BOARD STAFF INTERROGATORIES

References:

1. **AMPCO Evidence: "Potential efficiencies from improving transmission rate design in Ontario" , August 26, 2010**

2. **AMPCO Expert Evidence of Anindya Sen "Will greater load shifting by industrials result in lower electricity prices for all? Evidence from Ontario, Canada", August 2010**

3. **Exhibit H1 / Tab 5 / Schedule 1 / Attachment 1: Power Advisory "Assessment of AMPCO's High 5 Proposal for Establishing Network Charge Determinants", July 6, 2010**

Board Staff Interrogatory #1

Ref: # 1, p. 3

AMPCO quotes A.E. Kahn, The Economics of Regulation, in an excerpt at p. 89 from Chapter 4, "The Application of Long- and Short-Run Marginal Costs".

a. Please file a copy of the paragraph that follows the one quoted in AMPCO's evidence. In light of the third caveat expressed in that paragraph, does AMPCO suggest that it would be practical for Hydro One to set its Network charge at marginal cost in 2012, or the foreseeable future?

b. Please file a copy of Kahn's text pp. 106-107. If necessary please file any additional excerpts that AMPCO considers would be helpful in understanding the second paragraph on p. 107 and assessing its applicability to AMPCO's High 5 proposal.

Response

a) It is important to understand the context for Kahn's treatment of marginal costs. For example, at the time of the writing of his text (in 1970; the version we cite is the 7th printing, in 1998), restructuring electricity markets had hardly commenced. Indeed, in the introductory chapter to his book (at page 10) he recites a list of public utilities, among which is "the generation, transmission and distribution of electric power", i.e., a vertically integrated monopoly. In this context, the application of marginal cost pricing in setting appropriate public utility rates finds extensive scope.

   In considering an application where the assets concerned are exclusively network transmission assets, however, the question of marginal cost is more narrowly defined. In fact, apart from transmission-related losses, which already are recovered from customers on a marginal cost basis, there are no costs which could clearly be defined as marginal on a short-run basis, i.e., that vary with marginal changes in demand. The costs associated with network transmission service, therefore, can fairly be characterized as capacity costs or long run marginal costs in which case the application of the principle is clear.

The third caveat, referenced above, deals with circumstances of decreasing costs, in which, as Kahn points out, the marginal cost may be less than the average cost and marginal cost pricing would cause the utility to collect insufficient revenues. These circumstances may once have pertained in the world of vertically integrated utilities, where the premise was one of continually increasing economies of scale, but it is not a circumstance that is associated with network transmission costs, each incremental investment in which is apparently increasingly expensive, at least in Ontario.

b) Please find attached Chapter 4, "The Application of Short- and Long-Run Marginal Costs", pp. 87-122, incorporating the paragraphs requested by Board Staff.

CHAPTER 4

The Application of Long- and Short-Run Marginal Costs

Having established in Chapter 3 that economically efficient prices of public utility services would be based on some *a priori* unspecifiable mixture of short and long-run marginal costs, in this chapter we consider two major contexts in which the best mixture needs to be discovered and applied: in determining which customers should pay the capital costs, and in deciding how and to what extent rates ought to be changed over time, as marginal costs change.

THE DISTRIBUTION OF CAPACITY COSTS

In industries as capital-intensive as public utilities, the costs of providing the capacity to serve—depreciation, property, income taxes, and return on investment—are very large. Yet, we have asserted, capacity costs are not part of SRMC and therefore, in principle, should not be reflected in price (except to the limited extent that they are in fact variable).

However, what if a price that covers only the variable operating costs elicits a demand for the service so great that it cannot be supplied with existing capacity? Economists have long been bemused by Dupuit's and Hotelling's historic example of the bridge and the strong case they made against charging tolls, on the ground that operating, maintenance, and capital costs do not vary significantly with the rate of utilization.[1] But what if charging a zero toll would, at least at certain hours of the day, produce such an increase in traffic that cars lined up for miles at the bridge entrance and a crossing took an hour instead of a few minutes? In that event, the SRMC of bridge crossings, at those times, is not zero. It can be envisaged in terms of congestion: the cost of every bridge crossing at the peak hour is the cost of the delays it imposes on all other crossers. Or it can be defined in terms of opportunity cost: if A uses the bridge at that time, he is taking up

[1] [illegible] "On The Measurement of the [illegible] Works," *Annales des Ponts et* [illegible] series, VIII (1844), and reprinted [illegible] *Economic Papers*, No. 2 (New York: Macmillan, 1952), 83–110; Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica* (July 1938), VI: 242–269.

space that someone else could use; therefore, the cost of serving him is the value of that space or capacity to others who would use it if he did not.[2]

The Shift to Long-Run Marginal Costs

Suppose now that for any one or more of the following reasons, we decide that it is either infeasible or undesirable to base tolls on SRMC

1. It would mean an unacceptable fluctuation in rates over time, depending on the changing relation of demand to capacity;
2. It would be too difficult, annoying, or expensive to compute the changing marginal congestion or opportunity costs just described and to base prices on them;
3. Pricing on this basis might not cover ATC over the life of the bridge, and therefore might require public subsidy.[3]

Suppose, therefore, we decided to base tolls on long-run marginal costs. Then we would have to recognize that satisfying additional demand at times of congestion may sooner or later call for construction of additional capacity. In these circumstances, LRMC includes capital or capacity costs: and efficiency (of a kind of "second-best," however[4]) requires that each potential bridge-crosser be confronted with the price that reflects those marginal opportunity costs of serving him.

But notice that a shift from SRMC to LRMC does *not* mean that prices should be set on the basis of current variable costs plus a gross return (including depreciation) on past investment, however valued. Marginal costs look to the future, not to the past: it is only future costs for which additional production can be causally responsible; it is only future costs that can be saved if that production is not undertaken. If capital costs are to be included in price, the capital costs in question are those that will have to be covered over time in the future if service is to continue to be rendered. These would be the depreciation and return (including taxes) of the future investments that will have to be made. These incremental capital costs per unit of output will be the same as average capital costs of *existing* plant only in a completely static world and under conditions of long-run constant cost. As for the former and by far the more important qualification, in a dynamic economy, with changing technology as well as changing factor prices, there is every reason to believe that future capital costs per unit of output will not be the same as the capital costs historically incurred in installing present capacity.

Here, then, we encounter a major discrepancy between the economist's

---

[2] See our discussion for the separability of joint costs on this basis, pp. 79–83, above. For an interesting example in the case of communications satellites, see Johnson, in Shepherd and Gies, *Utility Regulation*, 119, note 5 and 120. And for a strong demonstration of the inefficiencies caused by our failure to impose charges for the use of the radio spectrum reflecting these opportunity costs—measured by the value of any particular allotted channel to the next-excluded potential user—see Harvey J. Levin, "The Radio Spectrum Resource," *Jour. Law & Econ.* (October 1968), XI: 433–501.

[3] On the first two problems see pp. 83–86 and 103–109; the third is the subject of Chapter 5.

[4] "First-best" rates would equal LRMC only by chance at some instant of time: at certain times (when capacity is ample—for example, right after the new or additional bridge has been built) they would be far below, then (as demand grew) they would rise gradually to and above it, as congestion increased, to whatever point necessary to cover congestion costs and ration the limited capacity, until construction of yet more facilities was justified. LRMC would instead represent an average over time of estimated total additional costs. (See pp. 107–108.)

prescription for optimal pricing and the traditional and still generally followed approach of public utility regulation. The latter, preoccupied with assuring a reasonable gross return on the existing investment, cannot possibly, except by accident, be basing its permitted rates on marginal costs, long-run or short-run.[5] The one conceptual merit—in contrast with its crippling administrative infeasibility—of the use of the reproduction cost instead of the original cost rate base was that it sought to bring the computation of capital costs closer to current, and away from historic costs. But, as we shall see later in this chapter, its manner of doing so was defective.

Does the shift to LRMC mean that all users of the bridge should pay a price that includes capacity costs? No. The off-peak users impose no such costs on society, *provided* their demand is sufficiently slight and inelastic that even at a zero toll no congestion occurs at the time they cross over. The incremental costs of serving them—in the long-run, not just the short—are still zero, and may remain so indefinitely. This is the case even if the off-peak demand grows over time and continues to be satisfied without congestion only because the bridge's capacity is being expanded. The necessity for expansion is imposed by the customers at the peak hours. It remains true that if one or all of the off-peak users ceased to cross the bridge, briefly or permanently, society would be saved no costs whatever.

Notice how the intensity and elasticity of demand help determine the level of marginal costs. For those hours of the day at which demand is insufficiently strong or responsive to a toll covering only operating expenses, long-run marginal costs include only those operating expenses; for those times of day at which demand is strong or so responsive to a lower toll as to cause congestion, LRMC necessarily includes capital costs as well.[6]

Peak Responsibility

The economic principle here is absolutely clear: if the same type of capacity serves all users, capacity costs *as such* should be levied only on utilization at the peak. Every purchase at that time makes its proportionate contribution in the long-run to the incurrence of those capacity costs and should therefore have that responsibility reflected in its price. No part of those costs as such should be levied on off-peak users.

The principle is clear, but it is more complicated than might appear at first reading. Notice, first, the qualification: "if the same type of capacity serves all users." In fact it does not always; in consequence, as we shall see, off-peak users may properly be charged explicitly for some capacity costs. Second, the principle applies to the explicit charging of capacity costs, "as such." Off-peak users, properly paying *short-run* marginal costs, will be making a contribution to the covering of capital costs also, if and when SRMC exceeds average variable costs. Third, the principle is framed on the assumption that all rates will be set at marginal cost (including marginal capacity costs). Under conditions of decreasing costs, uniform marginal-cost

---

[5] [illegible] example, Troxel, *Economics of Public* [illegible] 305–306, and in Shepherd and Gies, [illegible]

[illegible] other state commissions, the Michigan [illegible] relies mainly on legal and accounting [illegible] of thought. In a general-rate case, for instance, attention is focused primarily on past revenues, past costs, and the value of an existing plant."

[6] This should not be surprising. Except when marginal costs are constant, it has to be the intersection of the demand and the cost functions that determines the equilibrium level of MC.

pricing will not cover total costs. Lacking a government subsidy to make up the difference, privately owned utilities have to charge more than MC on some of their business. In these "second-best" circumstances, some (or all of the difference between average and marginal) capacity costs might better be recovered from off-peak than from peak users. We will illustrate all these facets of the principle presently though reserving systematic consideration of the decreasing costs situation for Chapter 5.

First, to establish the basic principle, it is wisest to simplify. Consider first a uniform type of capacity, serving both peak and off-peak users; assume that marginal costs, both short- and long-run, are constant, so that SRMC cannot ever exceed AVC and there can be no difference between marginal and average capacity cost; and assume, finally, that the peak is fixed—that is, that demand at one fixed time or period always presses hard on capacity (after making allowance for reserve capacity held in standby for emergency) and at "the" other period never does so even if the former bears all and the latter none of the capacity costs.

The problem of apportioning capacity costs between these two classes of customers is, precisely, the problem of costing joint products—the solution for which we have already described, using the example of cotton fiber and cottonseed oil. In the present instance, the same production capacity is available to provide two separate services, in fixed proportions: every kilowatt of electricity capacity, every cubic inch of natural gas pipeline space, every telephone circuit available for service in January is available also in July. As we have already seen, the respective supply prices of the joint services (as far as the joint portion of the production process—in this case, the provision of capacity—is concerned) depend on the relative elasticities and intensities of the two demands. The competitive solution requires, first, that the combined prices of the two services add up to no more than the marginal cost of producing the two together, and, second, that the price of each be set at a level at which the quantities demanded and the quantities supplied will be equated.

The case of the fixed and unchanging peak is the case illustrated in our Figure 2. The peak demand there is for cotton; the demand for cottonseed oil is irrevocably "off-peak" and must bear none of the capacity costs. Any attempt to shift capacity costs to the off-peak demands, by raising prices for that service above its own separate, incremental cost (MC_{cso}), will cause available production capacity at that time (cottonseeds) to be wasted, and would cut off purchasers willing to pay the additional cost of serving them. Any reduction of the peak (cotton) price below the full joint cost, P_c, would stimulate additional purchases at the peak, requiring additions to capacity that would not be made if buyers had to pay the full opportunity costs of the additional resources required to supply them. Similar to the cotton seeds in Figure 2, the capacity available off-peak is a free good and should be priced that way: it has a zero marginal cost at the point of intersection of competitive supply and demand.

---

[7] Actually a utility has some discretion about the times and seasons when it will close down various units of capacity for repair and maintenance, and will try to concentrate those shutdowns in off-peak periods. It will have identifiable peak and off-peak uses, then, only if the fluctuations of demand are wider than the fluctuations in plant availability due to maintenance. See Ralph Turvey, "Peak-Load Pricing," *Jour. Pol. Econ.* (January–February 1968), LXXVI: 103.

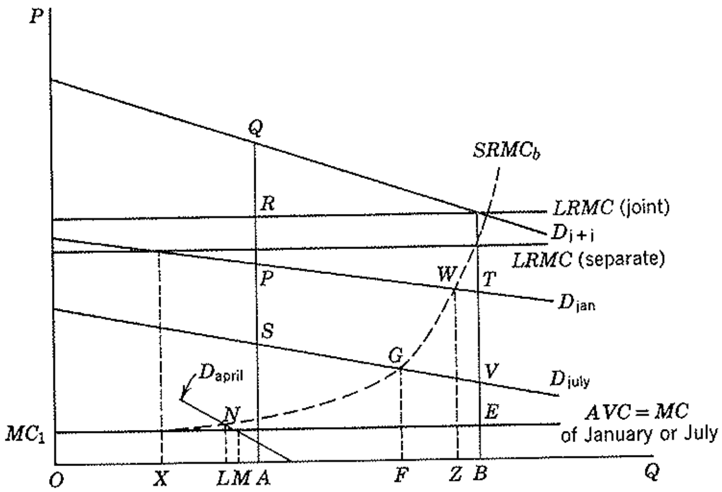
In the real world, demand peaks do not necessarily stay fixed: their location may shift for two reasons. First, if the elasticities of the separate demands are great enough, imposing all of the capacity costs on the peak customers and none on the off-peak may give rise to excess capacity at the former time (the previous peak) and congestion or shortages at the latter. And, second, the pattern of demand may change over time. The first phenomenon is static: it exists because the relationship between the demand functions is such that the proportion of total capacity costs imposed on each will determine at which time demand presses on capacity. An example would be the way in which public utility promotion of residential air-conditioning has apparently contributed in some areas to a shift in the peak demand for electricity from winter to summer. The second source of the shift is dynamic: it occurs as a result of changes over time in the respective demand functions. The increased use of electricity for summer air-conditioning has almost certainly reflected, above all else, dynamic factors such as the general rise in incomes, the perfection and reduced prices of air-conditioning equipment and the inclusion of air-conditioned summer comfort in the American standard of living. We confine our attention in this section to the static phenomenon; the second clearly belongs in our discussion, below, of what utility companies should do if long-run marginal costs change over time.

The demand situation in Figure 3 is the one that corresponds to the shifting peak. We reproduce it in Figure 4 with captions relevant to a public utility situation—for example, the demand and supply of natural gas in January and in July[8]—with the added complications of (1) recognizing that each of these services will have its own, separate set of variable costs (AVC), (2) introducing (in a dashed line, $SRMC_b$) an alternative short-term marginal cost, embodying the assumption of short-run increasing costs (for simplicity, we do so for a plant designed to produce OB units at lowest possible cost), (3) adding a line (AQ) that enables us to show how the competitive norm would apply in a situation of long-run disequilibrium, and (4) introducing a third, much weaker demand—for April.[9] Let us begin with the assumption that the variable costs are constant—that is, that the $SRMC_b$ does not apply. In this event we can ignore the April customers: their demand is unchangeably off-peak; they should pay MC_1, consume OM, and contribute nothing to the joint costs.

Suppose, initially, that capacity is OA. Clearly, it would be wrong at this point to levy all the capacity costs on the January customers: at such a price (LRMC separate), which includes the total joint capacity costs plus the

[8] In this event the x-axis would represent cubic feet, the y-axis cents per cubic foot; MC and $LRMC$ would be the marginal costs of supplying various quantities of cubic feet *per month*. The assumption here is that those costs would not differ from one month to another; it is not a necessary assumption but it simplifies the presentation. And it is assumed that only January and July are potential peaks, depending on the allocation of the joint capacity costs between them.

[9] The figure and exposition are based on Hirschleifer, *op. cit.*, *Q. Jour. Econ.* (August 1958), LXXII: 452, slightly reformulating those of Peter O. Steiner, "Peak Loads and Efficient Pricing," *Q. Jour. Econ.* (November 1957), LXXI: 588. For a slightly different presentation of these same solutions see Ronald L. Meek, "An Application of Marginal Cost Pricing: The 'Green Tariff' in Theory and Practice," Part I, *Jour. Ind. Econ.* (July 1963), XI: 224–230 and the famous article by Marcel Boiteux, *op. cit.*, in Nelson, *Marginal Cost Pricing in Practice*. For an elegant, more generalized statement of the solution, see Oliver E. Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraints," *Amer. Econ. Rev.* (September 1966), LVI: 810–827.



LRMC (joint) includes long-run capacity costs plus two long-run variable costs, one for January, one for July. Both are assumed constant. This is the amount that the sum of January and July prices must cover, in long-run equilibrium.

LRMC (separate) includes long-run capacity cost plus one long-term variable cost, representing the long-run costs of serving January or July.

AVC: the long-run variable costs of serving January or April or July.

$SRMC_b$: an alternative short-run marginal cost of producing from a plant of *OB* capacity.

Figure 4. Allocation of capacity costs, shifting peak.

variable costs of serving the January customers, they would demand only the quantity *OX*; *XA* capacity would remain idle in January. And the July users, being charged only their own separate variable costs (MC_1), would experience shortages. For the limited available supplies to be effectively rationed, while fullest economic use is made of capacity, the July users would have to be charged *AS* and the January customers *AP*. At such prices, the supplier would be earning excess profits: a combined price of *AQ* for one unit of sales in each month (remember that the demand curve D_{j+j} represents a summation of D_{July} and $D_{January}$)[10] compared with a combined unit cost of *AR*. The long-run competitive solution would be to increase capacity to *OB*, which is at the juncture of the combined demands and the joint long-run marginal costs of supplying both (capacity costs plus one set of long-run average variable

[10] The addition of the two demands (and the entire discussion in the text) assumes that they are independent of one another. This was a reasonable assumption for cotton fiber and cottonseed oil. But as between peak and off-peak power, gas or telephone service, there is likely to be some cross-elasticity of demand: a reduction in the rate on night long-distance telephone calls, for example, is likely to induce some users to shift their telephoning from day to night. To this extent, adding together the separate demand curves as though they were independent will exaggerate the elasticity of the joint demand curve. (In Figure 4, for example, $D_{January}$ shows how the quantities of power purchased in January would vary if the January price *alone* were changed and correspondingly for D_{July}. But if their prices were reduced simultaneously, from *AP* to *BT* and *AS* to *BV*, respectively, and if the elasticity of each was in part a reflection of cross-elasticity, the combined quantities purchased would not increase by the full *AB* indicated.) Rate cuts on either of the two services would take some business away from the other, and would therefore not increase total sales by the amount indicated by the separate curves. (An attempt to sell the additional quantities *AB* both in January and July would thus depress the respective prices below *BT* and *BV*.)

In practice, public utility companies will, of costs for each), charging the January users BT, the July users BV. At this point, both efficiency requirements are met: the combined prices do not exceed the marginal cost of producing the two services, and the price of each is set at a level that clears the market. There is no other set of prices at which these equilibrium conditions would be satisfied.

The introduction of the shifting peak does not alter the fundamental principle of peak-responsibility. The point is, simply, the January and July users both represent the peak, and they must therefore share the costs. On what basis? On the basis of the respective intensities and elasticities of the two demands.[11] It remains true that the long-run marginal cost of supplying purchasers in, for example, April and October includes no capacity costs as such; nor, ideally, should their prices.

Whether confronting a fixed or a shifting peak, the principle of peak responsibility is a relatively simple one in the presence of simplifying assumptions. One of these assumptions is that variable costs are constant to

---

course, have to take into account the impact of changes in some of their rates on revenues from other parts of their business in planning both their price structures and their decisions with respect to capacity. But this consideration in no way vitiates the conclusions we have reached with respect to socially optimal pricing; it merely suggests that in our Figure 4, the presence of cross-elasticities of demand will produce a long-run equilibrium capacity of something less than OB and rates for January and July sales somewhat higher than BT and BV, respectively.

[11] Steiner concluded for this reason that his solution to the problem of apportioning capacity costs in the case of the shifting peak involved price discrimination: the separate marginal costs of serving the January and the July customers are the same (AVC in Figure 4), yet the prices charged them differ. Hirschleifer contends, correctly I believe, that the solution is not discriminatory. His exposition follows two alternative lines.

1. The fact that the correct solution would involve the nondiscriminatory result of carrying on production in each market to the point where short-run marginal cost equaled price is obscured, in Figure 4, by the convenient, simplifying assumption of constant short-run marginal costs up to the limit of physical capacity. If the more conventional costs were assumed, increasing before the limit of producing capacity is reached (as in $SRMC_b$), it would be clear, as we point out in the text immediately following, that the efficient solution would involve producing in each market up to the point (F and Z respectively) where that short-run marginal cost was equated with demand price.
2. But even if the short-run marginal cost were indeed horizontal, then discontinuous, as at EVT, in Figure 4, so that in both January and July production was carried on to the physical limit of capacity, the efficient prices would still be equated with the respective marginal opportunity costs. As we have already pointed out in discussing the separate costing of joint products (pp. 82–83), the economic cost of supplying an additional unit to any single customer in each of the two markets is measured by what the next customer, the first one not served, would have been willing to pay for that service: and, in Figure 4, that would be (infinitesimally less than) BV and BT in the respective markets. Therefore, according to this conception of cost as well, the correct solution involves equating price in each market to $SRMC$, and hence no discrimination.

