Appendix: Price Caps

Recently, regulators in various settings have begun to utilize "price caps" as a means of regulating firms. Acton and Vogelsang (1989) specify four characteristics of price-cap regulation:

1. The regulator sets a price, called the price cap. The regulated firm can set a price below or equal to this cap, and it is allowed to retain whatever profits it earns at that price.

2. In multi-output situations, the regulator might define an aggregate cap for a basket of related products. This aggregate cap takes the form of a price index or a weighted average of prices. The firm is allowed to change prices for the goods (raising some and lowering others) as long as the index, or weighted average of prices, does not rise.

3. The regulator might specify that the price cap will be adjusted over time by a preannounced adjustment factor that is exogenous to the firm. For example, the cap may be tied to an index of input prices.

4. At longer intervals, the price cap is reviewed by the regulator and possibly changed. This review is expected to consider the cost, demand, and profit conditions of the firm.

This form of regulation has been adopted for AT&T's interstate service, for telecommunication services in California, and for natural gas and electricity in Great Britain. It is also being considered for many other settings.

The emergence of this new form of regulation provides an excellent opportunity to apply the concepts described in this book. To what extent and in what ways can we expect price-cap regulation to attain the optimal price, output, and input levels? Even if full optimality is not attained, can we expect price-cap regulation to be preferable to traditional rate-of-return regulation, which the analysis in chapters 1 and 3 indicates induces the firm to produce inefficiently?

We will address these questions within a sequence of increasingly complex settings. Starting with a fully static situation in which demand, costs, and the price cap do not change over time, we will then move to a situation in which costs and demand change and the regulator adjusts the price cap on the basis, at least partially, of the firm's profits in previous periods. This method of analysis allows a clear identification of the benefits and difficulties of price-cap regulation. The conclusions of this analysis can be summarized as follows.

1. If the price cap is fixed (or if it changes in a way that is exogenous to the firm), the firm will produce with the cost-minimizing input mix, invest in cost-effective innovation, and adjust optimally to changes in cost. The reason is clear: The firm is allowed to retain as profit any cost reductions that it achieves and, consequently, it will choose to produce efficiently. Total surplus therefore increases relative to a situation in which the firm does not cost-minimize (such as under rate-of-return regulation). However, unless the previous regulation resulted in higher prices than the firm would have charged without any regulation, the increased surplus accrues entirely to the firm: consumers do not benefit from the production efficiency.

2. If the firm produces more than one good and the price cap is expressed as an Laspeyres index of the prices for all the outputs, then the firm will adjust outputs and prices in a way that increases profit without decreasing consumer surplus. Total surplus will therefore rise. For marginal changes in prices, the entire gain will accrue to the firm with no gain for consumers. However, for larger-than-marginal changes, consumers will also benefit.

3. Over time, the regulator will review the price cap and adjust it up or down based on the profits of the firm under the existing cap and other factors. This periodic review is the heart of the issue regarding price caps. It provides the means by which consumers also benefit: the regulator can lower the price cap to require that at least some of the reduced costs be passed on to consumers. However, the existence of the review also introduces the possibility of strategic behavior by the firm, which could prevent the attainment of cost reductions. The form of the reviews has not been specified by regulators. Depending on what the firm expects the regulator to do in the review, the firm could be induced to incur higher costs than necessary as a strategic move, similar to that discussed in chapter 5 regarding the Vogelsang-Finsinger mechanism. Furthermore, if, as can perhaps be reasonably expected, the review of price caps is conducted like the price reviews under rate-of-return regulation, then the distinction blurs

between price-cap regulation and rate-of-return regulation. More work is needed on the form and timing of the price-cap reviews before an definitive statement can be made about whether, or the extent to which, price caps induce optimality or improve upon traditional rate-of-return regulation.

