maximize} \quad N \cdot F + P \cdot Q(P) - C(Q(P)) \qquad (4)$$
$$\text{Subject to} \quad N \cdot F \leq \overline{R} - P \cdot Q(P).$$

If the restriction is binding, i.e. if the revenue cap prevents the unrestricted monopoly solution (F^*, P^*), we get the decision problem

$$\text{Maximize} \quad N \cdot F + P \cdot Q(P) - C(Q(P)) \qquad (5)$$
$$\text{Subject to} \quad N \cdot F = \overline{R} - P \cdot Q(P).$$

If we substitute $N \cdot F$ with $\overline{R} - P \cdot Q(P)$ in the objective function and simplify, we get

$$\text{Maximize} \quad \overline{R} - C(Q(P)) \qquad (6)$$
$$\text{Subject to} \quad N \cdot F + P \cdot Q(P) = \overline{R},$$

which has the same solution (F, P) as

$$\text{Minimize} \quad C(Q(P)) \qquad (7)$$
$$\text{Subject to} \quad N \cdot F + P \cdot Q(P) = \overline{R}.$$

Thus, the solution to the problem when the revenue cap is binding is to choose the specific tariff combination (F, P) that minimizes cost among the possible combinations that yields the maximum allowed total revenue. Since cost is assumed to decrease when demand is decreased, and demand in turn is assumed to decrease when marginal price P is increased, the monopolistic firm will maximise profit by setting the highest possible marginal price that yields total revenue of $\overline{R}$.

Hence, moving from linear pricing to two-part pricing under a fixed revenue cap does not eliminate the Crew-Kleindorfer effect for the regulated firm. In general, it is easy to see that the effect actually is reinforced under a two-part tariff. Oi (1971) showed that an unrestricted monopoly solution under a non-discriminatory two-part tariff is characterized by a marginal

price lower than the unrestricted linear monopoly price and thus provides a higher efficiency. When a binding revenue cap is applied, the profit maximizing solution for the firm is to set the highest possible marginal price which still allows the firm to extract enough consumer surplus in order to exactly reach the revenue cap. Since some of the revenue in this situation is derived from consumer surplus, the firm reaches the revenue cap at a higher marginal price, and thus at a lower social efficiency level, than in the linear pricing situation.

A simple example can be used to provide better understanding of this phenomenon. Assume e.g. that the monopolistic firm has the cost function $C(Q) = 10000 + 60Q$, demand function $Q(P) = 200 - 0.5P$ and the fixed revenue cap $\overline{R} = 14400$. The unrestricted monopoly solution under linear pricing is then to use the price $P = 230$ which leads to the demand $Q = 85$. Monopoly profit will then be 4450. However, total revenue will be 19550, which violates the revenue cap. The profit maximizing solution under linear pricing given this revenue cap is to use the price $P = 306$, which leads to a profit of 1562. The Crew-Kleindorfer effect induces the monopolistic firm to raise its price.

Now assume that the monopolistic firm uses a two-part tariff. If the consumers are e.g. 10 homogeneous actors, the unrestricted profit maximizing tariff will be $F^* = 2890$ and $P^* = 60$, yielding a profit for the firm of 18900. But if the firm is subject to fixed revenue cap regulation with $\overline{R} = 14400$, the profit maximizing tariff will be $F = 160$ and $P = 320$ yielding a profit of 2000. The revenue cap combined with a two-part pricing structure provided a relatively larger efficiency loss than with simple linear pricing. Thus, the Crew & Kleindorfer (1996) efficiency problem with revenue caps exists and may even be reinforced by the fact that the monopolistic firm uses two-part pricing.

TWO-PART PRICING UNDER A HYBRID REVENUE CAP AS A FUNCTION OF QUANTITY

The single-period decision problem of the monopolistic firm under a hybrid revenue cap as a function of quantity becomes

$$\text{Maximize} \quad N \cdot F + P(Q) \cdot Q - C(Q) \tag{8}$$
$$\text{Subject to} \quad N \cdot F + P(Q) \cdot Q \leq \overline{R} + B \cdot Q.$$

where B is the variable part of the cap and Q is assumed to be the decision variable.