This second line of argument, as Steiner points out, does seem to obscure the difference between the demand function and the cost function, which are supposedly equated at the margin in each market if there is to be no discrimination: the "marginal opportunity *cost*" is defined as equal to the market price at *whatever* level the latter happens to be set. It might seem therefore to define away the possibility of discrimination. "Reply," *Q. Jour. Econ.* (August 1958), LXXII: 467–468. But in fact it does not. The critical consideration, Hirschleifer responds, is that when markets are artificially separated, in the familiar price discrimination model, the marginal opportunity costs in the two markets are really the same—namely the price that the next-unsatisfied customer *in the higher price market* would be willing to pay: so, since their MC's are the same and their prices differ, genuine discrimination is practiced in that case. In the January–July model, instead, the joint products are not the same; the marginal cost of supplying the July customer is the lost opportunity of supplying the next-unsatisfied purchaser in *July*, not in the physical limit of capacity, as in the horizontal *AVC* of Figure 4.[12] But in point of fact the production of public utility services is at times subject to short-run increasing costs. This is particularly clear of the generation of electricity. At any given time, generating plant will vary widely in age, type, location, and efficiency—hence in the level and slope of its *SRMC*. The common—and entirely rational—practice of electric companies, therefore, is to hold the less efficient generating units in reserve and phase them into operation from one moment to the next, according to the level of demand, in ascending order of their marginal costs.[13]

In this event some such alternative as $SRMC_b$ of Figure 4 becomes the relevant marginal cost function. How does its introduction alter the solution? Now, ideally, production in July should not exceed *OF*, and July users should pay not *BV* but *FG*.[14] Clearly the July users, though charged only the *SRMC* of serving them, will end up paying a much larger contribution to joint costs than before in each MCF of gas they buy. Even more interesting, note what happens to the April customers. They ought now to be charged *LN*. But this price exceeds the average variable costs. Therefore, although they pay only *SRMC*, consume far less gas than the January and July customers, and their purchases are certainly off-peak, they make some contribution (over and above *AVC*) to joint capacity costs—and correctly so.[15]

January; and correspondingly for the January customer. So *their* marginal opportunity costs do differ, and, if so, their prices *should*; and, if they do, there need be no price discrimination. *Ibid.*, 459. The classic discussion of this very issue was that of Frank W. Taussig and A. C. Pigou, on the subject of "Railway Rates and Joint Costs", *Q. Jour. Econ.* (February, May, and August 1913), XXVII: 378–384, 535–538, and 687–694. A central issue in their debate was whether railway services to different customers are in fact homogeneous, in which event it would be proper to characterize rate differences as discriminatory. Their underlying difference of opinion was whether the costs of serving different railroad customers may properly be regarded as joint, as we have used the term (in which event the rate differences, they agreed, would not be discriminatory), or common.

However, in practice, for reasons we have already suggested, it has usually been infeasible for utility companies to differentiate and to vary their rates with the ever-shifting balances of capacity and demand at various times of day, year, and planning period in such a way as to equate price at each moment to short-run marginal opportunity costs. Therefore, they typically—in wholesale and large industrial sales—charge separately for variable and for capacity costs (as in the two-part tariff, discussed below), and do engage in considerable price discrimination in distributing the latter burden among various classes of customers on the basis of their respective elasticities of demand (see pp. 95–98, on the 2-part tariff, and Chapter 5).

12 Another simplifying assumption that we make in this chapter is that LRMC are constant. We turn to the decreasing cost situation in Chapter 5. For the solution to the problem of apportioning capacity costs in the case of the shifting peak under conditions of decreasing costs, a case that unequivocally requires price discrimination if revenues are to cover total costs, see Johnson, in Shepherd and Gies, *op. cit.*, 131–132.

13 For a more precise statement see the discussion of integration and power pooling, at note 51, Chapter 2, Volume 2. Also Meek, *op. cit.*, Part II, *Jour. Ind. Econ.* (November 1963), XII: 46–47; Bonbright, *Principles of Public Utility Rates*, 320–322.

14 January users would now pay *ZW*, instead of *BT*, and consume correspondingly less. We make no effort here to relate the case of rising SRMC to determining the correct size of plant or ascertaining at what level of capacity total revenues collected at the new January and July prices (equated to the *SRMC* of that plant) would cover total costs. See the sources cited in note 9. It might appear that *OB* was still the proper size, since the combined demand prices just suffice to cover total *LRMC* at that point. But the *LRMC* curves represent the costs *per unit* of output, on the assumption that whatever plants are built are operated to capacity in January and July. If in fact, with rising *SRMC*, only *OF* and *OZ* are now consumed, the *LRMC* per unit for a plant of *OB* capacity will be higher than those shown on the diagram.

15 See, for example, Meek, *op. cit.*, Part II, *Jour. Ind. Econ.* (November 1963), XII: 47–48.

Public utility companies do employ peak-responsibility pricing to some degree. The telephone companies charge lower rates for night than for daytime long-distance calls; electric companies frequently have low night rates for hot-water heating; both they and natural gas companies—local distributors and interstate pipelines alike—offer at lower rates service that the customer will agree may be interrupted if capacity is being taxed by other users and try to promote off-peak sales in numerous ways;[16] railroads charge lower rates for return-hauls of freight, when the greater flow is in the opposite direction; airlines offer special discount fares—family plans, youth fares, and so forth—for travel on unfilled planes or in slack seasons or days of the week.[17]

The two-part tariff, generally credited to John Hopkinson, an English engineer, and almost universally used by electric and gas utilities for large-volume sales at wholesale and to industrial users, represents an effort to apply just such a principle. The first part—the energy, commodity or "running" charge—embodies the variable costs, properly charged to all customers, and is levied on a per unit of consumption basis (per kwh or per MCF of gas). The second part—the demand or capacity charge—is a charge for the utility's readiness to serve, on demand. This readiness to serve is made possible by the installation of *capacity*: the demand charge, therefore, distributes the costs of providing the capacity—the fixed, capital costs—on the basis of the respective causal responsibilities of various buyers for them. And the proper measure of that responsibility is the proportionate share of each customer in the total demand placed on the system at its peak. (Sometimes the tariff will have three instead of two parts—the third, "customer" charge reflecting the costs of services such as meter-reading and billing that vary on a per customer basis instead of with different amounts purchased.)[18]

Unfortunately, the principle has usually been badly applied, in several important ways. First, if the demand charge were correctly to reflect peak responsibility it would impose on each customer a share of capacity costs equivalent to his share of total purchases at the time coinciding with the

---

[16] A particularly illuminating example is provided by the case of a combination company—that is, one distributing both electricity and gas—the two major portions of whose business had noncoincident peaks. The Chairman of the Board of Directors of the Public Service Electric and Gas Co. reported to his stockholders:

> "In our sales promotion programs we are stressing the selling of 'off-peak loads', such as electric heating, to increase the winter use of electricity, thus helping to offset the summer air-conditioning peak; and gas air-conditioning and interruptible gas service to induce greater use of gas in the off-peak summer period." Annual Meeting of Stockholders, April 18, 1966.

Note that the company was competing with itself—pushing the off-peak sales of each product in competition with the other in periods of the latter's peak demand.

Our discussion of peak-responsibility has run entirely in terms of pricing policies. As the Public Service example suggests, the same considerations would justify public utility companies using various other sales promotional devices, such as intensive advertising or the sale of the relevant appliances at cost, or less, to increase off-peak sales. On the general question of the proper treatment of selective promotional expenditures, see pp. 149 and 164, note 10.

[17] For a decision sustaining reduced railroad rates for coal shipped during the slack season, provided those rates were available nondiscriminatorily to all shippers, see *ICC v. Louisville & N.R. Co.*, 73 F. 409 (1896) and another disallowing a similar seasonal reduction by a motor carrier on household goods because it did not meet the condition of nondiscrimination, ICC, *Reduced Seasonal Household Goods Rates*, Report and Order, 332 ICC 512 (1968).

[18] More often the customer costs will be recovered by specifying a minimum bill, or in sufficiently high per unit charges for the first block of electricity or gas purchased.

system's peak (a "coincident peak" demand charge). Instead, the typical two-part tariff bases that rate on each customer's *own* peak consumption over some measured time period, regardless of whether *his* peak coincides with that of the system (hence the designation "noncoincident" demand charge). That is, the peak (for example, half-hour) consumption of all customers, regardless of the time of day or year in which each falls, is added up, and each then is charged a share of total system capital costs equivalent to the percentage share that his peak consumption constitutes of that total. The noncoincident demand method does have some virtue: it encourages customers to level out their consumption over time, in order to minimize their peak taking, hence their share of capacity costs. This, in turn, tends to improve the system's load factor—the ratio of average sales over the year to capacity—that is, the degree of capacity utilization. But it is basically illogical. It is each user's proportion of consumption at the *system's* peak that measures the share of capacity costs for which each is causally responsible:[19] it is consumption at *that* time that determines how much capacity the utility must have available. The system's load factor might well be improved by inducing individual customers to cut down their consumption to a deep trough at the *system* peak and enormously increase *their* peak utilization at the system's off-peak time: yet the noncoincident demand system would discourage them from doing so.[20]

Second, the charges have typically been based on average instead of marginal costs. Therefore, the energy charge has generally ignored the fact that electricity is produced under conditions of short-run increasing cost; and the demand charge has tended to embody the opposite error.

Third, the two-part tariff has applied only to bulk sales. Retail sales of gas and electricity to households typically contain no such differentials based on time of consumption (with specific exceptions such as special night rates for water-heating). Instead, they usually carry block rates, with diminishing charges for larger blocks of consumption: for instance, 6¢ for the first 30 kwh, 4¢ for the next 50, 3¢ for the next 100, 2¢ for the next 570 and 1½¢ for anything above 750 kwh—regardless of the time of taking.[21] Since household utilization typically has a marked peak that coincides roughly with that of the system (whether because of air-conditioning on hot summer days, or for home heating, lighting, and cooking in the early evenings of short and cold winter days), the use of diminishing block rates has a strong perverse tendency to underprice marginal sales at the peak.[22] Against this distortion, however, one must weigh the tendency of such declining block rates correctly to reflect the declining unit costs of electricity and gas distribution with increased intensity of use.

---

[19] This entire discussion continues under the assumption that capacity costs are constant, so that *average* capacity costs (which is what are measured by both coincident and noncoincident demand methods) are the same as marginal capacity costs. If instead the system is subject to decreasing costs (see Chapter 5), each user will be *marginally* responsible for less than his percentage of coincident peak demands multiplied by total capacity costs, because marginal cost is less than average.

[20] See W. Arthur Lewis, *op. cit.*, 50–53; Ralph K. Davidson, *Price Discrimination in Selling Gas and Electricity* (Baltimore: Johns Hopkins Press, 1955), 84–88, 133–134, 192–193.

[21] This schedule is taken from C. F. Phillips, *op. cit.*, 352, who identifies the preponderant uses of the successive blocks as lighting; refrigeration, washer, and dryer; cooking; water-heating and air-conditioning; and electric house heating, respectively.

[22] See Shepherd, "Marginal-Cost Pricing in American Utilities," *South. Econ. Jour.* (July 1966), XXXIII: 62.

In recent years, both England and France have taken important steps toward remedying some of these deficiencies of the Hopkinson tariff. The famous French "Tarif Vert," put into effect in 1956 (only for bulk and industrial sales), instituted rates varying with the time of day and season of the year in order to base demand charges on the system peak. The change recognized that energy charges too should vary with the level of demand because variable costs are not constant.[23] The British Central Electricity Generating Board (CEGB) went over in 1962–63 to the coincident peak for determining demand charges on its (wholesale) sales to the regional Area Boards and introduced a differential day-night "running" (that is, energy) charge.[24] In 1967–68, explicitly recognizing that the latter charges were erroneously based on average (day and night) instead of marginal operating costs, it introduced differential time-of-day, -week and -year energy charges reflecting the increasing SRMC function.[25]

The 1967–68 reforms reacted to another, even more interesting problem already alluded to briefly above: how should the principle of peak responsibility be applied if the same capacity does not serve all users? If capacity is not interchangeable, so that the same type of plant or equipment does not necessarily serve both peak and off-peak users, it is no longer true that peak consumption alone should bear all capacity costs. In electricity generation, it is economical for short periods of time to use gas turbine generating units, which have low capital costs but high operating costs. These are inefficient for continuous utilization, but are less costly than installing regular capacity for just the extreme peak demands.[26] In consequence, when the CEGB tried to incorporate the entire capacity costs in the demand charges, at about £10 a year per kw, it found that some of its Area Board customers began to install their own gas turbines, at a cost of about £4 per kw, and therefore cut down their peak purchases. The Board correctly recognized that the true incremental or avoidable costs of supplying capacity that would be used for peaks of comparatively short duration (it estimated this type of capacity would be economic if operated no longer than 250 hours of the year) were not £10 but £4 per kw, and that the £11 now estimated to be the capital costs per kw of basic capacity, such as would be economic for longer periods of operation (because of its far lower variable costs) should therefore be borne by

---

[23] The demand charge to industrial customers in the Paris region provides discounts ranging from 0% in winter peak hours to 98% in summer "empty" hours. Eli W. Clemens, "Marginal Cost Pricing: A Comparison of French and American Industrial Power Rates," *Land Econ.* (November 1964), XL: 391. See also Meek, *op. cit.*, Part II, *Jour. Ind. Econ.* (November 1963), XII: 45–63, and the articles by Marcel Boiteux and Pierre Massé in J. R. Nelson, *Marginal Cost Pricing in Practice*, 134–156.

[24] R. L. Meek, "The Bulk Supply for Electricity," *Oxford Econ. Papers* (July 1963), n.s. XV: 107–123.

[25] The Board settled for three running or energy rates:

"... one for *peak units*—now defined as those used between 8 and 12 A.M. and for 4:30 and 6:30 P.M. from Mondays to Fridays in December and January, except for Christmas and Boxing Days ...;
"... a second rate for *day units* used between 7:30 A.M. and 11 P.M. daily, but outside the peak ...;
"... a third rate for *night units* used between 11 P.M. and 7.30 A.M. ..." "Puncturing the Power Peak," *The Economist*, May 14, 1966, 734.

The consequence of moving to increasing marginal charges for operating costs was to cause the operating charges to make some contribution to capacity costs as in our model, p. 94, above; the French Green Tariff has the same effect.

[26] For a general, diagrammatic statement of the conditions for such a choice, see M. A. Crew, "Peak Load Pricing and Optimal Capacity: Comment," *Amer. Econ. Rev.* (March 1968), LVIII: 168–170.

consumption during the longer-period, "winter plateau" of demand.[27] Similar qualifications of simple-minded peak responsibility pricing would clearly be appropriate to the extent storage capacity instead of basic pipeline capacity served the peak needs of natural gas consumers.[28]

Although most public utility executives and regulators recognize that peak responsibility pricing has some validity, probably most would also vigorously resist its wholehearted acceptance. William G. Shepherd's survey disclosed that the majority of American electric utilities practice little or no explicit marginal cost pricing, and among those that do, the main emphasis is on raising off-peak sales, by charging them something less than average capacity costs, instead of purposefully imposing all the capacity charges on the peak users.[29] He found, moreover, that publicly-owned companies, if anything, follow marginalist and peak responsibility principles even less than private;[30] and that electric utilities in states with "tough" regulatory commissions, such as New York and California, similarly incorporate little marginalism in their rate structures.

An outstanding illustration of the resistance of strong regulatory commissions is provided by the Federal Power Commission's formula for natural gas pipeline rate-making specified in its famous *Atlantic Seaboard* decision of 1952.[31] The distinctive feature of the Atlantic Seaboard formula is that it requires that capacity costs be distributed 50–50 between the demand and commodity charges instead of incorporated exclusively in the former. Since the demand costs are distributed among customers in proportion to their shares in the volume of sales at the system's (three-day) peak, while the commodity costs are borne in proportion to their annual volume of purchases, the consequence of the 50–50 formula is to shift a large proportion[32] of capacity costs to off-peak users. This produces an uneconomic encouragement to sales at the peak (whose price falls short of the true marginal costs of peak

---

[27] Accordingly, it introduced two demand rates: an £11 "basic capacity charge" for consumption during the winter plateau, when it estimated that demand would be on the average no more than 90% of the maximum system demand, and a "peaking capacity charge" of £4 for the period, estimated not to exceed 250 hours a year, when demand would exceed the 90% plateau. See R. L. Meek, "The New Bulk Supply Tariff for Electricity," *Econ. Jour.* (March 1968), LXXVIII: 48–53 and *passim;* "Puncturing the Power Peak," *The Economist,* May 14, 1966, 734.

This complicating factor in peak responsibility pricing was pointed out by Melvin G. de Chazeau, "Reply," *Q. Jour. Econ.* (February 1938), LII: 357 and recognized—along with most other problems—by Bonbright, *op. cit.*, 354 note.

[28] For an analysis of the ways in which the introduction of gas storage requires a modification of the simple charging of all capacity costs to peak users, see R. K. Davidson, *op. cit.*, 138–147.

[29] *Op. cit., South Econ. Jour.* (July 1966), XXXIII: 61–65. Effective earlier critics of the failure of electricity as well as gas distribution companies to employ marginal costing, in particular with respect to the allocation of capacity costs, were I. M. D. Little, *The Price of Fuel* (Oxford: Clarendon Press, 1953), 54–76 and R. K. Davidson, *op. cit.*, especially 81–97, 111–147.

[30] See also Richard L. Wallace, "Cost and Revenue Associated with Increased Sales of TVA Power," *South. Econ. Jour.* (April 1967), XXIII: 526–534; and, for an Australian example, H. M. Kolsen, "The Economics of Electricity Pricing in N. S. W.," *Economic Record* (December 1966), XLII: 564–565.

[31] *In the Matters of Atlantic Seaboard Corporation and Virginia Gas Transmission Corporation,* Opinion No. 225, 11 FPC 43 (1952).

[32] This is not wholly 50%, because peak users also pay their proportionate share of the commodity charge, which includes half of the capacity costs. But the point is that in deciding to what extent to cut their purchases at the peak relative to off-peak, peak customers are influenced by only the 50% of capacity costs incorporated in the demand charge; the other 50% does not affect that calculation because they pay it equally whenever they take the gas.

service[33]) and an uneconomic discouragement of off-peak.[34] (In fairness, it should be pointed out that the FPC has permitted departures from this strict formula when it appeared that the pipelines would suffer large losses of interruptible, off-peak sales at the inflated commodity charges it produced—permitting them instead to "tilt" the rate schedule downward on the commodity side of the balance.[35] Among other alleged harmful consequences of *Atlantic Seaboard* has been a tendency to discourage distribution companies from installing storage capacity: demand and commodity charges more fully reflecting the true respective marginal costs of peak and off-peak purchases would have increased their incentive to "shave" their purchases at the former

---

[33] This is so, as we have already pointed out, only to the extent that the pipeline function is subject to constant costs. Since pipelines do have some tendencies to long-run decreasing costs (see the section on "Natural Gas Transmission, Economies of Scale," Chapter 4, Volume 2), so that LRMC may be lower than ATC, the arbitrary 50–50 allocation tends to produce a less harmful result, that is, less of an understatement of the true *marginal* costs of peak service, than would otherwise be the case. Laurence C. Rosenberg concludes, however, that some considerable distortion remains. See his *Natural Gas Pipeline Rate-Making Problems*, unpublished Ph.D. dissertation, Cornell University, June 1963, 176–184 and *passim*. See also Stanislaw H. Wellisz, "Regulation of Natural Gas Pipeline Companies: An Economic Analysis," *Jour. Pol. Econ.* (February 1963), LXXI: 33, who contends that the constant cost assumption is not unreasonable.

It should be reemphasized, too, that the fact of gas storage may justify imposition of some capacity costs on off-peak customers—to the extent that, by using pipeline space that could otherwise be used to pump gas into storage, they create a need for more capacity than would otherwise be required. (See p. 98.)

[34] The formula discourages off-peak sales not only by the pipeline companies but also by their distribution company-customers, since *their* variable costs are inescapably inflated by the 50% allotment of fixed costs to the commodity charge. Wellisz demonstrates that the pipeline companies would in any event have an incentive to exploit their off-peak customers, charging them a monopoly price and using the supernormal profits thereby earned to subsidize peak sales, in this way "justifying" an uneconomic expansion of capacity. See *op. cit.*, 35–36. This tendency is discussed at greater length in Chapter 2, Volume 2.

[35] See, for example, C. F. Phillips, *op. cit.*, 624–628 on this and for a description of the Atlantic Seaboard formula; also Garfield and Lovejoy, *Public Utility Economics*, 181–185 and Wellicz, *op. cit.*, 30–43. The difference between the two rates is large: in 1968 the demand charges of one pipeline company on one schedule ranged between $2.79 and $4.28 per MCF, while its commodity charges ran from 22.1¢ to 26.2¢.