A.1 One-Output Firm in a Fully Static World

Suppose the demand and cost functions facing the firm are fixed and that the regulator has established a price cap, denoted P_{cap}. Under the terms of the regulation, the firm can choose any price equal to or below P_{cap} and can retain any profits it earns at that price. The behavior of the firm in this situation can be represented with the concepts developed in chapters 1 and 2. Figure A.1 gives, in the space of input combinations, the expansion path and the zero-profit contour of the firm. Point M is the top of the profit hill and as such is the point that the firm would choose if it were not regulated. Denote by Q_{cap} the level of output demanded at the price cap P_{cap}. The firm, by the terms of the regulation, can choose any price at or below P_{cap}, which means that it can choose any output at or above Q_{cap}. That is, the firm is constrained to choose an input combination on or above the isoquant for Q_{cap}.

There are two possibilities: the price cap is either below or above the price that the firm would charge if it were not regulated. Panel (a) depicts the more standard situation in which the price cap is below the unconstrained profit-maximizing price. Since P_{cap} is less than the price at M, Q_{cap} is above the output at M, meaning that the isoquant for Q_{cap} passes above M. The firm will choose point C, the point on

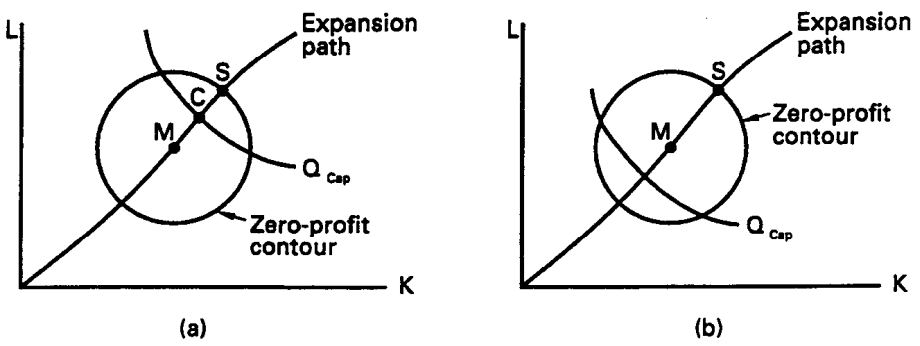

Figure A.1
Choice of firm under price caps

the expansion path associated with output level Q_{cap}. (The reason should be clear: at any output level, profits are higher on the expansion path than off, and profits decrease as the firm moves along the expansion path down the profit hill from M, the top.) The firm will charge the price cap, sell the output demanded at that price, and produce with the least-cost input combination. Essentially, the firm cost-minimizes because it is able to retain as profit any cost reductions that it attains. And the firm prices at the cap because this price allows it to be as close as possible to the unregulated profit-maximizing price.

Panel (b) depicts the other possibility: the price cap is above the price that the firm would choose if it were not regulated. The firm chooses point M, the top of the profit hill. The firm prices below the cap, sells more output than would be produced at P_{cap}, and uses the cost-minimizing inputs. In this case, price-cap regulation is equivalent to no regulation at all.

In both cases, the firm cost-minimizes but produces less output than is optimal. Recall that if the firm cannot be subsidized, then the optimal (second-best) output is that produced at point S, where the expansion path intersects the zero-profit contour. Ideally the regulator could set the cap at the optimal price (that is, set the cap such that the isoquant for Q_{cap} intersects the expansion path at S such that the firm chooses S). However, without knowing the cost and demand conditions facing the firm, the regulator cannot determine the optimal price. More fundamentally, if the regulator knew the optimal price, the regulator could simply mandate this price. If the cap is set so low that the firm loses money (that is, if the isoquant for Q_{cap} falls outside the zero-profit contour), the firm would go out of business eventually. Because the regulator cannot identify the optimal price, and a cap that allows only negative profits is infeasible, price will inevitably be higher than optimal and output will be lower than optimal. Stated succinctly: under price-cap regulation, the firm will choose the cost-minimizing inputs, but will produce less output than is optimal. The firm will price below the cap only if the cap is so high that price-cap regulation is equivalent to no regulation.

Consider now a comparison of ROR and price-cap regulation. In particular, suppose that ROR regulation has been applied in the past, and that the regulator decides to switch to price-cap regulation. It can be shown that total surplus will rise when ROR regulation is replaced with price caps. However, except in rare circumstances, all of the benefits of switching to price caps will accrue to the firm in the form

of extra profits; consumers will not benefit from the change in regulation.