If the restriction is binding, i.e. if the revenue cap prevents an unrestricted monopoly solution (F^*, P^*), we can express the decision problem as

$$\text{Maximize} \quad N \cdot F + P(Q) \cdot Q - C(Q) \tag{9}$$
$$\text{Subject to} \quad N \cdot F = \overline{R} + B \cdot Q - P(Q) \cdot Q.$$

Substituting $N \cdot F$ with $\overline{R} + B \cdot Q - P \cdot Q$ in the objective function, the decision problem becomes unrestricted

$$\text{Maximize} \quad \overline{R} + B \cdot Q - P(Q) \cdot Q + P(Q) \cdot Q - C(Q) \tag{10}$$

which simplifies to

$$\text{Maximize} \quad \overline{R} + B \cdot Q - C(Q). \tag{11}$$

This means that the first order condition for a profit maximizing monopoly that uses two-part pricing under a binding linear hybrid cap is

$$\frac{\partial[\overline{R} + B \cdot Q - C(Q)]}{\partial Q} = 0 \tag{12}$$

which simplifies to

$$\frac{\partial C(Q)}{\partial Q} = B. \tag{13}$$

Thus, the regulated firm maximizes profit under a binding linear hybrid cap at the quantity Q_R where its marginal cost equals the slope of the linear hybrid cap. Hence, the profit maximizing behaviour of the regulated firm when the cap prevents unrestricted two-part monopoly pricing is (see figure 1) to set its marginal price to ensure that Q_R units are demanded and to extract all consumer surplus with the fixed fee F_R so that total revenue reaches the cap, i.e.

$$F_R = [\overline{R} + B \cdot Q_R - P(Q_R) \cdot Q_R] / N \tag{14}$$

or

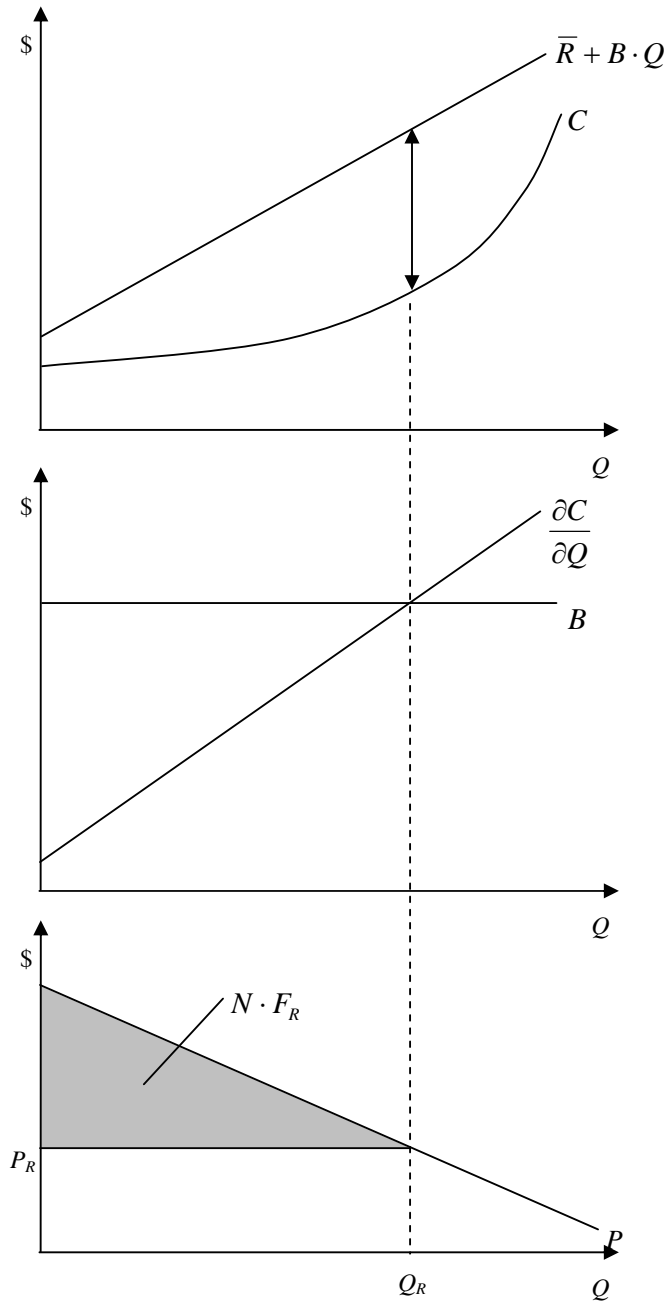$$F_R = [\overline{R} + Q_R (B - P(Q_R))] / N. \tag{15}$$

