8 Time-of-Use Prices and Riordan's Mechanism

8.1 Motivation

In many situations, demand varies over time more quickly than capacity can be adjusted. The classic example is an urban highway: demand is higher during rush hours than during other times of the day, and yet the size of the freeway cannot be adjusted hourly to accommodate these shifts.[1] Similarly, demand for electricity in many areas is greater in the afternoon, when people run their air conditioners, than during the morning and night; however, generation capacity is fixed, at least within the span of a day.

For many regulated firms facing demand that varies over the day, different prices are charged in different times of day. Telephone companies, especially long-distance carriers, charge higher rates during business hours than during evenings, nights, and weekends. Electric utilities often offer time-of-use rates, particularly for commercial and agricultural customers.

In situations such as these, important questions arise:

1. What is the first-best price to charge in each time period, given the available capacity? With a sufficiently low price, demand could exceed capacity, such that congestion occurs, or, in the case of electricity, possibly blackouts. However, with a sufficiently high price, demand could fall below capacity, such that the capacity is, in a sense, being wasted. Is the optimal price in each period the price that results in demand equaling capacity, such that there is no congestion or under-

1. Actually, many urban areas adjust highway capacity in each direction by moving the median barrier. This practice is an attempt to deal with the problem of fluctuating demand.

utilization; or is it perhaps optimal to have a degree of congestion and/or underutilization in certain periods?

2. Given the first-best price in each period, is there a mechanism that regulators can use to induce a firm with fixed capacity and time-varying demand to charge these prices? Riordan (1984) has proposed a mechanism that does just this.

3. Over time, capacity can be adjusted. What is the optimal capacity in situations where demand fluctuates? If sufficient capacity is constructed to handle the periods with high demand, then capacity in other periods will be underutilized. However, if less capacity is provided, there will be congestion in the periods with high demand. What is the optimal point in the trade-off between these two factors?

4. Finally, is there a mechanism regulators can use to induce the firm to provide the optimal level of capacity? Riordan's mechanism can, in certain circumstances, be used for this purpose as well as to induce optimal pricing.

These questions are addressed sequentially in the following sections. Section 8.2 identifies first-best prices given capacity. Section 8.3 describes Riordan's method for inducing these prices. Section 8.4 identifies the optimal capacity. And section 8.5 discusses the extent to which Riordan's method can induce optimal capacity.

The findings of this chapter can be summarized as follows:

• When demand fluctuates over time periods (e.g., rush and nonrush hours) and capacity is fixed, the first-best pricing rule is the following. In each period, price at marginal cost as long as doing so results in no more demand in that period than can be met with the available capacity. If demand exceeds capacity when price equals marginal cost, raise price until demand equals capacity.

• At these first-best prices with fixed capacity, there may be either excess capacity or congestion (i.e., insufficient capacity) in any period. That is, the existence of excess capacity and congestion are both consistent with first-best pricing under fixed capacity.

• The firm's profits may be positive or negative at the first-best prices with fixed capacity.

• Under Riordan's mechanism, the firm is subsidized on the basis of the price it charges in each period. The subsidy in each period consists of (1) the fixed costs of production in that period minus (2) the amount by which price exceeds marginal cost in the period times the capacity of the firm. With this subsidy, the firm earns zero profit at the first-best prices and negative profit at any other prices. The firm therefore chooses the first-best prices.

- The optimal capacity is the level at which the average amount that customers are willing to pay for extra capacity (averaged over all periods) equals the cost of extra capacity. With this optimal capacity, first-best prices are obtained by the same rule as when capacity is fixed.

- Under Riordan's mechanism, the firm is indifferent between choosing the optimal capacity and any other capacity: the firm earns zero profit at any capacity level. The firm therefore has no reason *not* to choose the optimal capacity. However, the mechanism does not *necessarily* induce the firm to choose this capacity.

- If the regulator knows the optimal capacity, the subsidy under Riordan's mechanism can be calculated on the basis of the optimal capacity rather than the firm's chosen capacity. With this subsidy, the firm will necessarily choose the optimal capacity. However, it is unlikely that the regulator knows the optimal capacity so as to implement this subsidy.

8.2 First-Best TOU Prices Given Capacity

Consider first a particular stylized situation.[2] Suppose a firm has a plant with fixed capacity K, which is the maximum number of units that can be produced per period of time (say, per hour.) The firm incurs a fixed cost, F, for the plant; this fixed cost is expressed as a flow of expenditures over time, that is, as dollars per period. The variable costs of production consist of a constant marginal cost, c, for each unit of output produced. An example might be an electric utility with a coal-powered electricity generating plant. The lease or mortgage on the plant, or the opportunity cost of funds tied up in the plant, is F per period. The cost of coal, labor, and other inputs for producing an extra kilowatt is c, and the plant is capable of producing

2. The framework of this analysis, and for the analysis of optimal capacity in section 8.4, follows most closely that of Williamson (1966). Riordan used this framework to describe his regulatory mechanism, which is one of the reasons for adopting it in this section. Issues of optimal pricing and capacity with fluctuating demand, often called peak-load pricing, have been examined extensively over the years; seminal studies include those by Steiner (1957), Houthakker (1958), Boiteux (1960), Mohring (1970), and Keeler and Small (1977).

up to K kilowatts per hour but no more. Figure 8.1 gives the marginal cost curve in this situation: MC is flat at c up to quantity K after which no more output can be produced.

Consider first a situation in which demand does not fluctuate over time. Setting price equal to marginal cost assures, as always, that customers buy units if and only if the the value of each unit to the customer is greater than or equal to the cost of producing the unit. With p set to c, two events can occur: either the quantity demanded in each period can be met with capacity K, or more units are demanded each period than can be produced. Panel (a) of figure 8.2 depicts the first case. At marginal-cost pricing, quantity $q^* < K$ is demanded. In this case, the first-best price is clearly c: if price were lowered, the additional output demanded at the lower price would be valued at less than the marginal cost of producing the additional output; and if price were raised, units that are valued above their cost would not be produced. Either way there would be a loss compared to marginal-cost pricing. Note that the first-best price in this situation results in "wasted" capacity, in the amount of $K - q^*$. In the short run, with K fixed, there is nothing that can be done about this extra capacity.[3] The second possibility is that demand exceeds capacity when price is set equal to marginal cost. This case is depicted in panel (b). At $p = c$, demand cannot be met, such that rationing is necessary. The issue therefore becomes: what is the most efficient basis on which to ration the K units that can be produced. With p maintained at c, rationing could occur in any number of ways: by customers queuing up, such that the customers who are most willing to spend time in line get the goods;

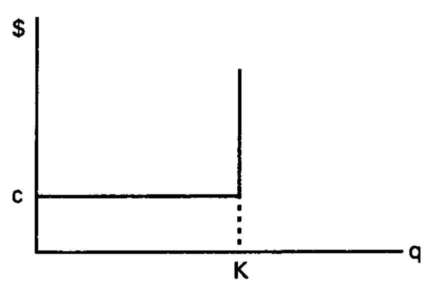

Figure 8.1
Marginal cost with fixed capacity

3. Lowering price in an effort to utilize a larger portion of capacity is counterproductive, because the additional units that are sold would be valued at less than the variable cost of producing them. Over time, the capacity of the plant should be reduced; this issue is addressed in section 8.4.

