12 Natural Monopoly Regulation and Electric Power

The theory of natural monopoly and alternative policy solutions were the main topics of the last chapter. The most common policy solution in the United States—regulation—is the subject of this chapter. Electric power will be used as our primary example, although some references to other regulated natural monopolies will be made, such as telephone service, natural gas distribution, and water supply.

In examining regulation, it is useful to keep in mind the benefits and costs. The benefit is to reduce deadweight losses in efficiency that would exist under unregulated monopoly. The costs are less obvious, but include the direct costs of regulatory agencies as well as unintended side effects of regulation. An important side effect is higher costs because of changed incentive structures of regulated firms.

The technology for centralized production and distribution of electric power was first put into operation in September 1882 in New York City, where Thomas Edison began producing electricity in the famous Pearl Street plant of the Edison Electric Illuminating Company. In 1876, Alexander Graham Bell received a patent on the telephone. Hence these two important utilities have been in existence only for about 120 years.

During the early years the common method of regulation was by the award of a franchise by the town or city. The city would normally grant a franchise for exclusive operation within the city in return for an agreement by the firm that it would provide a certain quality of service at certain rates. As the technology changed and firms began to serve larger regions and even entire states, community regulation became ineffective. It also became more efficient to have regulatory experts at the state level only, rather than having duplicate experts in every community. Finally, the franchise agreement was not very flexible in dealing with constantly changing economic conditions. All these factors led to the institution of the regulatory commission at the state level. The first such commissions began in 1907 in New York, Wisconsin, and Georgia. All states have regulatory commissions today.

Most of the discussion in this chapter will be concerned with state regulatory commissions, but we should observe that several federal regulatory agencies are also involved in natural monopoly regulation. For example, the Federal Energy Regulatory Commission (FERC) regulates interstate wholesale transactions of electricity and natural gas pipelines. The Federal Communications Commission (FCC) regulates interstate telephone service.

The regulatory commission is usually appointed by the governor or president, although in some states the commissioners are elected or appointed by the legislature. A typical commission consists of three to twelve commissioners who are assisted by a staff of up to 2,500 members. Many state commissions are smaller. The staff members are trained in accounting, engineering, economics, and the law.

The commissions typically focus on prices charged by the monopolist. Their mandate is usually somewhat vague, such as requiring that prices be "just and reasonable" and that there be no "undue discrimination." The procedure is that prices are set in rate cases and are generally fixed until the next case. Rate cases are similar to civil court cases. Expert witnesses on various topics are heard (for instance, economists testify on the cost of capital) with the final decision being made by the commissioners. In certain cases, appeal to higher courts is possible.

In the next section we will discuss the typical features of the rate case. The following sections focus on so-called rate-level and rate-structure issues. The rate level concerns the relation of overall revenues to costs, whereas the rate structure deals with how individual prices are set. The last part of this chapter deals with major changes now under way to deregulate electric power.

The Rate Case

The company being regulated usually initiates the rate case by applying for an increase in the prices (or rates) that it charges. A "test period" is selected—usually the last accounting period. Adjustments are made to reflect known changes that affect financial data. For example, it might be known that a natural gas distribution company will have to pay higher prices in the future for the gas it purchases from the pipeline. These changes are factored into the computation of the company's rate of return.

Accounting Equation

In essence, the following accounting equation describes the process:

$$\sum_{i=1}^{n} p_i q_i = \text{Expenses} + s(RB) \tag{12.1}$$

where p_i is the price of the ith service, q_i is the quantity of the ith service, n is the number of services, s is the allowed or "fair" rate of return, and RB is the rate base, a measure of the value of the regulated firm's investment.

The underlying idea, of course, is that the company's revenues must just equal its costs, so that economic profit is zero. Notice that economically efficient prices are not required by the equation, only prices that cover total costs.

As observed earlier, it is common to discuss the regulation of natural monopoly in two parts: the rate-level problem and the rate-structure problem. The rate-level problem is concerned primarily with finding s so that the company will have the appropriate level of earnings on its investment (or rate base). The rate structure problem deals with issues of price discrimination among customer classes and products, that is, the p_i on the left-hand side of equation (12.1).

To begin a rate case, a company will submit detailed financial exhibits that show, for example, that at current prices its rate of return on its rate base for the test period is too low. It probably will argue that its true cost of capital is such that it needs a higher return in order "to

Table 12.1
North Carolina Natural Gas Corporation Statement

	Year Ended Dec. 31, 19xx	Adjustments for Rate Increase	After Adjustments for Rate Increase
Revenues	$29,572,747	$2,832,332	$32,405,079
Expenses			
(1) Purchased gas	$19,411,430		$19,411,430
(2) Labor	2,968,387		2,968,387
(3) Depreciation	1,234,798		1,234,798
(4) Taxes	4,338,300	358,500	4,696,800
Total expenses	27,952,915	358,500	28,311,415
(5) *Net Operating Income*	1,619,832		4,093,664
Rate Base			
Plant less depreciation	41,871,387		41,871,387
Working capital	1,002,989		1,002,989
(6) Total	42,874,376		42,874,376
(7) Rate of return [(5)/(6)]	3.77%		9.54%

continue to attract capital." Basically, what the company is arguing is that the prices it is now charging—as set in the last rate case—are too low. The commission staff will probably argue that the company's requested rate of return is too high, and that prices need not be raised as much as the company wants.

Eventually, after much testimony, the commissioners must make a determination of what they believe the rate of return should be, that is, the value of s. Assuming that they choose an s value that is higher than the company is now earning, the prices on the left-hand side in equation (12.1) will be adjusted to yield the new rate of return allowed by the commission. (Raising the prices will change the quantities bought, so correct regulatory decision making requires information on demand elasticities.)

Regulatory Lag

Once the new prices are set, they remain unchanged until the next rate case. Hence the period during which prices remain fixed provides an incentive for the company to be cost efficient. The company is able to earn higher rates of return than allowed if it can reduce its costs and, of course, earn lower rates of return if its costs rise. This incentive for cost efficiency is often referred to as the result of regulatory lag. That is, if the commission were somehow able to continuously adjust prices to keep the company's rate of return always equal to s, there would be no lag and, importantly, no incentive for cost efficiency.

Before turning to more detailed discussion of the rate level and rate structure problems, it should be instructive to examine Table 12.1. This table presents an abbreviated exhibit used in a past rate case by North Carolina Natural Gas Corporation. It shows the effect of a proposed rate increase that the company requested.

The first of the three columns shows the financial data for the year ended December 31, 19xx. Hence the company had total revenues (the left-hand side of equation [12.1]) of over \$29 million. After subtracting its expenses for purchasing gas, labor, depreciation, and taxes, it had a net operating income (row 5) of \$1.6 million. The company's net investment or rate base is shown in row 6 as \$42.9 million. This yields a rate of return of only 3.77 percent (that is, net operating income divided by the rate base).

The second column shows the adjustment for an increase in prices that is expected to raise an additional \$2.8 million in revenues. These higher revenues require higher taxes, as is also shown in this column. Other adjustments could also be shown in this column—such as for higher costs for purchased gas—but we have suppressed additional adjustments for clarity.

The third column shows the way the company would like their financial results to look. That is, the company has estimated that the allowed rate of return s should be 9.54 percent, and so it has built in price increases necessary to increase its rate of return to that figure. As we will see in the next section, the rate level problem is almost completely that of determining the appropriate value of s.

The Rate Level

The rate-level problem, as we observed earlier, is concerned with determining the values of the variables on the right-hand side of equation (12.1). That is, what are the legitimate expenses of the firm, including its required return on investment? The expenses (fuel costs, wages and salaries, taxes, and depreciation) usually account for about 80 to 85 percent of the total costs of the firm, with the remainder being the return on investment. Although these expenses are large, commissions typically do not spend much time monitoring them. They sometimes question advertising expenses (which might be viewed as trying to persuade the public that the company needs a rate increase). Sometimes they examine the salaries of top management. If the company purchases its inputs from a wholly owned subsidiary, it certainly needs to decide if the prices charged are reasonable.

Rate Base Valuation

Some commissions have taken a tough approach on whether to allow certain investments to become part of the rate base. This, and regulatory lag, appear to be the main forces that provide incentives for regulated firms to be cost efficient. Consider, for example, the following excerpt from an annual report of the California Public Utilities Commission:

In May 1987, the PUC Public Staff Division's Diablo Canyon Team recommended that of the \$5.518 billion that PG&E spent before commercial operation of Diablo Canyon Nuclear Power Plant, the utility should only be allowed to collect \$1.150 billion in rates. . . . The Public Staff Division alleges that unreasonable management was to blame for a large part of this cost overrun.[1]

Generally, the largest portion of a rate case is devoted to the issue of what the proper return to investment should be—or what should be the values of s and RB in equation (12.1). In terms of Table 12.1, the company proposed that s should be 9.54 percent and that the rate base should be \$42.9 million.

One point that should be obvious is that what really matters is the product of the two variables, that is, 0.0954 times \$42.9 million, or \$4 million. In fact, much controversy used to surround the determination of both variables. At the present time it appears that many commissions have turned to an original cost method of valuing the rate base, and thus turned most of their attention to the appropriate value of s. That is, original cost valuation is simply the amount that the company originally paid for their plant and equipment, less depreciation. There can be little debate about the actual numbers in original cost valuation (aside, of course, from the issue of imprudent investment). Other valuation methods are much more subject to judgment calls. For example, valuing the rate base by reproduction cost means to estimate the current cost of reproducing the plant, even though some of the plant may be twenty years old.

One concern about using original cost is that in periods of inflation the reproduction cost rate base will exceed the original cost rate base. That is, an electric utility might have built much of its capacity at much lower plant prices twenty years ago. To reproduce the plants at current prices might cost five times original cost. And, because the economically correct prices should reflect current marginal costs, it might be thought that original cost leads to a setting of prices that are too low. This is certainly a possibility if the commission determines price by simply dividing the sum of "Expenses" and $s(RB)$ by quantity.

However, in principle the commission is not bound by this method. Ideally, the prices should depend on current marginal costs, and if these prices yield total revenues that are too high or too low in terms of the rate-level "solution," it is necessary only to adjust fixed fees in an appropriate manner. (Recall from Chapter 11 that economic efficiency generally requires only that prices per unit equal marginal costs—fixed fees, independent of output, can be adjusted to cover deficits or return additional revenues.)

There are other valuation methods that we should also mention. Replacement cost refers to what it would cost to replace the capacity with plants embodying the newest technology—as opposed to simply reconstructing the older-technology plants at today's prices. Another valuation method is simply to add up the value of all of the company's outstanding stocks and

1. State of California, *Public Utilities Commission Annual Report 1986–1987*, p. 13. Later, a negotiated agreement led to a disallowance of "only" about \$2 billion, but with the company having the opportunity to lower it further by good performance. See Yeon-Koo Che and Geoffrey Rothwell, "Performance-Based Pricing for Nuclear Power Plants," *Energy Journal* 16, No. 4 (1995).

bonds, as given daily in the *Wall Street Journal*. This method has the defect that it is circular. The purpose of finding a rate base is so that the prices and returns can be determined; but the market-value method just described takes as the rate base a value that depends on prices and returns set by the commission in the past. (As we shall discuss shortly, the value of a share of common stock is a function of the firm's earnings per share.) In any case, we will proceed to discuss the determination of the allowed rate of return s under the assumption that the rate base is valued at original cost.