7

Figure 1: Profit maximizing two-part tariffs under a hybrid revenue cap

This means that a reduction of $\overline{R}$ means theoretically nothing when it comes to social efficiency, it only reduces the profit potential of the firm. Thus, model incentives are determined by B, the variable part of the hybrid cap.

Note that the condition for profit maximization in the above section is not possible to apply in the specific case where the total cost function of the firm is strictly linear, since the condition in (13) will not be valid for any quantity (or valid for all quantities in the even more specific case where $C(Q)$ is strictly linear and has the same slope as the hybrid cap). However, if the cap is binding in this case, the decision problem of the regulated firm is closely related to the situation where a monopoly using linear pricing is regulated by a binding fixed revenue cap (see Crew & Kleindorfer, 1996). Then, the monopoly will always choose the inefficient

solution Q_1 of the two possible solutions Q_1 and Q_2 (where $Q_1 < Q_2$) that provide the same total revenue, since the inefficient solution Q_1 by definition is characterised by a lower total cost.

To provide an incentive for the monopolist to choose the efficient one of the two possible solutions that exactly satisfies the revenue cap, total revenue must obviously be allowed to increase faster than total cost when quantity increases from Q_1 to Q_2. In other words, the value of the hybrid cap must increase faster than firm total cost when quantity increases. Thus, the formal condition that must be fulfilled in order to avoid the inefficient incentive described above is

$$\frac{\partial C(Q)}{\partial Q} < \frac{\partial [\overline{R} + B \cdot Q]}{\partial Q} \tag{16}$$

or simply

$$\frac{\partial C(Q)}{\partial Q} < B \,.^{4} \tag{17}$$

Thus, in the specific case of a strictly linear total cost function, the variable part of the hybrid cap cannot fall below firm marginal cost in order to avoid inefficient incentives. (If the variable part is equal to marginal cost, the firm will of course be indifferent between an inefficient and an efficient solution.) As in the general case, the level of the fixed part of the cap has nothing to do with social efficiency – it only determines the profit level of the regulated firm. The behaviour of firm in terms of output quantity and thus social efficiency is a function of only the variable part of the hybrid cap.

TWO-PART PRICING UNDER A PRICE-REVENUE HYBRID CAP

The single-period decision problem of the monopolistic firm under a price-revenue hybrid cap becomes

$$\text{Maximize} \quad P \cdot Q(P) - C(Q(P)) \tag{18}$$
$$\text{Subject to} \quad P \cdot Q(P) \leq \overline{R} - B \cdot P$$

where P is the decision variable. This hybrid cap was analyzed by Comnes *et al* (1995), who claimed that the slope of the hybrid cap must be steeper than the revenue curve in order to prevent the Crew-Kleindorfer effect. This is, as we shall see, not necessary.