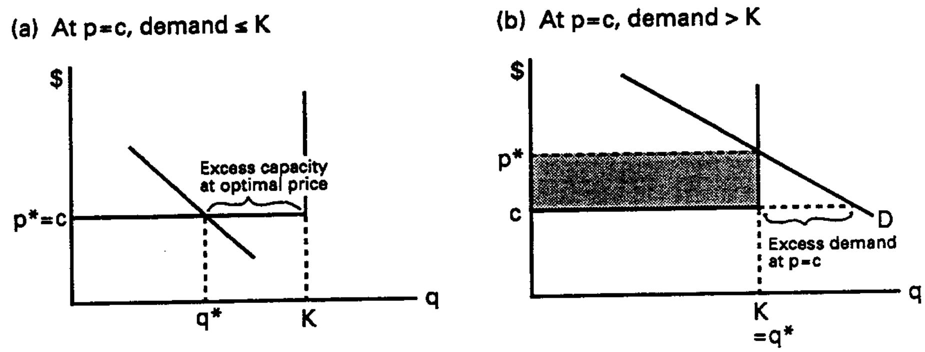

Figure 8.2
Optimal price with nonvarying demand

through a lottery; through force, such that the strongest customers obtain the units; and so on. An alternative form of rationing is to raise p until the quantity demanded drops to K; in the figure, this constitutes raising price to p^*. With price at p^*, customers that value the good at p^* or more obtain the good, and those that value it at less than p^* do not. That is, raising the price to p^* rations the K goods on the basis on customers' willingness to pay for the good. From an efficiency perspective, this basis for rationing is clearly optimal.[4]

The general rule for pricing under fixed capacity is: set price at marginal cost unless doing so results in more units being demanded than can be produced, in which case raise price until the quantity demanded drops to capacity.

Under this first-best pricing, the firm could end up making positive or negative profits. With price at marginal cost, the firm covers its variable costs but not its fixed costs, such that it incurs losses of F per period. When price is raised above marginal cost to eliminate excess demand, the firm obtains revenues in excess of its variable costs, de-

4. Under other forms of rationing, a customer that is willing to pay more that p^* for the good might not get the good, while another customer that values it at less than p^* might obtain it. Both of these customers would benefit from a transaction in which the customer with the higher value buys the good from the customer with the lower value at a price that is between their two values. Because mutually beneficial transactions are available after rationing on the basis of something other than price, nonprice rationing does not, by definition, provide an outcome that is Pareto optimal. This fact can be seen in another way. If any nonprice form of rationing were utilized and then followed by a series of voluntary, mutually benefiting barters among all customers, the customers that are willing to pay at least p^* would end up with the K units and those valuing it less would end up with payments but not the good. The result would be the same as pricing the good at p^* originally (except for transfers, which do not affect total surplus).

picted as the shaded area in the figure. Depending on how high price must be raised to equate quantity demanded with capacity, this extra revenue may exceed, fall short, or just cover the firm's fixed costs. In cases where the firm would lose money under the first-best prices, the firm must be subsidized to remain solvent. This subsidy can be provided in three ways: (1) directly, (2) by adding an access charge without changing the usage price, if access demand is fixed, or (3) by resorting to Ramsey prices.

These concepts can be readily translated to situations in which demand fluctuates over time periods. Suppose each day consists of two periods called peak and off-peak, with demand being greater in the peak. For convenience, suppose the two periods are the same length (twelve hours)[5] and that demand is constant over all the hours in each period.

Figure 8.3 depicts the three possibilities for the relation of demand in each period to capacity when price is at marginal cost. The subscript p refers to peak and the subscript o refers to off-peak. In panel (a), demand in each period can be met with existing capacity when price is set at marginal cost. In this case the first-best price is the same in both periods, namely marginal cost. The quantity demanded is q_p^* in the peak and q_o^* in the off-peak, for a total daily output of $q_p^* + q_o^*$. There is excess capacity throughout the day, and the firm loses money if it is not subsidized. In panel (b), the quantity demanded exceeds capacity in the peak but not in the off-peak when price is set at marginal cost. The optimal price in the off-peak is marginal cost, and the peak-period price must be raised, for optimality, until demand equals capacity in the peak. The firm earns revenues in excess of variable cost from the peak-period customers, but not from the off-peak customers. Hence, to the extent that fixed costs are covered, peak-period customers bear these costs. In panel (c), demand exceeds capacity in both periods when price equals marginal cost. For optimality, price is raised in each period until the excess demand is eliminated. Revenues in excess of variable costs are earned in each period. This excess is the rectangle *HGJE* in the peak and rectangle *ABJE* in the off-peak. Because the peak and off-peak are of equal length, the average revenue

5. This assumption is convenient for determining whether revenues in both periods cover fixed costs, because it allows average revenues per period to be the simple average of revenues over the two periods. With periods of unequal length, a weighted average is required, with the weights being proportional to the length of each period. The concepts are the same, but the notation and language is easier with equal lengths.

