5 The Vogelsang-Finsinger Mechanism

5.1 Introduction

Vogelsang and Finsinger (1979) have suggested a regulatory mechanism that induces the regulated firm to move, over time, to Ramsey prices and outputs. This mechanism is intriguing for several reasons. First, it is very simple, both conceptually and in implementation. Given the complex nature of the Ramsey rule (namely, that the percent deviation of price from marginal cost multiplied by the "net" elasticity is the same for all goods), the simplicity of the mechanism that attains these prices is surprising. Second, the mechanism explicitly accounts for the asymmetry in information that necessarily exists in regulatory settings—namely, that the firm knows more about its operations than does the regulator. The regulator using the Vogelsang and Finsinger (or V-F) mechanism need not know the firm's demand or cost functions, which the firm itself knows. Nor does the regulator need to know beforehand what the Ramsey prices and outputs are. The mechanism will induce the firm to move to them even without the regulator knowing what they are. A large literature has arisen on the issue of optimal regulation under different forms of asymmetric information. The V-F model serves as an excellent introduction to this literature. Third, the mechanism is dynamic, in that actions of the firm in one time period determine the options that the regulator allows the firm in the following period. More precisely, the information revealed to the regulator through the firm's actions in one period is used by the regulator to constrain the firm in the following period. This sequence of revealed information followed by a constraint based on that information leads, in the situation described by Vogelsang and Finsinger, to the Ramsey outcome. A large literature has arisen on dynamic regulation of this kind. The V-F model, and especially

comments on the model by Sappington (1980), elucidate some of the fundamental issues within this literature—including the important issue of when the firm will have an incentive to "reveal" incorrect information.

Section 5.2 describes the mechanism proposed by Vogelsang and Finsinger. Sections 5.3 and 5.4 demonstrate that the mechanism induces the regulated firm to choose, in equilibrium, the second-best prices in one-good and multi-good situations, respectively. Section 5.5 delineates conditions under which the firm has an incentive, under this mechanism, to inflate its actual or reported costs.

The chapter can be summarized as follows. Under V-F regulation, the firm's prices, output, and costs in one period determine the prices that the regulator allows the firm to charge in the next period. In particular, the firm is able to charge any prices in the next period, as long as those prices, when multiplied by the firm's output in the *current* period, do not exceed the firm's costs in the *current* period. Successive application of this constraint over several periods results in second-best pricing. In particular, the one-output firm prices at average cost in equilibrium, and the multi-output firm charges Ramsey prices.

Under V-F regulation, the firm might have an incentive to report costs that are higher than those actually incurred (i.e., misreport costs) and/or to incur costs in excess of the minimum required for production (i.e., waste). In both cases, the firm is allowed to charge higher prices in the next period because its reported costs are higher in the current period. Both of these possibilities are explored.

To prevent, or reduce, misreporting, the regulator can audit the firm's cost. If the regulator is able (legally) to levy a sufficiently high penalty for misreporting, the firm can be induced to report costs truthfully. The regulator performs audits very infrequently, but does not tell the firm when the audit will occur. To avoid risking the large penalty, the firm reports truthfully in *each* period. Furthermore, the cost of auditing is essentially zero because very few audits are actually performed. (With a high enough penalty, the frequency of audits can be reduced nearly to zero and still induce the firm to be honest.) If audits are not able perfectly to determine the true costs of the firm, there is a chance that the firm will be levied a penalty even if it reports truthfully (i.e., if the audit "finds," incorrectly, that costs are lower than reported). However, this chance is nearly zero because the fre-

quency of audits is nearly zero and the probability of an incorrect penalty, given an audit, is low because the firm reports truthfully.

If the maximum size of the penalty is limited, the frequency of audits cannot be nearly zero and still induce honesty. A higher frequency of audits is required. In this case, the cost of auditing and the chance of an incorrect penalty are not nearly zero, but rather are non-negligibly positive.

When the firm is audited appropriately, it will not misreport costs. However, it might waste. We show that, if the firm knows beforehand that V-F regulation will be imposed, it might waste in one period so as to be allowed higher prices in the next period. This waste might occur for several periods. However, the firm will not waste in equilibrium. Furthermore, if the firm does *not* know beforehand that V-F regulation will be imposed, it will not waste either in the movement toward, or in, equilibrium.

5.2 The V-F Mechanism

The process can be started in any time period, with a time period being a year, month, or whatever is feasible. In the first period of regulation, the regulated firm charges some price for each good, sells some quantity of each good, and incurs some cost in producing these outputs. The regulator observes these prices, outputs, and costs. The regulator uses this information to constrain the firm in its choice of prices in the second period. In particular, the regulator tells the firm that the firm can charge whatever prices it wants in the next period—so long as the next period's prices, when multiplied by the first period's quantities, do not exceed the first period's costs. For example, suppose the firm produces only one good and in the first period prices the good at \$10, sells 150 units, and incurs costs of \$1,200 (thereby earning a profit of \$300 in the current period). The regulator tells the firm that it can charge any price in the second period that, when multiplied by 150, does not exceed \$1,200. Clearly, the first-period price is not permissible for the next period: \$10 times 150 equals \$1,500, which exceeds \$1,200. However, \$6 is permissible, because \$6 times 150 is \$900, which is less than \$1,200. A price of \$8 dollars is also permissible for the next period, because \$8 times 150 is \$1,200, which does not *exceed* the first period's costs of \$1,200.

In the second period, after the firm has chosen its new prices within

the regulator's constraint, the firm will sell a different quantity of each good A and \$6 for good B, sells 100 units of good A and 50 units of good B, and incurs costs of \$600. In the second period, the firm would be allowed to charge, say, \$3 for good A and \$5 for good B [because $(\$3 \cdot 100) + (\$5 \cdot 50) = \$550$, which is less than \$600]. Prices of \$3.25 pose our one-output firm chose a price of \$8, which was permissible as described above. At this price, suppose the firm sells 180 units (because a lower price increases quantity demanded) and incurs costs of \$1,400 in producing this output (because costs rise with quantity), earning a profit of \$40. The regulator observes the price, quantity, and costs and tells the firm that now, in the third period, it can only charge a price that, when multiplied by 180, does not exceed \$1,400. A price of \$7 is permissible in the third period, as is a price of \$7.50 or \$7.77; however, the price of \$8 is no longer permissible.

This process continues from period to period, eventually inducing the firm (as shown below) to charge the second-best price.

The mechanism operates analogously with two or more goods. Suppose a two-output firm in the first period charges a price of \$4 for good A and \$6 for good B, sells 100 units of good A and 50 units of good B, and incurs costs of \$600. In the second period, the firm would be allowed to charge, say, \$3 for good A and \$5 for good B [because $(\$3 \cdot 100) + (\$5 \cdot 50) = \$550$, which is less than \$600]. Prices of \$3.25 and \$5.50 would also be permissible, as would many other price combinations. The current prices of \$4 and \$6, respectively, would, however, not be permissible [because $(\$4 \times 100) + (\$6 \times 50) = \$700 > \$600$]. After the firm chose its prices for the second period, the regulator would observe the quantities it sells at these new prices and its costs for producing these outputs. The regulator would then require that for the third period, the firm adjust its prices to meet the constraint based on these second-period quantities and costs.