A later decision, interestingly enough involving the Atlantic Seaboard Company itself, demonstrates how far the "tilting" process has gone, under the pressures of competitive necessity. The case involved a protest by the Lynchburg Gas Company against special "partial requirements" rates that the FPC had permitted Atlantic Seaboard and other subsidiaries of the Columbia Gas Company to institute—higher, penalty rates imposed on those customers that purchased their supplies in part from suppliers other than the Columbia companies. In justification of these rates, the Commission had accepted the pipeline companies' justification that any losses of sales that they suffered by virtue of such diversions made it necessary for them to raise the rates that they charged their more loyal, full requirements customers. The Circuit Court of Appeals sent the case back to the Federal Power Commission, holding that these special rates had been inadequately supported in the record. See *Lynchburg Gas Co. v. FPC*, 336 F. 2d. 942 (1964). Upon rehearing, the FPC Presiding Examiner found that the Columbia System companies had so drastically departed from the original Atlantic Seaboard formula as to undermine their previous justification of the partial requirements rates. As the Circuit Court of Appeals explained in 1968, after the case had returned to it for a decision, one reason for the alleged necessity of protecting the full requirements customers from any loss of sales by their suppliers was that "not all of Columbia's fixed costs are recovered by the demand charges." At the time of the original *Lynchburg* decision, roughly half of Seaboard's fixed costs were recovered by the commodity rate. Thereafter, the Columbia companies had departed drastically from the original formula, in order to be better able to compete for off-peak business, with the result that only 6% of the company's fixed costs were now being recovered in the commodity rate. *Atlantic Seaboard Corporation et al. v. Federal Power Commission et al.*, 404 F. 2d. 1268, at 1270–1271 (1968).

by installing storage, which they could fill by low-cost purchases off-peak and draw on at the peak.[36]

We present two last examples of the pervasive uneconomic departure from peak responsibility pricing. First, commutation books and other such devices that give commuters quantity discounts on passenger trains and toll bridges have the consequence that occasional travelers, who usually travel off peak, pay a higher rate than commuters, who concentrate their traveling in the rush hours.[37] Second, airplane landing fees do not reflect the enormous variations in airport congestion, from one time of day, day of the week, or one airport to another. These variations themselves doubtless tend to induce air travelers and airplane companies to rearrange their traveling plans and schedules to avoid peak hours and locations and make fuller use of off-peak time; equivalently varying landing fees could make a further contribution.[38]

There are often very good reasons of expediency and practicality for these widespread departures from economically efficient pricing, to some of which we shall allude below. But objections are sometimes made to the principle itself. Prominent among these are:

1. It is unfair and discriminatory to charge peak-utilization alone with the fixed costs, since the capacity obviously serves all users at all times;[39]
2. The utilities or regulators have a special responsibility to protect the ordinary, unorganized householder, and should try to keep down his rates;

---

[36] See Homer Ross, "How Practical Is The Seaboard Formula?" *Public Utilities Fortnightly* (January 3, 1963), LXXI: 32 and Wellisz, *op. cit.*, 41. For (1) a reminder that the Commission's methods are defective also because the capital charges they allocate are historic instead of future costs, (2) an interesting and persuasive application of the peak responsibility principle to the problem of allocating the demand costs among customers located in different geographic markets, and (3) a clear demonstration that the Commission's methods of doing so fail by a wide margin to reflect customers' respective marginal responsibilities for the capacity of the various segments of the pipeline, see Laurence C. Rosenberg, "Natural-Gas-Pipeline Rate Regulation," *Jour. Pol. Econ.* (April 1967), LXXV: 159–168.

[37] See William Vickrey, "Some Implications of Marginal Cost Pricing for Public Utilities," *Amer. Econ. Rev., Papers and Proceedings* (May 1955), XLV: 619.

[38] See William D. Grampp, "An Economic Remedy for Airport Congestion: the Case For Flexible Pricing," *Business Horizons* (October 1968), XI: 21–30.

[39] As Garfield and Lovejoy state, "The fact that the [Atlantic Seaboard] formula permits no free ride on the line is its greatest strength. . . ." *Op. cit.*, 183. Here is the Commission's justification, in that decision:

"A pipeline would not normally be built to supply peak service, that is to say, service on the peak days only. We know . . . that pipelines are built to supply service not only on the few peak days but on all days throughout the year. In proving the economic feasibility of the project in certificate proceedings, reliance is placed upon the annual as well as the peak deliveries. Stated another way, the capital outlay for the pipeline facility is made—and justified—not only for service on the peak days but for service throughout the year. Both capacity and annual use are important considerations in the conception of the project and in the issuance of certificates of public convenience and necessity. Both capacity and volume, therefore, are what are known as cost factors or incidences in respect to the capital outlay for a pipeline project. It follows that reasonably accurate results can be achieved only by allocating the fixed expenses flowing from the capital outlay to both operating functions, viz., capacity and volume" *Atlantic Seaboard*, Opinion No. 225, 11 FPC 43, (1952).

A similar, further illuminating example of this argument was offered by the National Association of Railroad and Utilities Commissioners (NARUC) in support of a procedure, adopted by the Federal Communications Commission in 1967, for allocating the interexchange plant of the Bell System between interstate and intrastate service—a necessary procedure for FCC rate regulation since it has jurisdiction only over the former. The issue was whether it was appropriate

3. The utiliti
subject on
costs—goa

Justificatio
refutation, sir
economist car

1. In econo
between p
ences not
costs mor
is to refle
actually *i*
capacity c
off-peak
whether
discrimin

to lump together plant of AT
Department, which is used
interstate service, with other
both inter- and intrastate ser
cating the total between the tv
basis of the number of messag
each type. NARUC argued
two on the ground that "The
been designed as an entity a
benefits every other portion. . .
American Telephone and Teleg
Decision and Order, 9 FCC
AT&T argued, correctly, tha
facilities used exclusively by c
(in this case the Long Lines
assigned directly to that serv
common plant be allocated be
It also pointed out that co
discriminated against long-d
since it obscured the fact t
investment cost (per circuit
Lines Department is markedl
the Associated Company pla
Indeed, since it is the Long
that has the greater tend
decreasing costs, the margina
more than the average. Here
that all users *benefit* from the p
was used to obscure the mar
of the two services, which
reflected in their respective
the Commission were later p
the AT&T position. *In the M*
of procedures for separating
investment etc., Docket No.
Order, January 29, 1969, par
discussion of these separatio
pp. 152–153, Chapter 5.
[40] This is the implicit ass
Ponsonby, when he seems wi

ost purchases off-peak

nomic departure from
and other such devices
trains and toll bridges
no usually travel off
entrate their traveling
ot reflect the enormous
y, day of the week, or
ubtless tend to induce
ir traveling plans and
e fuller use of off-peak
further contribution.[38]
l practicality for these
cing, to some of which
made to the principle

ization alone with the
sers at all times;[39]
sibility to protect the
ry to keep down his

: is to say, service on the
w . . . that pipelines are
ot only on the few peak
nroughout the year. In
asibility of the project in
reliance is placed upon
e peak deliveries. Stated
l outlay for the pipeline
justified—not only for
but for service through-
city and annual use are
in the conception of the
iance of certificates of
necessity. Both capacity
are what are known as
in respect to the capital
project. It follows that
lts can be achieved only
xpenses flowing from the
perating functions, viz.,
tlantic Seaboard, Opinion
52).

ninating example of this
by the National Asso-
Utilities Commissioners
f a procedure, adopted
ications Commission in
terexchange plant of the
nterstate and intrastate
ocedure for FCC rate
risdiction only over the
ether it was appropriate

3. The utilities should promote the maximum extension of their services, subject only to the condition that aggregate revenues cover aggregate costs—goals that may well conflict with peak responsibility pricing.[40]

Justifications such as these are for the most part not susceptible to scientific refutation, since basically they involve nonscientific value judgments.[41] An economist can only cite the following counterconsiderations:

1. In economic terms, peak-responsibility pricing is not discriminatory between peak and off-peak users. Discrimination consists in price differences not corresponding to cost differences. It is an objective fact that it costs more to supply users at the peak than off-peak, and the proposal is to reflect that cost difference in the respective prices. Every peak user actually *imposes* on society, in the long run, the incremental cost of the capacity on which he draws. There is no such causal connection between off-peak utilization and capacity costs: the capacity would be there whether or not the off-peak user made demands on it. It would be discriminatory to levy any of these costs on the off-peak user.[42]

to lump together plant of AT&T's Long Lines Department, which is used exclusively for interstate service, with other facilities used for both inter- and intrastate service, before allocating the total between the two services on the basis of the number of message-minute-miles of each type. NARUC argued for lumping the two on the ground that "The toll network has been designed as an entity and every portion benefits every other portion. . . ." *In the Matter of American Telephone and Telegraph Co.*, Interim Decision and Order, 9 FCC 2d 30, 71 (1967). AT&T argued, correctly, that the costs of any facilities used exclusively by one of the services (in this case the Long Lines plant) should be assigned directly to that service, and only the common plant be allocated between the services. It also pointed out that combining the two discriminated against long-distance messages, since it obscured the fact that the (average) investment cost (per circuit mile) of the Long Lines Department is markedly less than that of the Associated Company plant (*Ibid.*, 96–97). Indeed, since it is the Long Lines Department that has the greater tendency to long-run decreasing costs, the marginal costs diverge even more than the average. Here again, an argument that all users *benefit* from the presence of capacity was used to obscure the markedly different *costs* of the two services, which should have been reflected in their respective rates. NARUC and the Commission were later persuaded to accept the AT&T position. *In the Matter of Prescription of procedures for separating and allocating plant investment etc.*, Docket No. 17975, Report and Order, January 29, 1969, pars. 6–19. For further discussion of these separations procedures, see pp. 152–153, Chapter 5.

[40] This is the implicit assumption of G. J. Ponsonby, when he seems willing to see off-peak users of city buses charged more than the LRMC of serving them, in order to permit a greater improvement of service at the peak than what that traffic would itself be willing to pay for. "The Problem of the Peak, with Special Reference to Road Passenger Transport," *Econ. Jour.* (March 1958), LXVIII: 78–82, 87. For a suggestion that these goals may also be in the interest of the public utility companies themselves, see note 34.

[41] However, Wellisz does point out that the *Atlantic Seaboard* formula is of dubious efficacy even as a means of achieving the goal of subsidizing household consumption (much of which is at the peak). The householder buys his gas from the distribution company. The latter, he points out, will have an incentive, for the reason just suggested (note 34), to maximize profits on its off-peak sales, because this will enable it to subsidize peak sales. But when its own marginal costs of supplying off-peak gas are inflated by the incorporation of capital costs in the commodity charge, it will be forced in turn to price such sales above the point that would maximize their contribution to the overhead costs of the system as a whole, hence to the subsidization of on-peak sales. *Op. cit.*, 37–38. When seller B buys some of his inputs from producer A at prices in excess of marginal costs, B's own profit-maximizing price will in turn exceed the price that would maximize the profits of A and B together. On the reason for this, see the section "Financial Integration," Chapter 6, Volume 2.

[42] It might be argued that peak-responsibility pricing involves no discrimination, also, because peak and off-peak power are not the same service, in economic terms, any more than are Tuesday and Saturday evening tickets to the same theatrical performance; and price discrimination occurs only when different prices are charged for

2. In this sense, it is inequitable to make off-peak users pay some share of capacity costs, for which they are not themselves causally responsible.[43]
3. Moreover, such a policy would be economically inefficient. To the extent that off-peak demand has any elasticity at all, a charge to these users that incorporates any capacity costs will cause them to give up satisfactions, the true social costs of which they would be perfectly willing to pay. And some productive capacity is left wastefully idle. Conversely, and subject to the same condition, if peak users do not have to pay the full (marginal) costs of being supplied, they will induce society to provide them with a capacity that uses resources that would have given greater satisfaction if directed to other employments.
4. In these circumstances, off-peak users would be subsidizing peak users. Now this may be something that society is willing to do; but such a policy will make sense only if the membership of each group is clearly identified, so that the decision in effect to transfer income from the one to the other is a conscious one. The mere identification of the subsidizing group as consisting of commercial and industrial customers and of the subsidized group as householders, for example, does not in itself demonstrate that the abandonment of economically efficient pricing is justified as a means of promoting a more equal distribution of income. The higher-than-marginal costs imposed on business and industrial customers must ultimately be paid by *their* customers or the customers of their customers; and these, too, are people, who may or may not, on the average, have higher incomes than the direct household customers of the utilities.[44]

---

the same service. As students of the Robinson-Patman Act will recognize, this argument is a treacherous one. What makes the Tuesday and the Saturday evening tickets of "unlike grade and quality," in the terms of that Act, is not just the physical difference in the service, but also the fact that buyers are willing to pay more for the latter than for the former. Since discrimination always requires that some customers be willing to pay more than others, if that difference in willingness to pay is then used in turn to prove that the services are economically different, there never can be any discrimination. However, the fact remains that the two services are in this case objectively separate: the use of capacity to supply gas in July for cooking could not be transferred to January and used instead to cut the costs of home-heating; the two markets are not artificially, but physically and inescapably separate. See note 11.

[43] It might appear that peak users are responsible not for the entire capacity, but only for that portion by which their consumption exceeds off-peak consumption—that is, that efficiency requires that they pay the entire costs only of the "peak" or protuberance of the mountain above the surrounding plateau, not of the entire mountain. The answer is to be found in the case for equating price to *marginal* cost. *Every* peak kilowatt hour consumed is marginal in the sense that capacity costs would be less in its absence. The purchaser of every kilowatt hour at the peak must therefore be confronted with a price that, by including its share of incremental capacity cost, makes him decide whether it is worthwhile to him to impose on society the cost of constructing his share of the entire mountain. *No* off-peak use is marginal in this sense.

[44] The *Atlantic Seaboard* formula, for example, has the effect of imposing higher-than-marginal cost rates on, among others, electricity generating companies that purchase natural gas on an interruptible basis for their steam generating plants; in effect, therefore, it tends to impose lower-than-economic costs on residential purchasers of natural gas (who buy heavily at the peak) and higher-than-economic costs on residential purchasers of electricity. It would be difficult to demonstrate that the formula produces a more acceptable distribution of income than would full peak-responsibility pricing. Again, the Federal Communications Commission is often urged to exercise its regulatory authority over interstate telephone rates to give particular protection to the householder instead of the business subscriber. But the average income of the (weighted) average household user of interstate (that is, long distance) telephone service could well be considerably higher than that of the customers of most of the commercial and industrial users of Bell System services.

ers pay some share of
ausally responsible.[43]
fficient. To the extent
charge to these users
iem to give up satis-
be perfectly willing to
ılly idle. Conversely,
not have to pay the
uce society to provide
ld have given greater

ibsidizing peak users,
do; but such a policy
ıp is clearly identified,
n the one to the other
subsidizing group as
and of the subsidized
self demonstrate that
is justified as a means
ne. The higher-than-
trial customers must
ers of their customers;
on the average, have
s of the utilities.[44]

ilowatt hour at the peak
onted with a price that,
of incremental capacity
whether it is worthwhile
ety the cost of construct-
e mountain. *No* off-peak
ense.
ormula, for example, has
gher-than-marginal cost
s, electricity generating
ıse natural gas on an
their steam generating
ore, it tends to impose
osts on residential pur-
who buy heavily at the
an-economic costs on
f electricity. It would be
that the formula pro-
e distribution of income
k-responsibility pricing.
munications Commission
e its regulatory authority
e rates to give particular
eholder instead of the
the average income of
ge household user of
ng distance) telephone
onsiderably higher than
most of the commercial
Bell System services.

5. Even if society were to make a conscious decision to transfer income from off-peak to peak users, such a policy would not be an intelligent one if it did not take into account the fact that departures from efficient pricing are an economically inefficient way of effecting such a transfer. The "proof" of this, which depends only on peak demand having some elasticity, is that if the transfer were made as a money grant to those users, instead of in the form of prices below cost, they would not use all that money to purchase the public utility service in question, but would spend some of it for a variety of other goods and services.[45]

THE APPROPRIATE TIME PATTERN OF RATES

Efficient pricing of public utility services calls for as fine a differentiation as practical of rates for the various services provided, in various locations, so as to reflect the different marginal costs of each. To the extent that differences in marginal cost can be ascribed also to the particular times at which service is taken, rate schedules should incorporate time-differentials as well. All this would be true even in a purely static situation, with unchanging cost and demand functions.

In the real world, costs and demands are constantly changing over time. These dynamic changes give rise to at least three major types of problems of efficient pricing. First, the ever-changing relationship of capacity to demand raises the question of whether it is feasible to price on the basis of short-run marginal cost, and what mixture of SRMC and LRMC would provide the optimum combination of feasibility and efficiency. Second, how should prices be varied over time in reflection of the changes in cost functions, because of ever-changing input prices, productivity, and technology? Third, if one departs from pricing on the basis of SRMC, what is the proper time pattern for the recovery of fixed costs such as depreciation, in the face of fluctuating demand on the one hand and unpredictably changing technology on the other?

Changing Relationship of Capacity to Demand

System demand peaks shift not merely because the quantities demanded respond to different prices—a static, elasticity phenomenon; they shift also because demands at any given price change over time. Promotional pricing of air-conditioning equipment by electric companies, like the pusher's free distribution of addictive drugs, may have played a role in spreading the habit; but the habit spread for other reasons as well, moving demand sufficiently to the right and making it sufficiently price-inelastic to give rise in some areas to a fixed, summer peak. If the peak shifts, or becomes shiftable, or *may* shift five, ten, or fifteen years in the future, how should that shift or shiftability be reflected in peak responsibility pricing today?

---

[45] For a more formal demonstration, see pp. 190–191, Chapter 7. This "proof" is of course no better than the assumptions underlying the economic model whose conclusions it adopts: notably that the subsidized purchasers would be the best judges of what most satisfies them and that society *ought* therefore to be more willing to give them cash grants than, for example, cheap electricity. See pp. 67–69, Chapter 3. Even if a governmental body rejects the second precept, the economist could still observe that the social decision is more likely to be rational and the economic costs less if the subsidized service is paid for openly by appropriation of taxpayer funds instead of covertly in charges levied on off-peak users.

Apart from peak shifts, changes in demand from one moment, month, or year to the next will involve movements along existing SRMC functions; and unless those costs are constant (horizontal), this will mean corresponding changes in marginal costs. We have already seen how volatile the SRMC of individual airplane trips must be—ranging between zero and levels far in excess of average total costs, depending on what price it takes to ration the fixed number of places—that is, to fill the plane while turning away no traveler willing to pay the price.

Similarly, on the supply side, capacity is constantly changing. Typically, public utility companies must build in advance of demand in order to be in a position to meet unexpected peak requirements and simply because the investment process is a lumpy one: additions to capacity are most economically made in large units. Therefore, at any given time, there is almost certain to be excess capacity,[46] which will remain idle if customers are charged long-run marginal costs. What, in these circumstances, is the proper measure of marginal costs? Or, to put the question another way, how far into the future should the calculation of long-run marginal costs extend?

As we have already seen, there is a strong economic case for letting price rise and fall as demand shifts along the rising SRMC curve. If SRMC is at times zero and at other times discontinuous (because an absolute physical limit of capacity has been reached), the price should fluctuate—down to zero, if necessary, because in the presence of excess capacity, no matter how temporary, no business should be turned away that covers the SRMC of supplying it; and up in periods of shortage to whatever level is necessary to ration limited supplies among customers. Once the new bridge is built, it is wasteful to keep people from crossing it; the time to charge for crossings is when congestion sets in.[47]

But, we have also pointed out, this is a counsel of perfection. It may well require modification in a world where (1) buyers and sellers make mistakes, (2) perpetual price fluctuation can be expensive for sellers to administer and buyers to keep track of and respond to intelligently, (3) capital and labor are incompletely mobile, (4) many other prices are highly inflexible, and (5) there is a business cycle.[48]

---

[46] For a demonstration that economies of scale make it rational to have excess capacity see Hollis B. Chenery, "Overcapacity and the Acceleration Principle," *Econometrica* (January 1952), XX: 1–10. Chenery shows that the optimum amount of the excess is comparatively insensitive to the interest rate, which determines how costly it is to build capacity in advance of the time it is utilized.

[47] It might appear that no customer whose continued patronage would eventually require additions to capacity should ever be charged a price that completely excludes those capital costs; the economic ideal, it might appear, would be to include them, but discounted back to present value, to reflect the fact that continued service of the customer in question would require their incurrence only sometime in the future. Such a prescription ignores the fact that buyers whose continued patronage *could* require the incurrence of additional capacity costs are not in fact responsible for them if they drop out of the market when the time comes for the supplying company to make the decision whether to make the additional investment. It is only at *that* time, when there is a possibility of resources being used to expand capacity, that *all* (peak) customers should be confronted with a price that once again incorporates those costs, forcing them (as before) to decide whether the benefit to them of continued service justifies society's incurring the marginal opportunity costs of serving them.

[48] See pp. 83–86, Chapter 3. Consider how vexatious it would be to price each airplane flight at *SRMC* by auctioning seats off at whatever price it takes to clear the market—how difficult and costly to administer, how much time of passengers and airplane employees (and computers) would be used in the auctioning process itself, how difficult it would make it for

Consider what it means, for example, about the proper method of recovering the gross cost of capital—depreciation, interest, profits, income and property taxes—over the business cycle.

Except for the portion of depreciation that varies with the extent to which the facilities are used (that is, that represents user cost), these costs are a function of time—so much per dollar of investment per year—instead of a function of the rate of output from given facilities. There are three possible ways in which they might be recovered from customers:

1. In equal amounts for equal periods of time—for example, a certain amount per year. This method would seem to be recommended by the fact that these costs are a function of time. This means they would be recovered in prices that fluctuated inversely with the fluctuations in demand—capital charges per unit being low in years of large sales and high in years of low sales.
2. In equal amounts per unit of sales, on the theory that no purchase or purchaser should pay more than any other, per unit of purchase, over the life of the investment. This would mean recovery in constant prices over the cycle, which is the tendency under regulation. Depreciation, taxes, and return are calculated at a particular amount of money per year.[49] Usually, the companies are then instructed to propose a schedule of rates that will cover the total charges on the average over some period of years in the future. This means that in practice the capital costs will be recovered unequally over the cycle—in larger amounts when sales exceed the average, in smaller amounts when they fall below the average.[50]
3. In prices that fluctuate directly with the business cycle. This would be the tendency of pricing at SRMC: prices would be low and revenues would tend to cover only variable costs in periods of weak demand in recognition of the fact that the demand at such times puts no burden on capacity. Price would have to move up far enough above that level and remain there long enough in periods of recovering and strong demand not only to cover ATC at that time but to make good the losses of previous periods.