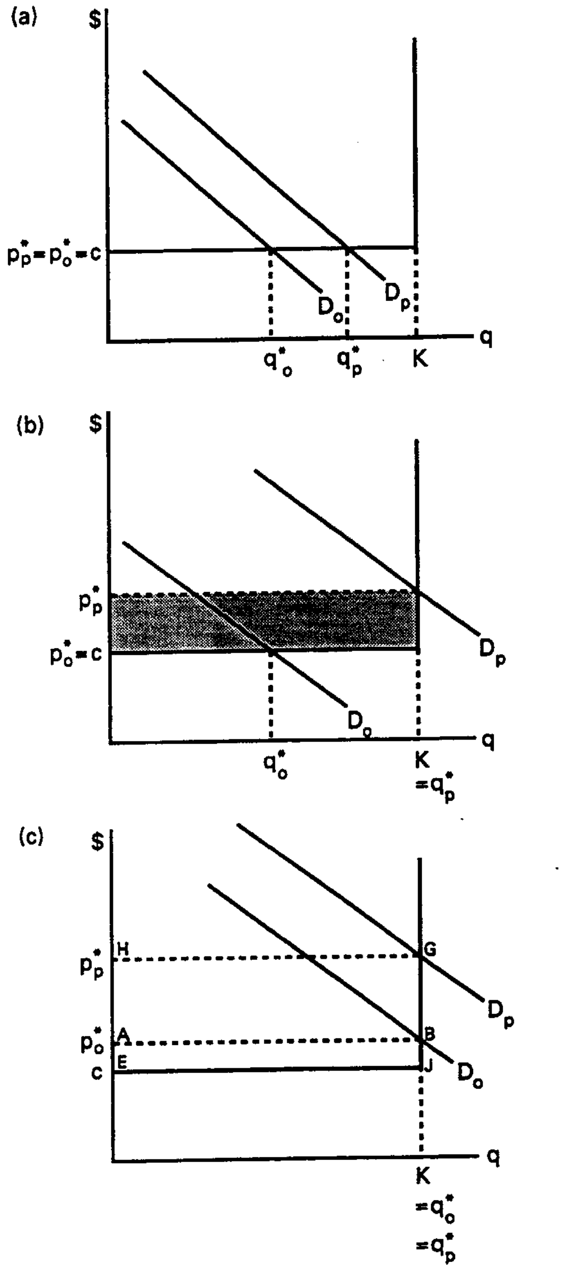

Figure 8.3
Optimal prices with fluctuating demand

per period in excess of variable cost is the average of these two areas; this average is compared to F, fixed costs per period, to determine whether the firm breaks even without subsidization.

As in the case of nonfluctuating demand, the first-best pricing rule with fixed capacity and fluctuating demand is: in each period, price at marginal cost unless doing so results in more quantity demanded in that period than can be met with the available capacity, in which case raise price until demand equals capacity.

As we have seen, this pricing rule can result in "wasted" capacity, the existence of which is not suboptimal in the short run when capacity is fixed. (It simply denotes the need to reduce capacity in the long run.) The rule can also result in congestion (as defined below), the existence of which is also not suboptimal.

This latter point requires elaboration. With fixed capacity, congestion usually occurs before output reaches full capacity. The classic example is freeway traffic. As more cars enter a freeway, traffic slows down. Such congestion imposes costs on drivers in terms of longer travel times. From a social perspective, the marginal cost of output includes both the cost to the firm of providing the output plus the extra cost borne by consumers through increased congestion.

The marginal cost curves in figures 8.1–8.3 do not, by their shape, permit congestion. In these graphs, marginal cost is constant until capacity is reached, at which point greater output is not possible. With congestion, marginal cost usually increases gradually as output approaches capacity. For example, it is not the case that traffic on a freeway flows at a constant speed as more and more cars are added and then immediately stops completely when the freeway's capacity is reached. The slowdown in traffic is more gradual, with speeds dropping as congestion gets worse and worse. Figure 8.4 contains a marginal cost curve for this type of situation. The marginal cost of one extra car, including the cost imposed on other drivers through the extra congestion this car produces, is constant for low levels of traffic (when there are so few cars that no congestion occurs) and increases gradually as more cars are added; eventually, capacity K is reached and no additional cars are possible. This curve is conceptually similar to that in the previous figures, except that the lower-right corner of the marginal cost curve in the previous figures is smoothed to allow a gradual increase in marginal cost as capacity is approached.

The concepts of optimal pricing given capacity are the same under

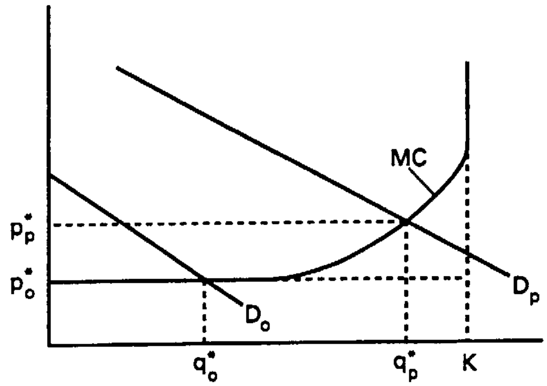

Figure 8.4
Optimal output can entail congestion

this marginal cost curve: the first-best price is marginal cost unless this price results in demand exceeding capacity, in which case price is raised to equate demand with capacity. In the off-peak, the first-best price is p_o^* and there is no congestion. In the peak, the first-best price is p_p^*, which results in some congestion. While the absolute capacity K has not been reached, the marginal-cost curve has started to rise indicating, in the case of a freeway, that there is congestion and that each additional driver is making the congestion worse. This example shows that the first-best prices can result in congestion when capacity is fixed.

8.3 Riordan's Mechanism for Inducing First-Best TOU Prices with Fixed Capacity

To describe Riordan's proposal, we return to the more stark cost situation where marginal cost is constant at c up to capacity K, beyond which extra output is impossible. Suppose the regulator observes c and K, but does not know the demand curves that the firm faces in the peak and off-peak and consequently does not know the optimal prices. Riordan has devised a subsidy mechanism that the regulator can impose on the firm under which the firm is induced, through its pursuit of profit, to charge the first-best price and sell the first-best output in each period. The mechanism consists of the following. In *each* period, the regulator pays the firm a subsidy $S(p)$ that depends on the price that the firm charges in that period. The particular form of the subsidy is:

$$S(p)=\begin{cases}F-(p-c)K & \text{if } p\geq c\\ 0 & \text{if } p<c.\end{cases}$$

That is, if the firm prices below marginal cost c in the period, the regulator pays the firm nothing. If the firm prices at or above marginal cost in the period, the regulator pays the firm its fixed costs of capacity (F) minus the amount by which price exceeds marginal cost for each unit of output that can be produced *at capacity*.[6] For example, if the firm has a capacity of 5,000 units and prices at \$6 when its marginal cost is \$4, the regulator pays the firm its fixed costs minus \$10,000 (that is, (6–4) · 5,000).