The regulatory mechanism can be stated algebraically. Given the output and cost of the firm in any period t, the firm must choose prices in the next period, period $t + 1$, that satisfy the following constraint. For one-output firms, the constraint is

$$P^{t+1}Q^t \leq C^t. \tag{5.1}$$

That is, the price in the next period, P^{t+1}, when multiplied by current-period output, Q^t, cannot exceed current-period costs, C^t. For two-output firms

$$P_a^{t+1}Q_a^t + P_b^{t+1}Q_b^t \leq C^t, \tag{5.2}$$

where a and b denote the two goods. The constraint is defined analogously for firms with more than two outputs. In each case the constraint is applied repeatedly, period after period. Eventually, as shown below, the constraint induces the firm to choose the second-best prices.

Before demonstrating the outcome of this process, two notes are important. First, the quantity $P^{t+1}Q^t$ (and its counterpart in the two-good case) appears to be a revenue calculation because it takes the form of price times quantity. However, this product is not revenues in either period t or in period $t+1$. In period t, revenue is period t's price times period t's quantity (that is, P^tQ^t); and revenue in period $t+1$ is price in period $t+1$ times the quantity that is demanded in period $t+1$ (that is, $P^{t+1}Q^{t+1}$). The quantity $P^{t+1}Q^t$ is price in period $t+1$ multiplied by the quantity in period t, which is neither period's revenues. Rather, it is a calculation that the regulator defines in order to implement the regulatory mechanism; because it takes the form of price times quantity (though with price taken from one period and quantity from another) it is often called "pseudorevenue" for period $t+1$. It is the revenues that the firm would earn in the next period under a new price if demand remained the same as under the old price. However, with price-sensitive consumers, a change in price from one period to another induces a change in quantity demanded, such that pseudorevenue in period $t+1$ does not generally equal the actual revenues in period $t+1$.

The second aspect of the mechanism that requires note is the informational requirements of the mechanism. To specify the constraint on the firm, the regulator must, in each period, observe the prices, quantities, and costs of the firm in that period. This information can reasonably be assumed to be available to the regulator, because the prices and quantities sold can be observed from the bills sent by the firm to its customers, and the costs that are actually incurred can be observed from wage payments, invoices for expenses, and capital accounting. The regulator does *not* need to know the demand and cost functions of the firm. That is, the regulator must be able to observe the demand and costs that actually occur in each period, but does not need to know the demand and cost curves. This is important because cost and demand relations are generally inaccessible to the regulator.

5.3 Demonstration for a One-Good Firm

Consider the situation depicted in figure 5.1. The firm is assumed to be a natural monopoly, such that its average cost curve is downward sloping. Suppose in the first time period the firm charges a price of P^1 and sells quantity Q^1 (how the firm arrived at this point is not relevant). At this quantity of output, the firm's average costs are AC^1, such that its total costs are AC^1Q^1.

The regulator observes price, quantity, and total costs. It then tells the firm that in the next period (period two) the firm can charge any price as long as the new price, when multiplied by Q^1, does not exceed AC^1Q^1:

$$P^2Q^1 \leq AC^1Q^1. \tag{5.3}$$

In a one-good situation, this constraint can be expressed more readily in terms of average cost. Dividing both sides of equation (5.3) by Q^1 gives

$$P^2 \leq AC^1.$$

That is, the firm can charge any price in period two as long as that price does not exceed period one's average cost.

The firm will choose the highest price that it is allowed to charge in

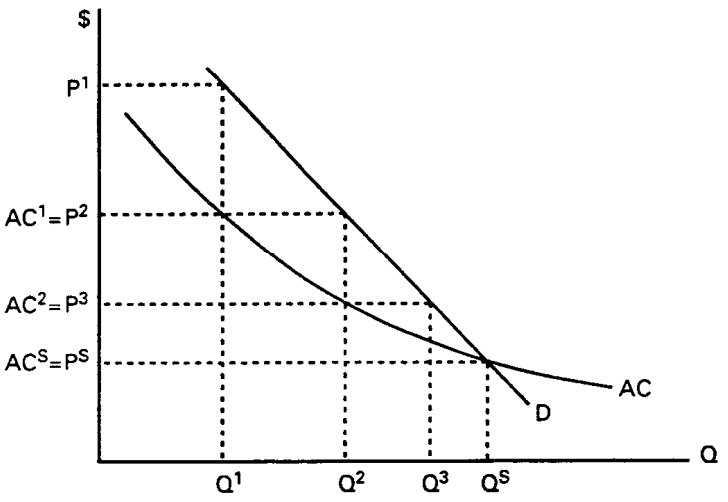

Figure 5.1
V-F regulation with one good

period two.[1] Consequently, the firm will charge a price in period two that is equal to its average cost in period one. This price is represented as P^2 on the graph.

At its new price, the firm sells a higher level of output: Q^2. Its average cost drops to AC^2 (with total costs becoming AC^2Q^2.) The regulator observes in period two the firm's price, quantity, and costs. The regulator then tells the firm that, in period three, the firm can only charge a price that is no greater than the firm's average cost in period two. In response to this directive, the firm lowers its price to AC^2, which is the highest permissible price for period three.

This process continues, with the firm lowering its price and increasing its output in successive period until it reaches P^S and Q^S, which are the second-best price and output. When this outcome is reached, the regulatory mechanism induces the firm to continue choosing this price rather than changing its price any further. The regulator requires that the firm charge in the next period a price no higher than AC^S. However, since the firm is already charging a price equal to AC^S, its current price is the highest permissible price. The firm chooses not to change its price, staying at P^S.

The optimal outcome is reached even if the firm starts out charging a price below average cost, that is, if the firm starts out earning negative profits. Figure 5.2 depicts the situation. In period one, the firm charges price P^1, sells output Q^1, and incurs average cost of AC^1, which exceeds price. The regulator tells the firm that it can charge any price in the second period as long as the new price is below AC^1. The firm raises its price in period two to AC^1. Quantity demanded drops in period two and average cost rises, such that the firm loses money in the second period as well (though its losses are smaller). This process continues until price becomes P^S, the optimal outcome consistent with nonnegative profits.

It is interesting to note that the mechanism can also identify whether marginal cost pricing is feasible. Recall that the first-best price is mar-

1. The firm will charge the highest permissible price if marginal revenue is negative, such that raising price and reducing output increases revenues. More generally, the firm will choose the permissible price that provides the greatest profit. If marginal revenue is positive, this price might be the highest permissible price. However, the firm will still be pushed toward the optimal price in the same way as under the assumption that the firm chooses the highest permissible price. The assumption allows a less cluttered graph and a more intuitive explanation of the movement toward optimality.