If the restriction is binding, i.e. if the revenue cap prevents the unrestricted monopoly solution, we get the decision problem

---

[4] In the Crew and Kleindorfer (1996) base case, we have $\partial C(Q) > B$ by definition since they consider a fixed revenue cap. This explains why the monopolist always chooses an inefficient solution when it is regulated by a fixed revenue cap.

$$\text{Maximize} \quad P \cdot Q(P) - C(Q(P)) \tag{19}$$
$$\text{Subject to} \quad P \cdot Q(P) = \overline{R} - B \cdot P.$$

Substituting $P \cdot Q(P)$ with $\overline{R} - B \cdot P$ in the objective function, the decision problem can be expressed unrestrictedly

$$\text{Maximize} \quad \overline{R} - B \cdot P - C(Q(P)). \tag{20}$$

Thus, the first order condition for profit maximization under a binding hybrid cap of this kind is

$$\frac{\partial[\overline{R} - B \cdot P - C(Q(P))]}{\partial P} = 0 \tag{21}$$

or

$$\frac{\partial[\overline{R} - B \cdot P]}{\partial P} = \frac{\partial C(Q(P))}{\partial P} \tag{22}$$

or simply

$$-B = \frac{\partial C(Q(P))}{\partial P} \tag{23}$$

Hence, when the hybrid cap is binding, it is the relation between the slope of the cap and the slope of the cost function that determines the profit maximizing behaviour of the firm. The optimal solution to the decision problem is to choose the price that most closely corresponds to (23).

It may be possible to provide a better understanding for this by graphical means. If the slope of the hybrid cap is steep enough, it will intersect with total revenue only once for a unique price and thereby behave like a price cap. In this case, the regulation provides no incentive to raise price since this unique price is also the profit maximizing price if the cap is binding. In figure 2, we see such a cap.
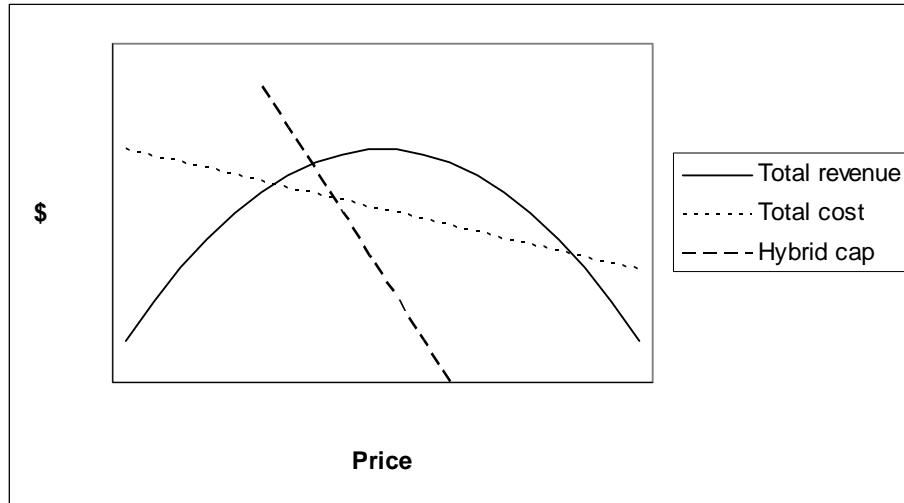
Figure 2: A hybrid cap that is steep enough to work like a price cap