To calculate this subsidy, the regulator needs to know the fixed and marginal costs of the firm, its capacity, and the price it charges in the period. Demand information is not required. Note that $S(p)$ might be negative. If the firm prices high enough above marginal cost, the quantity subtracted [namely, $(p-c)K$] could exceed the fixed costs (F) such that the "subsidy" is negative. In this case, the regulator takes money away from the firm; that is, S(p) is actually a subsidy if positive and a tax on the firm if negative.[7]

The firm's profit is the sum of its profit, including subsidy, in the peak and its profit, including subsidy, in the off-peak:

$$\pi=(p_p-c)q_p-F+S(p_p)+(p_o-c)q_o-F+S(p_o), \tag{8.1}$$

where q_p and q_o are the quantities sold in the peak and off-peak, respectively. Note that fixed costs F are defined on a per-period basis rather than per day, such that F is incurred in both the peak and off-peak. This approach is not a restriction: if fixed costs are actually incurred on a daily basis, then our F is simple one-half of the daily fixed costs.

The firm cannot sell more than its capacity K in each period. Therefore, q_p is equal to capacity if quantity demanded in the peak exceeds capacity and is equal to quantity demanded if quantity demanded in the peak is less than capacity; and similarly for q^o and quantity de-

6. Note that the difference between price and marginal cost is multiplied by capacity, not output, which may be less than capacity.

7. The regulator need not actually subsidize or tax the firm directly. If access demand is fixed, the regulator can implement the subsidy/tax by raising or lowering the access fee. An access fee would be required to cover the firm's fixed costs at the first-best prices anyway; Riordan's scheme simply adjusts the level of the fee in a way that depends on the price that the firm charges in each period.

manded in the off-peak.[8] The question is: what price will the firm charge in each period and what quantity will it sell in each period to maximize profits?

The firm will clearly never choose to price below marginal cost in either period. With $p < c$, the firm does not cover its variable costs [that is, $(p - c)q$ is negative] and it receives no subsidy. It therefore loses its entire fixed costs plus a portion of its variable costs. It would clearly be better off raising price at least to marginal cost, at which point it would cover its variable costs and receive a subsidy for its fixed costs, thereby earning zero profits.

With price at least as high as marginal cost in each period, the subsidy in each period is $S(p) = F - (p - c)K$. Substituting this into equation 8.1, the total profits of the firm are:

$$\pi = (p_p - c)q_p - F + F - (p_p - c)K + (p_o - c)q_o - F + F - (p_o - c)K$$
$$= (p_p - c)(q_p - K) + (p_o - c)(q_o - K).$$

Consider the profits made in the peak, namely $(p_p - c)(q_p - K)$; the arguments regarding the off-peak are analogous. Because the firm does not price below marginal cost, the term $(p_p - c)$ is either zero or positive. However, since output cannot exceed capacity, the term $(q_p - K)$ is either zero or negative. There are four possibilities for the magnitudes of these two terms:

	Output is below capacity $(q_o - K) < 0$	Output equals capacity $(q_p - K) = 0$
Price equals marginal cost $(P_p - c) = 0$	*A*: profit = 0	*B*: profit = 0
Price exceeds marginal cost $(P_p - c) > 0$	*C*: profit < 0	*D*: profit = 0

8. We are implicitly assuming that the firm must meet demand in each period. The possibility that the firm leaves some demand unmet is discussed in the next footnote. We also assume, following Riordan, that demand in each period is not affected by the price in the other period. In reality, demand in each period usually depends on the prices in both periods. For example, if price in the peak is raised, some customers might shift their consumption to off-peak times, such that off-peak demand increases with the peak price. The specification can be generalized to allow for this possibility; however, the demonstration that the firm charges optimal prices becomes less transparent.

The most profit the firm can earn under this subsidy scheme is zero, which occurs in cases *A*, *B*, and *D*. The firm will clearly not choose to price above marginal cost and sell less than capacity because doing so will result in negative profits (case *C*). The firm will price at marginal cost if the quantity demanded at that price is less than or equal to capacity (cases *A* and *B*, respectively). Or, the firm will raise price above marginal cost and sell an amount equal to capacity (case *D*). This last case is possible only if the quantity demanded when price equals marginal cost exceeds capacity (such that raising the price results in demand equaling capacity).[9]

Stated succinctly: The firm will price at marginal cost in the peak as long as doing so results in demand less than or equal to capacity. If demand exceeds capacity at marginal-cost pricing, the firm will raise price until demand equals capacity. As described in section 8.2, this pricing rule is optimal.

The decision process of the firm under this subsidy can also be shown graphically. Figure 8.5 depicts the two possibilities: demand is less than capacity when price is set at marginal cost (panel a) or demand exceeds capacity at marginal cost pricing (panel b). Consider panel (a) first. The first-best price is $p_p^* = c$. At this price, the firm covers the variable costs of producing q_p^* units of output and receives a subsidy for its fixed costs, such that its peak-period profit is zero. That is: $S(p_p^*) = F - (p_p^* - c)K = F$, such that profit in the peak is $(p_p^* - c)q_p^* - F + S(p_p^*) = -F + F = 0$.

Suppose the firm were to raise its price above marginal cost, to say r, in an attempt to earn more profit. Its profit from operations would

9. If the firm can choose not to meet some demand, the firm could price at marginal cost even though demand at that price exceeds capacity: the firm would simply sell as much output as its capacity allows and leave the excess demand unsatisfied. That is, the firm could choose to be in case *B* by not meeting demand, rather than case *D*. Because the firm earns zero profits whether it (1) charges marginal-cost price and leaves excess demand unmet, or (2) raises price above marginal cost until demand equals capacity, the firm is indifferent between these two courses of action. Consequently, there is no guarantee that the firm would choose the latter, which is optimal, instead of the former, which is not. This problem can be solved by having the regulator levy a stiff penalty if the regulator obtains evidence that the firm is not meeting demand. Even if the chances are very low that the regulator would discover that demand is not being met, the firm will choose to raise price to choke off excess demand rather than risk the penalty. That is, zero profits without any risk is preferable to the firm to zero profits with the risk of a penalty, no matter how slight the possibility that this penalty would actually be levied.