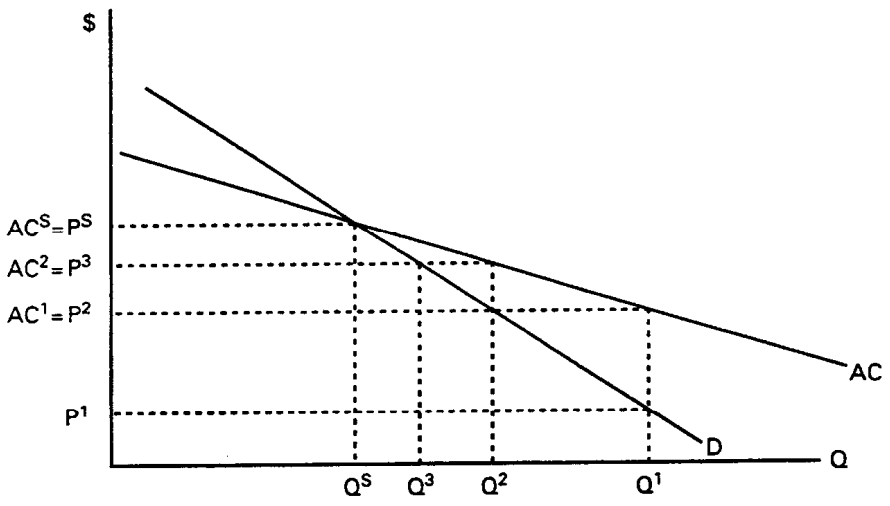

Figure 5.2
V-F regulation with one good and price starting below average cost

ginal cost. If average cost is decreasing, as in figures 5.1 and 5.2, then marginal cost is below average cost, and pricing at marginal cost would result in negative profits. However, if average cost is increasing in the relevant range, marginal-cost pricing is feasible. In this case, V-F regulation will still induce the firm to price at average cost; however, the movement toward average-cost pricing will provide information that allows the regulator to observe that marginal-cost pricing would be feasible. Consider figure 5.3. Note that the firm could operate at the intersection of the marginal cost and demand curves, because price at this point is above average cost. Suppose, however, that the regulator did not know this fact (because the regulator does not know the demand and cost curves of the firm) and applies V-F regulation to the firm. The firm starts at, say, P^1, where it is earning a positive profit. The regulator requires that its price in period two not exceed AC^1. The firm therefore sets price in period two equal to AC^1. Its demand increases to Q^2 and, since its average cost is upward sloping in the relevant range, its average cost increases to AC^2. Because P^2 is below AC^2, the firm loses money in period two. The regulator continues to apply V-F regulation, requiring that the price in period three not exceed AC^2. Because average cost in period two is greater than that in period one, this new constraint is less stringent than the previous one and allows the firm to *raise* its price in period three, to AC^2. In period

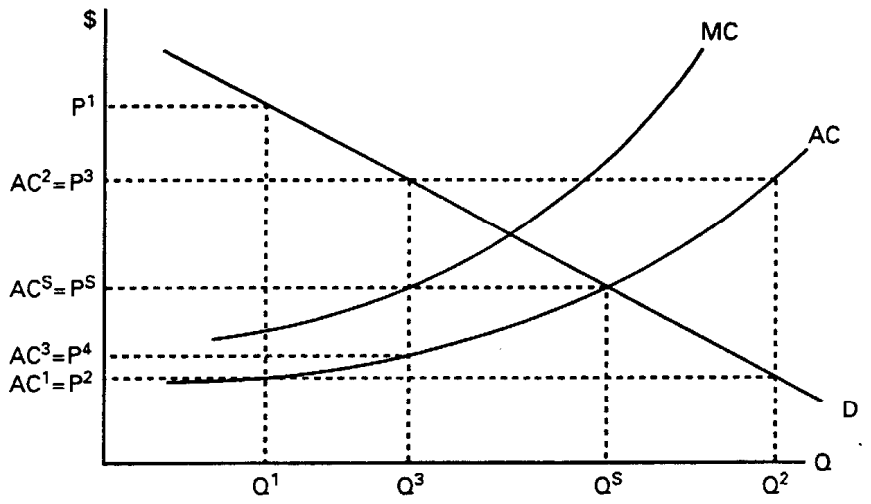

Figure 5.3
V-F regulation with upward-sloping AC

three, the firm earns positive profit. This process continues until the firm reaches P^S and Q^S, where price equals average cost. Notice, however, that the process results in cyclical movements of price and profits. That is, price is alternately raised and lowered by the firm in successive periods, and profits are alternately positive and negative. This pattern occurs because average cost is upward sloping. When average cost is downward sloping, prices and profits move continuously in one direction over time.

The regulator can use this information to determine whether marginal-cost pricing is possible, without needing to know the firm's demand and cost functions. If the regulator implements V-F regulation and observes that the firm's price and profit move continuously in one direction over time, the regulator can infer that the firm's average-cost curve is downward sloping. In this case, marginal-cost pricing would result in negative profits, such that average-cost pricing is the optimal feasible outcome. The V-F regulation will induce the firm eventually to move to this second-best outcome. If, however, the regulator observes that prices moves up and down over time, and profits are alternatively positive and negative in successive periods, the regulator can infer that average cost is upward sloping. Marginal-cost pricing is therefore feasible and should be pursued. In this case, the regulator should discontinue the V-F regulation, because it induces

the firm to move to average-cost pricing instead of marginal-cost pricing.[2]

5.4 Demonstration for a Two-Good Firm

In a one-good situation, only one price results in zero profit, namely, price equaling average cost. The second-best price is therefore attained simply by pushing the firm toward zero profit, as the V-F mechanism does. With two or more goods, many price combinations result in zero profit. Consequently, pushing the firm to zero profit is not sufficient to attain the second-best prices. A demonstration of V-F regulation with two goods is therefore not a trivial generalization of the one-good case. Rather, it introduces concepts that are critical to the meaning and function of the mechanism.

In any period t, the firm has a price for each good, sells a certain quantity of each good, and incurs certain costs. The regulator observes these amounts and constrains the firm's choice of prices for period $t + 1$. The constraint takes the form

$$P_a^{t+1}Q_a^t + P_b^{t+1}Q_b^t \leq C^t, \tag{5.4}$$

where a and b denote the two goods.

The price combinations that meet this constraint can be depicted graphically. Consider figure 5.4. Suppose that in period t, the firm is charging the price combination designated by point X in the graph. The question to be addressed is: what price combinations is the firm permitted to charge in period $t + 1$? The permissible price combinations can be depicted on the graph using three pieces of information, each of which is discussed in detail below: (1) the permissible price combinations form a line in the graph, (2) the slope of this line is the same as the slope of the isobenefit contour at X, and (3) the y-intercept of this line is an identifiable point. With the shape, slope, and intercept determined, the set of permissible prices can be located on the graph.

---

2. Two caveats are required, however. First, if the firm's profits fluctuate between positive and negative levels in the movement to equilibrium, it is possible that the present value of the stream of profits is negative. In this case, the firm would choose to shut down rather than submit to V-F regulation. Second, suppose the firm knows that the regulator will discontinue V-F regulation if prices fluctuate. The firm would perhaps have an incentive, then, to alternately raise and lower its price as a means of eliminating V-F regulation. This strategic behavior is separate from that described in section 5.5. (I thank David Sappington for pointing out these caveats.)

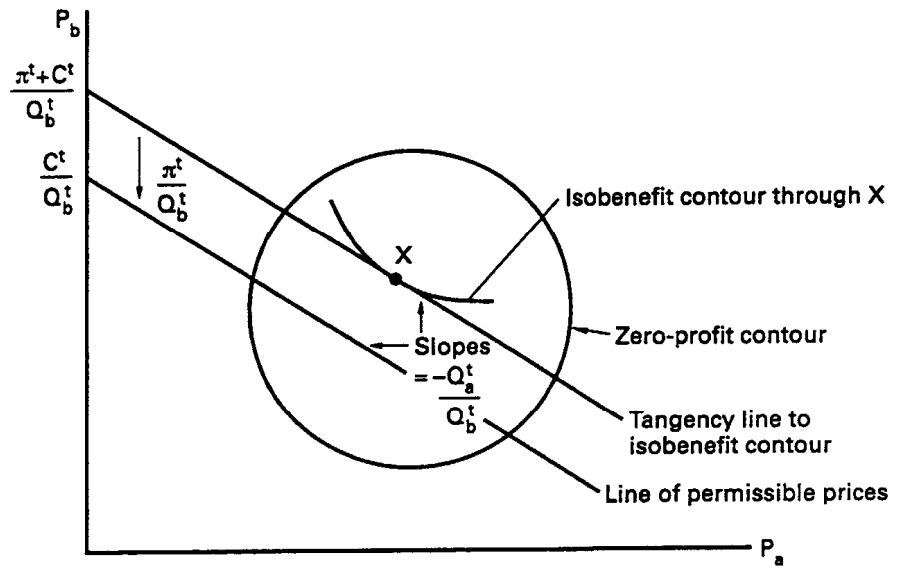

Figure 5.4
Permissible prices for two goods

1. *The permissible price combinations form a line.*

The graph in figure 5.4 places the price of good A on the x-axis and the price of good B on the y-axis. Inequality (5.4) is rearranged such that P_b^{t+1} is expressed as a function of P_a^{t+1}:

$$P_b^{t+1} \leq C^t/Q_b^t - (Q_a^t/Q_b^t)P_a^{t+1} \tag{5.5}$$

That is, the highest price for good B that is permissible in period $t + 1$ depends on the price that is charged for good A in period $t + 1$. The equality part of equation (5.5) is a line with slope $-(Q_a^t/Q_b^t)$ and y-intercept C^t/Q_b^t. The graphical interpretation of this slope and intercept is given below.