There is no settled economic theory of the ideal behavior of the general price level over the cycle (Indeed, cyclical fluctuations have become so much more modest since World War II that the subject itself is of reduced interest.) Economists would almost certainly reject the first as a general rule. In view of the general tendency of prices in the economy to move up and down with the cycle (although the latter tendency especially has become faint in recent

---

passengers to plan their travel. Or suppose the same free pricing system were used for taxicabs: consider how their rates would have to fluctuate from sunny to rainy days, from one time of day and from one section of the city to another, and with what costs of time and annoyance to passengers and drivers.

49 Straight-line depreciation—writing off an investment in equal amounts over its estimated life—remains the usual procedure. When accelerated depreciation is taken for purposes of calculating income tax liability, the taxes incorporated in the cost of service are often "normalized," that is, set at a stable annual figure over the life of the asset, although actual taxes will be below that normalized level in the earlier years and above it in later years. See pp. 32–34.

50 This is only a tendency. It may be offset by a tendency of *variable* costs to fluctuate directly with the level of operations—for example, because the utility puts inefficient plant into service only when necessary—while the variable component of the cost of service is an average estimated for a period of years.

decades), it would probably result in uneconomic distortions to have public utility prices going the other way. Moreover, it would tend to have precisely the wrong effect in rationing consumption—discouraging utilization when demand is low relative to capacity and encouraging it when marginal (congestion) costs are high.[51] Economists would be less clear in their response to the third way of recovering costs. It would probably be fair to say that most believe it serves no useful purpose to have the general price level fluctuate widely with the cycle, either. In view of the comparative rigidity of the general price level it would probably do more harm than good to have the prices of public utility services alone fluctuate as widely as the third rule would dictate.[52]

Therefore, the usual practice of charging capital costs on a per unit of sales basis and recouping them *on the average* over the cycle—that is, the second rule—is probably the best available general rule as far as the determination of general rate *levels* is concerned.[53] The rate stability that it provides is not just a pragmatic compromise between the extremes produced by rules 1 and 3: it also has the positive virtue of making it easier for customers to make the type of long-run commitments that consumption of a utility service usually involves (to install a certain furnace, to locate industry in a particular area), on the basis of reasonably stable expectations about the prices they are likely to have to pay.

This does not mean that SRMC, or fluctuations in the relationship of demand to capacity, should be ignored. The key is the necessity for reasonable predictability on the part of supplier or customer. Where not only *differences* but *changes* in the SRMC of supply at different times can be predicted with reasonable assurance, it is economically efficient to embody corresponding differentials in rate *structures*. In the presence of excess capacity, utility companies ought to make every effort to design rates, down to SRMC, to put it to use. We have recognized that the "ideal" of rates fluctuating with SRMC could have highly inefficient consequences on the buyers' side of the market. Household customers induced to shift to electricity or natural gas or industrial users persuaded to locate somewhere by rates approximating short-term marginal costs in periods of excess capacity would have a legitimate complaint if, having made the switch, they were then faced unpredictably with steadily rising rates as SRMC came increasingly to cover capacity costs as well. Therefore, the essential proviso would have to be attached that the proffer of any such temporarily low rates be accompanied with the warning that service would be interrupted as demand caught up, or rates increased to whatever extent necessary to ration demand until additional capacity was once again constructed.

---

[51] See the same observation in an analogous context, by Seymour Smidt, "Flexible Pricing of Computer Services," *Management Science* (June 1968), XIV: B-582.

[52] However, see the reference to J. R. Nelson, note 67. Additionally, Bonbright points out that if public utility rates were free to fluctuate like purely competitive prices over the cycle, it would in principle be necessary for regulatory commissions to play a much more active role than they now need to in the investment decision process—specifically, to compel companies to expand capacity when necessary. Private managements would have a strong temptation to delay capacity expansion in time of strong demand, hoping instead to enjoy the high profits resulting from the high prices required to ration customers. *Op. cit.*, 99.

[53] The Civil Aeronautics Board has explicitly recognized that it would be futile to try to iron out year-by-year fluctuations in rate of return in an industry so subject to fluctuation in its financial fortunes. See its *General Passenger Fare Investigation*, 32 CAB 291, 294–309 (1960).

Such a notice would protect the utility against charging customers less than the true and full eventual MC of serving them, while also giving the latter the predictability of future rates necessary if they are to avoid making irrational commitments.

The same sort of warnings are necessary with peak responsibility pricing. The ideal, once again, would be to price on the basis of *current* MC: this means that currently off-peak users would pay zero capacity costs today and full capacity costs in the event of a shift to a new fixed peak at their time of consumption tomorrow. But fluctuations of this kind would be impractical, for the reasons already given. The key once again must be reasonable predictability. Therefore, capacity costs ought to be shared in varying proportions, from 100% down to zero, by purchasers in periods that have correspondingly varying likelihoods of being or becoming peaks in the foreseeable future. The farther ahead the utility can see with reasonable assurance that certain sales will remain off-peak, the greater is the justification for offering this service on a firm (noninterruptible) basis at rates that incorporate no capacity costs. Conversely, if there is a strong possibility that within a very few years the service that is now off-peak will in fact put a strain on capacity, the off-peak tariffs are justified only if they are interruptible[54] and/or the buyers are put clearly on notice that the rates may in time have to be sharply increased.[55]

On the other hand, customers have no absolute right to perpetual protection against drastic rate changes if cost changes counsel such measures. Considerations of fairness will join with considerations of efficiency in calling for reasonable compromises between the interests of different classes of customers: offering moderate assurances of rate stability in the face of unanticipated cost changes unfavorable to certain customers, yet refusing indefinitely to burden other customers with the necessity of paying prices disproportionately *above* the MC of serving them in order to shelter the former group permanently against the consequences of change or of errors, their own or the company's.[56]

We return to some of these practical considerations in Chapter 7. But there is no simple solution available. The proper balance will ordinarily have to make very large concessions in the direction of rate stability: the efficiency advantages of having rates vary over the life-cycles of particular increments to capacity are typically outweighed by the numerous disadvantages.[57] And even the advantages are probably diminishing over time: as the electricity generating industry and natural gas pipelines (the latter far more slowly than the former) approach participation in regional or national grid systems, in which additions to capacity in the various regions of the country are synchronized over time, the problem of temporary excesses of capacity on the one hand, and the efficiency advantages of varying prices over time because of them, will become correspondingly less important.

For these various reasons, growing public utility industries that are constantly adding to capacity generally must attempt to set their rates, as

---

[54] Standby, no-reservation airplane service, to which the youth fares apply, is by definition this type of service.

[55] On most of these problems, see Bonbright, *op. cit.*, 360–366.

[56] See *ibid.*, 129.

[57] For examples of electricity rates remaining absolutely unchanged for 10 and 15 year periods and a consideration of the reasons for this, see Raymond Jackson, "What Others Think, Rigidity in Electric Rates," *Pub. Util. Fort.* (June 20, 1968), LXXXI: 42–44.

stably as possible, on the basis of some estimated average cost level over some more or less arbitrarily selected planning period—of perhaps five years.[58] In this event, the appropriate benchmark must be some estimated average level of long-run, not short-run marginal costs[59]—and closer approximations to SRMC must be confined largely to (1) incorporation in rate schedules, insofar as prediction is possible,[60] (2) the offer of special rates for interruptible service, and (3) exemption of clearly off-peak pricing from capacity charges.[61]

Yet there all sorts of possible situations in which stability and predictability may be less important than the efficiency advantages of flexible SRMC pricing. With sufficient ingenuity, it is often possible to find ways of practicing it. For example, Smidt has proposed that computers be programmed so as to vary the price of their own service from one five-minute interval to the next, as well as over the life of the equipment, depending on the balance of demand and capacity, with users given considerable choice in specifying in advance

---

[58] An indirect confirmation of the reasonableness of an offer of stability over a more or less arbitrary planning period is provided by the special, short-term arrangements under which the Tennessee Valley Authority and the Bonneville Power Commission sold surplus power to private utilities, in one case under a five-year contract, in the other subject to cancellation on five-year notice if the power were required by municipal distributors and co-ops. See Glaeser, *Public Utilities in American Capitalism*, 499–500, 557. According to officials of AT&T's Long Lines Department, they build capacity when and where needed in the expectation that it will come to be fully used within four to six years. They base prices on average estimated costs for such a period, and on a nationwide basis, because they feel it is entirely infeasible to price flexibly within that period or with reference to local differences in excess capacity. Their system is national in the sense that communications are automatically switched from one route to another depending on which circuits are busy and where there is excess capacity. (See the section on "the national telecommunications network," Chapter 4, Volume 2.) For a finding of ten- and twenty-year (or more) planning horizons for capital budgeting by electric power companies—projections over which are "not . . . taken very seriously"—with five years being "the standard long-term forecast in the industry," and three years "the point at which important practical consequences follow from projections of demand," see Michael Gort, "The Planning of Investment: A Study of Capital Budgeting in the Electric-Power Industry," Part I, *Jour. of Bus.* (April 1951), XXIV: 81–82.

Vickrey defines the proper planning period for which marginal costs are appropriately measured as being determined by the period during which it is anticipated rates will be stable: "The proper time horizon for the cost determination is the probable interval between rate adjustments."[62] And he would solve the question of the proper size of the incremental block of output, for which LRMC are to be estimated, in a similarly pragmatic way:

"The increment in traffic for which the cost increment is to be estimated should have a composition similar to the increment induced by the rate change under consideration. . . . Thus unless a policy is being contemplated of suppressing a class of service entirely, the traffic increment for which the cost is being ascertained should never be an entire class of traffic, but only a final increment in that traffic corresponding to a realistically contemplated rate change." Testimony in FCC Dockets 16258 and 15011, *In the Matter of American Telephone and Telegraph Company*, Networks Exhibit No. 5, July 22, 1968, mimeo., 23–24.

[59] See Bonbright, *op. cit.*, 331–336.

[60] See some of the challenging suggestions of William Vickrey in Lyle Fitch and Associates, *Urban Transportation and Public Policy* (San Francisco: Chandler Publishing Co., 1964), esp. pp. 146–156; "Pricing in Urban and Surburban Transport," *Amer. Econ. Rev., Papers and Proceedings* (May 1963), LIII: 452–460; and "Pricing Policies," *International Encyclopedia of the Social Sciences* (New York: The Macmillan Co. and The Free Press, 1968), XII: 457–463.

[61] The second and third examples are really only one illustration of the first. Moreover, they are equally examples of LR as of SRMC pricing: the LRMC of definitely off-peak business (and interruptible service is by definition off-peak) includes no capacity costs either. And, it should be noted, (2) and (3) both are aspects of given rate schedules instead of examples of rate *changes* over time. (It is not really the long-distance telephone rate that "changes" at 6:00 P.M.: it is time that changes, moving subscribers from one prescribed rate schedule to another.)

the priority they wish assigned to their jobs and for which they are willing to pay.[62] William Vickrey, one of the most assiduous and ingenious proponents of "reactive pricing," has suggested it could feasibly be applied to long-distance telephoning, tickets for theatrical performances, sporting events, long-distance airplane travel, and electricity. For the first, he would have computers programmed to inform subscribers, immediately on dialing, how much—depending upon the availability of circuits—it would cost them to place that particular call at that particular time, giving them the opportunity then to decide whether the call should be completed. He has proposed that the prices of the theater and airplane tickets might be gradually reduced as the time of performance or departure nears, depending on how much of the space has been sold to customers willing to pay the premium for advance reservation. For electricity, he observes,

> ". . . the same load signal used by Électricité de France to switch rates according to time of day could be used to vary rates on a reactive basis, simultaneously encouraging the switching on and off of deferrable demands such as water heaters and refrigerators."[63]

He is certainly correct in pointing out that there is no reason in logic or fairness for public authorities to impose tolls on bridges immediately on their completion and take them off when the investment is fully paid off—a practice likely to produce a time pattern of rates just the opposite of what the relation of demand to capacity requires.[64]

Changing Cost Levels: Reproduction versus Original Capital Costs

Most of the time and energy expended in regulatory proceedings is taken up with recomputing aggregate company revenue requirements, with a view toward adjusting the general rate level to changes in total costs.[65] There is no question of economic principle about the necessity for these efforts: ideally, prices should reflect marginal cost at the time of sale—not at some time in the past.

This consideration constitutes the strongest economic argument for the use of reproduction instead of original cost as the basis for computing capital charges. As proponents of that method of valuation point out, prices in competitive markets will tend to be set at the level that covers current, not past, capital costs. This is a statement of long-run tendency only, to be sure; in periods of rising reproduction costs, it will be achieved only as demand presses to the limits of existing capacity and new investments need to be attracted or old capacity replaced. Still, it represents the competitive ideal, and departures from it in periods of long-run inflation or deflation involve inefficiencies, to the extent that there is any elasticity in demand. These

---

62 *Op. cit.*, B-581-600.

63 *Op. cit.*, *International Encyclopedia of the Social Sciences*, XII: 460. See also his "The Pricing of Tomorrow's Utility Services," paper presented at Occidental College, June 1966, processed, and note 29, Chapter 3.

64 As a result, he observes, in New York City a great deal of traffic over the East River is diverted from the new toll bridges, with comparatively large capacities and ample access facilities, to the clogged, ancient bridges that no longer carry tolls, and dump traffic right in the middle of the most congested areas of the city. *Op. cit.*, *Amer. Econ. Rev.*, *Papers and Proceedings* (May 1963), LIII: 455.

65 To some extent this is accomplished automatically, as in the purchased-gas and fuel-adjustment clauses incorporated in the tariffs of many distributors of gas and electricity. See Garfield and Lovejoy, *op. cit.*, 146.

distortions may take various possible forms (for simplicity, we assume the situation is one in which reproduction costs exceed book costs):

1. Current costs reflect marginal social opportunity costs. Since under original cost valuation the buyers pay not these, but some lower average as the cost of new, increasingly expensive plant is blended in with that of the old, the result is excessively large purchases of the public utility service and correspondingly excessive flow of resources into its supply.

   This tendency is accentuated when the service competes with unregulated commodities whose prices may behave more nearly according to the competitive norm—railroads with trucks, buses, and private cars, electricity and gas with oil, common carrier transportation and communication with private microwave. Since the competition of these substitutes increases the elasticity of demand for the regulated services, this is only another way of saying that the greater the elasticity of demand, the greater the waste consequent on holding price below marginal cost.

   It might appear that basing prices on historic capital costs would cause offsetting distortions in investment incentives—for example, in periods of rising capital costs discouraging capital expenditures that would otherwise be made. If this were true there would be an underallocation instead of overallocation of resources into regulated industries in periods of inflation. But, as long as the permitted rate of return covers the cost of capital, neither management nor investors need be deterred from making whatever additional investments are required to satisfy the artificially inflated demand. Whatever current dollars are required for the incremental investment enter the original cost rate base, on which the required rate of return is then permitted: as long as additional investments bring in enough additional net revenue to cover the cost of capital, it is in the interest of the company and of its existing stockholders to make the additional investments.[66] The distortion is on the demand, not the supply side.
2. When customers have a choice to buy from one utility company or another, that choice will be determined not solely, as it should be, by their respective marginal costs, but, quixotically, by differences in the average age of their plants, which will produce different average rates. Thus, industry may be impelled to locate where the suppliers of electricity or transportation have a rate base of comparatively old vintage, even though long-run marginal supply costs may actually be higher there than elsewhere.
3. Original rate-base costing tends to produce a perverse cyclical behavior of prices, holding down the charges for utility services when commodity prices generally are rising, and holding them up in periods of general

[66] Walter A. Morton, who supports reproduction cost valuation on grounds of fairness, concedes that an original cost earning base need not involve economic inefficiency on this score. "Rate of Return and the Value of Money in Public Utilities," *Land Econ.* (May 1952), XXVIII: 117–118. The statements by M. J. Peck and J. R. Meyer to the contrary in "The Determination of a Fair Return on Investment for Regulated Industries," in *Transportation Economics*, A Conference of the Universities—National Bureau Committee for Economic Research (New York: National Bureau of Economic Research, 1965), 202–203 seem to be incorrect. Actually a company could raise additional capital, within limits, even if the permitted rate of return were below the cost of capital. But since this would dilute the equity of existing stockholders, the company presumably would be reluctant to do so. See note 64, Chapter 2.

price decline. In addition to the inefficient substitution effects of the resulting relative price movements already mentioned, there are adverse macroeconomic consequences because of the high capital-intensity of these industries: holding down their prices and encouraging consumption of their services in periods of general inflation forces them to expand their disproportionately large investment outlays, thus contributing to the excessive levels of aggregate demand. The opposite happens on the downturn.[67]

Although arguments such as these have traditionally been directed toward the question of the proper valuation of the rate base, the same economic principles apply to the computation of depreciation expense and rate of return. What gross cost of capital (depreciation plus return) should be entered into the economically efficient price? Is it the historic cost of the dollars invested in the enterprise at various times in the past? Or the current or future cost of capital? Setting aside the consideration that capital costs as such do not enter at all into the computation of short-run marginal costs, clearly, it is an average of future costs of capital over the planning period that properly belongs in LRMC. Yet the usual formula for capital cost calls for a heterogeneous mixture—composed of depreciation computed by applying some conventional length of time based roughly on past experience to historic investment; a current cost of equity capital, usually estimated over some period in the recent past, and applied to an original-cost rate base; and an actual, historic cost of debt capital, as embodied in existing bond obligations, adjusted to include the cost of any debt planned for incurrence in the near future—all of these plus an inescapable element of "judgment."

This practice is rationalized essentially on the ground that it is the function of regulation to permit companies simply to cover their revenue requirements—to recover the money capital actually invested in the service of the public, to earn a return on investment sufficient to meet their actual debt service obligations, and to attract new equity capital. And it does do these things. It avoids conferring windfall gains on stockholders, such as they would earn in a period of inflation and high interest rates under a system that incorporated in the cost of service the current instead of the (lower) historic cost of debt capital; and it protects them in turn from the windfall losses they would otherwise suffer when interest rates turn down. In so doing, it also protects the credit standing of the company and its ability therefore to serve the public, both of which could be impaired under a reproduction cost system, when current interest rates were much lower than those the company had actually incurred in the past.[68]

But it is precisely the characteristic of competitive markets that they expose stockholders to the possibilities of earning unanticipated, windfall gains and losses, of getting back much more or much less than the dollars they originally invested, or earning much more and much less than the cost of capital on

---

67 See James R. Nelson, in Shepherd and Gies, *op. cit.*, 71–76. See the classic statement of these various arguments by Harry Gunnison Brown, "Railroad Valuation and Rate Regulation," *Jour. Pol. Econ.* (October 1925), XXXIII: 505–530 and "Railroad Valuation Again: A Reply," *ibid* (August 1926), XXXIV: 500–508; also Willard J. Graham, *Public Utility Valuation*, Studies in Business Administration, Vol. 4 (Chicago: Univ. of Chicago Press, 1934).

68 See Bonbright, *op. cit.*, 186–187, 245, 248–249, 278–280. We consider below the opposing argument that such a system does not really treat stockholders fairly for the very opposite reason: that it fails to take into account the changing purchasing power of the dollar.

those dollars. Such markets necessarily expose buyers to prices that vary and fluctuate correspondingly.

These arguments against the use of historic costs can be appraised at various levels. Some of the opposing considerations concede the theoretical validity of the criticisms but deprecate their practical importance; others strike at their theoretical validity as well. Some responses (in addition to the very important considerations of administrative feasibility, already suggested in Chapter 2) argue for the superiority of an original-cost rate base; others emphasize that to fix prices principally by applying a gross cost of capital to a rate base, however valued, is economically unsound and in any case constitutes the minor part of the task of efficient rate making:

1. The actual importance of the case for reproduction costs depends not on principle but on fact. How serious a distortion is created by the lag of prices behind reproduction costs depends on(a) the size of the lag and (b) the elasticity of the demand for public utility services. As for the first, all utility company rate bases are a mixture of vintages; all of them, in growing industies, are heavily weighted by recent expenditures.[69] The difference between the prices produced by original and reproduction cost valuation therefore can easily be exaggerated. As for the second, defenders of original costing have tended to argue that demand is comparatively price-inelastic.[70]
2. The seriousness of the distorting effect also depends on how promptly prices in unregulated markets, and particularly of substitute services, adjust to the long-run competitive equilibrium level. The pervasiveness of market imperfections in the nonpublic utility sectors of the economy (consider for example the cyclical price behavior of such competitors as trucks, cars, buses, and petroleum) suggests that any attempt to fix public utility rates at the purely competitive equilibrium level would produce distortions in the opposite direction. This observation, the reader will recognize, raises once more the problem of the "second best." It suggests the necessity of looking to the prices of specific public utility services in the light of the elasticities of their particular demands, which necessitates in turn a consideration of the price of substitutes and the relationship of those prices to *their* costs, before deciding whether to try to move any of the former closer to the purely competitive level. Since this is an argument as much against marginal-cost pricing in general as against reproduction cost rate bases, we return to it briefly in Chapter 7. But most economists would almost certainly reject any general attempt to make prices in only one sector of the economy highly flexible cyclically.
3. As Justice Brandeis pointed out decades ago, the "reproduction cost" to which prices in purely competitive markets tend to correspond is not the current cost of reproducing the existing plant, brick by brick, but the current cost of producing the *service* with the most modern technology available. It has been the former, not the latter, that public utility

[69] See B. W. Lewis, in Lyon, Abramson and Associates, *op. cit.*, II: 689, including the reference in note 134 to Bernstein, *Public Utility Rate Making*.

"For instance, the 1963 A.T.&T. Annual Report shows a net telephone plant investment of more than $23\frac{1}{2}$ billion dollars in 1963 . . . as against a mere $7\frac{1}{4}$ billion in 1950. All of the difference, as well as a considerable part of the $7\frac{1}{4}$ billions, consists of warm *new* dollars. . . ." Lewis, in Shepherd and Gies, *op. cit.*, 237.