Actually, utilities are financed by investors in bonds, preferred stock, and common stock. For this reason, the allowed rate of return s that we seek is a weighted average of the costs of these three types of securities. For example, one rate case included the following estimate:

	Percent of Capitalization	Percent Cost	
Bonds	48	9.34	
Preferred stocks	14	8.22	
Common stocks	38	12.5	
	100	10.4	(weighted average)

The first two types of securities, bonds and preferred stocks, are not controversial because they have easily determined costs. The 9.34 percent figure for bonds is the so-called embedded cost; that is, it is the actual cost in interest payments that the company incurred when it issued all of its bonds in the past. Preferred stock also has a known yield. Hence the really controversial issue is what value should be used for the cost of common stock, which we shall denote by k. In the example above $k = 12.5$ percent, and therefore $s = 10.4$ percent. [That is, $s = (0.48)(9.34) + (0.14)(8.22) + (0.38)(12.5)$.]

Cost of Equity Capital

The estimation of the cost of equity capital, k, is the subject of an extensive literature in finance.[2] We can indicate only briefly here how it might be estimated. Although alternate conceptual methods exist, one popular method is the discounted cash-flow method. Consider a share of common stock of Commonwealth Edison currently traded on the New York Stock Exchange at a price P. How is this price determined? One view is that it is equal to the present value of the stream of dividends expected by investors. In equation form,

$$P = \frac{D_1}{1+k} + \frac{D_2}{(1+k)^2} + \cdots + \frac{D_i}{(1+k)^i} + \cdots \tag{12.2}$$

2. The book by A. Lawrence Kolbe and James A. Read, Jr., with George R. Hall, *The Cost of Capital: Estimating the Rate of Return for Public Utilities* (Cambridge, Mass.: MIT Press, 1984), contains a good description of five methods of estimating the cost of capital.

where P is the price of stock, D_1 is the expected dividend in year 1, D_2 is the expected dividend in year 2, D_i is the expected dividend in year i, and k is the cost of equity capital.

Hence, k is simply the discount rate used by investors—it is the rate of return the investors can obtain on their next best opportunity at the same degree of riskiness. Now if we assume that investors expect dividends to grow at some constant rate g, where g is less than k, equation (12.2) can be solved for the unknown k:

$$k = D_1/P + g. \tag{12.3}$$

Assume that the current dividend yield D_1/P is 8 percent and investors expect that dividends will grow over time at 7 percent. Then equation (12.3) would yield 15 percent as the cost of equity capital.

A moment's thought reveals that this method of determining k is subject to great uncertainty. How should the dividend yield be measured? Should it be the yield security analysts expect? Or should it be the last year's dividend divided by the average stock price? Perhaps even more difficult is how to choose g. Should it be the average growth rate over the last three or five or ten years? In fact, past growth rates may be totally unrelated to what happens in the future. Also, should the estimates of k be based on the particular company under review, or should they be based on averages of a group of companies with about the same degree of risk for investors? Clearly, the possibilities for differences of opinion on the value of k are numerous.[3]

A somewhat subtle unintended cost of regulation can be noted here. The value of k depends upon the riskiness of the firm as perceived by investors. Simply put, a typical risk-averse investor in security A with a certain return of 8 percent would require, perhaps, 10 percent from security B if its return were subject to some uncertainty. That is, the investor would require a risk premium to compensate for the uncertainty. To the extent that the uncertainty is due to the behavior of the regulatory agency itself—for example, by setting rates in a very unpredictable manner—the cost of equity capital will be higher than necessary.

It is appropriate to step back at this point and consider the economic function of the rate of return. In competitive markets the function of the rate of return is to attract capital into the industry (that is, by paying for capital) and to provide an incentive for efficiency. Firms in competitive markets are rewarded by high profits if they are efficient, and punished by low

3. One article reported that in rate cases over the 1980–1984 period, the average k requested by firms was 17.02 percent and the average k recommended by the commissions' staffs was 15.11 percent. Interestingly, the average k allowed by the commissions was 15.39 percent—a figure much closer to the staffs' recommendations than to the firms' requested values. S. B. Caudill, B. Im, and D. L. Kaserman, "Modeling Regulatory Behavior," *Journal of Regulatory Economics*, September 1993.

profits or losses if not. Generally, commissions seem to regard the first function as the more important. This has a legal basis. A famous Supreme Court decision in 1944, *Federal Power Commission v. Hope Natural Gas Co.,* held that rates should be set to enable the company to operate successfully, to maintain its financial integrity, to attract capital, and to compensate its investors for the risks involved.

Recently, however, there have been some steps taken by commissions to reward efficient utilities and penalize the inefficient ones. Although these new steps do not necessarily operate by adjusting the allowed rate of return, it is instructive to consider the possibility of tying s to the firm's performance. There are undoubtedly serious problems in measuring a firm's performance, but one often-heard problem is not true—namely, that the firm must have a value of s that is not less than its cost of capital or it cannot attract capital.

Suppose for simplicity that a firm is known to have a cost of capital of 10 percent and the commission allows it a rate of return of $s = 10$ percent. The firm could raise capital by selling stock and using the proceeds to invest in new plant, earning the necessary 10 percent rate of return. Now suppose that the commission elects to cut its s value from 10 to 9 percent to penalize the firm for some inefficient behavior. Would the firm now be unable to raise capital to build needed new capacity? Clearly, the answer is that the firm *could* raise new capital—but its stock price would fall, inflicting losses on its existing owners. It is also true that the company would resist raising capital under the conditions cited, but it could do so. The point is that temporary adjustments in the allowed rate of return could be made without fear of the company being unable to attract capital.

The Sliding Scale Plan and Yardstick Competition

Joskow and Schmalensee have analyzed a variety of proposals for performance-based regulation of electric utilities.[4] One method is the *sliding scale* plan. It is interesting in that it has been used in practice (although rarely), and it also has some features that theoretical work has suggested as desirable. The essential property is that it permits the sharing of risks and rewards between the owners and consumers.

Let r^* be the target rate of return and let r_t be the actual rate of return at the prices that prevail initially in year t. The sliding scale would adjust prices so that the actual rate of return, r_a, at the new prices would be

$$r_a = r_t + h(r^* - r_t) \tag{12.4}$$

where h is a constant between zero and unity. Notice that if $h = 1$, regulation is essentially "cost-plus." That is, prices are always adjusted to give the firm a rate of return of r^*. The firm would neither benefit from being efficient nor be hurt by being inefficient.

4. Paul L. Joskow and Richard Schmalensee, "Incentive Regulation for Electric Utilities," *Yale Journal of Regulation* 4, No. 1 (1986).

Similarly, if $h = 0$, then regulation is basically "fixed-price." All gains from efficiency accrue to the firm, and all unexpected cost increases beyond management's control also affect the firm alone. An h value of 0.5, however, would share unexpected benefits and costs equally. The "optimal" value of h is unknown; however, one would think that neither extreme is likely to be best.

We turn briefly to a second proposal that would seem to be possible in the electric power industry—*yardstick competition.* The reason is that there are hundreds of private electric utilities in the United States. Suppose a regulator in Illinois selects electric utilities in other states that face production and demand functions comparable to those of firm X in Illinois. The regulator determines the average cost per kilowatt-hour for all comparable firms, say AC, and then sets firm X's price equal to AC. If regulators in other states followed the same procedure, then this approach would eliminate the cost-plus character of regulation. Each firm's prices would be completely independent of its own costs, and cost reductions would lead to profit increases.

Unfortunately, as Joskow and Schmalensee have pointed out, finding truly comparable utilities is probably not possible. "Utilities differ from one another in so many dimensions, not only because of current market conditions but also because of past investment decisions, that we are unlikely to find a large number of truly comparable utilities." They also argue that traditional regulatory accounting practices also raise serious problems for the yardstick approach.

Price Caps and Performance Standards

Beginning in the 1980s, "price cap" regulation for telephone companies has been used by the FCC and some states in lieu of rate-of-return regulation. It is a form of regulation used in Britain for their recently "privatized" industries—such as telephones, gas, and water. ("Privatized" refers to the fact that government-owned industries were sold to private investors.)

The price cap is set so that the firm is free to raise its prices at the rate of inflation minus some amount selected to reflect expected productivity. For example, the productivity might be 3 percent and if inflation is 5 percent, the firm could raise its prices at $5 - 3 = 2$ percent per year. Of course, the company would be free to charge lower prices if doing so should appear to be more profitable. Price cap regulation is viewed as providing incentives for the firms so regulated to be cost efficient. In a sense, it builds regulatory lag into the process in a nonaccidental way. And, of course, it is quite similar to the sliding scale plan described earlier inasmuch as consumers are permitted to share in productivity gains.

A 1989 study by Mathios and Rogers has found evidence that favors price cap regulation in comparison with rate-of-return regulation.[5] They examined the thirty-nine states that have

5. A. D. Mathios and R. P. Rogers, "The Impact of Alternative Forms of State Regulation of AT&T on Direct-Dial, Long-Distance Telephone Rates," *The Rand Journal of Economics*, Autumn 1989.

more than one Local Access and Transportation Area (LATA). The 161 LATAs in the United States were created as a result of the AT&T antitrust case that was settled in 1982. Basically, local telephone companies are allowed to provide long-distance service within LATAs, but not from one LATA to another. Hence, AT&T and other long-distance companies provide the inter-LATA, intrastate service. It is these intrastate, inter-LATA markets that the study investigated.

According to Mathios and Rogers, twenty-eight of the thirty-nine multi-LATA states moved to some form of price cap regulation of this long-distance service between 1984 and 1987. Kansas, for example, permits increases in rates by 4 percent and decreases by 7 percent without the need for a rate case. The authors found that "states that allowed pricing flexibility had lower 1987 prices than other states for all mileage bands."

In a 1993 article, Braeutigam and Panzar reported that some twenty-two states have moved to combined sliding-scale–price-caps (SS-PC) incentive regulatory schemes for telephone companies. "A common form allows the telephone company to retain all earnings under PC rates as long as the rate of return is less than some specified amount, typically in the neighborhood of about 13 percent." Then, for example, the firm may be allowed to retain half of any additional earnings for a rate of return between 13 and 15 percent, and refund all additional earnings over 15 percent.

Braeutigam and Panzar conclude:

> The limited U.S. evidence available supports the view that PC regulation is an effective means of controlling the *prices* of dominant firms when the control of their *profits* is left to the competitive marketplace. Thus, as has been observed many times, PC regulation is probably most effective as a transitory step on the path toward total deregulation and full competition.[6]

Finally, to round out our discussion of incentives to create efficiency, we note the use of performance standards for electric power plants. For example, a specified norm for plant performance is defined. Penalties are imposed for performance below the norm, and rewards are given for above-norm performance.