2. *The slope of this line is the same as the slope of the isobenefit curve at X.*

As indicated by equation (5.5), the line that designates the permissible prices for period $t + 1$ has slope equal to the ratio of outputs in period t. This fact is convenient, because it relates to the isobenefit contour at X. Recall from chapter 4 that an isobenefit contour is a set of price combinations that result in the same level of consumer surplus. The slope of an isobenefit contour at any point (that is, at any price combination) is the ratio of outputs that are demanded at that price combination. This fact was established in section 4.3. For ex-

ample, suppose that, at given prices, demand for good A is 100 and demand for good B is 50. Then, if the price of good A is raised by \$1, then the price of good B must be lowered by \$2 (that is, $-100/50$) in order for the person to afford the original consumption levels and hence to attain the same surplus.

Point X in figure 5.4 is the price combination the firm charges in period t. The isobenefit contour through this point is drawn in the graph. The slope of the isobenefit contour at price combination X is the ratio of outputs that are consumed in period t, namely:

$$-Q_a^t/Q_b^t.$$

The line designating the permissible prices in period $t + 1$ also has this slope. Consequently, the line of permissible prices must be parallel to the tangency line of the isobenefit contour at X.

We know that the line of permissible prices is parallel to the tangency line of the isobenefit contour at X. However, we do not yet know the location of this parallel line. By determining its y-intercept, its location is established.

3. *The y-intercept of this line is an identifiable point.*

By the equality part of equation (5.5), the line of permissible prices has a y-intercept of

$$C^t/Q_b^t.$$

The question is: where is this y-intercept relative to the intercept of the tangency line of the isobenefit contour? Consider the tangency line of the isobenefit contour through X. This line has slope

$$-Q_a^t/Q_b^t,$$

as described above. It therefore takes the form

$$P_b = k - (Q_a^t/Q_b^t)P_a. \tag{5.6}$$

The value of k, the y-intercept, can be determined by knowing that the line passes through point X. The profits of the firm in period t are, by definition,

$$\pi^t = P_b^t Q_b^t + P_a^t Q_a^t - C^t.$$

Rearranging:

$$P_b^t = (\pi^t + C^t)/Q_b^t - (Q_a^t/Q_b^t)P_a^t. \tag{5.7}$$

Because equation (5.6) holds for all points on the tangency line, and equation (5.7) holds at point X, which is on the tangency line, then k in equation (5.6) must be equal to

$$(\pi^t + C^t)/Q_b^t.$$

That is, the y-intercept of the tangency line is

$$(\pi^t + C^t)/Q_b^t.$$

This fact is important because it allows us to locate the line of permissible prices. The y-intercept of the line of permissible prices is C^t/Q_b^t, while the y-intercept of the tangency line is $(\pi^t + C^t)/Q_b^t$. That is, the y-intercept of the line of permissible prices is π^t/Q_b^t lower than the y-intercept of the tangency line. Therefore, by starting at the y-intercept for the tangency line and moving down an amount equal to the profits in period t divided by the quantity of good B sold in period t, the y-intercept of the line of permissible prices is located.

In summary, the permissible price combinations for period $t + 1$ form a line that is parallel of the tangency line of the isobenefit contour at the price combination charged in period t, and yet is lower than the tangency line by the amount of profits earned in period t divided by the quantity of good B sold. This line is given in figure 5.4. Taking into account the "less than" part of equation (5.5), any price combination on *or below* this line is permissible in period $t + 1$.

In period $t + 1$, the firm will choose the permissible price combination that provides the greatest profit. That is, the firm will choose the point on or below the permissible line that provides the greatest profits. The regulator will then use the prices, output, and costs in period $t + 1$ to establish permissible prices for period $t + 2$. And so on, until equilibrium is reached.

A sequence of two price changes is depicted in figure 5.5. The firm starts at price combination X in period t. The set of permissible prices for period $t + 1$ is denoted by the lower solid line. The firm chooses the profit-maximizing point on or below this line; say this is point Y. With the firm charging price combination Y in period $t + 1$, the regulator places a constraint on the prices that the firm can charge in period $t + 2$ based on the quantities and costs in period $t + 1$. The set of permissible prices for period $t + 2$ is denoted by the lower dotted line. The firm chooses the profit-maximizing price combination on or below this dotted line, say, point Z. And the process continues.

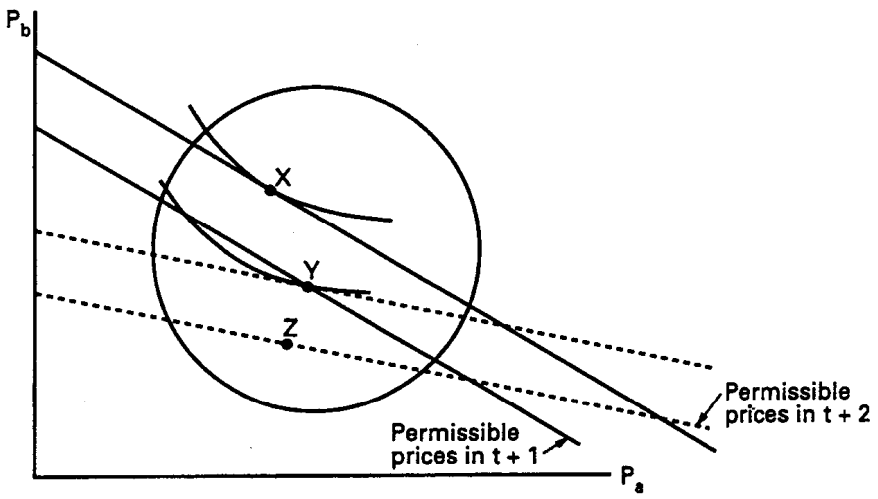

Figure 5.5
Sequence of price combinations

Note that the lower dotted line is not parallel to the lower solid line. That is, the line of permissible prices in one period does not have the same slope as the line of permissible prices in the next period. This is because the quantities of goods sold changes from one period to the next as the firm changes its prices. The line of permissible prices for period $t + 1$ has a slope equal to the ratio of outputs sold in period t — that is, Q_a^t/Q_b^t. However, the line of permissible prices for period $t + 2$ has a slope equal to the ratio of outputs sold in period $t + 1$, not t (that is, Q_a^{t+1}/Q_b^{t+1}). Because prices are different at X than Y, the ratio of outputs is generally different.