[70] See Bauer and Gold, *op. cit.*, 405–413.

commissions have typically been involved in laboriously estimating in reproduction-cost proceedings.[71] In view of the rapid technological progress that has characterized some public utility industries and their tendencies toward long-run decreasing costs (see pp. 124–130), it is by no means clear that reproduction costs correctly defined are typically higher than original costs in periods of moderate inflation.

4. If the reproduction-cost rate base were correctly defined to embody the most recent technology, it would still be anomalous to add together, as is the typical regulatory procedure, capital costs for such a hypothetical new plant and operating expenses actually incurred in some test year in the plant that actually did the producing. If the competitive norm is conceived to be the average total cost of a new plant, using new technology, it is the operating cost of *that* plant that would have to be incorporated in the cost of service.[72]

5. Of course, the proper economic standard is not current average total cost but either short- or long-run marginal cost. The entire concept of determining rates by incorporating some average necessary rate of return on *total investment*, however valued, is a misleading one except in the circumstance that the industry operates under conditions of long-run constant cost. Under any other conditions, the level of cost depends on the level of output, and the latter in turn depends on the price that is set, if demand has any elasticity at all. In these circumstances, the typical method of basing average prices on average current *or* past costs of producing current or past levels of output in some test year[73] becomes

---

[71] See, for example, the listing of the typical methods of estimating reproduction costs in C. F. Phillips, *op. cit.*, 241–242, all of which ,it will be noted, ignore technological progress. For the Brandeis observation, see his famous dissent in *Southwestern Bell Telephone Company of Missouri v. Public Service Commission*, 262 U. S. 276, at 312 (1923); see also J. M. Clark, *Social Control of Business*, 2d ed. (Chicago: Univ. of Chicago Press, 1939), 306–308; and compare H. G. Brown, *op. cit.*, *Jour. Pol. Econ.* (October 1925), XXXIII: 505–530 with John Bauer, "Rate Base for Effective and Non-Speculative Railroad and Utility Regulation," *Jour. Pol. Econ.* (1926), XXXIV: 494–495. As J. R. Nelson aptly observes, "if particular assets are really to be replaced in kind, there must be something wrong with allowing for *any* obsolescence in the annual depreciation charge." Shepherd and Gies, *op. cit.*, 72.

[72] See Charles W. Smith, "Public Utility Depreciation," *Pub. Util. Fort.* (October 23, 1952), L: 630. However, this defect in the application of reproduction cost would disappear if the allowance for depreciation deducted were just sufficient to reflect the obsolescence of the old plant and hence to offset its excessive (by new plant standards) operating costs. See Bonbright, *op. cit.*, 229 and our fuller discussion in the next section of this chapter.

[73] An interesting departure from reliance on a single past test-year has been the informal acceptance by the Federal Communications Commission staff of an accounting system for the Communications Satellite Corporation in which revenue requirements are estimated on the basis of *anticipated* cost experience over a five-year period in the future. This innovation was dictated by the fact that Comsat's rate schedules had to be developed before the company had accumulated any operating experience with its revolutionary new method of communication, and—of particular significance at this point—the elasticities of demand and the prospective future behavior of unit costs made it evident that the company would suffer high operating losses during its initial period of operations under any conceivable system of rates. Congress had instructed the FCC to develop a global system of space communications as rapidly as possible. Had Comsat attempted to set rates high enough to cover the high initial costs of doing so, it would have found itself without customers. Therefore, it and the Commission properly decided that some of these high initial costs were chargeable to later users. (On this principle, see pp. 121–122.) They agreed to amortize some of the developmental costs and preoperating expenses over a ten-year period, using the *reverse* sum-of-the-digits method—that is to say, with heavier depreciation allowances taken in later years than in earlier—in order better to match

> hopelessly circular: it offers no indication of what average costs *would be* if some other level of rates were set, leading to some other volume of sales.[74]

Of course, competitive prices do move up and down in correspondence with costs. But they are not determined solely by costs, and certainly not by average total costs, reproduction or historic.[75]

---

depreciation expenses with the anticipated sharply rising flow of revenues. They envisaged, similarly, that the annual rate of return would be substantially below the ordinary range of reasonableness in the early years of service, and substantially above it in later years. These various understandings, they believed, would permit Comsat to charge rates low enough to induce a rapidly increasing utilization of its large initially installed capacity, thus drastically reducing unit costs over the life of the satellite. See the interesting paper by A. Bruce Matthews, "Problems Posed by Current Regulatory Practices to the Rapid Introduction of Communications Satellite Technology," delivered at a Symposium on the Rate Base Approach to Regulation at the Brookings Institution, June 7, 1968.

[74] In a survey of 90 public utility commission decisions over the period from 1937 to 1946, Troxel found that only two indicated any allowance for buyer responses to price changes. "Demand Elasticity and Control of Public Utility Earnings," *Amer. Econ. Rev.* (June 1948), XXXVIII: 372–373. "To achieve better regulatory effects, commissions need studies of demand behavior—any studies." *Ibid.*, 382.

Troxel and, following him, Phillips both argue also that regulation should pay closer attention to the marginal instead of the average return on investment—pointing out that efficiency requires firms to invest up to the point at which the return on *incremental*, not average, investment is equated with the cost of capital (k). See Troxel, *Economics of Public Utilities*, 391–395 and C. R. Phillips, *op. cit.*, 300–302. (On the possibility of firms being faced with an increasing cost of capital, where the MC of capital exceeds the average, and the implications of profit-maximizers equating the marginal return with the former instead of the latter, see note 30, Chapter 2, Volume 2.) I confess to great difficulty in following their argument.

I can think of three tendencies (in addition to those suggested in the text, above) to which they may be referring when they imply, as they seem to do, that the concentration of regulatory attention on the average return on total investment (however valued) may produce inefficiencies. First, there are times when permitting regulated companies to earn an average return equal at least to k may result in excessively high prices and underutilization of capacity. This would be so in a period of inadequate demand, when the prospective marginal return on investment is in any event below k, so that the companies would not be making any investments anyhow and would therefore have no economic need or justification (on SRMC-grounds) for earning such returns. This may be what Troxel has in mind when he says:

> ". . . the marginal rate of return is not the same thing as the cost of capital. Yet utility commissions use current costs of capital to determine the fair rate of return. Either the utility companies pay these borrowing costs, the commissioners say, or no borrowing can be done. True; but the utility company does not borrow unless the marginal rate of return is above the market rates of interest." *loc. cit.*, 392.

The second possible resulting distortion is that regulated companies may undertake investments the marginal return on which is *less* than k—investments that are, therefore, socially undesirable—where they have reason to believe regulation will permit them to recoup the difference in other markets, in order to keep their average rate of return at the legally permissible level. To prevent such investments, regulatory commissions might have to investigate the return on each investment, in order to disallow those that fell short of k. (We discuss this "A–J–W" tendency at length in Chapter 2, Volume 2.)

The third possibility is that the traditional policy may discourage regulated companies from undertaking very risky investments—risky because they offer a strong possibility of heavy losses, but worth undertaking because they offer the possibility also of very high returns—by threatening to take away the gains from a successful venture if it raises the company-wide average return too high. (See pp. 53–54.)

Apart from these three possibilities (and it is by no means clear that these are what they intend) I do not see what the authors have in mind. It is certainly *not* true that regulation prevents companies from attracting whatever additional capital they need for investments on which the marginal return exceeds the cost of capital. See our discussion of this point, p. 110.

[75] For an argument that rate base calculations are irrelevant to efficient pricing in the cases of railroads, natural gas, and urban transport, as well as an excellent analysis of the entire reproduction cost rate base issue, see Bonbright, *op. cit.*, 224–237.

Quasi-competitive pricing can be achieved for public utilities only by an explicit and separate consideration of the short- and long-run marginal costs on the one hand, and the intensities and elasticities of demand on the other for each one of their services in each of their markets. The economic and constitutional requirement that investors be given some assurances of a sufficient average return on their investment must, to be sure, exert some influence on these individual pricing decisions, and especially on the level of the entire structure. But, as Melvin de Chazeau has eloquently argued, the valuation of property, which is an essential part of the process of determining an "*earnings* base" for purposes of regulating the return to investors, is of very little use as a "*rate* base"—that is, for the determination of rational individual prices.[76]

This much remains valid in the economic case for a reproduction cost rate base and gross rate of return as an approach to price making: that unless second-best considerations dictate otherwise, and to the extent that prices are to be based on LR instead of SRMC, it is definitely the current and future—not the historic—capital costs that are relevant. But the use of reproduction cost valuation itself makes small contribution to efficient pricing compared with the immense resources that have gone into its support and application. Indeed, to the extent that LRMC are below ATC, because of economies of scale, its contribution could well be negative, even in a period of long-run inflation. The reason for this is that both original and reproduction cost rate base valuation are relevant only to average-cost pricing. If average costs are higher on the latter than on the former rate base, because of inflation, moving to reproduction costs may compound the inefficiency inherent in such pricing, and prices based on average historic costs may therefore come closer to the proper level.[77]

It should be emphasized that the case for reproduction cost does not rest on economic considerations alone. At least equally influential, particularly as a result of the general inflation since 1940, has been the noneconomic argument that reproduction cost valuation is much fairer to utility company stockholders than original cost. To base depreciation charges and return on investment on historic costs during or after a period of inflation is to return to the investors dollars of much lower purchasing power (measured in terms of the cost either of consumer goods and services or of replacing the old capital goods with new, that is, of keeping their capital investment intact in real terms) than the dollars they originally invested.[78]

The consensus of most economists in this matter would seen to be the following:

1. As Ben Lewis has put it, "any scheme of compensation is fair provided only that it was reasonably anticipated at the time of investment."[79] The

[76] *Op. cit. Q. Econ. Jour.* (February 1938), LII: 346–359; see, also, essentially in agreement, Bryan and Lewis, *ibid.*, 342–345.

[77] That is, efficiency could require two corrections of prices based on average historic costs in these circumstances—upward because of inflation and downward because LRMC are below ATC. Reproduction cost valuation does only the first of these. In so doing it may push prices far above LRMC; whereas prices based on average historic costs may come much closer to that level.

[78] See the excellent survey of the arguments on both sides of this issue in Glaeser, *op. cit.*, 315–331, 393–402; and for a strong presentation of the view just summarized, see Morton, *op. cit.*, pp. 91–131.

[79] In Lyon, Abramson and Associates, *op. cit.*, II: 688.

argument here is that as long as investors are informed in advance of whether they will be explicitly protected against inflation (or, by use of an original cost rate base, against deflation), they can in fairness be left to take that fact into account in the prices they pay for the stock at the time of purchase. If, for example, they anticipate inflation, they will presumably pay a lower price per dollar of current earnings for the stock of company A, which promises them no protection, than for company B, whose rate base and/or depreciation are determined on the basis of reproduction cost. In this way they will demand—and get—a percentage yield on their actual investment in A sufficiently higher to compensate for their poorer treatment.

2. By this reasoning, it is *impossible* to compensate *future* stock purchasers for past inflation—they will simply bid up the price of the stock and thereby offset that compensation; or to protect them against future inflation: they will simply compete to pay a higher price for the stock when they buy it, in reflection of this better treatment. And to change the regulatory rules in order to give such compensation to *existing* stockholders would be simply to confer on them a windfall, a higher return on their investment than they had reason to expect when they made it.[80]
3. If the desire is, rather, to compensate existing stockholders because such inflation as has occurred or may occur in the future has exceeded or may exceed their expectations—protecting them against their mistakes—what ethical reason is there to do so for stockholders and not for bondholders? It is only the former who would benefit by increasing the total number of dollars allowed for depreciation or included in the rate base. This seems particularly anomalous when it is stockholders who typically demand and receive the higher return, precisely in order to compensate them for the greater risks they are supposed to bear.
4. If, nonetheless, the government does want to adjust stockholder returns, in the interest of fairness, it can do so just as well and with far less damage to the efficiency of the regulatory process by varying (their part of) the permissible rate of return, or by applying some sort of price index number to the total dollars of permitted net income.[81]
5. Finally, to return to our main theme of whether revaluation of property or investment is necessary in order to assure fair *earnings* to existing stockholders, it makes economic sense as the basis for fixing *prices* only as some sort of average for all services taken together and over a number of years.[82]

---

[80] "The yield on securities cannot be determined [that is, fixed or set] by regulatory fiat in the same manner as the rate of return on invested capital or equity. The best commissions can do with the market is arbitrarily influence the prices of the securities by altering investor expectations and generating windfall gains or losses to those who hold the stock coincidentally with the effectuation of those influences." Morris Mendelson, "The Comparable Earnings Standard: A New Approach," paper presented to the Bell Telephone Co. of Pennsylvania Seminar on the Economics of Public Utilities, June 9, 1967, mimeo., 4–5.

[81] It is difficult to quarrel with Bonbright's observation that employing a reproduction cost rate base is "an absurdly crude device" for remedying this situation, if indeed it calls for a remedy. *Op. cit.*, 189–191.

[82] See de Chazeau, "The Nature of the 'Rate Base' in the Regulation of Public Utilities," *Q. Jour. Econ.* (February 1937), LI: 298–316, the illuminating comment by Robert F. Bryan and Ben W. Lewis, "The 'Earning Base' as a 'Rate Base,'" and the "Reply" by de Chazeau, *ibid.* (February 1938), LII: 335–359. Also see Bonbright, *op. cit.*, 266–276.

Depreciation Policy and Technological Progress

One of the most difficult and interesting problems of rate making in the face of cost changes over time has to do with the appropriate reflection of technological change in determining the depreciation component of cost of service.[83] In view of the immense importance of technological progress for economic welfare it becomes especially important to see to it that cost-of-service determinations are compatible with the optimum adoption of new technology.

The purpose of including an allowance for depreciation in price is to ensure recovery of invested funds over the economic life of the physical capital in which they have been embodied; and of course to see to it that price reflects this authentic economic cost. (We assume, as is the case in most jurisdictions, that the original vs. reproduction cost issue has been resolved in favor of returning the dollars originally invested, no more and no less.) The principal limits on that economic life are wear-and-tear (a user cost) and obsolescence; we confine our attention here to the latter, since, in principle, the former obviously should be included among the other variable costs of production.[84]

It is equally correct to say that the *total* of depreciation charges is supposed to reflect the total decline in the value of the physical asset, from original cost to scrap value—that is to the point where it is just as valuable in the form of scrap as installed production capacity. A familiar question in the public utility literature has been whether the *periodic*—for example, annual—depreciation charges should have the same function—that is, whether they should also reflect as closely as possible the year-by-year decline in that economic value; or whether, instead, they can be nothing more than a conventional and arbitrary mechanism for prorating the total amount to be recovered over the total estimated economic life. We make no effort to resolve that controversy, although it is clear that the latter is surely more accurate than the former as a description of actual practice, considering that (1) rate bases are now typically stated in original costs instead of "fair value,"[85] (2) straight-line depreciation is the method almost universally used,[86] and (3) the rate of decline in economic or market value depends primarily on trends in replacement cost[87] and technological change, whose year-by-year rates are surely irregular and unpredictable. However, some of the most interesting economic questions arise when the rate of decline in market value differs significantly from the depreciation rate actually employed.

What happens if technological change has been unexpectedly rapid? (It is

[83] Is it unbearably repetitious to remind the reader again that rate making on the basis of SRMC can ignore the depreciation that is not part of variable cost, but that pricing on the basis of LRMC or ATC must take it explicitly into account?

[84] See pp. 71–73, Chapter 3.

[85] As Bonbright points out, it is anomalous to think of depreciation as measuring decline in market value when one uses an original cost rate base, making no effort to adjust it up or down with the current value of the asset. *Principles of Public Utility Rates*, 194–201.

[86] Eugene F. Brigham, *op cit.*, *National Tax J.* (June 1967), XX: 210.

[87] In a period of inflation the market value of the asset may remain stable or actually rise, despite wear and tear and obsolescence: presumably "economic depreciation" would have to take this offsetting factor into account, so that the book value of the asset (original value less depreciation) would correctly reflect market value at the end of each accounting period. For simplicity we assume constant price levels in this discussion.

this possibility instead of the opposite one that most troubles economists and public utility companies, partly because regulatory commissions have typically been very conservative in the depreciation rates they allow[88] and partly because the discrepancy of inadequate depreciation can have a more seriously distorting effect on pricing and replacement policy.) There is a real danger in this event that replacement of old plant and equipment with new will be uneconomically discouraged. To understand this danger, we must have a brief and simplified look at the economics of replacement.

The way for a company to decide whether to replace a piece of machinery (or plant or other equipment) is to compare the average *variable* cost of producing with it (AVC_0) with the average *total* cost of production with new equipment (ATC_n). Only the variable costs of the old can be saved by turning to the new; the choice therefore is between continuing to incur those AVC, on the one hand, or incurring the ATC—including the capital costs as well—involved in purchasing a new machine. If the AVC_0 are smaller than the ATC_n it is economical to continue to use the old capital goods. But if, *regardless* of the fixed costs on the old, the AVC_0 are the greater, it is foolish not to scrap; every moment of continued production with the old means a greater drain on the company's resources, a greater avoidable cost of production, than would be involved in replacement.[89]

In either event, the continuing, fixed costs on the old equipment—the depreciation that may not yet have been fully recovered, the return on the net investment not yet fully written off, interest on the debt already incurred—are irrelevant to the decision. Sunk costs such as these are bygones, unchangeable past history, and best forgotten. The way to maximize profit is to minimize the variable, or incremental, or avoidable costs of production (since the others are fixed anyway); and that means the variable costs for existing plant and the total costs for new. This is just as true for a monopolist as for a firm operating under pure competition.

But it need not be true for a regulated company. That company cannot ignore the fixed costs on existing assets, because the regulatory commission may or may not choose to include them in its cost of service once the assets have been replaced. Suppose, for example, that the average variable costs under the old process are a constant \$7 per unit, the average fixed costs (depreciation and return on the unamortized part of the investment) \$3, and the regulated price is \$10. Suppose then a new process becomes available with the same capacity as the old, with average variable costs of \$4.50 and average fixed costs of \$2. Such an investment would be economically efficient; every unit that continued to be produced under the old process (at an avoidable cost of \$7) would be involving society as well as the firm in the unnecessary expenditure of 50 cents worth of resources (since the *total* unit

---

[88] In 1964, depreciation expense for large privately owned electric utilities seems to have run at 2 to $2\frac{1}{2}$% of gross book investment. Federal Power Commission, *Statistics of Electric Utilities in The U. S., 1964, Privately Owned*, March 1966. The typical depreciation rate for interstate gas pipelines is 3 to $3\frac{1}{2}$%. Richard W. Hooley, *Financing the Natural Gas Industry* (New York: Columbia Univ. Press, 1961), 66. The FCC prescribed an increase in the rate for the telephone industry from 5.1 to 5.4% effective January 1, 1968.

[89] This statement ignores the effect on these calculations of the expectation that in the future some even more efficient plant or machine may become available. Such an expectation might justify a company practicing what Fellner has termed "anticipatory retardation"—stalling the replacement of an old machine with a new, in order to await the next, even lower ATC_n that will be available. See note 91.

costs under the new are only \$6.50). In an unregulated industry, even a monopolist would make the investment: at the very least, he could produce at the same rate and sell his product at the same price as before, and simply pocket the 50-cent per unit cost saving. Suppose, however, that the company was a regulated utility and that its regulatory commission insisted that the cost savings be more or less promptly translated into price reductions. A price cut to \$9.50 would raise no difficulty, apart from possible considerations of risk: the company could continue to obtain the \$3 of capital costs on the old equipment plus the \$6.50 full unit cost of the new. But the commission might well insist that the old assets be removed from the rate base, once they had been replaced, even though depreciation on them had not yet been fully recovered. It might insist, that is, that the price be reduced to \$6.50, the new unit cost of service, thus forcing the stockholders to bear the loss of the unamortized portion of their investment in the old equipment. In this event the company would find itself in a position of having incurred *additional* capital costs of \$2 a unit, and yet had its gross return on capital (depreciation and profit) reduced from \$3 a unit to \$2 a unit. It obviously would have been better for the company to postpone the new investment and continue to take in the \$3 per unit of depreciation and return on the old assets until the latter had been completely written off.

What happened in this example was that technological progress had outrun the allowances for depreciation: it reduced the economic value of the old plant to zero (or to its value as scrap) before those assets were wholly written off in the books and the original investment fully recovered from customers. And the moral would seem to be that when this occurs, a regulated company will be deterred from replacing old assets with economically more efficient new ones unless it is permitted to continue to charge customers the capital costs of the unamortized portion of previous investments.[90] These customers may complain, with justice, that they are being made to pay more than the marginal, or indeed the total cost of serving them; that the company is being permitted to recoup from them sunk costs that should have been charged against customers in the past. But they are still better off than if the company refused to install the new, lower-cost equipment for serving them.[91]

---

90 See Troxel, *Economics of Public Utilities*, 356–369.

91 There is no well-developed "economics of error"; in the present instance there is no perfect solution of a problem that arises because mistakes have been made in the past. But it should also be emphasized that price will not instantaneously fall to the ATC of the newest and lowest-cost available processes even under perfect competition. As William Fellner has pointed out, rational firms will practice "anticipatory retardation" in the face of a continuous flow of cost-reducing innovations over time. Even pure competitors will not instantly adopt a new technology as soon as the ATC of the latest available process falls below market price. With correct anticipations they will recognize that, since technological progress is continuous, such an investment policy would produce continual disappointment; with further improvements in technology, price would be perpetually slipping below (or, to the extent that the gains of improved productivity are passed on in higher incomes to the factors of production, costs would be perpetually rising above) the levels at which the calculations were made, and investors would therefore continuously fail to make the anticipated return on investment incorporated in the ATC_n. They will therefore systematically delay the introduction of new processes, introducing not the first improvement whose ATC is below current market price but one later on in the flow of improvements, waiting until the return from cost savings promises to be sufficiently high in the early years of life of the new equipment to offset the eroding away of those gains as still later techniques become available—until, that is, it appears they will be able to earn the anticipated depreciation and return over the life of the new plant. So purely competitive price remains on

It could be argued, instead, that the costs of mistakes such as these ought to be borne by the stockholders. It is their function, not that of consumers, to bear the risks of unanticipatedly rapid obsolescence; their rate of return ought to be high enough to compensate for such risks. The argument would not be wrong, in principle. But by the same reasoning the allowable rate of return of public utilities is kept typically below that in industry generally precisely because stockholders share these risks with consumers; what consumers would gain by a different treatment of depreciation in these circumstances they would lose by having to pay a higher return.[92] Moreover, allowable depreciation is usually determined by the regulatory authorities: if it proves *ex post* to have been inadequate, it is not clear that the burden is properly borne by stockholders.[93] Here again is reflected the conception that regulation should in the face of change and uncertainty permit public utility companies to cover their authorized revenue requirements—not more and not less—rather than treat them as they would be treated by a competitive market. Finally, there remains the basic problem that putting the burden on stockholders would discourage economically efficient replacement of obsolete assets.[94]

---

the average sufficiently above the total unit costs under the latest available technique to permit investors on the average to write off old plant and earn the required return on its undepreciated portion. William Fellner, "The Influence of Market Structure on Technological Progress," in Amer. Econ. Ass'n, *Readings in Industrial Organization and Public Policy* (Homewood: Richard D. Irwin, 1958), 287–291. This is precisely what the recommended public utility commission treatment of depreciation and return on undepreciated, replaced equipment would accomplish: by holding price above ATC_n it would permit recovery of the fixed costs of the old. The danger, then, would not be that utility companies would be unduly discouraged from introducing new techniques but that they would be encouraged in this manner to inflate their rate bases, being permitted by their commissions to recover investment in the old and to earn a return on the new even though the latter was unneeded. On this "A–J–W" danger, see Chapter 2, Volume 2.