Joskow and Schmalensee have described a standard that the Arizona Public Service Company's nuclear plant, Palo Verde 1, had imposed on it. The target is the plant's capacity factor—the actual amount of electricity the plant generated divided by the amount it could generate running continuously throughout the year. Ideally, because of the low running costs of a nuclear plant, it should have a high capacity factor. The incentive provision is that if the capacity factor is between 60 and 75 percent, no penalty or reward is actuated. However, if the factor is between 75 and 85 percent, the company's reward equals 50 percent of the fuel-cost saving from running the plant more. Conversely, capacity factors between 50 and 60 percent

6. R. R. Braeutigam and J. C. Panzar, "Effects of the Change from Rate-of-Return to Price-Cap Regulation," *American Economic Review*, May 1993.

result in a penalty equal to half of the additional fuel costs incurred by running more costly plants.

Averch-Johnson Effect

We turn now to an example of how rate-of-return regulation can create perverse incentives. The model we describe is an analysis of rate-of-return regulation published in 1962 by Averch and Johnson.[7] Their work led to a large outpouring of both theoretical and empirical research. Using what some today regard as very strong assumptions about how regulation constrains the firm, Averch and Johnson found that firms would choose too much capital relative to other inputs. As a result, the output would be produced at an inefficiently high cost. The key idea is that because allowed profit varies directly with the rate base (capital), the firm will tend to substitute too much capital for other inputs.

In mathematical terms, the problem is one of maximizing profit subject to a rate-of-return constraint. We will not develop the complete analysis, but it should be instructive to formulate the problem and provide the solution.

Hence the problem is to choose the quantities of labor and capital to maximize profit—that is, revenue minus the costs of the inputs, labor and capital. Maximize

$$\Pi = R(K, L) - wL - rK \tag{12.5}$$

subject to

$$\frac{R(K, L) - wL}{K} = s \tag{12.6}$$

where Π is profit, R is revenue function, K is quantity of capital, L is quantity of labor, w is wage rate, r is cost of capital, and s is allowed rate of return.

The rate-of-return constraint, equation (12.6), implies that the firm is continuously restricted to a rate of return equal to s. The numerator equals total revenue minus the cost of labor, divided by capital—that is, the rate of return on capital. This, of course, is not strictly correct because the firm's prices are fixed from one rate case to the next, and therefore the firm's rate of return can be greater than or less than s during these periods of regulatory lag.

Another key assumption is that s is greater than r. In other words, it is assumed that the regulatory agency permits the firm to earn a higher rate of return on capital than the true cost of capital. Of course, the opposite case of s less than r would imply that the firm would prefer to shut down if this were to be a long-term situation. And if s and r were equal, the firm would be indifferent among the quantities of K and L inasmuch as its profit would be zero for all choices. Hence, Averch and Johnson argued that $s > r$ is the interesting one.

7. H. Averch and L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *American Economic Review*, December 1962.

Using a standard mathematical solution technique (the Lagrangian multiplier method), Averch and Johnson found that

$$\frac{MP_k}{MP_l} = \frac{r - \alpha}{w} \tag{12.7}$$

where

$$\alpha = \frac{\lambda(s - r)}{l - \lambda} > 0.$$

MP_k is the marginal product of capital, and MP_l is the marginal product of labor.

The α variable is positive because $s - r > 0$ and λ, the so-called Lagrangian multiplier, can be shown to be between 0 and 1. (The economic interpretation of λ is that it measures the increase in profit of a \$1 increase in allowed profit; hence its value between 0 and 1 is sensible.)

This result can be explained by reference to Figure 12.1. The figure shows the isoquant for the level of output chosen by the regulated firm, Q^*. The axes show the quantities of the inputs, K and L, that can be used to produce output Q^*.

The economic theory of production requires that to minimize the cost of producing Q^*, it is necessary to equate the slope of the isoquant, that is, the ratio of marginal products, to the ratio of input prices. Equation (12.7) implies that the Averch-Johnson regulated firm would meet this requirement if $\alpha = 0$. However, $\alpha > 0$, and the firm acts as if the cost of capital is cheaper than it actually is. That is, the firm acts as if its cost of capital is $r - \alpha$.

In Figure 12.1, let the slope of lines MM and NN equal r/w and the slope of line PP equal $(r - \alpha)/w$. Then, cost minimization requires operation at point E, where the slope of NN equals the slope of the isoquant. Note, however, that the regulated firm will choose to operate at point F, which equates the slope of PP [or $(r - \alpha)/w$] with the slope of the isoquant. The result shows that the regulated firm uses too much capital, K^*, and too little labor, L^*, as compared to the least-cost solution, K' and L'. The excess cost can be measured in units of labor by the distance MN on the vertical axis. That is, the actual cost of producing Q^* is OM units of labor inasmuch as MM passes through point F. However, the least-cost production of Q is ON units of labor.

A less rigorous explanation is as follows. The key point is that the regulated firm perceives that its cost of capital, $r - \alpha$, is less than the true cost r. For simplicity, take $s = 10$ percent, $r = 8$ percent, and $r - \alpha = 6$ percent. The regulated firm can earn a "bonus" of 2 percent on each dollar of new capital (costing 8 percent) because it is allowed to earn 10 percent. This "bonus" of 2 percent per dollar can be interpreted roughly as a 2 percent discount—making its perceived cost of capital only 6 percent.

It is very difficult to test for the existence of an Averch-Johnson (A-J) effect empirically. Some economists have argued that it is common knowledge that utilities often choose capital-

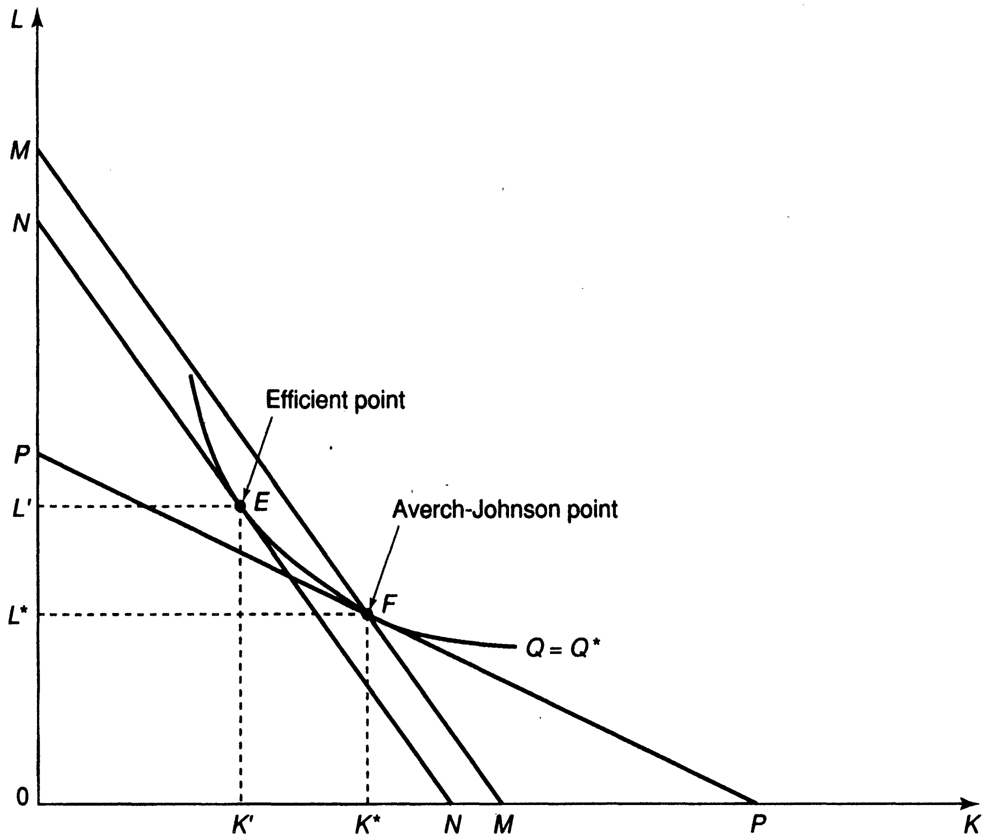

Figure 12.1
The Averch-Johnson Effect versus Least-Cost Production

intensive production, and this supports the A-J effect. One example is the resistance by electric utilities to highly integrated power pooling. The idea here is that utilities prefer to have their own capacity to meet peak demands rather than purchasing power from nearby companies. Although it might be least-cost to purchase power, these expenses are simply reimbursed under rate-of-return regulation. Investing in peak capacity, on the other hand, expands the rate base and thereby increases allowed profit.

A number of econometric tests for the A-J effect have also been performed. Using statistical analysis—usually in the electric power industry—the results have been mixed. Some have found support for the effect, and others have not.[8]

One reason for these mixed results might be that the regulatory environment in the United States was changing over the period of the statistical studies. Joskow, for example, has

8. For example, H. C. Petersen, "An Empirical Test of Regulatory Effects," *Bell Journal of Economics*, Spring 1975, found support for the A-J effect. W. J. Boyes, "An Empirical Examination of the Averch-Johnson Effect," *Economic Inquiry*, March 1976, did not find support for it.

observed that until the early 1970s, electric power regulators did not really monitor rates of return carefully.[9] They were rather concerned with nominal electricity prices, because it is these prices that cause concern to consumers. Hence, if, as appears to be the case, nominal prices tended to stay constant or fall prior to the early 1970s, electric utilities could earn high rates of return without fear of having price decreases imposed. This finding clearly conflicts with the assumptions of the A-J model.

We should mention one possible beneficial effect of the A-J bias toward capital intensity. For most regulated industries, technological change takes place through the substitution of capital for other inputs. For example, direct long-distance dialing replaced many operators with automatic switching equipment. Hence one might argue that the A-J effect has possibly stimulated innovation. (Other characteristics of regulation, however, can be argued to retard innovation, so the net effect of regulation on innovation is unclear. For example, profits created through innovation can be expected to be reduced through price decreases at the next rate case.)

Finally, consider the two-product monopolist and the incentive to increase capital to increase profits. Noll and Owen have explained this problem for regulators succinctly:

> Suppose that at all feasible prices, one market has elastic demand, while another has inelastic demand. If a regulated firm lowers the price in the former, it will increase sales by a relatively large amount, requiring that it commit substantially more capital to that market. But if it increases prices in the latter market, it will suffer a relatively small reduction in sales, and hence a small reduction in capital requirements. Thus, changing both prices simultaneously this way increases the total required capital. . . . This, in turn, increases the firm's allowed profits.[10]

Rate Structure

The rate structure has to do with how prices vary across customer classes and products. In the preceding chapter we described the prices that are economically efficient under various conditions. Of course, economically efficient prices (for instance, prices equal to marginal costs) are often not the prices set under regulation. However, peak-load pricing (a type of marginal cost pricing) has become important in electric power. We will examine this topic in depth later in this chapter.