Note also that the line of permissible prices for each period is below the tangency line of the isobenefit curve by an amount that depends on the profits of the firm in the previous period. That is, the lower solid line is below the higher one by an amount that is the profits earned in period t divided by the quantity of B sold in period t, π^t/Q_b^t. And the lower dotted line is below the higher one by the profits earned in period $t + 1$ divided by the quantity of good B sold in period $t + 1$, π^{t+1}/Q_b^{t+1}. Because profits and quantities sold change when prices change, the amount by which the line of permissible prices is below the tangency line changes from period to period.

As suggested in figure 5.5, the regulatory process pushes the firm closer and closer to the zero-profit contour. (This occurs because the line of permissible prices is lowered each period whenever profits are

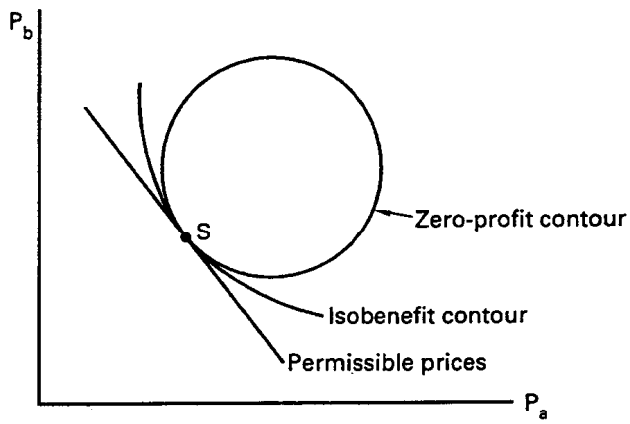

Figure 5.6
Isobenefit contour tangent to zero-profit contour

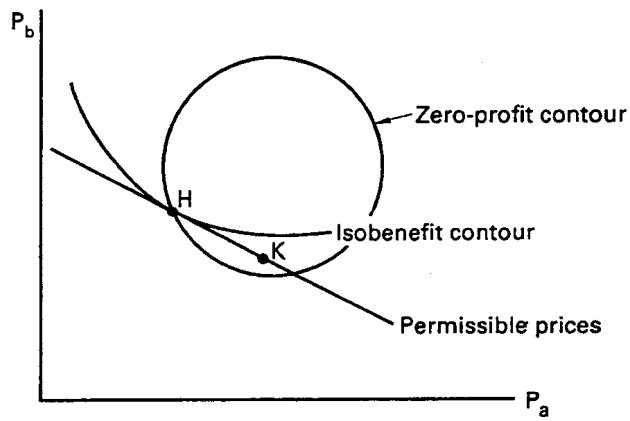

Figure 5.7
Isobenefit contour not tangent to zero-profit contour

positive.) The process continues until the firm earns zero profit—that is, until the firm's chosen price combination is on its zero-profit contour. When this occurs, one of two situations must arise. Either (1) the isobenefit contour through the firm's chosen price combination is tangent to the firm's zero-profit contour, as in figure 5.6, or (2) the isobenefit contour is *not* tangent to the zero-profit contour, as in figure 5.7. In the first case, as demonstrated below, the firm is charging Ramsey prices and will continue to do so indefinitely. In the second case, the firm will change its price, eventually moving to a situation like that in figure 5.6, that is, to Ramsey prices.

In figure 5.6 the firm is charging price combination S, at which the isobenefit contour is tangent to the zero-profit contour. This point is the Ramsey price combination. Recall from chapter 4, and especially figure 4.9, that Ramsey prices are those at which the isobenefit contour is tangent to the zero-profit contour (such that the highest possible level of consumer surplus is attained consistent with non-negative profits). The firm in figure 5.6 is therefore charging Ramsey prices.

Once at Ramsey prices, the firm will continue to charge these prices indefinitely over time. This fact can be demonstrated in two steps.

Step One

Consider the set of permissible prices for the following period, given that the firm is charging price combination S in the current period. The line of permissible prices is parallel to the tangency line of the isobenefit contour and is lower by the amount of profits earned by the firm divided by the quantity of good B sold. Note, however, that at S, the firm earns zero profits. The line of permissible prices is therefore lower than the tangency line by zero. Because it is parallel to and the same level as the tangency line, the line of permissible prices is the same as the tangency line.

Step Two

Since the isobenefit contour is tangent to the zero-profit contour, the tangency line, and hence the line of permissible prices, touches but does not cut the zero-profit contour. This means that the line of permissible prices is outside the zero-profit contour at all points except S, where it is on the zero-profit contour. Of all the points on the line of permissible prices, the firm chooses S, because at S it earns zero profit, whereas at any other permissible point it earns negative profit.

Stated succinctly, once the firm is charging Ramsey prices in one

period, the V-F mechanism will induce it to continue doing so in the next period. The firm will therefore stay indefinitely at Ramsey prices once it has reached them.

Consider now the possibility that the firm reaches the zero-profit contour at a point where the isobenefit contour is not tangent to the zero-profit contour, as point H in figure 5.7. In this case the firm will move away from point H in the following period. Given that the firm is at H, the line of permissible prices for the following period is the line of tangency of the isobenefit contour at H, for reasons discussed in relation to figure 5.6. Because the isobenefit contour is not tangent to the zero-profit contour, the line of permissible prices cuts the zero-profit contour. A range of permissible prices lies inside the zero-profit contour. In the following period, the firm will choose one of the points inside the zero-profit contour, say K, because these result in positive profits, rather than staying at H earning zero profits.

Once the firm moves to K, the situation becomes as before, with the line of permissible prices being lowered in each successive period until the firm is back on the zero-profit contour. If, this time around, the firm is at Ramsey prices, it will remain there. If, however, the firm is at prices that provide zero profits but are not Ramsey, then it will move in the next period, once again, to prices that result in positive profits. The process continues until the firm is at Ramsey prices. Stated another way, the process of price changes only stops when Ramsey prices are attained; consequently, Ramsey prices are the equilibrium of the V-F mechanism.

There are two fundamental reasons why the V-F mechanism results in Ramsey prices. First, the mechanism pushes the firm toward zero profits by lowering the prices that are permissible in the next period whenever positive profits are being made in the current period. Second (and this is the more subtle and more significant reason), the mechanism induces the firm to operate under a trade-off between prices that is the same trade-off that consumers are willing to undergo. Consider the consumer's perspective. When one price is raised, the other price must be lowered for consumer surplus to remain unchanged. As discussed, the amount that this other price must be lowered is equal to the ratio of quantities consumed at the original prices. That is, consumers are willing to "trade" a rise in one price for a lowering of another, as long as the rate of this trade is the ratio of quantities consumed. Under V-F regulation, the firm is induced to take the same perspective. From any permissible price combination,

the firm can move to another permissible price by raising one price and lowering the other, thereby staying on the line of permissible prices. The slope of the permissible price line determines the amount by which the price of a second good must be lowered when the first price is raised. Under V-F regulation, this slope is the ratio of quantities consumed. That is, the firm is permitted to "trade" a rise in one price for a lowering of another, as long as the rate of this trade is at the ratio of quantities. Because the rate at which the firm is permitted to trade off price increases and decreases is the same as the rate at which consumers are willing to trade off such increases and decreases, consistency is established between the choices of the firm and the surplus of consumers. As in the discussions of chapters 1 and 2, the establishment of such consistency leads to optimality.

5.5 Strategic Issues in V-F Regulation

A firm subject to V-F regulation might have an incentive to inflate its costs of production. This incentive arises from the fact that, when reported costs are higher in one period, the regulator allows the firm to charge higher prices in the next period. This inflation of costs can take either of two forms. First, the firm can simply report costs to the regulator that it does not actually incur; that is, it can lie about its costs. Alternatively, the firm can incur unnecessary costs. In this latter case, the firm reports its incurred costs accurately, but engages in some form of waste to increase its costs above that minimally needed for production.