Under ROR regulation, price is presumably above the optimal level due to cost inefficiencies. Suppose the regulator sets the cap at the price the firm has charged under ROR regulation. Under price-cap regulation, the firm will eliminate the cost inefficiencies, retaining as profit the cost savings. The effect on price and output depends on the relation of the price under ROR regulation to the unconstrained profit-maximizing price. Two situations are possible, as depicted in the two panels of figure A.2. In both panels, the constraint curve is shown for the firm under ROR regulation with the fair rate of return exceeding the cost of capital. (For graphical simplicity, the zero-profit contour is not shown; however, the zero-profit contour is known to encircle the constraint curve because all points on the constraint curve result in positive profit.) As described in chapter 1, the firm under ROR regulation chooses point R. The firm produces output Q_r at this point and charges price P_r. When the regulator switches to price caps, the cap is set equal to P_r: the firm is constrained to produce at any point on or above the isoquant that goes through point R. The two panels differ in whether this isoquant passes above or below M.

In panel (a), the isoquant intersects the expansion path above M. The firm chooses point C under price-cap regulation. That is, when ROR Regulation is replaced with price-cap regulation, the firm moves from R to C. Output and hence price do not change. However, the firm reduces its costs, producing its output with the cost-minimizing inputs.

Note that only the firm benefits from the switch from ROR to price-

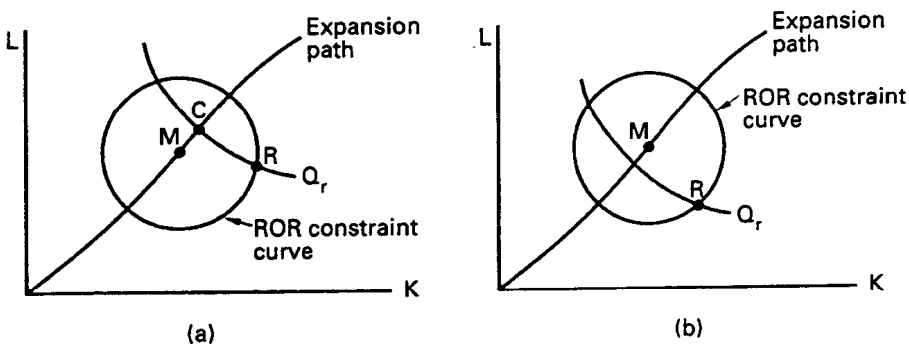

Figure A.2
Comparison of ROR regulation with price caps

cap regulation in this case. The firm reduces its costs, and yet none of these savings are passed on to consumers in the form of a lower price. Because price and output are unchanged, consumer surplus is constant. Profits rise by the amount by which the firm is able to reduce its costs (that is, by the amount of inefficiency it incurred under ROR regulation).

There is one circumstance in which consumers will benefit, depicted in panel (b). In this case, the price under ROR regulation exceeds the price that the firm would charge if it were unregulated. Recall from chapter 1 that this situation is possible: ROR regulation can, depending on the shape of the profit hill, induce the firm to produce less output and charge a higher price than if it were not regulated. In this case, the firm will choose point *M*, the unconstrained profit maximum, when ROR regulation is replaced with price-cap regulation. That is, the firm will lower its price from the level under ROR regulation to the unconstrained profit maximum. The firm earns more profit, but consumers also benefit because price is dropped.

Note that the situation in panel (b) arises only when ROR regulation is worse than no regulation (since it results in a price that is higher than the firm would charge without regulation) and price-cap regulation is equivalent to no regulation. The conclusions regarding a switch to price-cap regulation in a static world can therefore be summarized as follows. Total surplus necessarily rises when ROR regulation is replaced with price caps, because profits rise and consumer surplus does not fall. However, consumers benefit only if ROR regulation were worse than no regulation, in which case the switch to price-cap regulation is equivalent to eliminating regulation.