A common method of pricing used by regulatory commissions is to begin by allocating all of the utility's costs to various customer classes and services. Most utilities provide a variety of

9. Paul L. Joskow, "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation," *Journal of Law and Economics*, October 1974.

10. Roger G. Noll and Bruce M. Owen, "The Anticompetitive Uses of Regulation: *United States v. AT&T* (1982)," in J. E. Kwoka, Jr. and L. J. White (eds.), *The Antitrust Revolution*, 2nd ed. (New York: HarperCollins, 1994), p. 338.

services to different customer groups. They also have many facilities that are used in common by these customers and services. For example, power plants and transmission lines, telephone switching centers, and pipelines all represent common costs that apply to most customer classes or services.

FDC Pricing

To illustrate concretely one such pricing method, fully distributed cost (FDC) pricing, consider a simple two-product natural monopolist that sells electricity to two classes of customers. We denote the electricity sold to residential buyers by X and to industrial customers by Y. (Electricity sold to residential customers is usually at a lower voltage than industrial customers require, and therefore there truly are two different products.)

Assume the following cost functions:

$$\text{To produce } X \text{ alone:} \quad C_x = 700 + 20X \tag{12.8}$$

$$\text{To produce } Y \text{ alone:} \quad C_y = 600 + 20Y \tag{12.9}$$

$$\text{To produce both:} \quad C_{xy} = 1{,}050 + 20X + 20Y. \tag{12.10}$$

Note that the joint production of X and Y is subadditive. That is, least-cost production requires that X and Y be produced together because fixed cost is \$1,050, as compared to a total of \$1,300 if produced separately. The \$1,050 fixed cost also represents common costs that must be allocated to each product in order to implement FDC pricing.

Kahn has observed that utilities' common costs "may be distributed on the basis of some common physical measure of utilization, such as minutes, circuit-miles, message-minute-miles, gross-ton-miles, cubic feet, or kilowatt-hours employed or consumed by each. Or they may be distributed in proportion to the costs that can be directly assigned to the various services."[11]

The particular method may appear quite reasonable, but the essential point is that it is necessarily arbitrary. And more importantly, such cost allocations lead to prices that have no necessary relationship to marginal costs. To take an example, assume that some "reasonable" method leads to an allocation of 75 percent of the common costs to product X and 25 percent to product Y. Hence, FDC average costs would be

$$AC_x = 787.5/X + 20 \qquad \text{and} \qquad AC_y = 262.5/Y + 20. \tag{12.11}$$

That is, the average cost of X equals its 75 percent share of the \$1,050 common cost, divided by the units of X sold, plus the clearly attributable variable cost of X per unit of \$20.

11. Alfred E. Kahn, *The Economics of Regulation: Principles and Institutions*, Vol. 1 (New York: John Wiley & Sons, 1971), p. 151.

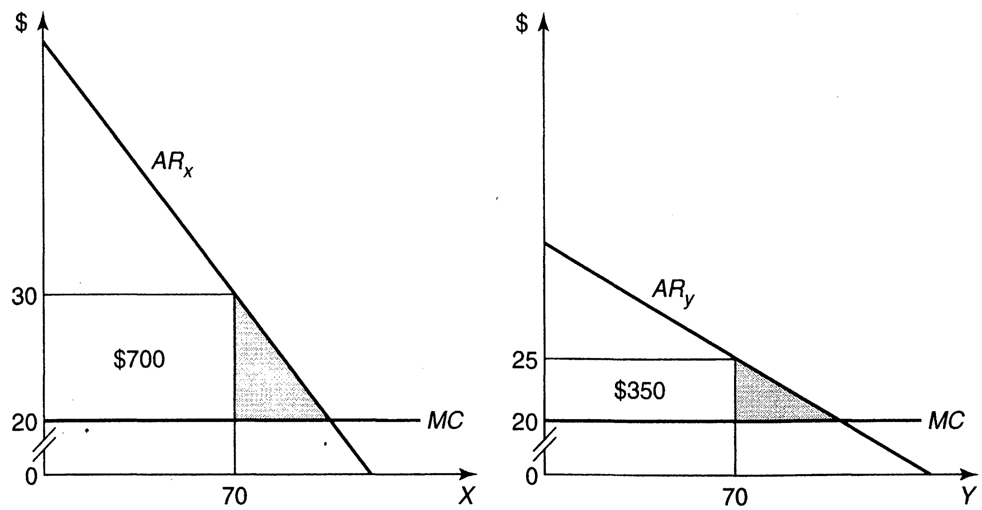

Figure 12.2
Ramsey Pricing

At this point, the demands for the two products must be specified. Assume that the demand functions are

$$P_x = 100 - X \qquad \text{and} \qquad P_y = 60 - 0.5Y. \tag{12.12}$$

With the demand information, the actual FDC prices can be found by equating equations (12.11) and (12.12). This method simply sets $P_x = AC_x$ and $P_y = AC_y$, ensuring that total revenues are equal to total costs. The result is

$$\begin{aligned} P_x &= AC_x = \$31.5 \qquad \text{and} \qquad & P_y &= AC_y = \$23.6 \\ X &= 68.5 & Y &= 72.8. \end{aligned} \tag{12.13}$$

Hence these FDC prices clearly satisfy the requirement that total revenues equal total costs. Again, however, there is no basis for expecting these prices to be the economically efficient prices. In general, such prices lead to deadweight losses.

It is easy to show that the efficient prices in this case are the Ramsey prices (as explained in Chapter 11):

$$\begin{aligned} P_x &= \$30 \qquad \text{and} \qquad & P_y &= \$25 \\ X &= 70 & Y &= 70. \end{aligned} \tag{12.14}$$

Figure 12.2 illustrates the Ramsey solution. By definition, the Ramsey prices have the smallest deadweight loss triangles (shaded in Figure 12.2) for all possible pairs that yield revenues equal to costs.

Although we have implicitly assumed that two-part pricing is not feasible in the example, notice that if it were feasible, the "first best" solution would be to charge marginal cost prices to each group; that is, each price would be \$20. Then the fixed fees would be set to just cover the fixed cost of \$1,050. And in this case, in achieving economic efficiency it would make no difference how the \$1,050 was allocated between the two groups. (This point was also explained in Chapter 11.)

One problem that FDC pricing raises is that "reasonable" allocations of common costs lead to disputes among customer classes. It is natural to expect residential customers to argue that their share should be lower than 75 percent and for industrial customers to argue that their share should be lower than 25 percent. Commissions have long been concerned with "undue discrimination" across customer classes, and we turn to that subject next.

Undue Discrimination

Undue discrimination is really not an efficiency issue; rather, it has to do with the fairness of the existing set of prices in the sense of whether one group may be "subsidizing" another group. It is clearly a controversial issue for commissions inasmuch as a rate case may find intervenors representing residential customers and industrial customers in total opposition to each other. Both groups may argue that they are paying too large a share of the common costs! This is basically an equity issue—which we have avoided for the most part in this book—but there are some clarifications that economic analysis can offer to the debate.

Economists argue that if one must examine cross-subsidization issues (assuming Ramsey prices are not used), the most logical tests are the so-called stand-alone average cost and the average incremental cost tests—which are, in fact, equivalent.[12]

Consider the stand-alone average costs for product X for an output of X of 70 units. That is, returning to equation (12.8), the average cost of X is \$30. The Ramsey price of \$30 for the same output therefore does not give an incentive for customers of X to break away and produce X alone. This test therefore would classify the Ramsey price as subsidy-free. Similarly, the Ramsey price of Y is also subsidy-free. (Note that the FDC price of X for 68.5 units is \$31.5, which exceeds the stand-alone average cost of \$30.2. It therefore fails the subsidy-free test.)

Another test is the average incremental cost test. Here, we compute the average incremental cost, AIC, of producing X in joint production with Y. Thus, subtract the cost of producing Y alone from the cost of producing X and Y jointly to get the incremental cost of X. The AIC of X is, therefore, $20 + 450/X$, or \$26.4 for $X = 70$. Similarly, the AIC for Y at $Y = 70$ is \$25. Here, the test for subsidy-free prices is that the prices equal or exceed their respective

12. For proof, see Stephen J. Brown and David S. Sibley, *The Theory of Public Utility Pricing* (New York: Cambridge University Press, 1986).

AICs. The logic is that if each product contributes to total revenue an amount that at least covers the extra costs it causes (when added to the production of the other products), then it should be viewed as a beneficial addition. To the extent that its incremental revenues exceed its incremental costs, the revenues required from the other products are reduced. The Ramsey prices of \$30 and \$25 also pass this test. In fact, these two tests always give the same answers. (Note that the FDC price of Y for 72.8 units is \$23.6 while the AIC of Y is \$24.8. Hence the FDC prices fail the subsidy-free test by this test too.)

It should be pointed out that FDC prices do not necessarily fail subsidy-free tests. They may pass the subsidy-free tests and still be economically inefficient.

Under certain conditions of subadditivity of cost (that is, when the natural monopoly is sustainable), it is true that Ramsey prices are subsidy-free in the sense that no outsider would find it profitable to enter. Hence the regulator need not be concerned about whether permitting entry would be socially beneficial. This argument assumes, of course, that the regulator permits the monopolist to charge Ramsey prices (rather than hold it to the FDC prices given previously). However, there are cases in which the cost function is subadditive and yet subsidy-free prices do not exist. This is another way of viewing the case of a natural monopoly that is unsustainable—least-cost production requires a single firm, but there are no prices that can keep all of the monopolist's products invulnerable to entry.

Here is an example. Three towns are in need of a well for water supply. One deep well could supply all three towns at a cost of \$660. This is the least-cost solution and would imply a price of \$660/3 = \$220 per town. Two towns could go together and dig a shallower well for \$400, and each town alone could dig an even shallower well for \$300. Clearly, \$660 is lower than any of the alternatives. If each town had its own well, the total would be \$900, and if two went in together for \$400 and the third went alone at \$300, the total would be \$700.

The problem is that \$220 per town would provide an incentive for two of the towns to join forces at \$400, or a price of \$200 each. One can think of it as (any) two towns, if they go along with the three-town project, subsidizing the third town in the amount of \$20 each. Clearly, there is no way to avoid the subsidies if one is going to achieve least-cost production.

In anticipation of the next section on peak-load pricing, we provide a final example of "unfair" subsidization. Consider two groups of electricity customers: day customers and night customers. A plant costing K dollars is necessary to meet the day customers' demand, which is larger than the night demand. Because electricity cannot be easily stored, the plant must be large enough to supply power on demand.

The plant, of course, can be used to supply the smaller night demand as well. Under certain assumptions to be discussed in the next section, the economically efficient solution is to charge day customers for the total cost of the plant, K! Even though both groups use the plant, the day customers pay for the entire plant cost, plus fuel costs for their output. Night customers pay only for their fuel costs. It is certainly "unfair" in certain senses—however, it is the demand of day customers that necessitates such a large plant, and it would be inefficient if the price they paid did not signal the cost of the larger capacity.