Each of these actions is a type of strategic response to V-F regulation: the firm engages in the activity for the *purpose* of affecting the decisions of the regulator. The potential for strategic behavior by the firm arises in any regulatory mechanism in which the regulator determines the constraints it places on the firm on the basis, at least partially, of the firm's actions. In the case of V-F regulation, the regulator bases its constraint (that is, determines the permissible prices) on the basis of the costs the firm reports for the previous period. The firm knows that the regulator behaves in this way and takes this fact into consideration when deciding what costs to incur and report.

Regulators often, if not always, base their decisions on actions of the firms they are regulating, and especially on the basis of information reported by the firms. Because the potential for strategic behavior is present in these contexts, it is important to determine whether (or,

more precisely, under what conditions) the firm will indeed choose to engage in strategic behavior that is not consistent with the goals of the regulator. In the paragraphs below, we address this issue in the context of V-F regulation, and, in particular, with respect to two forms of strategic behavior: misrepresentation of costs and purposeful waste. The discussion, however, is intended to serve a more general purpose as an illustration of the need and procedure for examining these issues under any regulatory mechanism.

5.5.1 Misreporting of Costs

If the firm is able to deceive the regulator as to its costs, then, under V-F regulation, the firm will clearly find it advantageous to do so. Any costs the firm reports but does not actually incur can be retained as profit. Overstatement of costs can slow the movement toward equilibrium: the regulator will think that the firm is earning less profit in each period than it actually is and consequently will require a smaller reduction in prices in the following period. More important, equilibrium can occur at a price above the true average cost, with the firm earning positive profits indefinitely. If the firm's reported profits are zero (based on its reported costs), the regulator will not require that the firm change its prices in the following period. Thus, by reporting costs that exceed actual costs by exactly the amount of profit the firm is earning, the firm can continue earning those profits indefinitely.

The incentive for misreporting of costs is pervasive in regulatory settings. A way for the regulator to prevent the practice is to audit the firm periodically and levy a penalty if the firm's actual costs are less than its reported costs. If audits are costless and, when performed, are able perfectly to determine the firm's true costs, optimality can easily be achieved. The regulator can simply audit the firm's costs each period. Because the firm knows that its true costs will be discovered, it has no incentive to misreport; in fact, it has a clear incentive to report truthfully to avoid the penalty it would have to pay when any untruth is (inevitably) discovered. Furthermore, because the audits are costless, society incurs no additional cost from this component of regulation.

In reality, audits do entail cost and, when performed, are not able to assess the firm's true costs perfectly. Because audits are not free, a trade-off exists between the benefits that audits induce, through more truthful reporting, and the costs of the audits themselves. Since au-

dits are not perfectly accurate, uncertainty enters the process in a way that can hurt both the public and the firm. The firm could misreport costs and the regulator not discover it; or the firm could report its costs truthfully and yet be penalized when an audit "finds" (incorrectly) that the firm overstated its costs.

Townsend (1979) has examined the issue of audits being costly, abstracting from the issue of uncertainty. That is, Townsend assumed that audits cost but are perfectly accurate. Within this context, his analysis suggests that the regulator can conduct audits in a particular way to obtain an outcome that is essentially the same as if audits were free. That is, the regulator can induce the firm to report its costs truthfully while incurring essentially no auditing costs. The procedure consists of auditing the firm very infrequently (that is, in only a few periods over a span of many periods), not letting the firm know beforehand whether or not it will be audited in a particular period, and levying a very high penalty if the firm is found to have misreported its costs. Auditing costs are maintained at a very low level because audits are seldom performed.

However, with a sufficiently high penalty, the firm is induced to report truthfully, even though the chance of an audit in any given period is low.[3] Because the firm does not know when an audit will be performed, it is induced to report truthfully *each* period in anticipation of an audit perhaps being performed in that period.

The fact that this procedure can obtain an outcome that is essentially the same as with free audits depends on there being very few audits, such that auditing costs are very low (essentially zero, as in the case with free audits). A very small number of audits can induce honesty when the penalty is sufficiently large. In many situations, however, there is a limit on the size of penalty that the regulator can impose. For example, the penalty might not be able to exceed the limits of corporate liability, under which a corporation cannot be held liable for more than its net assets. Or, the regulator might be bound to treat the firm "reasonably and fairly," which might be interpreted to mean that any penalty cannot be incommensurate with the mag-

3. The expected cost to the firm of not reporting truthfully is the probability of an audit times the penalty that is levied if any audit is performed and the untruth discovered. Even if the probability of an audit is low, the expected cost of misreporting can be high if the penalty is sufficiently high. In fact, the penalty can be set high enough such that the firm will always find the expected cost of misreporting to exceed the expected benefit it would obtain (in the form of higher prices) from any misreporting.

nitude of the offense (such that a misreporting of, say, $1 million cannot result in a $100 million penalty).

If there is a limit on the penalty, then the procedure does not necessarily result in the same outcome as with free audits. Specifically, auditing costs will not be essentially zero. The reason here should be clear. If the penalty is unlimited, the number of audits performed over a span of many periods can be reduced to nearly none: the firm can still be induced to report truthfully by raising the penalty sufficiently and not letting the firm know which period will be audited. Stated in probabilistic terms, the probability of an audit can be reduced arbitrarily close to zero if the penalty is raised sufficiently such that the expected loss of misreporting in each period is maintained at a high enough level to induce the firm to be truthful. If there is a limit on the magnitude of the penalty, the number of audits cannot necessarily be reduced nearly to none, since the penalty cannot necessarily be raised sufficiently to induce the firm to be truthful. In fact, the probability of an audit might need to be fairly high given the limited size of the penalty. The cost of auditing might therefore be significant, instead of essentially zero as in the case of a limitless penalty. As a result, there is a social cost involved in regulating when audits are not free and the penalty is limited.

In addition to being costly, audits are not perfect instruments for assessing the true costs of the firm. Because an audit provides only an estimate of true costs, a discrepancy between the audited cost and the costs that the firm reported does not necessarily mean that the firm was untruthful: the audit estimate could be incorrect while the firm's report is correct. The regulator needs to decide, therefore, how large a discrepancy between the audit's estimated costs and the firm's reported costs is enough to warrant a penalty (or, more generally, the regulator must determine a relation between the size of the penalty and the magnitude of the discrepancy). Errors can be made in either direction. If a small discrepancy results in a penalty, the firm could often be penalized even though it reports its costs truthfully. In fact, if a small enough discrepancy results in a penalty, the firm will have little incentive to be truthful, because it realizes that it will probably be penalized even if it is truthful. However, if the regulator requires that the discrepancy be large before assessing a penalty, the firm can misrepresent its costs without substantial risk of discovery.

Baron and Besanko (1984) derive pricing and auditing procedures for the regulator when audits are costly and not perfectly accurate in

determining the firm's true costs.[4] They demonstrate an important result, namely, if there is no limit on the size of the penalty the regulator can levy, auditing procedures can be established that attain the same outcome as with costless and perfect audits, even though audits are costly and imperfect. The firm is levied a penalty whenever the costs estimated by the audit are below the firm's reported costs. This penalty is levied even though the firm might have reported its costs accurately and the audit simply underestimated the true costs. (This procedure does not hurt the firm for reasons to be described shortly.) Audits are performed very infrequently, but without the firm knowing when. If an audit is performed and costs are estimated to be lower than reported, a very large penalty is levied.