[92] This consideration does not fully dispose of the argument. It might still be that efficiency would be better served by having risks of this kind borne in the overall rate of return than in continued amortizations of incompletely depreciated, obsolete assets: the incidence of these two methods would almost certainly differ, depending on how precisely the amortization was effected.

[93] But see note 94. "In a non-regulated competitive industry, market forces will punish those investors who select managers who have incorrectly foreseen the rate of technological advance. . . .

"In a regulated industry with only one or two suppliers, however, society can not afford the disruptive effects on supply which the market discipline enforces for inevitable errors of foresight. . . .

"Regulation should not, of course, provide an umbrella for all errors of managerial judgment; however, it appears to me that a consistently used current cost base might reduce the willingness of investors to provide capital funds . . . *unless* management slows down the rate of technological change to one that is more readily predictable and is in line with past investment decisions embedded in existing durable equipment." Testimony of Paul Davidson, in FCC, *In the Matter of American Telephone and Telegraph Co.*, Docket 16258, Western Union Exhibit 4, 1968, mimeo., 71–76.

[94] It is very largely the fear of unanticipatedly rapid technological obsolescence that apparently explains the recent tendency of regulated companies to press for higher allowable rates of depreciation—a tendency that might otherwise be difficult to understand, since higher depreciation expense means a more rapid diminution of the rate base. (See note 32, Chapter 2, Volume 2.) This fear is intensified where the utility companies face the competition of companies that have access to the newer technology and are unencumbered by the costs of older, incompletely amortized plant—a situation that has prevailed in communications in recent years. See the discussion in Chapter 4 of Volume 2, especially around notes 94–96.

On the other hand, the utilities are themselves responsible in part for these difficulties. Many of them have resisted the adoption of more rapid depreciation, with its attendant income tax advantages, precisely in order to avoid the more rapid decline in rate base that this would have

Therefore, if technological progress outstrips depreciation, and regulated companies are permitted to recover their as-yet unamortized investment in obsolete facilities, prices will exceed LRMC. This will be true even under conditions of constant costs, when LRMC equals ATC_n. The source of the discrepancy is the difference between ATC computed so as to include gross return on a historic rate base and ATC_n. If the rate of depreciation accurately reflects the year-by-year decline in the economic value of existing assets no such discrepancy can occur: ATC on a historic rate base will be the same as ATC_n. The reason for this is that the economic value of existing assets at any given time is, precisely, the current value of the differences between AVC_o and ATC_n over their remaining life (or their value as scrap, whichever is larger). As long as AVC_o is less than ATC_n, the plant clearly has positive value, measured by the cost-saving that continued use makes possible. Once those two are equal, the old plant has zero value (for purposes of production; it may have positive scrap value). If the economic value were correctly stated on the books, the addition of gross return on that net book value to the variable costs of operating the old plant would produce a cost of service exactly equal to that of a new plant.[95]

The same end would be achieved by using a true current value rate base. But it is not the calculation of a reproduction cost rate base, as such—with all the administrative travail and expense that this has traditionally involved—that is the goal. Instead the goal is to estimate the cost of reproducing the service, with current technology. In principle, this can be achieved just as well by following an economically realistic depreciation policy, applied to original cost. In practice, the task of predicting obsolescence is likely to be a difficult one: but so has been the use of reproduction cost.

Clearly the charging of depreciation raises interesting and difficult questions of who should pay what share of capital costs over time. We have already posed the question of the proper rate when a plant is built far in advance of total need—perhaps because there are great economies of scale. To charge depreciation in equal annual installments would be to impose a disproportionately heavy burden on customers in earlier years, when much of the capacity lies idle. Considerations of fairness—the idle capacity is really for the benefit of future, not present customers—and economic efficiency present a case for something similar to SRMC pricing, which would have the effect of concentrating the capital charges in later years.[96]

Precisely the opposite course is suggested with respect to an investment required to meet current needs, but which may be expected to become rapidly

---

entailed. They did so feeling secure in their monopoly positions and their ability to continue earning an acceptable return on the larger investment. They have resisted also for fear of being forced by regulatory commissions to flow through the resulting tax benefits in the form of lower rates—a practice that does expose them to the possibility of higher tax liabilities and consequently reduced earnings in the future, and the necessity of asking for rate increases at that time. See pp. 32–34, above, and the excellent "Comment" by William H. Melody in Trebing and Howard, *op. cit.*, pp. 164–175, which concludes: "The long-run viability of utilities in some markets that are subject to external competitive pressures may well depend upon the maintenance of a depreciation policy that properly reflects the rate of economic depreciation in an environment of rapid technological change."

95 The preceding discussion draws heavily on the testimony of Vickrey in FCC, *In the Matter of American Telephone and Telegraph Co.*, Docket 16258, mimeo., esp. 53–56.

96 See the example of Comsat, note 73, p. 113, note 4, p. 88, and p. 104.

outmoded by new technology already on the horizon.[97] In this instance, the investment should be written off rapidly, however long its physical life is likely to be, in reflection of the early anticipated decline in its economic value. The effect would be to put the heaviest capital charges on customers now and in the immediate future—and properly so, since it is for their benefit that the capacity is being built now instead of later, when it could embody the lower cost technology. Such higher charges might well restrict demand sufficiently to demonstrate that the investment in question would better be postponed until the new technology was perfected. The opposite course—to charge depreciation only at the modest rate dictated by average historic experience—would result in charging future users much more than LRMC: they would be stuck with the costs of writing off the inadequately depreciated and obsolete older equipment. And the effect in this event would be to discourage the introduction of the new technology, because demand at that later time would be restrained by the inefficiently high price for the services, "rolling in" the excessive ATC of the old, inadequately depreciated assets with the much lower ATC of the new.[98]

Manifestly, the rate at which depreciation is charged can have important effects on technological progress. And although it is an impossible task to estimate the proper rate in advance, as Vickrey states,

". . . even a rough approximation to the inclusion of such an analysis in the rate making process is to be preferred to sticking to a fundamentally erroneous approach."[99]

---

[97] This example also depends heavily on the Vickrey testimony, *ibid.*, 27 and 56–60.

[98] Considerations of this type were apparently central to the controversy within the FCC that eventuated in 1968 in its authorizing AT&T to lay a submarine cable between the United States and Spain. In the dissenting opinion of Commissioner Johnson the cable project played the role of supplying the additional capacity (questionably) required in the near future, with current technology, and the satellite the role of the superior technology of the future (indeed, he felt, of the present). Commissioners Cox and Loevinger asserted, in support of the FCC decision, that:

". . . satellites are not now, and will not for at least the next 5 to 7 years be, the most economic means of providing international communications service."

But, Johnson asserted:

"Of course, by depreciating the cable over *twenty* years it appears that the per-year cost of the cable is lower than the per-year additional satellite cost over its projected *five*-year life. The point is that . . . neither will be needed as insurance for more than five years." FCC 68-212, 12514, letter from Rosel H. Hyde, Chairman, to Richard R. Hough, Vice President, American Telephone and Telegraph Company, and accompanying Concurring Statement of Commissioners Cox and Loevinger and Dissenting Opinion of Commissioner Johnson, February 16, 1968.

[99] *Op. cit.*, FCC Docket 16258, 59.

Board Staff Interrogatory # 2

Ref: #1, p. 4

AMPCO quotes K Viscusi et al, Economics of Regulation and Antitrust, p. 352, to the effect that Ramsey pricing is economically efficient.

a. Please file the section titled "Ramsey Pricing", i.e. pp. 350 – 353.

b. Does AMPCO recommend that Hydro One should develop a Network rate structure with two (or more) prices within the peak period, based on differing elasticities of demand?

c. If so, does AMPCO suggest that this structure should have two rates (distinguishing between LDCs and Power Producers on the one hand and Directs on the other), a structure with multiple rates (for example, distinguishing amongst the industrial sectors such as those studied by Dr. Sen), or some other structure?

d. In light of the second from last paragraph in the requested excerpt, does AMPCO recommend that Hydro One adopt "value of service" as a principle in its Network rate design?

Response

a) Please find attached Chapter 11, "Theory of Natural Monopoly", pp. 337-360, incorporating the paragraphs requested by Board Staff.

b) No.

c) AMPCO does not recommend differentiating the network charge determinant or rate design among customers or customer classes. While Ramsey pricing provides the appropriate theoretical framework for rate design, the common problem—that of knowing elasticities perfectly and a priori—precludes a literal application of the theory. Instead, and as we have proposed, a rate design that is based on a critical peak demand-based charge determinant serves as the best proxy for Ramsey pricing. Critical peak pricing provides a good price signal. It approximates the long run marginal cost of network capacity. It gives customers an opportunity to modify their consumption behavior in a way which reduces both the long-term investment needs for network capacity and that customer's expenditures on network service. In other words, the more sensitive is a customer to the price of electricity, the more likely is that customer to reduce demand in response to a critical peak price for network services, effectively fulfilling the objective of Ramsey's theory, that is, that deadweight losses are minimized by setting prices higher for customers with lower elasticity, and lower for customers with higher elasticity.

d) While we have not considered the concept extensively, the example described by Viscusi et al. might not be directly transferable to the present case: examining freight and attributing values to shipments for the purpose of setting rail rates. While rail rates are designed to recover fixed (capacity) costs as well as variable (volumetric) costs, transmission network service provides capacity only.

   A second consideration concerns the capacity of the Board to discern and quantify in any empirical way the value a customer or class of customers might attribute to energy consumed to provide a broad range of services at each moment in time and over time. We suspect that the challenge in developing a defensible approach would be out of proportion to the benefit realized in terms of assisting the Board in any practical way in its duty of deciding efficient and effective rates for transmission network services.

11 Theory of Natural Monopoly

As we discussed in Chapter 10, there are a number of market-failure arguments for economic regulation. Perhaps the most important and widely accepted is natural monopoly, and it provides the rationale for regulating electric-power and natural-gas distribution, local telephone service, water supply, and some common-carrier transportation services. We begin this chapter with a discussion of the theory of natural monopoly. Actual regulation of natural monopoly will be the subject of the next two chapters.

We will be taking an economic efficiency view of natural monopoly here. In previous chapters we have discussed various explanations for the existence of regulation, including market failure and capture theory hypotheses. In this chapter we focus exclusively on the natural-monopoly market-failure argument and various theoretical and actual solutions.

This chapter is primarily theoretical, but it also serves as an introduction to the next few chapters. Chapter 12 will be concerned with the practice of natural monopoly regulation and an evaluation of its benefits and costs. Chapters 13–15 will discuss several alternatives to regulation that are introduced only briefly here.

The Natural Monopoly Problem

An industry is a natural monopoly if the production of a particular good or service by a single firm minimizes cost. The typical example is production of a single commodity where long-run average cost (LRAC) declines for all outputs. Such a case is illustrated in Figure 11.1. Because LRAC is declining, long-run marginal cost (LRMC) necessarily lies everywhere below it.

The case shown in Figure 11.1 makes clear the public-policy dilemma. Simply stated, the problem is how society can benefit from least-cost production—which obviously requires single-firm production—without suffering from monopoly pricing. The idea, of course, is that a single firm would eventually win the entire market by continuing to expand output and lowering its costs. Having won the market, it could then set the monopoly price.[1]

Shortly, we will turn to an analysis of the variety of solutions to this problem that have been proposed. Before we do so, however, we will examine more carefully the definition and characteristics of natural monopoly.

Permanent and Temporary Natural Monopoly

An important distinction is that of permanent versus temporary natural monopoly.[2] Figure 11.1 illustrates the case of permanent natural monopoly. The key is that LRAC falls

1. Entry, induced by the monopoly price, is usually assumed to be unlikely in natural monopoly situations.

2. The term permanent is perhaps misleading inasmuch as one can never rule out dramatic technological changes that could convert a natural monopoly into a competitively structured industry.

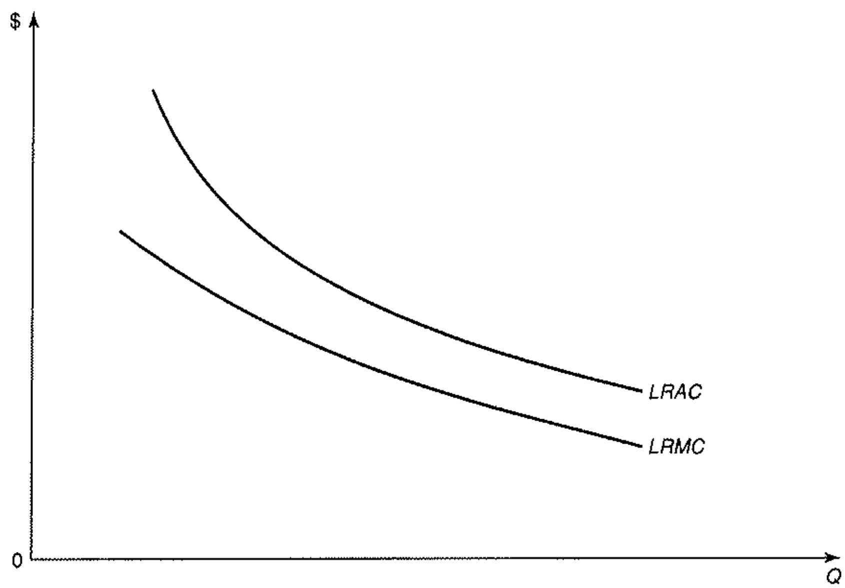


Figure 11.1
Cost Curves of Natural Monopolist

continuously as output increases. No matter how large market demand is, a single firm can produce it at least cost.

A temporary natural monopoly is shown in Figure 11.2. Observe that LRAC declines up to output Q^* and then becomes constant thereafter. Hence, as demand grows over time, a natural monopoly when demand DD prevails can become a workably competitive market when demand D_1D_1 holds.

One can argue that such a cost curve can be used to describe intercity telephone service. There are several factors that give rise to sharp unit-cost savings at low volumes of telephone calls, but they play out as volume increases.

For example, a microwave telephone system consists of a number of stations—about twenty to forty miles apart—that transmit signals of specific frequencies. Each station requires land, a building, a tower and antennas, electronic equipment, and so on. These inputs do not all increase proportionately with the number of circuits, and therefore as volume increases the fixed costs can be spread over more calls. This spreading effect becomes less and less significant, however, as volume grows.

As an example, long-distance telephone service between New York and Philadelphia required only 800 circuits in the 1940s. At this capacity, unit costs were falling and constituted a natural monopoly situation. In the late 1960s the number of circuits had risen to 79,000

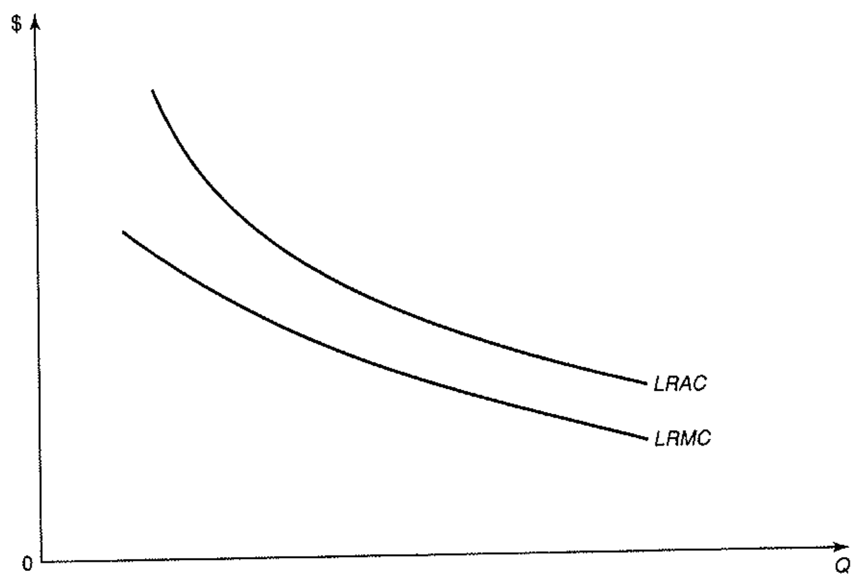


Figure 11.1
Cost Curves of Natural Monopolist

continuously as output increases. No matter how large market demand is, a single firm can produce it at least cost.

A temporary natural monopoly is shown in Figure 11.2. Observe that LRAC declines up to output Q^* and then becomes constant thereafter. Hence, as demand grows over time, a natural monopoly when demand DD prevails can become a workably competitive market when demand D_1D_1 holds.

One can argue that such a cost curve can be used to describe intercity telephone service. There are several factors that give rise to sharp unit-cost savings at low volumes of telephone calls, but they play out as volume increases.

For example, a microwave telephone system consists of a number of stations—about twenty to forty miles apart—that transmit signals of specific frequencies. Each station requires land, a building, a tower and antennas, electronic equipment, and so on. These inputs do not all increase proportionately with the number of circuits, and therefore as volume increases the fixed costs can be spread over more calls. This spreading effect becomes less and less significant, however, as volume grows.

As an example, long-distance telephone service between New York and Philadelphia required only 800 circuits in the 1940s. At this capacity, unit costs were falling and constituted a natural monopoly situation. In the late 1960s the number of circuits had risen to 79,000

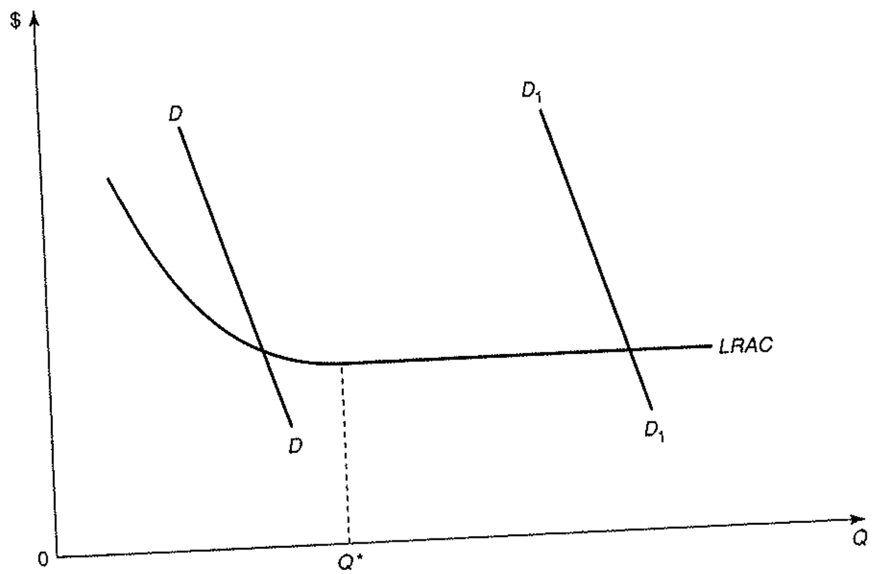


Figure 11.2
Temporary Natural Monopoly

(largely because of the requirements of television), and this volume was such that unit costs were essentially flat (beyond Q^* in Figure 11.2). Hence, by the late 1960s the temporary natural monopoly had disappeared.

This phenomenon is not rare. Railroads possessed significant cost advantages in the late 1800s, and these advantages were eroded considerably with the introduction of trucking in the 1920s. This example introduces a new element, namely, technological change.[3] That is, over long periods of time it is likely that the cost function will shift as new knowledge is incorporated into the production process. Hence, permanent natural monopoly is probably a rare category. Technical change can shift cost functions so as to render competition workable. And as we will see later, a serious deficiency of regulation seems to be that it often fails to "disappear" when the natural monopoly does.