In conclusion, the major point of this section is to make clear that the objective of economic efficiency may sometimes require pricing that conflicts with common notions of fairness. One justification for opting for efficiency, of course, is that the "size of the pie" is larger as a result, and authorities can in principle make everyone better off by appropriate taxes and/or subsidies.

Peak-Load Pricing

A major development in electric power has been the gradual movement toward the implementation of peak-load pricing. This term refers to the variation in prices by time of use—for example, in the middle of the day more electricity is demanded than in the middle of the night. The marginal cost of electricity is, as a result, much higher in the middle of the day than it is at night. Setting prices that vary over the day in proportion to the variation in marginal costs is a form of peak-load pricing. Until recently, in the United States most consumers paid electricity prices that did not vary over the day.

Costs of Power Production

It is useful to examine the cost structure of a typical electric power system before continuing with our discussion of peak-load pricing principles. A major point is that it is generally too costly (or impossible) to store electricity, and therefore sufficient capacity must be on hand to supply the demand at all times. This fact implies that capacity is determined by the amount of peak demand.

Demand for electric power typically varies in a reasonably predictable cyclical pattern—daily, weekly, monthly, and seasonally. The demand might follow the pattern in Figure 12.3 for a typical weekday. Hence, peak demand occurs at midmorning; demand at midnight is only 70 percent of that amount. Demand over the weekend might equal only 50 percent of the high during the week.

In a recent year, Duke Power Company produced some 55 billion kilowatt-hours of electricity. If it had experienced a constant rate of demand over the year, Duke Power could have produced that amount of electricity with a capacity of 6,300 megawatts (assuming, for simplicity, no downtime for maintenance or for other reasons). In fact, during one hour on January 11, Duke had its peak demand for the year, requiring a capacity of 11,145 megawatts! The company actually had some 13,234 megawatts of installed capacity. (Installed capacity is higher than expected peak demand to provide a reliability margin in view of the uncertainty in demand and to allow for unplanned outages of power plants.)

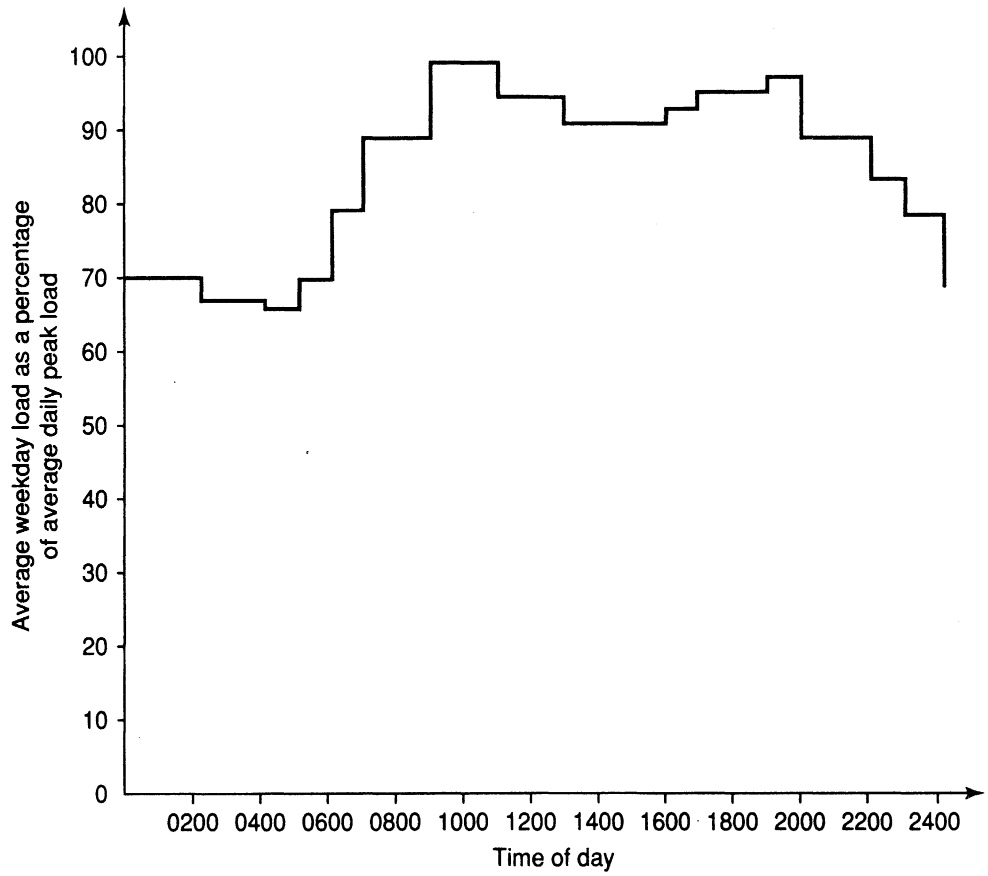

Figure 12.3
Average Daily Load Curve for Electricity

Duke Power's capacity consisted of

7,423 megawatts (MW) of coal-fired plants
3,760 MW of nuclear plants
1,452 MW of hydroelectric plants
599 MW of combustion turbine plants
13,234 MW total

A typical electric-power system has a mixture of plant types because it leads to a lower overall cost of supplying the variable pattern of demand. Nuclear plants have relatively low variable or "running" costs, but have relatively high fixed (capital) costs. They are, therefore, suited for running as the "base load" plants—as many hours per year as possible. Combustion turbines, on the other hand, have relatively high running costs but low fixed costs. They are used to meet peak demands that last for only a small number of hours per year.

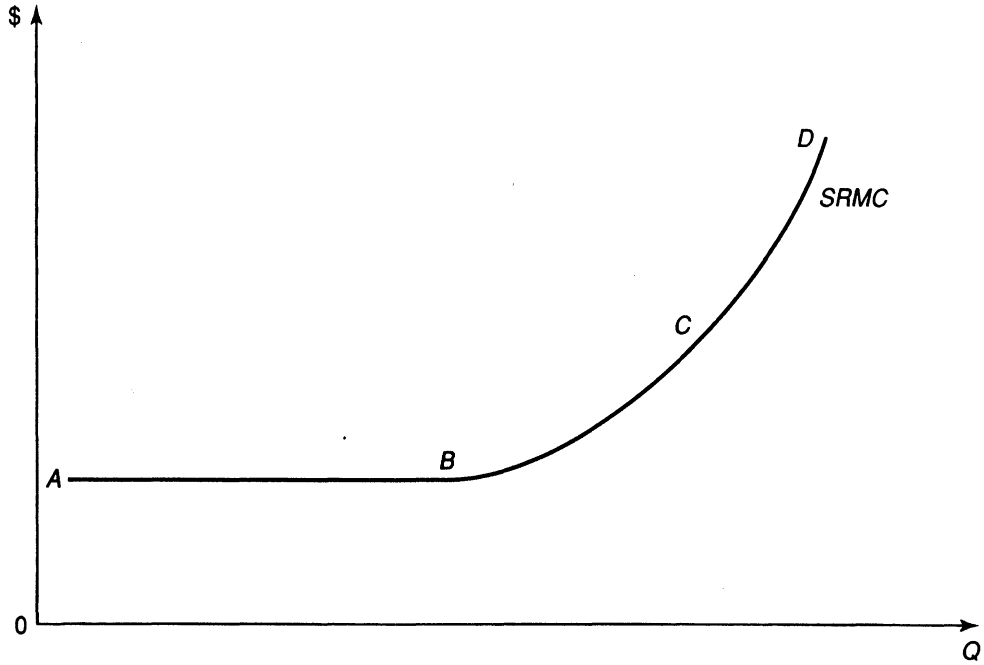

Figure 12.4
Short-Run Marginal Cost Curve for Electric Power System

The result is that the short-run marginal cost curve of a power system is similar to the rising curve shown in Figure 12.4. The costs given by segment AB might represent the base load nuclear plants' running costs; BC, the costs of coal-fired plants of varying ages and efficiency; and CD, the costs of the peaking plants (such as combustion turbines). In this context, it is easy to realize that, since demand varies continuously over time, charging a price equal to short-run marginal cost (SRMC) would require a continuously changing price.

In order to explain the principles of peak-load pricing most clearly, it will help to abstract greatly from the real-world complexity just described. Hence we turn now to a vastly simplified model.[13]

Peak-Load Pricing Model

In Figure 12.5 we make the assumption that demand is given by the peak demand curve for exactly half of the day, and by the off-peak demand curve for the other half of the day. For simplicity, it is assumed that the two demands are independent—the price in the peak period, for instance, does not affect the quantity demanded in the off-peak

13. The exposition follows Peter O. Steiner, "Peak Loads and Efficient Pricing," *Quarterly Journal of Economics*, November 1957.

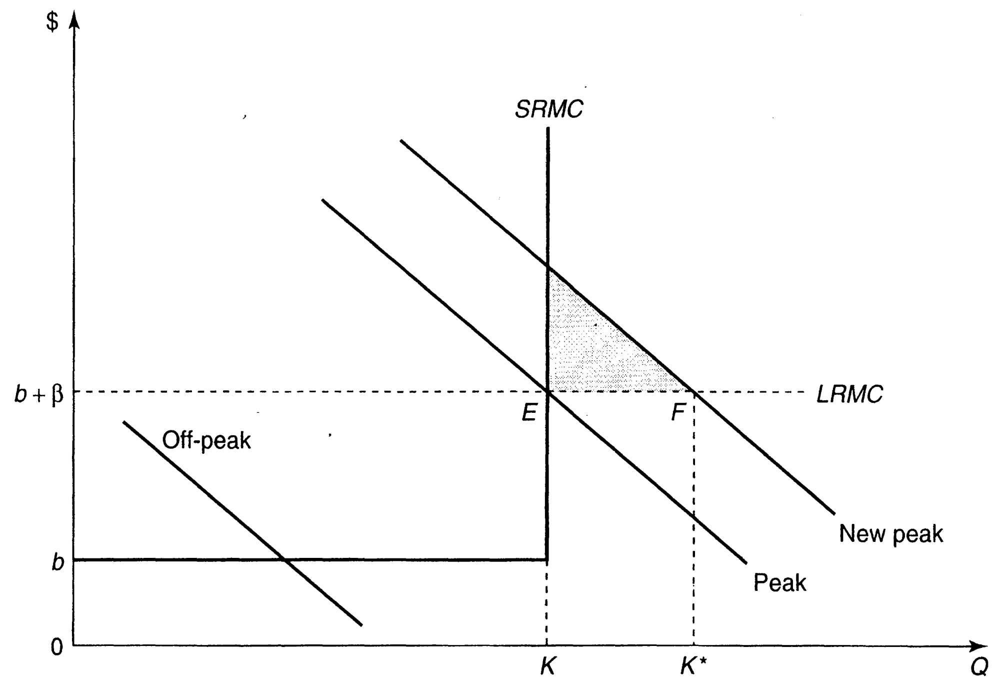

Figure 12.5
Peak-Load Pricing—Firm Peak

period.[14] Also, it is assumed that "running" costs—for example, fuel for electricity production—are constant at the level b until capacity is reached at K. (As we will explain, we have chosen K because it is the socially optimal plant size for the demands shown in the figure.) At the output K, no further output is possible, as indicated by the vertical line that is labeled $SRMC$. Hence we have a so-called "rigid" plant with the $SRMC$ curve being equal to b for outputs less than K, and then becoming vertical at the plant capacity. One can think of this as an approximation to the smoothly increasing $SRMC$ curve in Figure 12.4.