With a high enough penalty, the firm is induced to report its costs truthfully, because doing so minimizes its chance of being penalized (even though it could still be penalized when reporting truthfully). The number of audits is reduced to practically none by raising the penalty sufficiently to assure that the firm remains truthful in the face of fewer audits. With practically no audits, essentially no costs are expended on auditing. Furthermore, the firm's chance of being assessed a penalty even though it is reporting truthfully is essentially nil: hardly any audits are performed, and the chance of the firm being penalized, if audited, is as low as possible since the firm's report is honest.

If there is a limit on the size of the penalty, Baron and Besanko show that a loss is incurred compared to a situation with perfectly accurate and free audits. In particular, some (not arbitrarily low) cost will be expended on auditing, and the firm will face a (not arbitrarily

4. Technically, Baron and Besanko structure their analysis somewhat differently than described in the text; however, the substantive conclusions are the same. Baron and Besanko assume that costs depend on random factors (such as equipment failures). At the beginning of the period, the firm submits to the regulator information on the distribution of possible costs (e.g., the historical average number of breakdowns). The regulator sets prices for the upcoming period based on this information. During the period, some costs are actually incurred (e.g., some number of breakdowns actually occur). If the regulator audits the firm, the regulator is assumed to be able to determine accurately the costs the firm actually incurred. However, this information does not perfectly determine whether the firm's original report on the distribution of costs was accurate. The regulator calculates an estimate of the cost-distribution information on the basis of the costs observed in the audit. The regulator compares this estimate with the firm's original report to determine whether to assess a penalty. That is, the firm's report is compared with an estimate of the same information derived from an audit. In this sense, Baron and Besanko's model has implications for situations in which audits provide only an estimate of the firm's costs rather than perfectly assess the incurred costs.

low) chance of being assessed a penalty even while reporting its costs truthfully.

5.5.2 Purposeful Waste

If the regulator monitors costs appropriately, the firm will not be able to misrepresent its costs. However, the firm will still be able to affect the prices it is permitted to charge in future periods. In particular, rather than reporting costs that are not incurred, the firm can actually *incur* higher costs. This purposeful incurring of unnecessary costs can take many forms: an inefficient input mix (i.e., using a mix of inputs that is not cost minimizing), more inputs than minimally required (such as large offices and fancy boardrooms), paying more than necessary for inputs, excessive research and development efforts (such as research on topics with little chance of fruition or small potential benefit), and various other forms. For the purpose of discussion, all of these manifestations are collectively called "waste."

A firm under V-F regulation will engage in waste in some circumstances and not in others. The most critical factor is whether the firm knows in advance (that is, prior to being regulated) that V-F regulation will be imposed. As shown below, the firm could find it profitable to waste in some periods if it knows in advance that V-F regulation will be imposed. However, if there is no advance notice, the firm will not find it profitable to waste. The two situations are described separately below, with the analysis based largely on concepts of Sappington (1980).

1. Firm with Advance Notice

Consider first the situation in which the firm knows that V-F regulation will be imposed in the following period. Consider a one-output firm, depicted in figure 5.8, which is charging P^1 in period one and knows that V-F regulation will be imposed in period two. To focus on the central issues, average cost is assumed to be constant. (This assumption is not necessary for the results.) The question to be addressed is: will the firm waste in period one?

Without waste, the firm incurs costs of $AC \cdot Q^1$ in period one. Its profits are denoted by the rectangle *ABCJ*. When V-F regulation is imposed in period two, the firm must lower its price to P_n^2, equal to its average cost in period one. Its profits in the second period are therefore zero; and because price equals average cost in the second

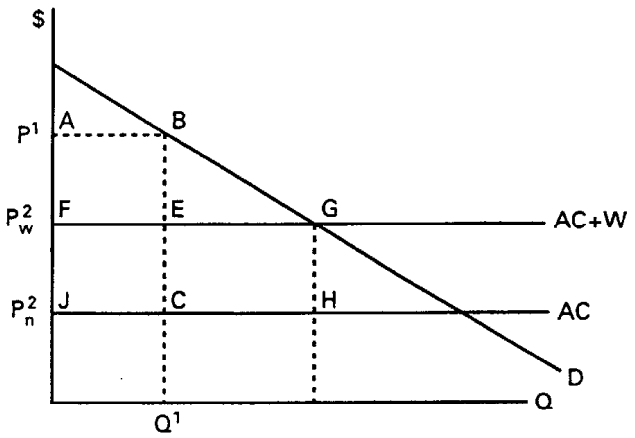

Figure 5.8
Firm wastes in period 1

period, the firm is in equilibrium and earns zero profits in all future periods. Its total profits from period one through all future periods is therefore the rectangle *ABCJ*.

Suppose now that the firm wastes an amount *W* for each unit produced in the first period. Its profits in the first period are then *ABEF*, which is less than if the firm did not waste. When V-F regulation is imposed, however, the regulator does not see the waste; the regulator only observes actual costs, which include both needed and unneeded expenses. The regulator therefore requires that the firm lower its price in the second period to $P_w^2 = AC + W$, which is higher than would be permitted if the firm did not waste in the first period. With the higher price, the firm has an opportunity to earn profits in the second period. In particular, suppose the firm does *not* waste in period two. Without wasting, it earns profits of *FGHJ* in the second period: the difference between price and average cost, over the number of units sold. In the third period, the firm is required to lower its price to the average cost incurred in the second period: $P_w^3 = AC$. The firm earns zero profit in the third and all subsequent periods. The total profits of the firm, over all periods, is therefore the area *ABEF* plus area *FGHJ*. Because these areas combined exceed area *ABCJ* (the firm's profits without waste), the total profits of the firm are greater if it wastes in the first period than if it operates efficiently. The firm therefore chooses to waste.

The driving force behind this waste is clear: by incurring unnecessarily high costs in one period, the firm is able to charge higher prices

in the subsequent period. The firm foregoes some profit in one period in order to earn greater profits in the next. Because demand is downward sloping, the extra profits earned in the next period exceed the reduction in profits in the first period.

The waste can be seen from an informational perspective as well. Essentially, the firm is providing incorrect information to the regulator in an effort to affect the regulator's decisions. If the regulator knew the minimum cost for producing the firm's output, the regulator would set each period's price equal to the minimum average cost in the previous period. In such a case, the firm would have no incentive to waste because doing so would only reduce its profits in the current period without raising its price in subsequent periods. Unfortunately, the regulator generally does not know the minimum cost of producing a given output and only observes the actual costs of the firm, which might include waste. By wasting, the firm induces the regulator to think that average cost is actually higher than it minimally need be. The regulator sets prices in the next period on the basis of this misinformation and thereby inadvertently allows the firm to earn greater profits than it would have if the regulator had known the true average costs.

Two important details have been omitted from the discussion thus far. First, when considering a stream of profits over time, a firm will discount (value less highly) profits earned in future years compared to profits earned in the current year. A dollar today is worth more than a dollar next year, because, among other reasons, a dollar obtained today can be invested to earn more money over the year. The standard way to represent this fact in the decisionmaking of firms is to calculate the "present value" of a stream on profits over time. Denote the discount rate as d, with $0 < d < 1$, to capture the fact that future-year profits are valued less than current-year profits. The present value of profits π_1 in year one, π_2 in year two, π_3 in year three and so on is

$$\pi_1 + d\pi_2 + d^2\pi_3 + \ldots,$$

where the discount rate reduces the value of each future year's profits relative to the current year, with the reduction being greater the further in the future the profits accrue. For example, at a discount rate of .9 (meaning that future profits are valued at 10% less than current profits), the present value of a profit stream of \$100 each year for two years is *not* \$200; rather, it is \$100 + .9 · \$100 = \$190. Similarly, if the

profit stream is \$100 for *three* years and zero thereafter, the present value is \$100 + .9 · \$100 + .9 · .9 · \$100 = \$271.