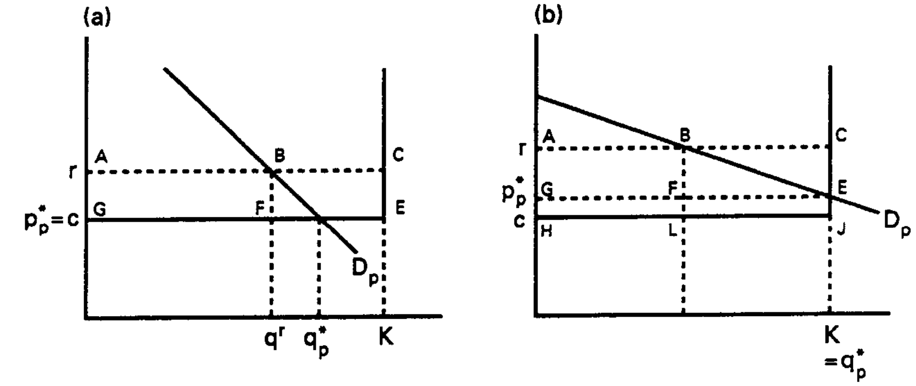

Figure 8.5
Firm's choices under Riordan's mechanism

increase by the area *ABFG*, namely, the amount by which price exceeds marginal cost times the quantity sold. However, the subsidy the firm obtains from the regulator would be diminished by the area *ACEG*, which is the amount by which price exceeds marginal cost times the *capacity* of the firm's plant. Because the subsidy is reduced by more than the firm's operating profits increase, the firm earns less profit at r than at p_p^*. The driving factor in this comparison is that the difference between price and marginal cost is multiplied by the quantity sold to obtain the increase in operating profits, and yet is multiplied by capacity to obtain the reduction in subsidy. Because the quantity sold cannot exceed capacity, the firm cannot make more money by raising price over marginal cost.

Consider now the possibility in panel (b) of demand exceeding capacity when price is set at marginal cost. The first-best price is $p_p^* > c$, at which demand equals capacity. At this price, the firm earns operating profits equal to the area *GEJH*, minus its fixed costs. The subsidy consists of the fixed costs of the firm minus the area *GEJH*. Thus the firm exactly breaks even. Suppose now that the firm were to raise its price to r in an attempt to earn additional profits. Its operating profits would increase by area *ABFG* minus *FEJL*. However, the firm's subsidy would be reduced by the area *ACEG*. The net effect is a loss for the firm. The basic reason again is the fact that operating profit is calculated on the basis of quantity sold and the subsidy is calculated on the basis of capacity. Because raising price above its optimal level p_p^* necessarily reduces demand below capacity, the increase in operating profit is dominated by the reduction in subsidy.

The same arguments hold for the portion of profit that is derived from the off-peak. The product $(p_o - c)(q_o - K)$ is negative if the firm acts nonoptimally (by pricing above marginal cost while selling less than capacity) and is zero if the firm prices optimally (at marginal cost as long as demand does not exceed capacity and above marginal cost only if necessary to reduce demand to capacity). The firm therefore chooses to act optimally, earning zero instead of negative profit.

8.4 Optimal Capacity

Over time, capacity can often be adjusted; or, when a firm is being established, the capacity for the firm's plant(s) is chosen. The question is: Given demand in each period, what is the optimal capacity? Following our stylized cost specification, suppose that capacity can be constructed at a constant cost of b per unit. That is, we assume that it costs b dollars more to build a plant with capacity $K + 1$ than to build a plant with capacity K.[10] For reasons that become clear later, we represent capacity costs in terms of a flow of expenditures, such as the mortgage payments on a loan for the funds to build the capacity, or the lease payments for renting the capacity. The quantity b is therefore the extra payment per period for an extra unit of capacity.

As before, given capacity, the variable cost of producing output is assumed to be a constant c per unit. Long-run marginal cost is therefore $b + c$: the cost of expanding capacity by one unit plus the cost of producing an extra unit with the extra capacity. Short-run marginal cost, given capacity, is c for output up to the capacity and can be considered either undefined or infinite for higher levels of output.[11]

---

10. Alternatively, one can think of b as representing the cost of increasing capacity by one unit given an existing capacity. This way of considering b is appropriate for examining adjustments in capacity, while the concept in the text is appropriate when original capacity is being constructed.

11. By definition, long-run marginal cost is the cost of an extra unit of output when capital (in this case, capacity) is adjusted optimally for each level of output. Suppose the firm is producing output q and has a capacity that is exactly equal to this q; this capacity is optimal since no more or less capacity is available than needed for q units of output. To produce an extra unit, the firm must increase its capacity by one unit, which costs b; given this extra capacity, the firm must also pay the variable cost of the extra unit, which is c. Long-run cost is therefore $b + c$. Short-run cost, by definition, is the cost of an extra unit when capital (in this case, capacity) is fixed at a given level. If the firm is producing less than its capacity, the cost of producing an extra unit is c. If, however, the firm is producing at capacity, then an extra unit simply cannot be produced (in the short run, when capacity is fixed). Short-run marginal cost is therefore c up to capacity and then becomes undefined or infinite.

Consider first the optimal capacity when demand does not fluctuate over time. Figure 8.6 depicts the situation. Long-run marginal cost *(LRMC)* is a constant $b + c$. The demand curve gives, at any quantity of output, the amount that consumers are willing to pay for an extra unit of output (or, stated alternatively, the demand curve gives the value that consumers obtain from an extra unit of output). For example, at output level q^1, consumers are willing to pay p^1 for an extra unit, which is the vertical distance *XZ*. This value to consumers can be decomposed into two parts to facilitate the analysis of optimal capacity. Part of this value is required to cover the variable costs of production given capacity; that is, cost c. The remaining portion, distance *YZ*, is therefore the amount that consumers are willing to contribute for additional capacity. In other words, *YZ* is the amount consumers would be willing to pay to have capacity expanded from q^1 to $q^1 + 1$, such that an extra unit of output could be produced.

Applying these ideas to all levels of output, the amount that consumers are willing to pay for additional capacity is the amount by which the demand curve is above c (that is, the vertical distance from c up to the demand curve). For high enough levels of output (beyond q^m in the graph), the demand curve is below c. This indicates that consumers are not willing to pay anything for additional capacity: they value an extra unit at less than even the variable cost of producing it and are therefore not willing to contribute anything to expanding capacity to allow more production. Stated completely: the amount that consumers are willing to pay for additional capacity is the differ-

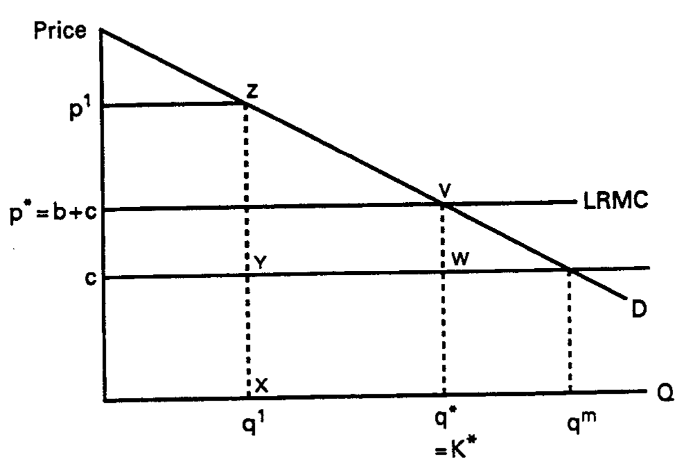

Figure 8.6
Optimal capacity with nonfluctuating demand

ence between the demand curve and c with a minimum of zero. Graphically, it is the the demand curve above c, with a kink, becoming flat (zero), at q^m.