The dashed horizontal line at the level $b + \beta$ is labeled long-run marginal cost ($LRMC$). The assumption here is that β represents the cost of an additional unit of capacity, and that it is possible to add capacity in increments of single units if desired.[15] The economically

14. This assumption is probably too strong. A well-known counterexample occurred in 1964 when AT&T began lower rates for long-distance telephone calls after 5 P.M. They found themselves deluged with calls from people who formerly called during the day. The interdependence of demands can be handled with an increase in mathematical complexity. See I. Pressman, "A Mathematical Formulation of the Peak-Load Pricing Problem," *Bell Journal of Economics*, 1970.

15. The production function underlying this model is one of fixed coefficients. Much of the early literature employed this assumption. While the suitability of this assumption to describe, say, electric power production is an empirical question, the alternative variable-proportions technology is somewhat more difficult to expound.

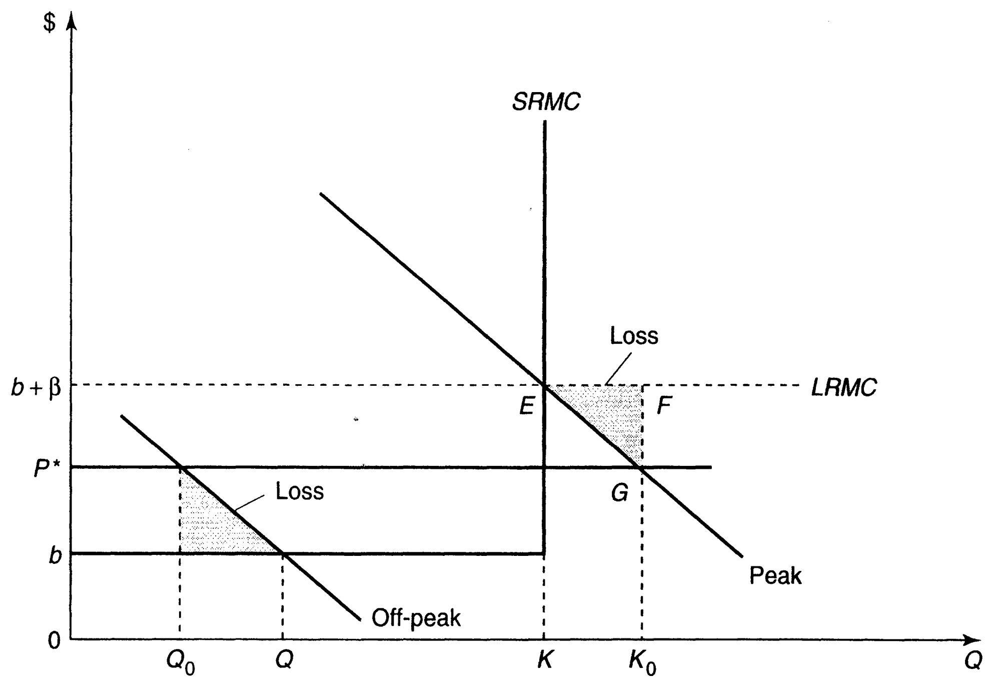

Figure 12.6
Deadweight Losses Due to Nonpeak Pricing

efficient solution is to charge a price equal to $SRMC$ in order to use the existing plant optimally. The $LRMC$ comes into play in order to decide whether the existing plant capacity is optimal. Hence, in Figure 12.5 the off-peak price should be b and the peak price should be $b + \beta$. Notice that the peak price is equal to both $SRMC$ and $LRMC$. This fact indicates that the capacity is in fact optimal. The reason is that the price can be interpreted as the "marginal willingness to pay," and $b + \beta$ represents the marginal cost to supply one more unit.

If, for example, the peak price exceeded $b + \beta$, it would pay society to increase capacity. This situation is shown by the demand curve labeled *new peak.* This new peak demand intersects $SRMC$ at a price higher than $b + \beta$. Therefore, an increase in consumer surplus can be had by expanding capacity out to K^*. This increase in consumer surplus is equal to the area under the demand curve between K and K^*—which represents willingness to pay—minus the cost of supplying the additional output, rectangle $EFKK^*$. Subtracting the cost from willingness to pay gives the shaded triangle that represents the increase in consumer surplus attributable to the capacity increase. Hence, at the new capacity K^*, price is again equal to $SRMC$ and $LRMC$—indicating that K^* is the optimal capacity.

Now, assume that the electric utility follows the practice of charging a single price that does not vary over the day, say, a price of P^*. This situation is shown in Figure 12.6. As we mentioned earlier, this would represent the pricing policy generally followed in the United States before peak-load pricing began to be implemented. In order to satisfy demand at the peak at this price, capacity of K_0 is required. Because optimal capacity is K where price equals $LRMC$, the single-price policy leads to too much capacity. The deadweight loss associated with this is shown as the shaded triangle EFG. It equals the difference between the cost of the excess capacity, rectangle $EFKK_0$, and the willingness-to-pay for that incremental capacity, $EGKK_0$. Intuitively, the peak demanders are not charged enough for the actual costs that they cause.

There is a second deadweight loss triangle in Figure 12.6, and it is associated with the nonoptimal use of the plant in the off-peak period. That is, with the price P^* charged in the off-peak period, consumption in the off-peak period is too low—at Q_0 rather than at Q where price would equal $SRMC$.

We should observe that the economically efficient prices that we have been discussing are what society should prefer, not necessarily what a profit-maximizing regulated utility would desire. In fact, the utility would have total revenues equal to total costs in either case (peak-load prices or single price). What gave the impetus to regulated electric utilities moving toward peak-load pricing was probably a combination of pressures—the energy crisis of the 1970s, high inflation, and other factors that made regulators seek alternatives to the traditional rate structures. In addition, Congress passed a law in 1978, the Public Utilities Regulatory Policies Act, that among other things required state commissions to study peak-load pricing for possible implementation in their state.

The cases of peak-load pricing described so far indicate that peak demanders pay $b + \beta$ whereas off-peak demanders pay only b. That is, peak demanders pay all capacity costs and off-peak pay none. This statement is true, however, only for the particular case shown—known as the *firm peak* case. An alternative case is the *shifting-peak* case, which has the property that the demands are "closer" together. Figure 12.7 illustrates this case.

To see why this case is known as the shifting-peak case, consider the effect of charging the peak demanders all the capacity costs and the off-peak demanders none. For simplicity, we assume that $b = 0$ in this case—this assumption makes the figure less cluttered and doesn't affect the key points. The result is easily seen in Figure 12.7. Peak demanders would demand R units of capacity (where price $= \beta$) and off-peak demanders would want S units (where price $= 0$), or a greater capacity than peak demanders! Intuitively, this result suggests that the prices are wrong. The correct set of prices can be found in conjunction with solving for the optimal capacity.

To obtain the optimal capacity, construct the *demand for capacity.* Conceptually, think of the plant as a public good—it can be used by both peak and off-peak demanders (though at different times). This statement implies that the total willingness to pay for the plant is obtained by adding vertically the demand curves for the two groups of demanders. This total

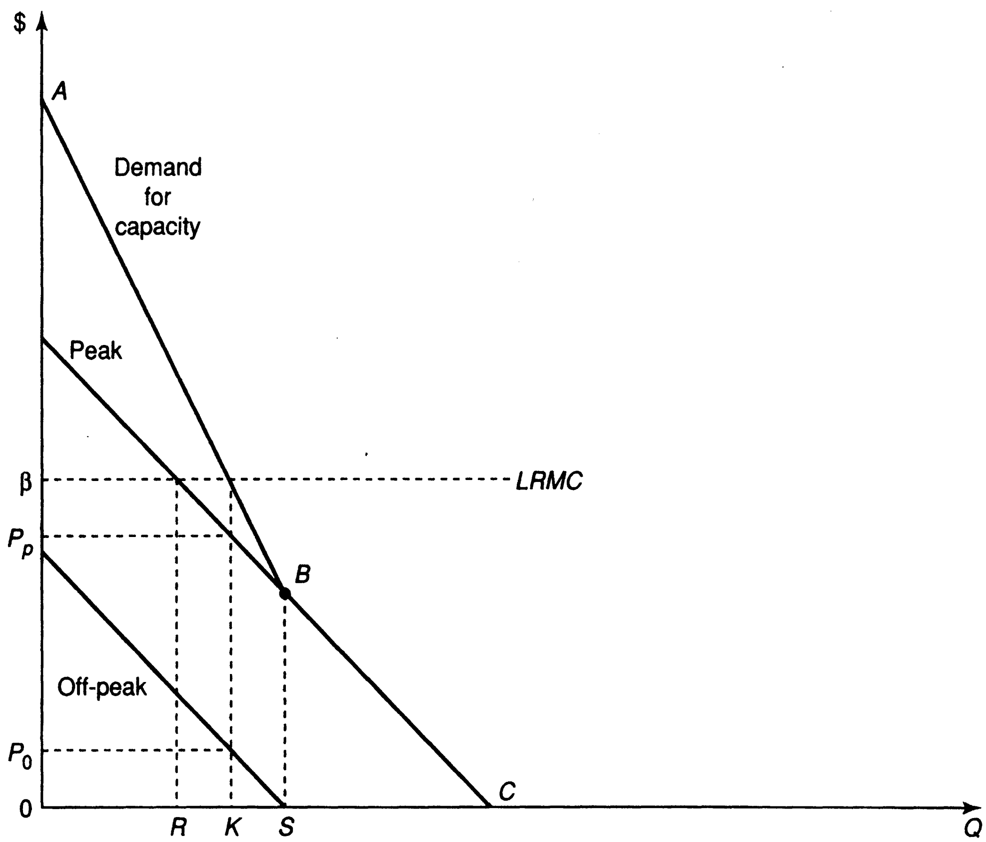

Figure 12.7
Demand for Capacity in Shifting-Peak Case

willingness-to-pay curve, the kinked curve ABC, is the demand for capacity. For example, at the output K the marginal willingness to pay is P_p by peakers and P_o by off-peakers, for a total of β. Because the capacity cost is β, K is the optimal capacity. The efficient prices for using this capacity are P_p and P_o, which, of course, add to β. Hence, in this case, the two groups share the capacity costs, unlike the firm-peak case. (If we had a nonzero b, the prices in each period would also include b.)

In practice, each day is not divided into two distinct periods of peak and off-peak demands. Rather, demand changes continually over the day, and there are also changes from week to week and season to season. This variability implies that the implementation of peak-load pricing requires judgment as to the metering of demand and how to fix periods during which prices are held constant. That is, one needs to recognize that it is not practical to have prices changing every few minutes—consumers would not be willing to respond to such constantly changing prices—and the costs of metering consumption in such short periods would be too high.