One might argue that the regulator could set the price cap below the price that the firm charged under ROR regulation, thereby assuring consumers some of the gains from the cost reductions. However, the regulator would only do this if it knew that the firm were engaged in some inefficiencies under ROR regulation. If the regulator knew that the firm's costs were too high, it would have disallowed these costs when the firm was under ROR regulation, thereby forcing the firm's price under ROR regulation to be lower. Since the regulator approved the firm's price under ROR regulation, the regulator must not know that the firm is not cost minimizing. (The regulator presumably suspects that the firm is not cost minimizing under ROR regulation—that is why the regulator shifts to price-cap regulation. However, the regulator does not know for sure whether, or the extent to which,

costs can be reduced and therefore cannot require that the firm charge a lower price.)

A.2 Multi-Output Firm in a Fully Static World

If there is more than one output, the regulator has a choice of how to apply price-cap regulation. A separate cap could be established for each product. In this case, the analysis for each good is essentially the same as in the one-output case. In many settings, however, regulators have applied an aggregate price cap on a set of interrelated products. The aggregate cap takes the form of an index, or a weighted average of prices. The firm is allowed to raise the price for some goods as long as the prices for other goods are lowered sufficiently that the price index does not rise. As we show below, the use of an appropriately defined aggregate cap enables the firm to change prices in a way that increases its profits without decreasing, and perhaps even increasing, consumer surplus. Total surplus can therefore rise under price-cap regulation because of a more efficient output mix in addition to the cost reductions described in the one-output situation.

Our analysis is motivated by the concepts advanced by Brennan (1989). Consider a two-good situation; generalization to more goods is straightforward. Let P_1^0 and P_2^0, and Q_1^0 and Q_2^0 denote the prices and quantities of the two goods prior to the imposition of price-cap regulation. Suppose the regulator placed the following constraint on the firm. The firm is allowed to charge any prices P_1 and P_2 that satisfy the following inequality:

$$Q_1^0 P_1 + Q_2^0 P_2 \leq Q_1^0 P_1^0 + Q_2^0 P_2^0.$$

That is, the firm can charge any prices as long as the prices, when multiplied by the quantities of the goods sold prior to the price cap-regulation, do not exceed the revenues that the firm obtained before the price caps.

This constraint can be expressed as an index. Rearranging the inequality, we have

$$(Q_1^0 P_1 + Q_2^0 P_2) \,/\, (Q_1^0 P_1^0 + Q_2^0 P_2^0) \leq 1.$$

The left-hand side is a Laspeyres index of prices, with the prices in the period prior to price caps taken as the base. The firm is allowed to change prices as long as the price index does not rise above one—that is, above the base level.

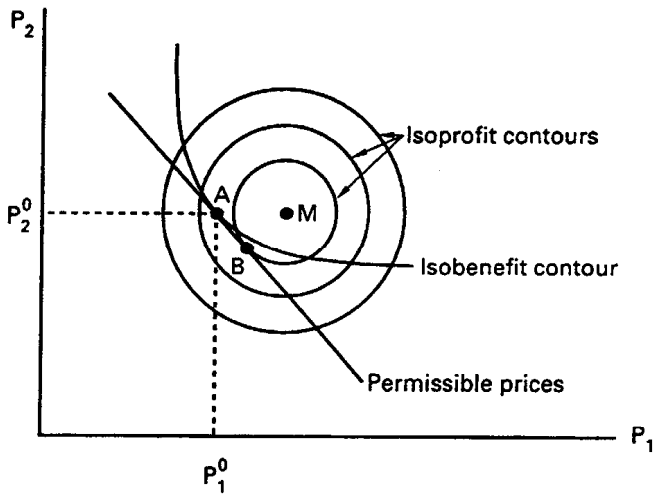

Figure A.3
Multi-output firm under price cap index

Using the concepts from chapter 5, we can represent this constraint graphically. Figure A.3 is a graph of price combinations. Originally prices are at A. The isobenefit contour through A gives the price combinations that result in the same consumer surplus as at A. Consumer surplus is higher for price combinations that are below this isobenefit contour (i.e., at lower prices). As discussed in chapter 5, the slope of the isobenefit contour at A is the ratio of outputs sold at those prices; i.e., $-(Q_1^0/Q_2^0)$. The tangency line is depicted in the graph; we can show that this tangency line is the set of permissible prices available to the firm.