As long as consumers are willing to pay more for additional capacity than the extra capacity costs, the capacity should, from a social perspective, be provided. Capacity costs b per unit; therefore, capacity should be expanded whenever consumers are willing to pay more than b for additional capacity. At $q^{1,}$ consumers are willing to pay distance YZ for extra capacity; since YZ exceeds b, extra capacity should be provided. As more capacity is provided, consumers' willingness to pay for additional capacity decreases, until, at q^*, the amount that consumers are willing to pay for extra capacity (distance WV) exactly equals the cost of extra capacity b. This is the optimal capacity: consumers are willing to pay no more or less than the cost of extra capacity. With capacity $K = q^*$, the optimal price is p^*, which equates quantity demanded with capacity.

This analysis could have been performed much more simply. From standard microeconomics, we know that the first-best price and output in the long run is where *LRMC* intersects the demand curve: at p^* and q^*. The optimal level of capital (in this case, capacity) is that which is cost minimizing for the optimal level of output. The least-cost way of producing output q^* is with a capacity of exactly q^*: any less capacity would be insufficient and any more would be unnecessary.

While this latter method for determining optimal capacity is more straightforward in the case of fixed demand, it does not generalize as readily to the case of fluctuating demand. The explanation based on identifying consumers' willingness to pay for capacity provides the key for examining capacity choice when demand fluctuates.

Return now to the situation with peak and off-peak demand. With two periods, optimal capacity is determined by comparing the cost of extra capacity in both periods with the willingness to pay of customers in both periods. The cost of additional capacity in each period is b, such that the cost over the two periods, peak and off-peak, is $2b$. (Recall that costs of capacity are expressed in terms of a flow of expenditures, as under a lease or mortgage.) An extra unit of capacity should be provided if the amount that consumers in the peak are willing to pay for extra capacity, plus the amount that off-peak consumers are willing to pay, exceeds $2b$. Or, stated in per-period terms, an extra unit of capacity should be provided if consumers' willingness to pay for capacity, when averaged over the two periods, exceeds b, the cost per period. As these statements make evident, extra capacity could be desirable even though off-peak consumers are not willing to pay as much as b for extra capacity, as long as the peak consumers are willing to pay sufficiently more than b to make up the difference.

We apply these concepts to the demand curves in figure 8.7. Consumers in the off-peak are willing to pay the distance between their demand curve and c for an extra unit of capacity, up to q_o^m; they are not willing to pay anything for extra capacity in excess of q_o^m. Peak consumers are willing to pay the vertical distance between their demand curve and c, up to q_p^m, and nothing beyond. Each of these two groups of customers pays in their periods only (the peak period consumers paying in the peak and the off-peak consumers paying in the off-peak). The average willingness to pay per period is therefore the willingness to pay of peak consumers averaged with that of off-peak consumers. We can construct a new curve that represents this average willingness to pay. For example, at q^1, off-peak consumers are willing to pay AE for an extra unit of capacity, and peak consumers are willing to pay AG. The average of these two amounts is AF, which becomes the point on the new curve associated with q^1. That is, at q^1, the average amount that consumers in the two periods are willing to pay for capacity is AF. This concept is applied to all levels of capacity up to q_o^m. Beyond this point, consumers in the off-peak are not willing to pay anything for additional capacity. At q^2, for example, consumers in the off-peak are willing to pay nothing; however, peak consumers

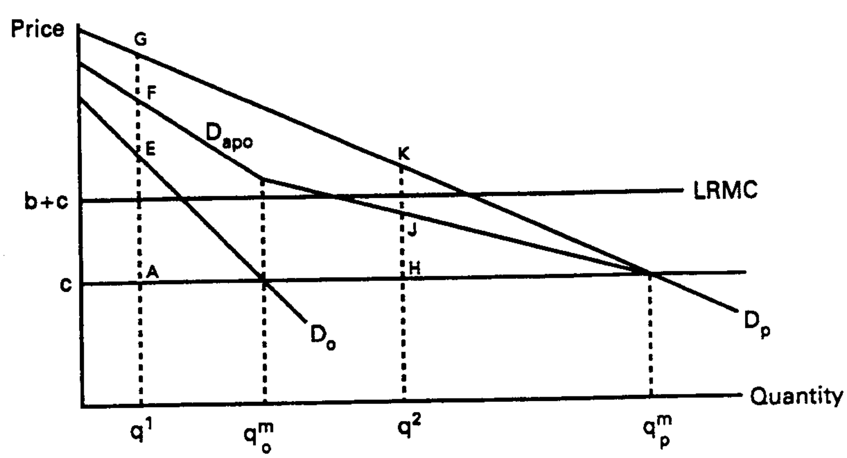

Figure 8.7
Average willingness to pay for capacity

are willing to pay *HK*. The average amount consumers are willing to pay is therefore half of *HK*; this is *HJ* and constitutes another point on the new curve. Stated succinctly, the new curve D_{apo} (where the subscript refers to the average of peak and off-peak) is the vertical average of D_o and D_p, each truncated at *c*. The vertical distance between this curve and *c* gives the average amount that consumers are willing to pay per period for extra capacity.

The optimal capacity is identified by comparing the average willingness to pay for extra capacity with its cost per period. Figure 8.8 contains the same demand and cost curves as figure 8.7 and also identifies the optimal capacity, prices, and outputs. At K^*, the amount that consumers are willing to pay for extra capacity, averaged over the peak and off-peak periods, is the distance *NM*. This amount exactly equals the cost of extra capacity per period, such that K^* is the optimal capacity. Given this capacity, the (short-run) marginal cost of output is c up to output K^*, after which no more output can be produced in the period. The rules derived in section 8.2 are used to determine optimal prices given this capacity. In the off-peak, the optimal price is $p_o^* = c$: with price set at the cost of producing an extra unit with the given capacity, the quantity demanded is less than capacity; hence that price is optimal. In the peak, pricing at *c* results in demand exceeding capacity. The optimal price is attained by raising price until demand equals capacity. This occurs are p_p^*.