For these reasons, the typical pattern of peak-load prices for a residential customer would be similar to the following:

Weekdays 6 P.M.–6 A.M.	4 cents/kwh
Weekdays 6 A.M.–2 P.M.	7 cents/kwh
5 P.M.–6 P.M.	
Weekdays 2 P.M.–5 P.M.	14 cents/kwh
Weekends	4 cents/kwh

In addition to the time-varying rates, there would normally be a fixed charge per billing period. The rates listed here indicate that for this particular electric utility the peak period is on weekdays between 2 P.M. and 5 P.M., and the off-peak period is from 6 P.M. to 6 A.M. and all weekend. It might also be the case that the summer peak would exceed the winter peak. Hence the utility might have the rates above in effect for, say, May through October, and another schedule with somewhat lower rates for November through April. If the costs of metering small residential customers is viewed as too high to charge rates that vary over the day, the summer-winter scheme might be the only peak-load pricing implemented. (The point here is that the cheaper metering equipment simply accumulates total electricity consumed by the month; more expensive equipment is needed to measure use by time of day.)

Regulation/Deregulation of Electric Power

In this section we consider two topics. The first topic is the question of whether regulation of electric power utilities has been effective. That is, what evidence is there on the basic question of whether electricity prices have been restrained relative to what they would have been in the absence of regulation? The second topic is that of deregulation of electric power, or the emergence of price competition in various parts of the electric power industry.

Effectiveness of Price Regulation

Has regulation of electric power made a difference? The best-known study is one published in 1962 by Stigler and Friedland (S-F).[16] Their idea was to compare the price of electricity in states with regulation to the price in states without regulation. Unfortunately, because state regulation now exists almost everywhere, S-F had to go back to the 1920s and 1930s to find

16. G. J. Stigler and C. Friedland, "What Can Regulators Regulate? The Case of Electricity," *Journal of Law and Economics*, October 1962.

states without regulation.[17] This data problem consequently makes their results of limited relevance to the present time.

To illustrate the nature of their study, S-F found that in 1922 the average price of electricity was 2.44 cents per kilowatt-hour in states with regulation and 3.87 cents in states without regulation. Of course, this simple comparison is invalid because prices vary for reasons other than the existence of regulation. However, once they controlled through multiple regression analysis for other variables (such as the percent of power generated in hydroelectric plants—which is less costly than power in coal plants), S-F claimed that no statistically significant difference in price remained.

Some critics have argued that S-F did find some beneficial effects of regulation at slightly lower standards of statistical significance. They have also observed that the study period covered a time when regulation was just getting started, and that regulators are more effective today. Finally, critics have argued that the threat of regulation (observed in an adjacent state) could cause unregulated firms to hold down their prices in the hope of avoiding regulation.

Other analysts have tried alternative methods to answer the fundamental question of regulation's effectiveness in electric power. Meyer and Leland used data for 1969 and 1974 in their study, which attempted to answer the following question:[18] To what extent do the prices charged by regulated firms differ from what unregulated profit-maximizing firms would charge? They made use of econometric estimates of demand and costs to find the hypothetical unregulated prices. They found that the regulated prices were significantly lower, but that even lower prices would have been preferred. Similar work by Greene and Smiley found that unregulated prices for electricity are 20–50 percent higher than actual regulated prices.[19] The appeal of this methodology is that it can be applied to current data. However, very strong assumptions are needed to calculate the profit-maximizing prices, and this requirement makes their findings somewhat less persuasive.

Of course, given the problems with full regulation discussed earlier, if changing technologies can make it possible for competition to work at least in certain sectors of the industry, that approach is likely to be preferred to regulation. We now turn to this possibility in electric power.

---

17. And, as Joskow and Rose have observed, the states without regulation cannot be considered to be completely unregulated. Prior to state regulation most companies were regulated to some extent at the municipal level. P. L. Joskow and N. L. Rose, "The Effects of Economic Regulation," in R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*, Vol. 2, North-Holland, 1989.

18. R. A. Meyer and H. E. Leland, "The Effectiveness of Price Regulation," *Review of Economic and Statistics*, November 1980.

19. W. H. Greene and R. H. Smiley, "The Effectiveness of Utility Regulation in a Period of Changing Economic Conditions," in M. Marchand et al. (eds.), *The Performance of Public Enterprises: Concepts and Measurement* (Amsterdam: Elsevier, 1984).

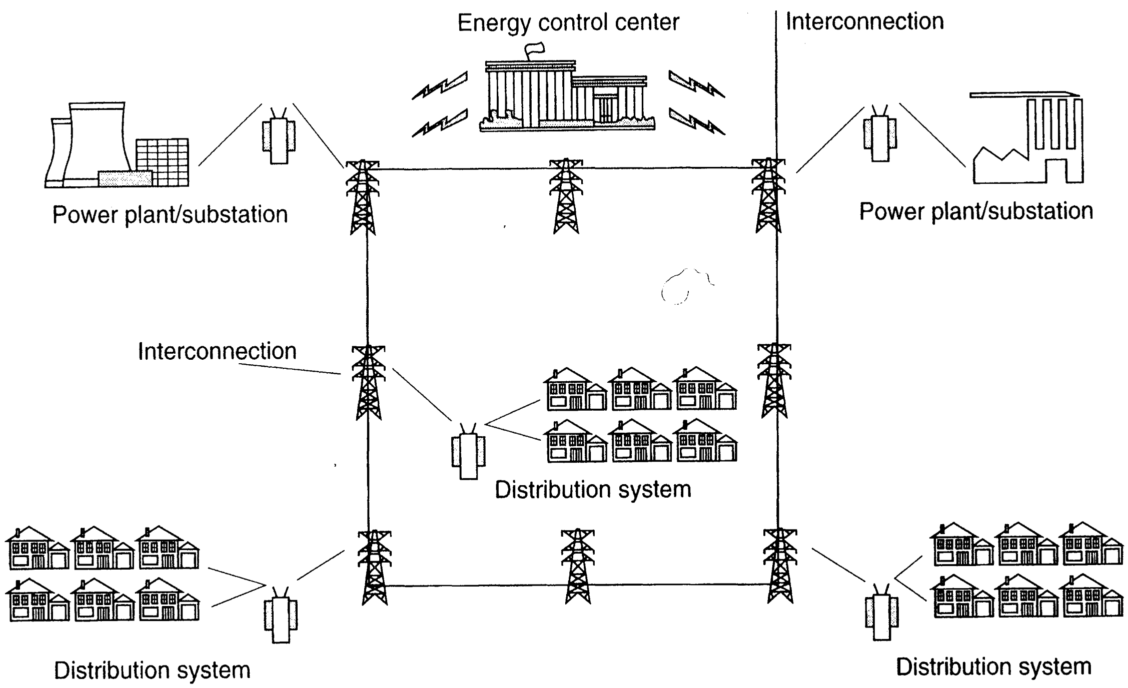

Figure 12.8
A Simple Electric System

Trend toward Competition

A major step toward price competition in electric power was taken by the passage of the Energy Policy Act of 1992. While the act had many provisions unrelated to electricity, the relevant provision for our purposes gave the Federal Energy Regulatory Commission (FERC) the right to order utilities to "wheel" power over their transmission lines. If utility A wants to sell power to nonadjacent utility B, and can only do so by using utility C's transmission lines, then utility C can assist in the transaction by wheeling the power from A to B.

To understand better the emerging competition in electric power, it is useful to consider a few basic facts about electric power systems. Figure 12.8 is a simplified power system. There were some 3,241 systems in the United States in 1990—most were small publicly owned systems. However, the 267 privately owned utilities accounted for 71 percent of the sales of electricity.[20]

Traditionally, most of the large privately owned electric utilities in the United States have been vertically integrated—owning power plants, substations, transmission lines, and distri-

20. Office of Technology Assessment, *Energy Efficiency: Challenges and Opportunities for Electric Utilities*, September 1993.

bution systems. The energy control center shown in Figure 12.8 coordinates the operation and dispatch of all power-system components within a geographic region. The control center is required because electricity must be generated as it is needed.

The interconnection shown in Figure 12.8 illustrates the fact that utilities are linked to others through a national grid. This linkage makes possible the sale of power from one utility to another—so-called wholesale transactions, or wholesale wheeling. Technological developments in high-voltage transmission have improved the feasibility of increasing numbers of these transactions. It should also be noted that wholesale power is also being sold by so-called independent power producers who are not subject to rate-of-return regulation.

One writer about the nature of the emerging electricity market has made a helpful analogy to a water system:

> Hundreds of utilities and wholesale power producers buy and sell electricity over the superhighway every day, in an increasingly competitive wholesale market. The system can be envisioned as a complex of water pipes, in which pressure must be maintained within a certain range. In a typical transaction, an electricity seller increases its level of power production, raising the "pressure" and causing electricity to flow into the network, while the buyer decreases its power production proportionately, lowering its "pressure" and thereby drawing a roughly equivalent amount out of the system.[21]

FERC is charged with regulating the price of these wholesale transactions, and has been moving away from traditional rate-of-return price setting toward market-based transactions. In 1994 a further important step was taken by some sixty utilities in an area covering approximately the western third of the United States plus British Columbia. They agreed to remove barriers to wheeling among themselves, and FERC agreed to permit market-based pricing within upper and lower limits.[22]

In addition to traditional utilities selling wholesale power to each other, other independent power producers who only generate power are becoming more important. The 1978 Public Utility Regulatory Policy Act (PURPA) required utilities to purchase power from "qualifying facilities" (QFs). These QFs are companies that install cogeneration equipment (e.g., companies that produce steam for heating and use the excess steam to generate power) and certain small power-production facilities that make use of renewable energy sources and a variety of waste fuels.[23] The Energy Policy Act of 1992 created another category of independent power producers not subject to regulation. These producers are termed "exempt wholesale

21. Matthew C. Hoffman, "Power Moves," *Reason,* June 1994, p. 52.

22. John Douglas, "Buying and Selling Power in the Age of Competition," *EPRI Journal*, June 1994.

23. P. L. Joskow, "Regulatory Failure, Regulatory Reform, and Structural Change in the Electrical Power Industry," *Brookings Papers: Microeconomics*, 1989, p. 163.

generators" (EWGs). A large fraction of the the first applicants for EWG status consisted of affiliates of existing utilities.

As one might expect, as wholesale wheeling has become increasingly important, large industrial buyers have begun to demand participation. They would like to have the right to buy power from any utility directly, not just from the utility that serves their geographic area. That is, "retail wheeling" would allow a retail customer to purchase power from utility *X* and have it delivered through its local utility. The local utility would be paid for the cost of delivering the power.