Subadditivity and Multiproduct Monopoly

In the real world a single-commodity producer is rare. Electric utilities supply high and low voltage, peak and off-peak power; telephone companies provide local and long-distance

3. Strictly speaking, technical change in lowering costs was also present in the telephone service example.

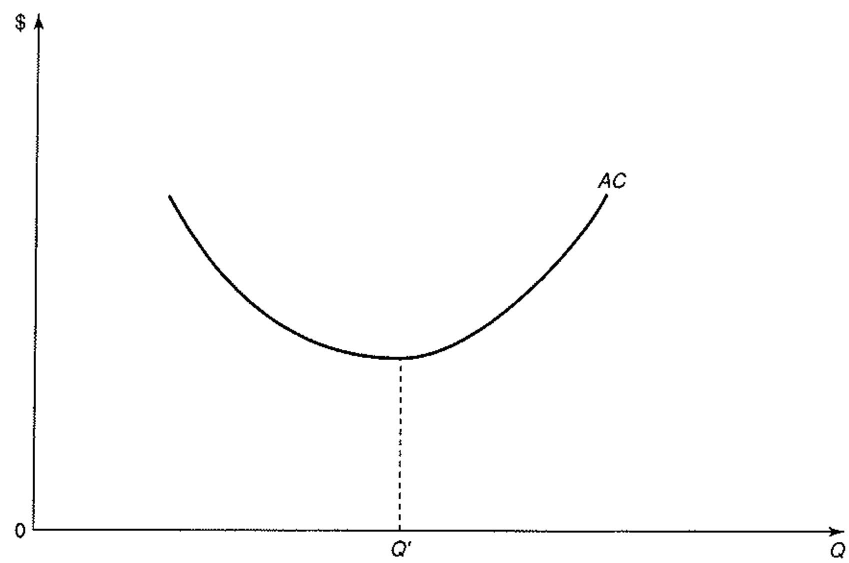


Figure 11.3
Economies of Scale up to Output Q'

service; and so on. It turns out that multiple-product natural monopoly is not only more realistic, but it also creates important theoretical issues that do not exist in the single-product case.

The definition of natural monopoly is that the cost function is subadditive.[4] We begin by explaining this concept in the single-product case because it can be illustrated graphically.

Consider the average cost curve shown in Figure 11.3. Average cost declines until the output Q' is reached, and then begins to increase. Economies of scale are said to exist at all outputs less than Q' and diseconomies at all outputs greater than Q'.

Subadditivity refers to whether it is cheaper to have one firm produce total industry output, or whether additional firms would yield lower total cost. For outputs less than Q', one firm is the least-cost solution, and therefore cost is subadditive for that range of outputs.

In order to examine the least-cost solution for outputs greater than Q', we introduce the minimum average cost function for two firms, AC_2. This curve and the single-firm AC curve from Figure 11.3 are both shown in Figure 11.4.

The curve AC_2 is obtained by construction from AC in the following manner. We know that for least-cost production, each firm must produce at the same output rate and thereby have the

4. An important article that defines natural monopoly this way is W. J. Baumol, "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry," *American Economic Review*, December 1977.

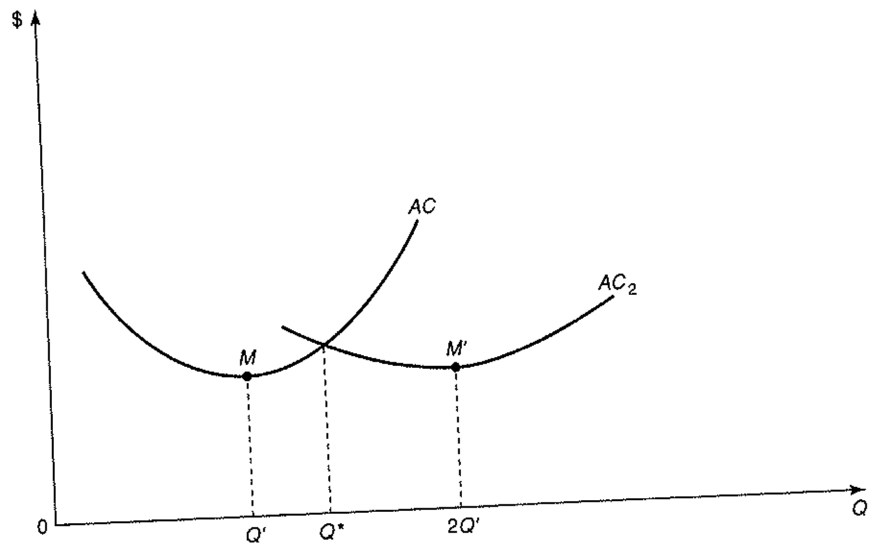


Figure 11.4
Minimum Average Cost Curve for Two Firms, AC_2

same marginal cost. Hence, for a given point on the AC curve, simply double the output rate to obtain a point on the AC_2 curve. For example, at the minimum average cost point M on AC, double Q' to get $2Q'$, which corresponds to the minimum point M' on AC_2.

The intersection of AC and AC_2 at output Q^* defines the range of subadditivity. For all outputs less than Q^*, a single firm yields least-cost production. Hence the cost function is subadditive for outputs less than Q^*. Notice that subadditivity is the best way to define natural monopoly. Even though diseconomies of scale obtain between Q' and Q^*, it would be in society's interest to have a single firm produce in that range. An important point is that economies of scale (declining average cost) are not necessary for a single-product natural monopoly (although they are sufficient).

When we turn to multiple-product natural monopoly, the distinction between subadditivity and economies of scale becomes even greater. Again, the proper definition of natural monopoly is that the cost function is subadditive. That is, whatever the combination of outputs desired (say, 85 cars and 63 trucks, or 25 cars and 78 trucks), it is cheaper for a single firm to produce that combination if the cost function is subadditive.

In the multiple-output case, it can be shown that economies of scale are neither necessary nor sufficient for costs to be subadditive! Economies of scale would hold, for example, if the total cost of producing, say, a 10-percent greater quantity of each commodity increased by some amount less than 10 percent. The reason that economies of scale are neither necessary nor sufficient for subadditivity is that in the production of multiple outputs, the interdependence among outputs also becomes important.

Although various ways have been proposed for measuring these interdependencies, the concept of economies and diseconomies of scope is appealing intuitively.[5] Economies of scope mean that it is cheaper to produce, say, 85 cars and 63 trucks within a single firm than it is for specialty firms to produce the required outputs. If you think of peak-period electric power and off-peak power as different commodities, then economies of scope are clearly present—the two commodities can share the same power plant and distribution system.

Sharkey has given an example of a cost function that possesses economies of scale for all outputs, but which is nowhere subadditive.[6] His example is

$$C(Q_1, Q_2) = Q_1 + Q_2 + (Q_1 Q_2)^{1/3}. \tag{11.1}$$

Notice that the total cost after increasing each output by 10 percent is

$$C(1.1Q_1, 1.1Q_2) = 1.1Q_1 + 1.1Q_2 + 1.1^{2/3}(Q_1 Q_2)^{1/3}$$

whereas the total cost increased by 10 percent is

$$1.1C(Q_1, Q_2) = 1.1Q_1 + 1.1Q_2 + 1.1(Q_1 Q_2)^{1/3}.$$

Because the former is less than the latter, economies of scale exist. Nevertheless, the function has diseconomies of scope that sufficiently outweigh the economies of scale to make cost nowhere subadditive.

To see this point, note that the third term in the cost function, equation (11.1), adds a positive amount to cost whenever both outputs are produced together. If, for example, all Q_1 was produced by firm A and all Q_2 was produced by firm B, then the sum of the total costs of the two firms would be less than if all production was carried out in a single firm, C. Specifically:

$$C_A = Q_1, C_B = Q_2, \quad \text{so} \quad C_A + C_B = Q_1 + Q_2$$
$$C_C = Q_1 + Q_2 + (Q_1 Q_2)^{1/3}.$$

Because $C_A + C_B < C_C$, production in the specialty firms, A and B, is cheaper than in a single firm, C. Thus, economies of scale are not sufficient for cost to be subadditive because of the diseconomies of scope.

In summary, the definition of natural monopoly in the multiple-output case is that the cost function must be subadditive. Subadditivity of the cost function simply means that the production of all combinations of outputs is accomplished at least cost by a single firm. It is a complex matter to specify the necessary and sufficient conditions for costs to be subadditive.

5. See, for example, J. C. Panzar and R. D. Willig, "Economies of Scope," *American Economic Review*, May 1981.

6. William W. Sharkey, *The Theory of Natural Monopoly* (New York: Cambridge University Press, 1982).

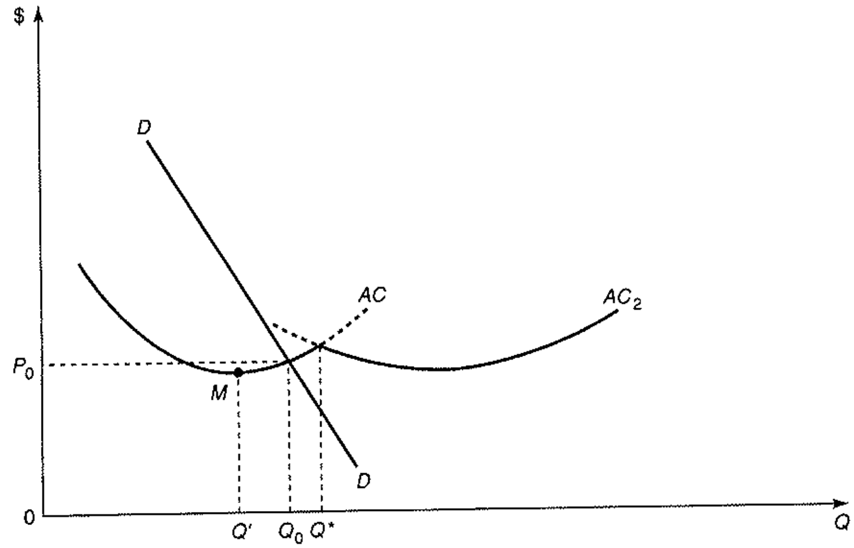


Figure 11.5
Sustainable Natural Monopoly up to Output Q'

We have shown through some simple examples, however, that it generally depends on both economies of scale and economies of scope. If both exist, then subadditivity will likely obtain.[7] Economies of scale alone, however, can be outweighed by diseconomies of scope. Thus, although economies of scale in the single-product case imply natural monopoly, this statement does not hold true for the multiple-product case.

Before turning to the various policy solutions to the natural monopoly problem, we shall briefly explain a related concept known as sustainability. It can be explained best by reference to Figure 11.5.

Figure 11.5 reproduces the cost function for the single-product case from Figure 11.4. Recall that the cost function is subadditive for outputs less than Q^*. Now consider a case in which market demand DD intersects average cost somewhere between Q' and Q^*, where AC is rising. If a single firm were to supply all output demanded at a price equal to average cost (at price P_0 and output Q_0 so that the firm would just cover all its costs), the natural monopoly would be termed unsustainable. That is, under certain assumptions, a potential entrant would have an incentive to enter the market and produce a share of total output even though doing so would increase the cost of producing the total industry output.

7. For a rigorous analysis, see W. J. Baumol, J. C. Panzar, and R. D. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt Brace Jovanovich, 1982).

The assumptions referred to in the preceding sentence are that the entrant expects the incumbent firm to keep its price unchanged for some period of time after entry, and that the incumbent will supply the residual output.[8] Under these assumptions, the entrant would perceive that it could profit by offering to sell output Q' in Figure 11.4 at some price above its minimum average cost (point M) but slightly less than the price P_0 being charged by the incumbent.

By contrast, a sustainable natural monopoly would be one where market demand intersects AC in Figure 11.5 to the left of Q'. In this case an entrant cannot undercut the incumbent and therefore has no incentive to enter. The concept of sustainability is relevant where a regulatory agency must decide whether to allow entry in a particular market of a multiple-product natural monopolist.

Alternative Policy Solutions

In this section we examine various alternatives that have been proposed (and, in some cases, implemented) to correct the natural monopoly inefficiency. These alternatives include "doing nothing"; various "ideal" solutions; competition among bidders for the right to the monopoly franchise; and, finally, actual regulation, as practiced in the United States, and public enterprise, as exemplified by the Postal Service.

The first alternative mentioned—doing nothing—might be appropriate if the potential monopoly power is not great. For example, a cable-television system might be viewed as a natural monopoly, but one with quite limited capacity for earning excess returns, for substitutes for cable television are rather close. Over-the-air broadcasting is one of them. Others are apparently becoming more important over time as new technologies are perfected.

We consider first a collection of "ideal" pricing solutions. The adjective "ideal" is employed to indicate that we are assuming that the firm is to be operated in the public interest and that the only issue is what prices produce economic efficiency.

Ideal Pricing

The most obvious candidate for the efficient price is, of course, marginal cost.[9] A natural monopolist that charges marginal cost for each product is said to practice linear (or uniform)

8. A further assumption is that the entrant perceives no entry barriers in the form of "sunk" costs. That is, the entrant believes that whatever investment is required can be recovered by transferring it elsewhere or by sale. All of these assumptions have been subject to controversy since the sustainability literature was introduced by Baumol, Panzar, and Willig.

9. See Chapter 4 for a detailed rationale. For a rigorous treatment of efficient pricing, see R. R. Braeutigam, "Optimal Policies for Natural Monopolies," in R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*, Vol. 2 (Amsterdam: North-Holland, 1989), and D. F. Spulber, *Regulation and Markets* (Cambridge, Mass.: MIT Press, 1989). For a more geometrical treatment, see K. E. Train, *Optimal Regulation* (Cambridge, Mass.: MIT Press, 1991).

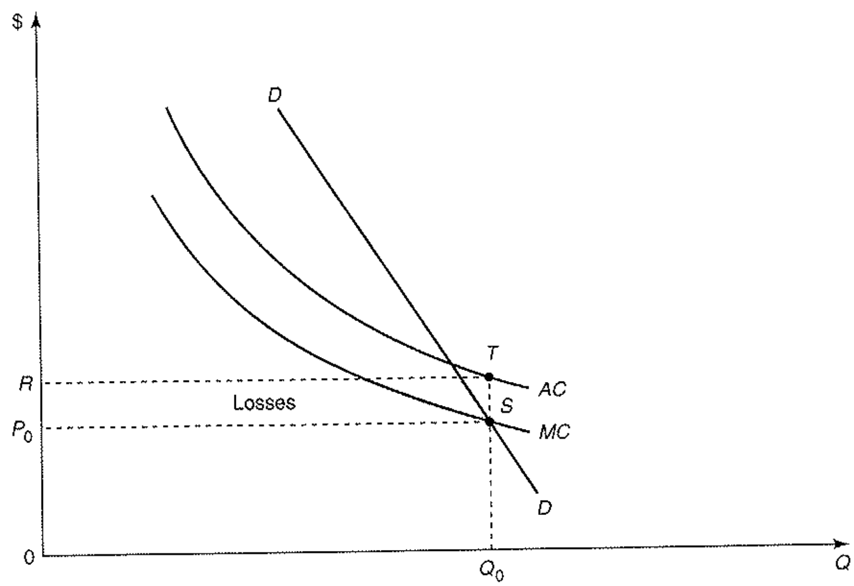


Figure 11.6
Marginal Cost Pricing Can Cause Losses

marginal cost pricing. In other words, a customer's expenditure for a product is a linear function of price and quantity sold, PQ. On the other hand, if the firm charges a fixed fee F, regardless of the amount bought, and also a per-unit charge P, nonlinear (or nonuniform) pricing would be in effect. Then the customer's expenditure would be a nonlinear function, $F + PQ$.

In our ideal pricing discussion, we begin with the linear marginal cost pricing solution. After considering nonlinear pricing we examine the so-called Ramsey pricing alternative, which applies to multiproduct cases. The section concludes with a discussion of a theoretical proposal by Loeb and Magat to induce profit-maximizing firms to price efficiently.

Linear Marginal Cost Pricing

Consider a single-product natural monopolist with decreasing average costs over the relevant output range. Figure 11.6 shows such a situation where market demand is DD.

The marginal cost price would be P_0 with output Q_0. The price does meet the well-known requirement for efficiency; however, on closer examination, several serious difficulties arise. An obvious difficulty is the loss, shown by the shaded rectangle RP_0ST.[10] Any enterprise

10. The loss is equal to the difference between price and average cost, multiplied by output.

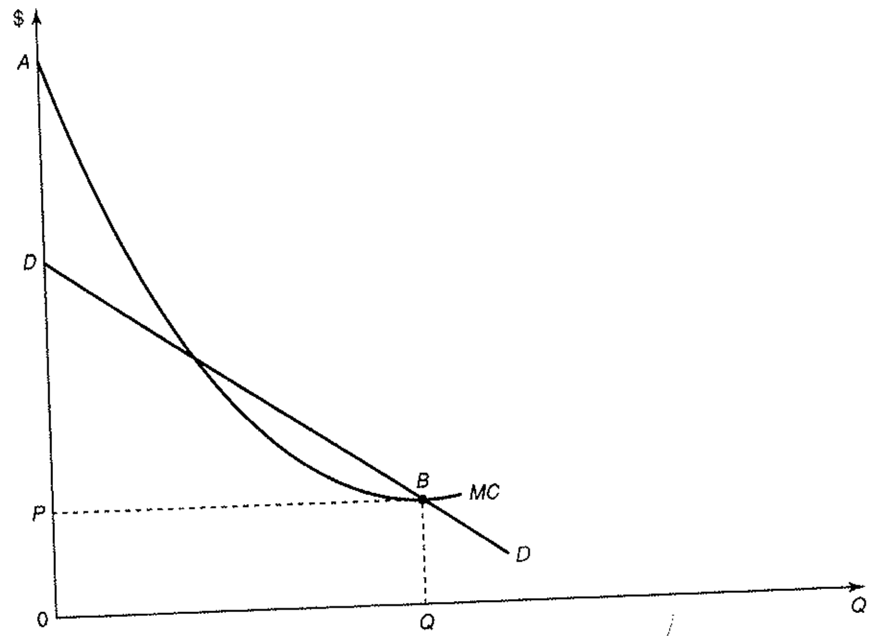


Figure 11.7
Natural Monopoly with Costs Exceeding Benefits

would need a subsidy to continue to operate at this output level, because price is less than average cost. The next question is to ask where the subsidy is to come from and what effect this will have on economic efficiency.

The only "correct" solution is for the government to raise the subsidy through a lump-sum tax, that is, a tax that would not distort other decisions throughout the economy. Such taxes are rarely, if ever, used in practice. Income taxes and sales taxes are unacceptable because they create inefficiencies themselves by introducing wedges between prices and marginal costs. Even this "correct" solution (lump-sum tax to pay subsidy) is subject to some rather persuasive opposing arguments. Three frequently mentioned arguments are as follows:

1. If total costs are not covered by consumer expenditures, it is possible that total consumer benefits (given by the area under the demand curve)[11] are less than total costs—which means the good should not be produced at all. Figure 11.7 provides such a case. Total costs $AOQB$

11. Throughout this chapter we make the common assumption that the area under the demand curve measures total willingness to pay by consumers. This requires one to assume that the income elasticity of demand is zero (or small enough to make the error unimportant). See R. D. Willig, "Consumer's Surplus without Apology," *American Economic Review*, September 1976.

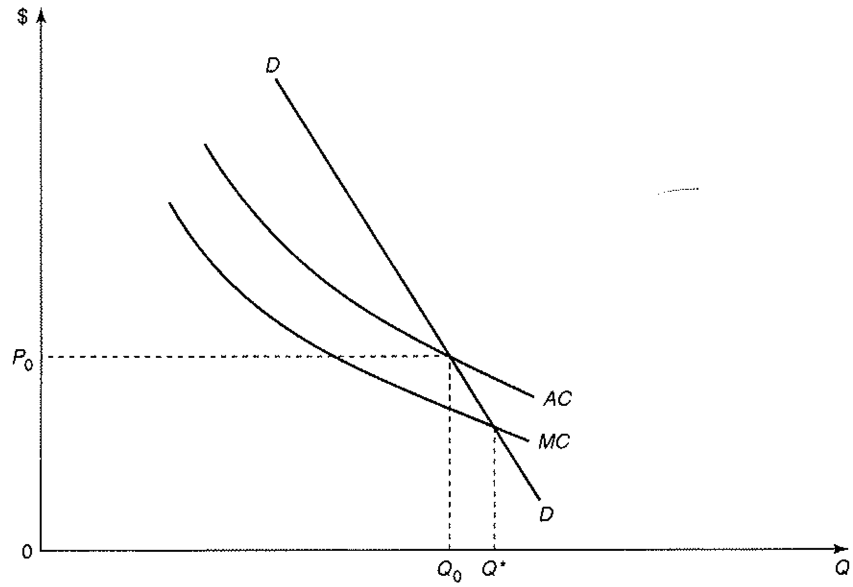


Figure 11.8
Welfare Loss with Average Cost Pricing

(the area under the MC curve) exceed total benefits $DOQB$. Only if consumers are required to actually cover total costs can we be sure that the good is socially beneficial.

2. Because the enterprise's management knows losses will be subsidized, the incentive and capacity to control costs is weakened. Postal Service employees, for example, have an advantage in bargaining with management, inasmuch as both sides know that the enterprise will not fail if revenues are less than costs. The Treasury can always be counted on to subsidize the Postal Service in a pinch. Steel industry labor unions do not have this advantage.

3. On distributional grounds, it can be argued that nonbuyers of the natural monopoly good should not be required to subsidize the marginal cost buyers. That is, why should the taxes paid by individuals without telephone service be used to subsidize individuals who purchase such service at a loss-creating price?

A major point of the preceding analysis is that enterprises should price so that their revenues cover costs. Furthermore, in the United States, because most public utilities are privately owned firms, it is politically unrealistic to imagine government subsidizing the losses of private firms. Hence we conclude that there are compelling reasons to accept the constraint that natural monopolies should operate such that total revenues and total costs are equated.

In the single-product case, linear pricing implies that price must equal average cost if total revenues must equal total costs. This relationship is shown in Figure 11.8 as price P_0 and output Q_0. This departure from marginal cost pricing leads, of course, to the welfare loss given by the shaded triangular area.[12]

This argument refers to linear pricing; that is, the buyer pays a single price per unit, and therefore the buyer's total expenditure is proportional to total consumption. An important alternative is nonlinear pricing.

Nonlinear Pricing

A two-part tariff is nonlinear and consists of a fixed amount or fee, regardless of consumption, plus a price per unit. If the price per unit equals marginal cost, then it is possible to have efficient pricing and have total revenues of the firm equal to its total costs.