Rearranging the inequality, we have

$$P_2 \leq k - (Q_1^0/Q_2^0)P_1,$$

where k equals $(Q_1^0P_1^0 + Q_2^0P_2^0/Q_2^0$. Considering the equality part only, we have an equation for a line. This line necessarily passes through A, since the firm can continue charging its original prices. The line has a slope of $-Q_1^0/Q_2^0$, which we have already stated is the slope of the isobenefit contour through A. The set of permissible prices therefore consists of the tangency line for the isobenefit contour through A, plus all price combinations below this line.

From the set of permissible prices, the firm chooses the price combination that provides the greatest profit. Note, however, that all the permissible prices provide at least as great consumer surplus as at A:

since the line of permissible prices is tangent to the original isobenefit contour, all permissible prices are on or below this contour. In the graph, the firm makes the most profit at B and hence choses these prices: the firm's profits increase, and consumer surplus also increases. In general, profits increase for any movement away from the original prices (that is, whenever the original prices do not provide more profit than any other permissible price). For marginal changes in prices (that is, infinitesimally small changes), consumer surplus does not increase: consumers stay on the same isobenefit contour at a point close to A. However, for larger changes, consumers, as well as the firm, benefit from the price changes.

The reason this form of multiproduct price cap operates to the benefit of consumers and the firm alike is essentially the same as the reason behind V-F regulation. The firm is required to trade off price increases and decreases at a rate that maintains or increases consumer surplus. In particular, the firm can raise the price of good 1 by one dollar only if it lowers the price of good two by at least Q_1^0/Q_2^0 dollars. Yet consumer surplus is unchanged for marginal price changes at this rate of trade-off (because the isobenefit contour has this slope) and increases for larger-than-marginal changes at this rate.

The results in this section are highly dependent on the particular form of the aggregate price cap placed on the firm. The use of other indexes for price-cap regulation with several products may allow profits to increase at the expense of consumer surplus. In fact, at an extreme, total surplus may decrease. The key to an appropriate index is to require price changes at the ratio of the original output levels.

A.3 Demand, Costs, and the Price-Cap Change over Time

Over time, costs and demand change. If the price cap does not change, the firm could end up either losing money or earning very large profits, neither of which can be feasibly or politically maintained. To account for these changes, the price cap must be revised. The crucial issue is how the cap changes over time.

In many settings the regulator has specified that the price cap will change in some predetermined way. For example, the price cap might be tied to a wholesale price index or an index of wages in the area, such that general increases in input prices translate into a higher cap. This procedure protects the firm from cost changes that are beyond its control.

If changes in the price cap are determined on the basis of factors that are outside the control of the firm, then the analysis of firm behavior is essentially the same as in a static situation. The firm will cost-minimize, because doing so increases its profits. As well as utilizing the efficient input mix and not wasting at each point in time, the firm will cost-minimize over time. In particular, the firm will adjust optimally to changes in input prices. The firm will also adopt cost-reducing new technologies and engage in research and development that can lead to such technologies whenever the present value of the expected cost savings to the firm exceeds the cost of the investment.[1] (Note, however, that the optimal amount of innovation will not necessarily be attained. First, the firm will not invest optimally in product innovations because the price cap prevents the firm from raising its price to capture the benefits to consumers of improved products. Second, the cost savings that result from new technologies depend on the output level of the firm. Because the firm produces less output than is optimal, technologies that would be cost-effective at the optimal output but not at the firm's output will not be adopted.)

The firm will price at the cap at each point in time unless the cap exceeds the unconstrained profit-maximizing price. Because the unconstrained profit-maximizing price changes over time when costs and demand change, the firm may charge at the cap in some periods and below the cap in others. As long as the firm earns positive profit, the firm's output is less than optimal. If costs and demand change in a way that preclude the firm from earning a positive profit at the prevailing cap, then the firm will request a formal review. In general, therefore, output will be lower than optimal and price higher than optimal.