Note that in this case, peak consumers pay the entire cost of the

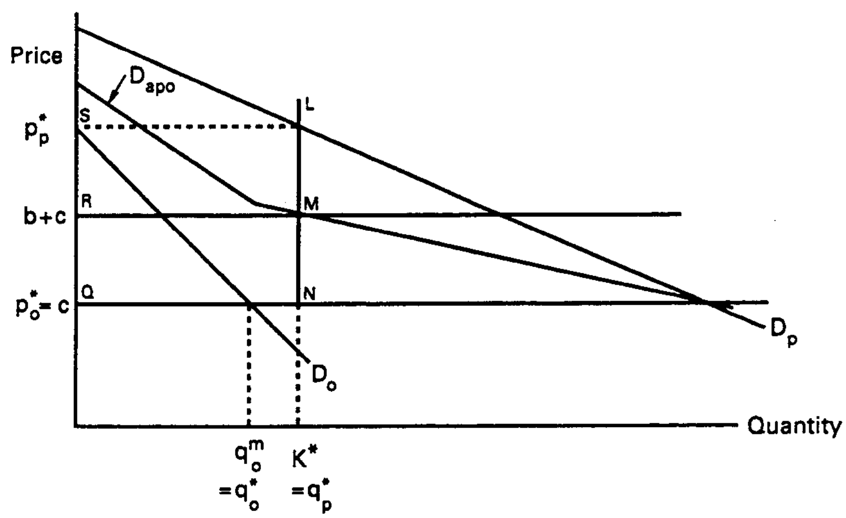

Figure 8.8
Optimal capacity with peak and off-peak demands

capacity. The total cost of capacity per period is b times K^* (the cost per unit times the number of units of capacity), which constitutes the area *NMRQ*. Over the two periods, peak and off-peak, the cost of capacity is twice this amount. Peak consumers pay a price p_p^*, which exceeds the variable cost of production c by the distance *NL*. With consumption at $q_p^* = K^*$, the peak consumers pay a total of the area *NLSQ* in excess of the variable costs of their consumption. By construction, the area NLSQ is twice as large as the area *NMRQ*: the distance *NM* (the average that consumers are willing to pay for extra capacity) is the average of NL (the amount peak consumers are willing to pay) and zero (the amount off-peak consumers are willing to pay at this level of capacity); hence *NM* is half of *NL* such that *NMRQ* is half of *NLSQ*. Peak consumers are, in this case, paying the entire cost of capacity over both periods. Off-peak consumers face a price that equals c and hence pay only the variable costs associated with their own consumption without contributing to the costs of capacity.

This outcome reflects the relative levels of the demand curves in the two periods. After capacity is sufficient to meet off-peak demand (that is, after a capacity of q_o^* has been provided), any additional capacity sits idle in the off-peak. For extra capacity to be warranted, the peak customers must value the extra capacity in the peak sufficiently to pay not only for the cost of the extra capacity in the peak but also for the cost of having the extra capacity in the off-peak, where it sits idle and provides no benefits.

If the difference between demand in the two periods is not as great as in figure 8.8, the optimal price can exceed c in both periods, such that off-peak consumers contribute along with peak consumers to the cost of capacity. Figure 8.9 illustrates such a case. D_o does not drop below c until after D_{apo} exceeds c by exactly b (i.e., intersects *LRMC*). The amount by which D_{apo} exceeds c is the average willingness to pay for extra capacity; this amount exactly equals b, the cost of extra capacity, at capacity K^*. At this capacity, pricing at c in the off-peak would result in demand exceeding capacity, such that the optimal off-peak price is above c; specifically, at p_o^*. Similarly, the optimal peak price is p_p^*. At the optimal capacity and prices, consumers in each period are willing to pay a positive amount for extra capacity (unlike the situation in figure 8.8 in which peak consumers were willing to pay for extra capacity but off-peak consumers are not). Consumers in each period pay a price in excess of the variable cost of production c and hence contribute to the cost of capacity.

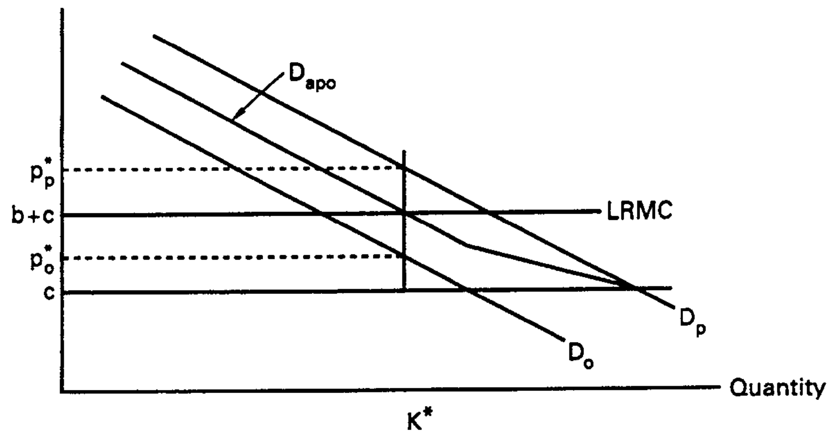

Figure 8.9
Optimal capacity with peak and off-peak demands: case 2

8.5 Riordan's Mechanism Applied to Capacity Choice

Riordan suggests two ways in which his subsidy mechanism can relate to the task of inducing the firm to choose the optimal capacity. Neither of the suggestions is a complete solution to the problem, as Riordan points out. We describe below each of the suggestions and discuss their potential and limitations.

First, suppose that the subsidy mechanism described in section 8.3 is applied with the firm choosing its own capacity. Recall the formula for the subsidy:

$$S(p) = \begin{cases} F-(p-c)K & \text{if } p \geq c \\ 0 & \text{if } p < c. \end{cases}$$

When capacity is fixed, as assumed in section 8.3, the firm can affect the amount of subsidy it receives only through its choice of prices. However, if the same subsidy formula is used and capacity is not fixed, the firm can affect its subsidy through its choice of capacity as well as its choice of prices. The firm's capacity choice determines K in the subsidy formula. It also determines F, because F is the cost of the capacity. In particular, given our specification that each extra unit of capacity costs b per period, F is actually bK per period.

Under this subsidy scheme, the firm earns at most zero profits at any level of capacity. Suppose the firm chooses some capacity, say, K^1. Once this capacity is given, the firm chooses the prices that maximize its profit, including the subsidy, with this capacity. As demonstrated in section 8.3, the firm earns zero profit if it chooses the prices that are optimal for its level of capacity and earns negative profit if it chooses any other prices. The firm of course chooses the optimal prices so as to avoid losses. Hence, if the firm chooses K^1 capacity, it earns zero profit.