If the slope of the cap is less steep, it will intersect with total revenue twice, which is a necessary but not a sufficient condition for the Crew-Kleindorfer effect to appear. Such a cap is illustrated in figure 3.
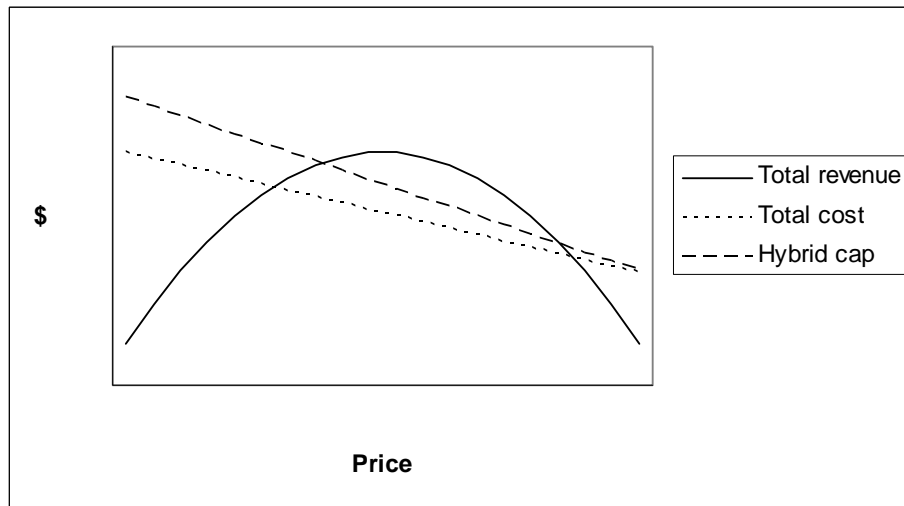

Figure 3: A hybrid cap where the slope prevents the Crew-Kleindorfer effect

The sufficient condition for the Crew-Kleindorfer effect in this setting, however, is that the higher price intersection between the hybrid cap and total revenue must correspond with a higher profit than the lower price intersection. In other words, for the Crew-Kleindorfer effect to appear when the hybrid cap intersects with total revenue twice, moving from the higher price to the lower price must correspond with a decrease in profit. Since it is the slope of the hybrid cap that determines the change in revenue between these prices, it must be the change in cost that determines if the higher or the lower price is profit maximizing. Thus, the sufficient condition to avoid the Crew-Kleindorfer effect in this context is that the slope of the hybrid cap must be more negative than the average slope of the cost function for the price interval where the cap is binding, i.e.

11

$$-B < \frac{C(Q(p_H)) - C(Q(p_L))}{p_H - p_L} \tag{24}$$

where p_H is the higher price and p_L the lower price at which the hybrid cap intersects with total revenue. Note that in the case of a linear cost function, as in the figures above, (24) can be expressed as

$$-B < \frac{\partial C(Q(P))}{\partial P} . \tag{25}$$

In figure 4, the slope of the hybrid cap is not steep enough to prevent the Crew-Kleindorfer effect. The firm will choose a price that is higher than the unregulated monopoly price.
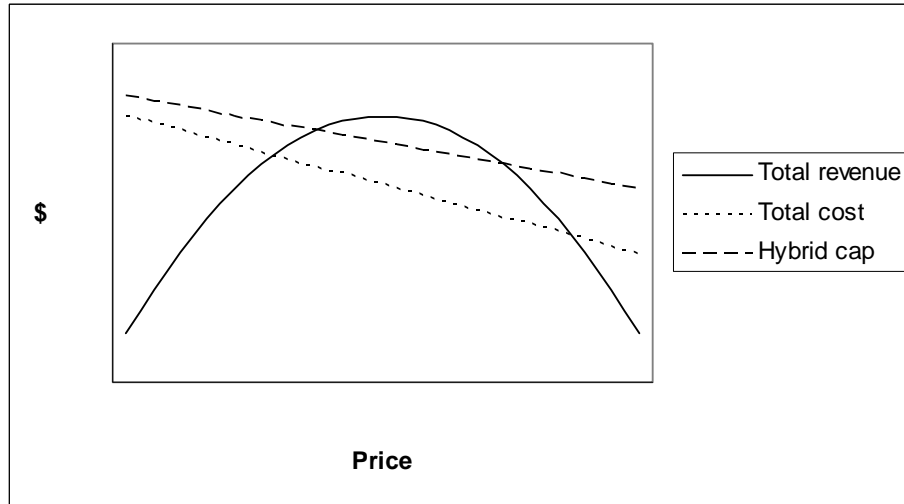