It is expected, and usually specified by the regulator, that a review of the price cap will be conducted periodically, even without the request of the firm. In fact, Acton and Vogelsang (1989) include this review as their fourth aspect of price-cap regulation. The form of this review—that is, the way in which the cap will be revised—has not

1. Cabral and Riordan (1989) show that if the cap that is prespecified for future periods is too low from the firm's perspective, the firm will *not* invest in cost-saving technologies. The argument hinges on the fact that the firm can, at any time, request a formal review as under ROR regulation rather than submit to a cap. If the firm knows that the future cap will be so low that it will lose money even with the cost-saving technologies, it has no incentive to invest in these technologies. The firm is better off by simply waiting until the cap is too low and then requesting a rate review, which it knows will result in a price that will cover its (inefficiently high) costs.

been established in the settings in which price caps have been imposed. However, it is reasonable to assume that the regulator will consider the profits the firm has earned in the past when determining the new cap. More to the point, it is reasonable to assume that the firm expects the regulator to act in this way.

When the price cap is changed on the basis of the firm's profits rather than only on exogenous factors, the analysis of the firm's behavior becomes far more complex. The possibility of strategic behavior must be considered, and the outcome can differ drastically from that derived above in a static setting or with exogenously determined changes in the cap. In particular, it is possible that the firm will waste so as to convince the regulator to allow a higher cap. Whether or not the firm engages in suboptimal strategic behavior depends on how the new cap is determined—or more precisely, on how the firm thinks the regulator will determine the new cap.

Numerous situations can be identified in which the firm has an incentive, for strategic reasons, not to cost-minimize. An incentive such as that identified in chapter 5 on V-F regulation might operate. For example, suppose the firm is currently under rate-of-return regulation and knows that the regulator will be switching to price-cap regulation. Under ROR regulation, the firm is allowed to recover its costs through higher prices. The firm would have an incentive to waste while under ROR regulation as a means of obtaining a higher cap when price-cap regulation is imposed. The waste would not reduce the firm's profits under ROR regulation because, by the terms of ROR regulation, its costs can be recouped. However, its profits would be higher after the switch to price caps. Total surplus would be reduced in two ways: (i) the waste of the firm in anticipation of the caps and (ii) the loss of consumer surplus (or more precisely, the deadweight loss) due to the higher price after price-cap regulation is imposed.

Consider another possibility, which could occur even after price-cap regulation has been imposed for a while. Suppose the regulator reviews the cap every three years and changes the cap on the basis of the firm's profits in the year immediately prior to the review (reflecting, perhaps, the idea that the most recent year is the most relevant). The firm would cost-minimize in the first and second years between reviews, but would have an incentive to waste in the third year. The reduced profit in this third year would translate into a higher cap than if the firm had not wasted. And this higher cap allows the firm to earn higher profit for the next *three* years. Unless the firm's

discount rate is very high, the present value of current and future profits will be higher if the firm wastes in the year before a review. This result can be obtained, though to a lesser extent, if the regulator considers all three years' profits but places more importance on the most recent year.

The review might be conducted like the price review under rate-of-return regulation. In this case, the distinction between price-cap and ROR regulation blurs. If reviews under price-cap regulation are conducted with the same frequency as under ROR regulation, the two forms of regulation become the same and the observed inefficiencies that have arisen under ROR regulation can be expected to carry over to price-cap regulation. The issue reduces therefore to the timing of the reviews. The more frequent the review, the more price-cap regulation will induce the inefficiencies of rate-of-return regulation. The less frequent the review, the greater the incentive of the firm to cost-minimize between reviews (absent any strategic behavior of the type described above).

These examples are not meant to imply that the firm will necessarily engage in suboptimal strategic behavior. It is possible that the form of the periodic reviews is such that the firm behaves the same as if the price caps were fixed or determined exogenously. The form that the review takes (or, more precisely, that the firm expects the review to take) is crucial to the behavior of the firm and hence to the advisability of price-cap regulation. Yet the review process has not been specified. And even if the regulator specified the procedure, the firm would not necessarily trust the regulator to adhere to the stated procedure when the time came. The issue of price caps rests, therefore, on what the firm believes the regulator will consider when adjusting the cap. Much more work is needed on this important area of investigation before a definitive statement can be made about whether, or the extent to which, price-cap regulation induces optimality or improves upon traditional ROR regulation.