Because future profits are valued less highly than current profits, the firm that knows that V-F regulation will be imposed will not *necessarily* choose to waste. Consider again figure 5.8. If the firm wastes, its profits in the first period are lower by *FECJ* (compared to the case in which it did not waste), but its profits in the second period are higher by *FGHJ*. If profits in each period were valued equally, the firm would always choose to waste because area *FGHJ* exceeds area *FECJ*. However, the firm generally discounts future profits. That is, it compares the *discounted* value of the extra profits it will earn in the second period with the *undiscounted* reduction in profits in the first period. It is possible that the discounted value of area *FGHJ* is less than the undiscounted value of area *FECJ*. If so, the firm will choose not to waste.[5] Stated generally, the firm that anticipates the imposition of V-F regulation will waste if the discounted value of the gains in the future period exceeds the reduction in current-period profits, and otherwise it will not waste.

Second, the firm will not necessarily waste in only one period. Our example has shown that the firm could make more profit by wasting in the first period and not wasting in the second period than by not wasting in either period. Similar logic can be used to show that the firm might make more profit by wasting in the first *and* second periods and not wasting in the third period than by wasting in only the first period and not in the second or third periods. And so on for subsequent periods.

The firm will never, however, choose to waste in equilibrium. Consider an equilibrium in which the firm is wasting. We can show that such an equilibrium is impossible. Suppose a firm is charging P^* and wasting an amount W per unit of output. The minimal average cost plus the amount of waste equals the price the firm charges, such that the firm is permitted to continue charging this price in each subsequent period. That is, the firm can choose to remain in equilibrium at this point if it so chooses. The situation is depicted in figure 5.9.

If the firm continues to charge P^* and waste W per unit, then its

5. For example, suppose a firm that has a discount rate of .9 would earn \$100 less profits in the first period from wasting and then would obtain \$108 extra profit in the next period. This firm would choose not to waste since the present value of its gain in the second period (.9 · \$108 = 97.20) is less than the \$100 reduction in profits in the first period.

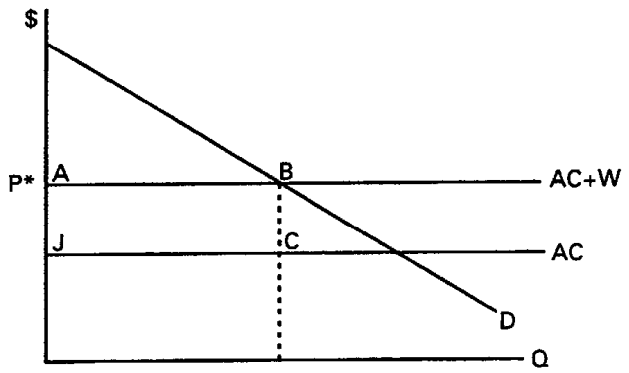

Figure 5.9
Firm does not waste in equilibrium

profits are zero in the current and all subsequent time periods. However, the firm can make positive profits by not wasting. In particular, if the firm chose not to waste in the current period, it would earn profits of *ABCJ*. It would then, of course, be required to lower its price to average cost and would earn zero profits in all subsequent periods. However, earning *ABCJ* in the first period and zero thereafter is clearly preferable to earning zero profits in the first period as well as thereafter. The firm will choose not to waste rather than remain in an equilibrium with waste. Eventually, the firm will move to an equilibrium with no waste at optimal prices. That is, the firm might waste on the way to equilibrium, but will not waste at equilibrium.

2. *Firm without Advance Notice*

Consider now the possibility that the regulator imposes V-F regulation without the firm knowing beforehand. In period one, the firm is unregulated and does not anticipate regulation. It charges, say, P^1 and maximizes profits by minimizing the cost of producing the output demanded at that price. That is, the firm does not waste in the first period. In the second period, V-F regulation is imposed on the firm. The regulator observes the firm's true average cost in the first period and requires that the firm reduce its price in the second period to this average cost. Because the firm did not know that the regulation would be imposed, it had no reason to incur waste in the first period to falsify its cost situation in anticipation of being permitted a higher price in the second period.

The question now is: will the firm choose to waste in the second period, after V-F regulation has been imposed? The answer is no.

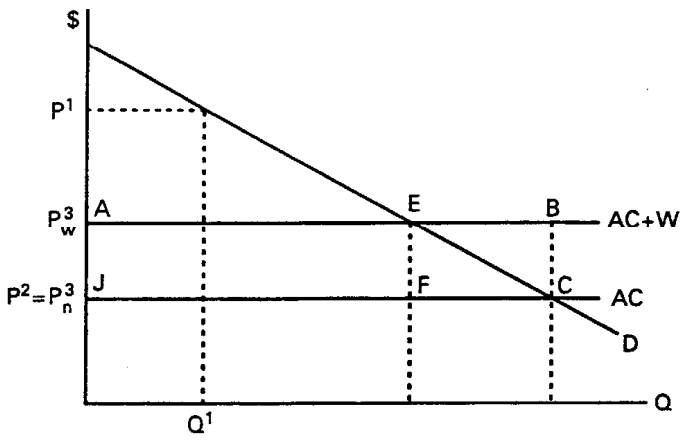

Figure 5.10
Firm without advance notice

Consider figure 5.10. The firm charges $P^2 = AC$ in the second period, because the regulator observed the firm's true costs in the first period. If the firm continues not to waste, it will earn zero profits in the second and all subsequent periods. The firm could, however, waste in the second period. If the firm wasted W per unit of output, it would earn negative profits in the second period: its losses would be area *ABCJ*. However, with higher costs in the second period, the regulator would allow the firm to charge a higher price in the third period. In particular, the firm, if it wasted in the second period, would be able to charge P_w^3 in the third period. Then, by not wasting in this third period, the firm could use its higher price to earn positive profits; in particular, its profits would be area *AEFJ*.

Note, however, that this gain in the third period is less than the loss in the second period: the firm would lose more than it would gain by wasting. The underlying reason is clear. Output is lower in the third period than in the second since price is higher. The price increase in the third period is applied to the lower third-period output. However, the inflated average cost in the second period is applied to the higher second-period output. The gain in the third period is therefore necessarily less than the loss in the second period. The firm would therefore choose not to waste.[6]

6. If demand is completely inelastic (that is, vertical), then area *ABCJ* is the same size as *AEFJ*. Even in this case, however, the firm would choose not to waste because the firm discounts future profits relative to current profits. The present value of the gain is smaller in magnitude than the present value of the loss, since the gain occurs after the loss.

The findings regarding waste can be summarized as follows. If the firm knows in advance that V-F regulation will be applied, it might have an incentive to waste so as to misinform the regulator as to the true costs of production, thereby inducing the regulator to permit higher prices in subsequent periods. Whether it is profitable for the firm in this situation to waste (that is, whether it does indeed waste) depends on a variety of factors, including the shapes of the demand and cost curves and the firm's discount rate. Waste, if it occurs, can continue over numerous periods while the firm is moving toward equilibrium. However, eventually the firm will reach equilibrium, at which it does not waste, and will charge optimal prices consistent with zero profits. Furthermore, if the regulation is imposed without the firm knowing in advance, the firm will not waste in its movement toward equilibrium or when it reaches equilibrium.