The same argument applies for any level of capacity: the firm earns zero profit by choosing the prices that are optimal for that level of capacity. In particular, the argument holds for the optimal capacity: the firm earns zero profit if it charges the optimal prices given the optimal capacity.

Because the firm earns zero profit at *any* level of capacity, it has no reason *not* to choose the optimal capacity. In this sense, the subsidy mechanism can be considered consistent with the firm choosing the optimal capacity.

While the firm has no reason not to choose the optimal capacity, it also has no reason to choose it. The mechanism therefore is consistent with optimal capacity choice, but does not necessarily induce it. This lack of a positive incentive is, of course, the difficulty of the mechanism in situations with variable capacity.

Riordan's second suggestion relies on the notion that if the regulator knows the optimal capacity, this information can be used to induce the firm to choose it. In essence, the indifference of the firm among capacities (all of which result in zero profit under the subsidy mechanism) can be broken by assessing the firm a penalty for not choosing the optimal capacity. The firm then chooses to earn zero profit with the optimal capacity rather than negative profit (zero profit minus the penalty) at any other capacity level.

Riordan suggests a revised formula for the subsidy that incorporates the regulator's knowledge of the optimal capacity. The revised formula implicitly levies a penalty on the firm if it chooses a nonoptimal capacity.

The revised subsidy formula is the following:

$$RS(p) = \left\{ \begin{matrix} F^* - (p-c)K^* \\ 0 \end{matrix} \right\} \begin{matrix} \text{if } p \geq c \\ \text{if } p < c. \end{matrix}$$

This is the same formula as earlier, but with the optimal capacity K^* replacing the firm's actual or chosen capacity (and the cost of the optimal capacity F^* replacing the cost of the firm's chosen capacity). The firm is allowed to choose any capacity it wants, but its subsidy is calculated on the basis of the optimal capacity. That is, the subsidy is calculated as if the firm chose the optimal capacity, even if the firm chooses some other capacity.

Under this revised subsidy, the firm earns zero profit if it chooses the optimal capacity, just as under the original subsidy. However, if the firm chooses any other capacity, its profit under the revised subsidy is negative. This fact is demonstrated in figure 8.10. At the optimal capacity and prices (K^*, p_o^*, and p_p^*), the firm earns zero profits after the subsidy. Suppose the firm were to reduce its capacity below the optimal level, to say K^1. We can show that the firm will earn negative profits at this nonoptimal capacity under the revised subsidy formula. With K^1, the firm prices at p_p^1 in the peak, at which demand equals capacity. Given the way the curves are drawn in this illustration, the chosen off-peak price is not affected by the reduction in capacity. Consider now the firm's profits over the peak and off-peak periods. The cost of capacity in each period is bK^1, which is the area *AEKL*. Over both periods, the cost is twice this amount: area *AFJL*. Revenues in the off-peak exactly cover variable costs. Revenues in the peak cover variable costs plus the amount *AGHL*. The firm, before the subsidy, therefore earns a profit of *AGHL* minus *AFJL*, which is area *FGHJ*. Consider, however, the subsidy. In the off-peak, the firm obtains subsidy equal to the cost of the optimal capacity, which is area *AENM* for the one period. Because price equals variable cost in the off-peak, nothing is subtracted from this amount. In the peak period the subsidy includes the cost of optimal capacity, *EFRN* (which is the same size as *AENM*). However, because the firm is pricing above

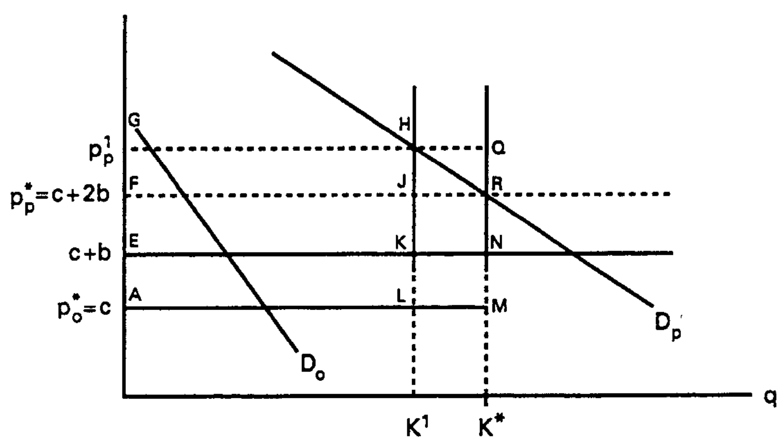

Figure 8.10
Choice of optimal capacity under revised subsidy

variable cost, the subsidy is reduced by the amount by which price exceeds variable cost times the *optimal* capacity, namely, area *AGQM*. The sum of the subsidy over both periods is therefore *AFRM* (the cost of the optimal capacity over both periods) minus *AGQM* (the excess of price over variable cost times optimal capacity), which is negative by the amount *FGQR*. That is, the subsidy becomes a tax in the amount of *FGQR*. The firm earns profits before the subsidy of *FGHJ*, but then loses more than this amount through the tax (negative subsidy), which is *FGQR* in size. On net, the firm loses *JHQR*. Clearly, the firm is better off choosing the optimal capacity and earning zero profits. Similar arguments can be made with other demand curves and with the firm increasing instead of decreasing capacity from the optimal level.

The revised subsidy mechanism indeed induces the firm to choose both the optimal capacity and the optimal prices. However, to implement the mechanism, the regulator must know the optimal capacity. This informational requirement is stringent, and, more fundamentally, conflicts with the purpose of establishing incentive structures for regulated firms. Regulatory mechanisms are established to induce the firm to act optimally when the regulator does not know exactly what the optimal outcome is. The revised subsidy mechanism assumes that the regulator knows the variable that the mechanism is supposed to induce the firm to choose, namely the optimal level of capacity. The regulator knows this not in conceptual terms only (as in knowing that price should equal marginal cost), but knows the exact number. If the regulator knows the optimal capacity, the regulator can simply mandate that the firm provide that capacity. A regulatory mechanism is not required.

Riordan's suggestions regarding optimal capacity are valuable, however, at a more fundamental level. The revised formula reflects the important concept that the regulator can use information on the optimal capacity to penalize the firm for not choosing it. While the formula as specified requires that the regulator have complete knowledge of the optimal capacity, the concept introduces the possibility of developing mechanisms that utilize partial knowledge. That is, the regulator might have some evidence of whether the firm's chosen capacity is optimal, without knowing the actual level of optimal capacity. Regulatory mechanisms can perhaps be devised that use this information to push the firm, if not to, then at least close to, the optimal capacity. This issue of inducing optimal capacity in the context of fluctuating demand is an important area for further research.