The incentive for retail wheeling is great in certain high-cost areas given the wide variation in power prices around the country. For example, in 1997 the average price per kilowatt-hour exceeded 10 cents in Massachusetts, Connecticut, and New York, and was 9.5 cents in California. Oregon's price was 4.7 cents, and Indiana and Wisconsin paid less than 6 cents. The U.S. average was 6.9 cents.[24]

Generating expenses account for slightly over half of the average prices, with transmission, distribution, and other costs accounting for the rest. It is generally believed that competition is feasible only in generation of electricity. The transmission function (high voltages) and the distribution function (low voltages) remain natural monopolies and will continue to require regulation.

Paul Joskow of MIT has pointed out that a large fraction of the variation in prices reflects differences in the sunk costs of generation and long-term purchase power contracts made during the 1970s and 1980s (when expectations about future costs were much higher than what actually transpired).[25] He also noted that there is a "price gap" of perhaps 3–4 cents in California and the Northeast between the price of generation service included in regulated retail rates and the lower current and projected wholesale market prices. If generation services were suddenly to be priced at market prices in those areas, the present value of losses to utilities would be on the order of $100 billion. This is the "stranded cost" problem. (There is currently a heated controversy over whether utilities should be compensated for these costs or simply suffer the losses when competition begins.)[26]

As noted, in high-cost areas of the country, the pressures for retail competition are especially high. Hence, it is no surprise that these states are already beginning such competition. California was the first state to initiate such a market, which began in April 1998. Hence, we shall briefly describe the California market, although it is far too early at this writing for meaningful studies evaluating the California market to have been made.

24. U.S. Energy Information Administration, *Electric Power Annual*, 1997.

25. Paul L. Joskow, "Restructuring, Competition, and Regulatory Reform in the U.S. Electricity Sector," *Journal of Economic Perspectives*, Summer 1997.

26. Gregory J. Sidak and Daniel F. Spulber, "Deregulatory Takings and Breach of Regulatory Contract," *New York University Law Review*, October 1996.

The California Power Market

The stranded-cost problem had to be "solved" to allow agreement on a new, deregulated electric power market in California. That is, the three large investor-owned utilities in California—Pacific Gas and Electric, Southern California Edison, and San Diego Gas and Electric—would not be able to recover their past high-cost investments (largely nuclear) and high-price purchase contracts if competition immediately lowered market prices far below the existing regulated prices. Hence, a four-year transition period was included in the deregulation law to allow the companies to recover the stranded costs through a "competition transition charge." In essence, small consumers were given an immediate small price cut, but then retail prices were frozen for the four-year period. At least partially in return for the ability to recover stranded costs, the companies agreed to sell about half of their generating capacity to seven other new firms so as to lower the possibility of market power in generation.

California's electricity market is centered on two newly created institutions. The Independent System Operator (ISO) has the responsibility to dispatch the generating plants in a least-cost manner, to manage the reliability of the transmission grid, and to provide open access to the transmission grid. Roughly, it is the control center shown in Figure 12.8, but for the whole state rather than a single system. While the transmission system is still *owned* by the utilities, it is to be completely under the control of the ISO. The ISO is a nonprofit organization regulated by the Federal Energy Regulatory Commission (FERC).

The other new institution is the Power Exchange (PX). It is an independent nonprofit corporation that seeks to provide an efficient, competitive energy auction open to all suppliers and demanders. It accepts day-ahead supply schedules from generators of power for each hour of the day. One might think of these schedules as similar to the one shown in Figure 12.4. Of course, they are not necessarily the firms' short-run marginal cost schedules. If a firm has some market power, it might bid prices greater than marginal cost. However, if the firm is sufficiently small, it would be more likely to bid marginal cost, or risk not selling any power. This topic will be discussed further in regard to the England and Wales market.

The PX then aggregates the supply functions and demand functions (price bids for various quantities by purchasers) and determines the market-clearing price for each hour. There are more complicated adjustments that we ignore here, but in essence, the results of the auction mechanism are then passed on to the ISO, which allocates the generation of electricity to those firms making bids below the market price.

A similar auction mechanism began in England and Wales in 1990.[27] Initially, unlike California, the market there was restricted to the 5,000 largest customers with maximum

27. Richard Green and David Newbery, "Competition in the British Electricity Spot Market," *Journal of Political Economy*, October 1992.

demands in excess of 1 megawatt. That is, only these 5,000 customers had the right to shop around for their preferred supplier. Of course, as in California, even though the local retail company's "wires" would still be used to deliver the power, the 5,000 customers could contract with other suppliers who would then pay a regulated price to the local company for the use of its wires. In 1994 competition was extended to an additional 50,000 customers. In 1998 the plan opened to the remaining 23 million small customers.

The England and Wales experience is discussed here because of horizontal market power problems that occurred there and the concern that these could happen in the United States. As mentioned earlier, part of the potential problem in California was mitigated by the sale of half of the three major utilities' capacities to other firms. Joskow has described the problem as resulting from the failure of the Thatcher government to divide the old state-owned generating assets into a sufficiently large number of new private companies. "Moreover, some generators have strategic locations on the grid and, from time to time, must run for reliability. Naturally, when generators know they will be called to run by the network operator to maintain network reliability (almost) regardless of what they bid, they submit high bids."[28] The government instituted temporary price ceilings and divestiture as a response.[29]

A further positive factor that should work for a more competitive market in California and elsewhere is the new generating technology, known as combined cycle gas turbine (CCGT). It is economical at small scales perhaps a fifth as large as the least-cost scale of the 1980s. It can also be built more quickly than the older, larger plants. Hence, if prices become too high relative to cost, entry should help to reduce them.

By July 1998, California was the only state with actual retail competition. However, various restructuring plans similar to that in California had been enacted in Connecticut, Illinois, Maine, Massachusetts, Montana, Nevada, New Hampshire, Oklahoma, Rhode Island, and Virginia. In addition, legislation at the federal level was being considered.

Summary

This chapter has discussed numerous problems of rate-of-return regulation. Cost-minimizing behavior is not encouraged by such cost-plus regulation. Firms that cut costs do not receive full cost savings, inasmuch as prices will be reduced to the new level of costs. However, the "accidental" institution of regulatory lag and the practice of disallowing certain expenses or additions to the rate base are factors that tend to offset this problem. The new price cap regulation in telephones and performance standards in electric power are also likely to encourage economic efficiency.

28. Joskow, 1997, p. 134.

29. David Newbury, "Power Markets and Market Power," *Energy Journal* 16, No. 3 (1995).

The Averch-Johnson effect leads to a capital bias and higher costs than are necessary. However, the applicability of the A-J model is not accepted by all economists because of the strong assumptions of that model.

Peak-load pricing is clearly an important development of the 1980s that has improved efficiency significantly. Although price regulation in electric power has probably resulted in lower prices than would have existed otherwise, a major current change that should be socially beneficial is the emergence of price competition at the wholesale level. Competition at the retail level began in California in 1998 and will likely spread quickly to a number of additional states.

Questions and Problems

1. The rate base can be valued in various ways. What, if anything, is wrong with the utility commission valuing the rate base at its value according to the stock market and the bond market? That is, use the market value of the firm rather than, say, the original cost of assets less depreciation.
2. Utility company executives are often quoted as saying that if their allowed rate of return is too low, they will be unable to attract capital to pay for capacity to meet increasing demands. Does this mean that regulators cannot use the allowed return as a device to provide incentives for utilities—raising the rate for good performance and lowering it for poor performance?
3. Consider the Edison Electric Company with a production function $Q = K^{0.5}L^{0.5}$, where Q is output, K is capital, and L is labor. The market rental rate of capital is \$0.50 and the wage rate is \$0.50 also. The utility commission has set the allowed rental rate at \$0.80. (Rental rates of capital are in dollars per unit of capital per year. With zero depreciation they are related to percentage costs of capital in the following way. Suppose that the utility must invest in a generator at a cost of \$5 per kilowatt of capacity, and 10 percent is its cost of capital; then the rental rate per year is 10 percent of the \$5 per unit, or \$0.50. Similarly, the percentage allowed rate of return would be 16 percent, since 16 percent of \$5 is \$0.80. Rental rates are therefore comparable to wage rates and other factor costs in applying standard static production theory.)

   Edison faces a demand curve with the constant elasticity of demand 2.857, or $Q = P^{-2.857}$. If Edison were unregulated, it would produce efficiently at a constant average and marginal cost of \$1. However, because of Averch-Johnson effects, it uses too much capital under regulation and produces at an average cost of \$1.01. Edison charges a price of \$1.35 and sells $Q = 0.42$.

   a. Find the price and quantity if Edison were an unregulated monopoly. Hint: Marginal revenue is $P(1 - 1/2.857)$.

   b. Find the sum of consumer and producer surplus for the case where Edison is regulated and where it is not. Hint: Using calculus, it can be shown that consumer surplus is $(0.54)Q^{0.65}$. Does regulation, even though imperfect because of Averch-Johnson effects, nevertheless result in an improvement over an unregulated monopoly case?

   c. Of course, the first-best case of price-equal marginal cost and efficient production is superior to regulation. Find the efficient solution. Draw a figure that shows the two types of losses that regulation causes as compared to the efficient solution.

d. Assume now that the utility commission decides to lower the allowed rental rate from \$0.80 closer to the market rate of \$0.50. Assume that it picks \$0.58. It can be shown that Edison will now choose to sell 0.67 units at a price of \$1.15. Its average cost of production rises to \$1.04. Compare this Averch-Johnson equilibrium with the earlier one in terms of total economic surplus. This, in fact, is the socially optimal allowed rental rate. Lower rates actually reduce total surplus. For further details, see A. Klevorick, "The Optimal Fair Rate of Return," *Bell Journal of Economics and Management Science*, Spring 1971.

4. How does the Averch-Johnson characterization of the regulatory process differ from reality?

5. Edison Electric Company's president has been arguing that residential electric rates need to be raised relative to industrial rates. His reason is that the rate of return that the company earns on its assets is higher from its industrial customers than from its residential customers. Is this a good reason? Hint: How can Edison determine its assets dedicated to the two classes of customers?

6. In a certain city where all parking is controlled by the city, it is possible to provide parking facilities in the downtown area at a constant marginal capital investment of \$10,000 per space. Costs of operation can be neglected. There are three equal periods during the day of eight hours each, and spaces are rented only for complete eight-hour periods. During the peak period of each of 250 days per year, the demand for parking is given by $P = a - bQ$, where P is the price per period for a parking space. During the other two off-peak periods of those 250 days, the spaces demanded are half that in the peak period, for each possible price. On other days demand is zero. Assume that the interest rate is 10 percent and the facilities do not depreciate.

   a. If $a = \$16$, $b = 0.08$, and existing spaces are 120, what would be the socially optimal prices during the three periods?

   b. What is the optimal number of spaces and what are the corresponding prices?

   c. This case is a so-called firm-peak case, with peak demanders paying all capital costs. Now suppose that $a = \$5$ and $b = 0.08$. If peak demanders pay all capital costs, what quantity is demanded by peak demanders? If off-peak demanders pay zero, what is their quantity demanded? (Fractions of spaces are legitimate.) This is the shifting-peak case.

   d. For the demand curves in part c, find the optimal number of spaces and the corresponding prices.