Figure 4: A hybrid cap where the slope does not prevent the Crew-Kleindorfer effect

To sum up, the hybrid cap must be steep enough either to intersect with total revenue only once, or, if it intersects twice, to provide a higher profit at the lower price intersection. Hence, the Comnes *et al* (1995) proposal that the slope of the hybrid cap needs to be more negative than the revenue curve in order to prevent the Crew-Kleindorfer effect is not true.[5]

SUMMARY AND CONCLUSIONS

In this paper, we have deepened the theoretical understanding of revenue capping as regulation regime in several directions. Firstly, "the Crew-Kleindorfer effect", i.e. the fact that a standard monopolist regulated by a fixed revenue cap will raise its price above the unregulated monopoly level, has been proven to be robust to two-part pricing. Secondly, when regulation of a two-part pricing monopolist is based on a hybrid revenue cap defined as

---

[5] This is essential because the steeper the hybrid cap becomes, the more it works like a price cap, and the less it provides incentives for Demand-Side Management (DSM). Such incentives are often very desirable in applied utility regulation, e.g. in the electricity industry. Thus, price-revenue hybrid cap regulation in practice may include a trade-off: The cap needs to be flat enough to provide large incentives for DSM, but steep enough to prevent the Crew-Kleindorfer effect.

a linear function of quantity, we have seen that it is the slope of the cap that determines its incentives for efficiency. The intercept only affects the profit level of the firm. This also holds if the cap is defined as a hybrid price-revenue cap. The general conclusion of this is that the slope of the hybrid cap needs to be steeper that the slope of the firm's cost function in order to prevent the Crew-Kleindorfer effect within a two-part pricing context. It does not, however, need to be steep enough to work like a price cap, as suggested elsewhere.

This highlights the main practical problem of revenue capping: The regulator needs a lot of detailed information about the firm to be able to determine the cap parameters in a way that ensures efficient incentives. Sound regulation should of course not rely on any kind of micromanagement. One way of handling this problem may be to create dynamic revenue caps where the cap parameters in one period are determined by observable actions of the firm in the previous period(s).[6] This could reduce regulatory burden and create better regulation while reducing the information requirement. Thus, regulators who use revenue caps may have a lot to learn from the design of anonymous regulatory mechanisms and other areas.

---

[6] Assume, for example, that the cap in period $t+1$ is defined as $\overline{R}_{t+1} - B_{t+1} \cdot P_{t+1}$ where $B_{t+1} = B_t \cdot P_t / P_{t-1}$. Then, the slope of the cap will be increased in period $t+1$ if $P_t > P_{t-1}$ and decreased if $P_t < P_{t-1}$. If this very simple regulation regime converges under a nonzero discount rate, it actually converges to a situation where B equals marginal cost. (Proof is available from the author.) However, conversion depends here on the initial parameter determination in period 0 and the cost and revenue conditions of the firm. In the general case, this simple model does not converge.

References

Comnes, G. A, Stoft, S., Greene, N. & Hill, L. J. (1995), "*Performance-Based Ratemaking for Electric Utilities: Review of Plans and Analysis of Economic and Resource-Planning Issues*", Lawrence Berkeley National Laboratory, University of California.

Crew, M. A. & Kleindorfer, P. R. (1996), "Incentive Regulation in the United Kingdom and the United States: Some Lessons", *Journal of Regulatory Economics*, 9:3, 211-225.

Jamasb, T. & Pollitt, M. (2001), "Benchmarking and Regulation: International Electricity Experience", *Utilities Policy*, 9:3, 107-130.

Oi, W. (1971), "A Disneyland Dilemma", *Quarterly Journal of Economics*, 85, 77-96.

Viljainen, S., Tahvanainen, K., Lassila, J., Honkapuro, S. & Partanen, J. (2004), "*Regulation of electricity distribution business*", Nordic Distribution and Asset Management Conference 2004.