For example, if the loss under linear marginal cost pricing is estimated to be K (the shaded rectangle in Figure 11.6), the fixed fee of the two-part tariff could be set so that the sum over all customers equals K. There are various ways for this equality to hold—the simplest is to set the fixed fee equal to K/N, where N equals the number of consumers.

There are possible problems with this nondiscriminatory two-part tariff. Because consumers usually vary considerably in terms of their demands for the good, it is possible for some consumers to be driven from the market if K/N exceeds their consumer surpluses at price equal to marginal cost. One might expect this outcome to be more likely for, say, telephone service than for such "necessities" as electricity and water. Hence, efficiency losses will occur if these excluded consumers would have been willing to pay marginal cost. It is also true that in some markets it is not feasible to enforce a fixed fee for the "right-to-buy" at a price per unit. Consumers would have an incentive to have one person purchase for all, thereby paying only one fixed fee. This is not a problem for most public utilities.

The obvious thing to do to avoid excluding consumers is to charge different fixed fees to different consumers, or classes of consumers. In short, discriminatory two-part tariffs could tailor the fixed fees to the consumers' willingnesses to pay where the sum of the fixed fees should add up to K. Although this solution is best in terms of efficiency, it may be illegal to so discriminate.

If all consumers must be charged the same fixed fee, it will still be more efficient to use a two-part tariff than to use linear pricing (which in the case of a single product implies average cost pricing). The reason is simply that by using a fixed fee to make a contribution to revenues, the price per unit can be lowered toward marginal cost—thereby reducing deadweight losses. (In principle, one can pick some fixed fee, no matter how small, that will not drive anyone from the market and will permit a lowering of the price.)

The next logical question is, What is the optimal two-part tariff? Here, we explain only the economic principle involved.[13] Suppose initially that the fee is zero and price equals marginal

12. For a discussion of welfare loss determination, see Chapter 4.

13. See Stephen J. Brown and David S. Sibley, *The Theory of Public Utility Pricing* (New York: Cambridge University Press, 1986), p. 93, for a formal analysis.

cost. The result is, of course, a deficit that must be covered by increasing either the fee or the price per unit, or both. In essence, the derivation depends on a balancing of efficiency losses because of exclusion of additional consumers as the fixed fee rises against the increased consumption losses as price per unit increases above marginal cost. Hence the optimal two-part tariff generally will involve a price per unit that exceeds marginal cost and a fixed fee that excludes some consumers from the market.

Multipart tariffs are often used by public utilities. Consider the following example of the type of tariff sometimes used for local telephone service (such tariffs are often referred to as declining-block tariffs).

Fixed fee per month—$5

+10 cents per call for up to 100 calls

+5 cents per call for all calls between 100 and 200

+0 cents per call for all calls above 200

Notice that the marginal price falls as one moves to successively larger calling "blocks"—from 10 cents to 5 cents to 0 cents. This multipart tariff is plotted in Figure 11.9 as the bold segmented line $ABCD$. (The reason for the extensions of these segments in Figure 11.9 will become clear shortly.) Hence the figure shows "total consumer expenditure" vertically as a function of total "calls per month" horizontally.

A rationale often given for the declining blocks is that utilities are characterized by economies of scale, and falling marginal prices stimulate consumption—in turn permitting the construction of larger, lower-unit-cost plants. An alternative rationale is to view the declining-block tariff as a *self-selecting* set of two-part tariffs, and a set of such tariffs can increase economic efficiency along the lines discussed earlier.

Recall that discriminatory two-part tariffs permit the firm to tailor the tariffs to fit the differences in willingnesses to pay across consumers. The efficient solution can be achieved if no consumers are excluded from the market and all pay marginal cost per unit. As an approximation to this "ideal," one can use the multipart tariff in Figure 11.9 to cause consumers to self-select a two-part tariff that they prefer—wherein consumers with high willingnesses to pay pay high fixed fees in return for low prices per unit.

The three "self-selecting" two-part tariffs are

Fixed Fee	Price/Unit
$5	10 cents
$10	5 cents
$20	0 cents

One can represent a two-part tariff by a vertical intercept (for the fixed fee) and a straight line with slope equal to the price per unit. The three such lines in Figure 11.9 represent the three two-part tariffs that we have referred to. (Notice that no consumer would wish to

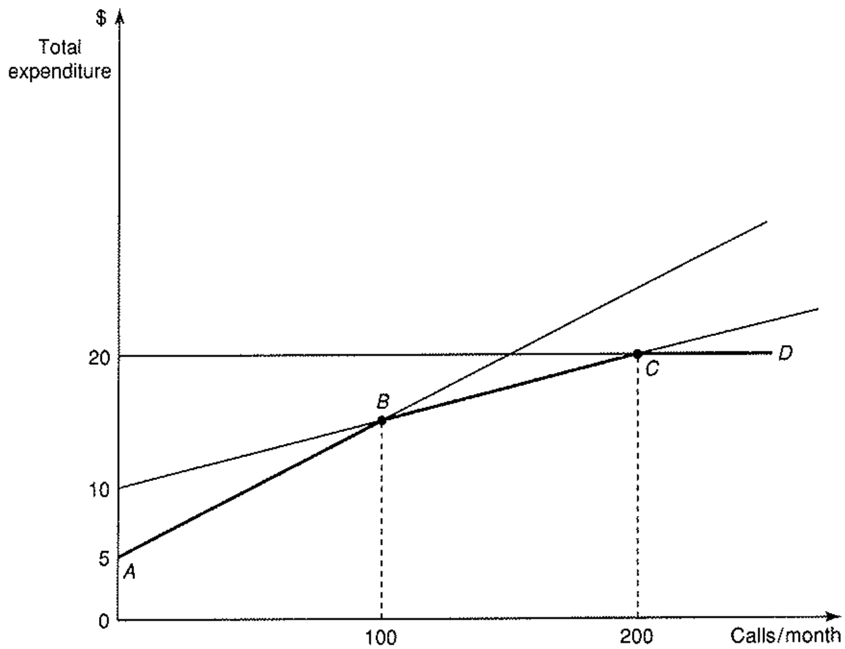


Figure 11.9
Multipart Tariff for Local Telephone Service

consume on portions of the tariffs other than the lower boundary $ABCD$. Hence it does not matter that these "dominated" portions of the two-part tariffs are not actually part of the declining-block tariff.) The point is that the declining-block tariff has the same effect as confronting consumers with two-part tariffs that are tailored to their demands. And, of course, all consumers are free to choose the particular tariff that they prefer, so that there is no discrimination involved that is likely to be disallowed.

Up to this point our discussion of ideal pricing has been limited to a single-product natural monopolist. We now turn to the case of a multiple-product natural monopolist and describe what has become known as Ramsey pricing.

Ramsey Pricing

In a famous article published in 1927, Frank Ramsey suggested the following pricing (and taxing) method.[14] It is applicable to a multiple-product natural monopolist that would generate losses if linear marginal cost pricing were used. In essence, Ramsey prices are those linear

14. Frank Ramsey, "A Contribution to the Theory of Taxation," *Economic Journal*, March 1927.

prices that satisfy the total-revenues equal-total cost constraint and minimize the deadweight welfare losses. Note that Ramsey prices are linear prices—one for each product—so that we are implicitly ruling out multipart tariffs.

It is useful to illustrate Ramsey pricing with a numerical example. Let the natural monopoly be a two-product firm with total cost

$$C = 1800 + 20X + 20Y.$$

The market demands for the two goods X and Y are given by

$$X = 100 - P_x$$
$$Y = 120 - 2P_y.$$

An important assumption that we will make for our example is that the demands are independent—the demand for X does not depend on the price of Y, and vice versa. The more general case of interdependent demands involves much more complex mathematics and is beyond the scope of the discussion here.[15]

It should be obvious that the marginal costs of X and Y are each \$20, and that marginal cost prices would exactly cover the variable costs but not the fixed cost of \$1,800. Because the firm must cover its total costs, it is clear that the prices will necessarily exceed their respective marginal costs. One possibility would be to raise the prices by the same proportion above marginal costs until total costs are covered. This is shown in Figure 11.10a.

The figure shows that prices would need to be raised from \$20 to \$36.1 to generate sufficient revenues just to cover total costs.[16] In particular, the contribution that product Y makes toward fixed cost equals the rectangle $CEFD$. This is just price minus the constant unit variable cost of \$20, multiplied by the output of 47.7. Similarly, the contribution that product X makes equals rectangle $CEKJ$. The sum of these two rectangles is \$1,800. (The fact that the demands intersect at the price equals marginal cost point for each is not necessary, and was chosen merely to make the graphical exposition simpler.)

Now consider the deadweight losses that this proportionate price increase method causes. The deadweight loss triangle for product Y is triangle DFH, and it is JKH for product X. The actual numerical values are \$260 and \$130, respectively, or a total of \$390. Hence, one way of summing up this method is to observe that it "costs" \$390 in deadweight welfare losses to generate the \$1,800 necessary for the firm to break even. The question becomes whether one can find another method for raising prices to generate the \$1,800 that entails a lower welfare cost.

15. The interested reader should consult Brown and Sibley, *The Theory of Public Utility Pricing*, p. 42.

16. Because P_x and P_y must be equal under the assumption that the marginal costs are both \$20, the \$36.1 value can be found by solving the equation that equates total revenues and total costs.

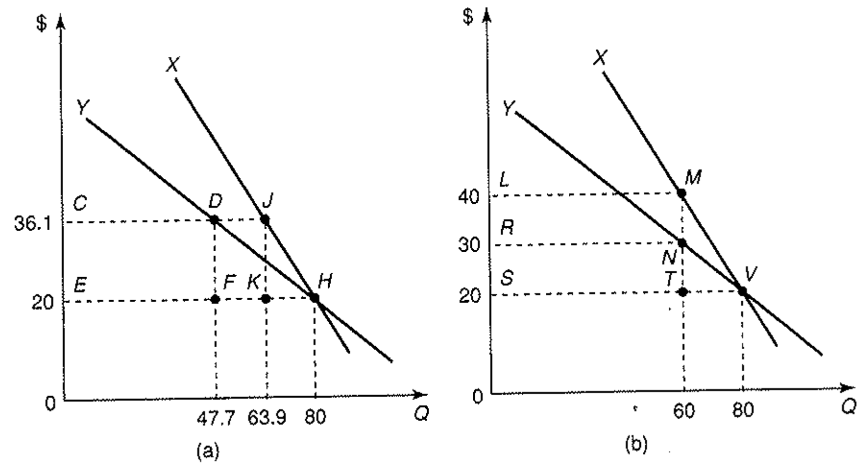


Figure 11.10
Proportionate Price Increase versus Ramsey Pricing

A bit of reflection while examining Figure 11.10a might suggest differential price increases. That is, it is clear that the same price increase produces a smaller contribution to fixed cost from product Y at a higher cost in terms of deadweight loss. This observation is not surprising when one realizes that product X has a more inelastic demand (at point H) than does product Y. This difference suggests that it would be better to raise the price of X more than the price of Y.

The Ramsey pricing "rule" that gives the prices that minimize the deadweight losses is to raise prices in inverse proportion to demand elasticities. Mathematically, the rule[17] is

$$\frac{P_i - MC_i}{P_i} = \frac{\lambda}{\eta_i}$$

where P_i is the price of good i, MC_i is the marginal cost of i, η_i is the absolute value of the elasticity of demand of good i, and λ is a constant. Using this rule, one can derive the actual Ramsey prices.[18] They are shown in Figure 11.10b. Hence the firm would minimize the welfare losses by charging \$40 for good X and \$30 for good Y. At these prices, the

17. See Brown and Sibley, 1986, p. 39, for a formal derivation.

18. Computations are made simpler by using the alternative rule for Ramsey prices that will be given shortly involving proportionate quantity changes. That rule implies that the two products will have equal outputs. Hence this fact together with the total-revenues-equal-total-costs equation yields the Ramsey prices.

demand elasticities are 0.67 and 1.0, respectively. The deadweight loss triangles are \$200 for good X (triangle MTV) and \$100 for good Y (triangle NTV) for a total of \$300. This is, of course, a lower "cost" in terms of welfare by \$97 than the proportionate method of Figure 11.10a.

Another interesting fact about Ramsey prices is apparent in Figure 11.10b. The proportionate decrease in output from the price-equals-marginal-cost output (outputs of 80 for both) is the same for the two goods. That is, both outputs are cut by (80–60)/80, or 25 percent. This is an alternative way of describing Ramsey pricing: cut output of all goods by the same proportion until total revenue just equals total cost. This way of stating the rule for Ramsey pricing is more general than the inverse elasticity rule, and holds true for the case of interdependent demands.

The Ramsey pricing rule can be viewed as providing theoretical justification for so-called *value of service* pricing that has been used for years in the railroad industry. It has been common for rail rates for shipping gravel, sand, potatoes, oranges, and grapefruits to be lower relative to shipping costs than for liquor, electronic equipment, cigarettes, and the like. The reason is that the elasticities of demand for shipping products that have low values per pound are higher than for products that have relatively high values per pound. (We are assuming that the actual costs of shipping are proportional to weight.)

In summary, all of the ideal pricing schemes discussed have problems (except for the two-part tariff with price equal to marginal cost and no exclusion of consumers by the fixed fee). It should be kept in mind that we have assumed away the very real difficulty of designing incentive systems that will induce enterprise managers to implement these pricing schemes. In short, managers of private firms are presumably interested in maximizing profits, not total economic surplus. Managers of public enterprises may also have objectives other than economic efficiency. Economists have recently begun to explore theoretical models of how regulatory agencies might provide incentives for natural monopolies to price efficiently. We will briefly describe the Loeb-Magat proposal in the next subsection.

Loeb-Magat Proposal

Of course, if regulators had perfect information as to the monopolist's costs and demands, the ideal pricing schemes that we have discussed could be put into effect by command. However, such is not the case. Although the monopolist may not have perfect information itself, most people would probably agree that the monopolist has much better knowledge of its costs than the regulators do. Because the firm's profits will increase with higher prices, the firm has an incentive to overstate its costs (which is the usual basis that a regulator uses to set prices).

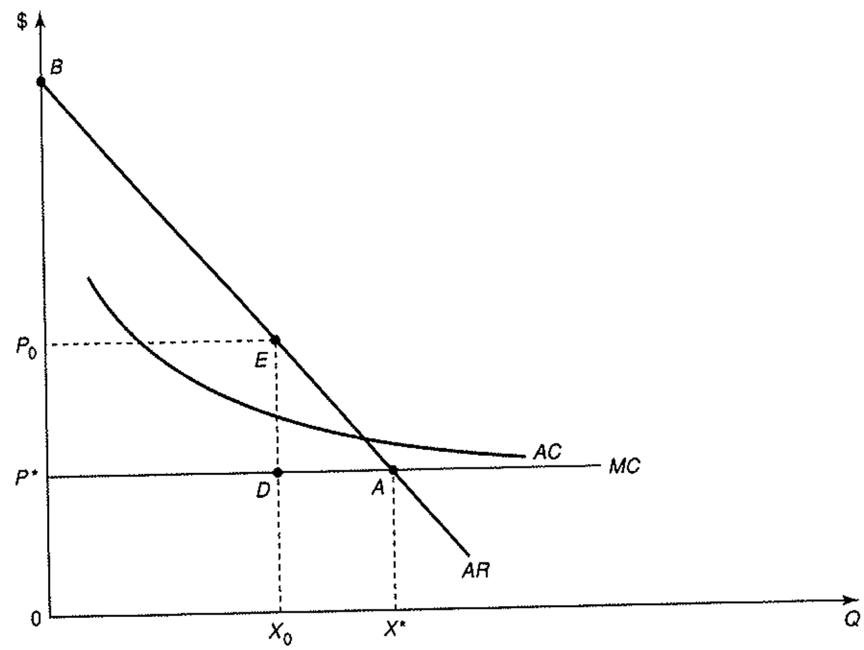


Figure 11.11
Loeb and Magat Incentive Scheme

Loeb and Magat (L-M) assumed that the monopolist knows costs and demand information perfectly, but that the regulator knows demand only.[19] Hence, given this asymmetry of information and the assumption that the monopolist's objective is to maximize profit, what might the agency do to induce efficient pricing? The L-M scheme can be explained easily with the aid of Figure 11.11, which shows a single-product natural monopolist.

The monopolist has declining average cost (AC) and demand curve (AR). For simplicity, we assume the total cost function is $K + vX$; hence, marginal cost (MC) is constant and equal to v. The L-M proposal is to allow the monopolist to choose its own price—this differs

19. Martin Loeb and Wesley Magat, "A Decentralized Method for Utility Regulation," *Journal of Law and Economics*, 1969. Some additional research on this same issue can be found in Ingo Vogelsang and Jorg Finsinger, "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," *Bell Journal of Economics*, 1979; D. P. Baron and R. B. Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 1982; D. Sappington, "Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities," *Bell Journal of Economics*, 1983. A comprehensive though difficult recent survey is D. P. Baron, "Design of Regulatory Mechanisms and Institutions," in R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*.

from the usual practice of the regulatory agency setting the price. However, they propose to have the agency subsidize the firm by an amount equal to consumer surplus at the selected price.

Suppose that the monopolist selects the price P_0. Its profits will be $P^*DEB - K$. The firm collects $0X_0EP_0$ from customers and P_0EB from the regulatory agency. Its variable cost is $0X_0DP^*$, leaving a variable profit of P^*DEB. Subtracting the fixed cost of K leaves the profit just asserted. Observe, however, that the firm can do better by lowering price. For example, if the monopolist selected P^*, it is easy to show that its profits will increase to $P^*AB - K$. That is, profits increase by the usual deadweight loss triangle DAE. This is, in fact, the profit-maximizing solution for the monopolist! Convince yourself that any other price will reduce profits. (Alternatively, note that the proposal causes the demand curve AR to become the monopolist's marginal revenue curve, and setting MC equal to marginal revenue is the profit-maximizing solution.)

The explanation for this price-equal-to-marginal-cost result is simply that the regulator has changed the firm's objective function by the subsidy. Now, in effect, the monopolist is maximizing total surplus—the total area under the demand curve minus costs.

The solution is economically efficient, but most people would find it objectionable on distributional grounds. The monopolist is appropriating the total economic surplus! To rectify this problem L-M suggest that a franchise bidding scheme (or a tax scheme) could recover some of the subsidy for the general treasury. In the case shown in Figure 11.11, the regulatory agency would auction off the right to operate the monopoly franchise. The key idea is that above-normal returns (of amount $P^*AB - K$) are available to the firm that operates the monopoly and that bidding for the franchise would continue until that amount is bid. Note that the subsidy is not completely recovered—there remains a net subsidy of an amount equal to fixed cost, K.[20]

Obviously, the L-M proposal is not the perfect solution to natural monopoly. Informational problems about the demand curve and the existence of a subsidy make it an unlikely substitute for the present regulatory process. It has, however, stimulated research by economists toward the goal of understanding how the regulatory process might be improved with respect to providing better incentive structures for natural monopolists.

In the next section we return to the discussion of alternative policy solutions to the natural monopoly problem. In contrast to the ideal pricing solutions that we have been examining heretofore, we now turn to actual solutions that have been used. The first is franchise bidding.

20. For a variation on the Loeb and Magat proposal that eliminates the net subsidy and the need of the regulator to know demand, see D. A. Graham and J. M. Vernon, "A Note on Decentralized Natural Monopoly Regulation," *Southern Economic Journal*, July 1991.

Franchise Bidding

Harold Demsetz has argued that the "theory of natural monopoly is deficient for it fails to reveal the logical steps that carry it from scale economies in production to monopoly price in the market place."[21] His point is that it may be possible to have bidding for the right to supply the entire demand (in effect, bidding for a franchise to serve a certain market). Even though only the single firm submitting the low bid would actually produce, there could be competition among potential suppliers. For example, given the situation shown in Figure 11.8, the low bid presumably would be a price of P_0 for Q_0 units.

Note that P_0 is not the efficient price. Nevertheless, P_0 would be an improvement over the natural monopoly price (a price above P_0). Then P_0 would be the lowest price bid for the right to supply the market inasmuch as any lower price would result in losses. At P_0 the winning bidder would just cover costs, including a normal return on investment.

This bidding for the franchise argument has stimulated a great deal of useful thinking about alternatives to natural monopoly regulation. However, the highly abstract example here oversimplifies many of the problems that such bidding would raise. A detailed discussion will be provided in Chapter 13.

Actual Solutions

In this section we briefly consider actual solutions that have been implemented in response to the natural monopoly problem. There are basically two distinct solutions: the regulatory agency and public enterprise. Extensive discussions of each will be presented in subsequent chapters; only a short treatment is given here.

Regulation

The typical natural monopoly in the United States is a private firm: Consolidated Edison, Bell Atlantic, and so on. The firm is controlled by a regulatory agency that must approve the prices the monopolist can charge. A key goal is that the firm's revenues just cover its costs.

The measurement of costs is obviously a major task for the agency. Indeed, the attempt by the agency to estimate the proper return on capital investment is perhaps its most time-consuming activity. For example, a typical regulatory hearing involves testimony by numerous experts as to the "true" cost of capital for the firm.

In contrast, relatively little of the agency's resources are expended on the issue of the correct pricing structure. However, this situation is changing and agencies are becoming more interested in, for example, marginal cost pricing. In short, regulatory agencies try very hard to

21. Harold Demsetz, "Why Regulate Utilities?" *Journal of Law and Economics*, April 1968.

ensure that the monopolist's revenues equal its costs, and historically have been less concerned with the pricing structure used.

As a result, there is no simple way to describe the pricing structures used under regulation. Price discrimination is often employed both across customer groups (industrial, commercial, residential, and so on) and within groups (declining block rates, for instance, 5 cents per unit for the first 300 units, 4 cents per unit for the next 500 units, and so on).

Richard Schmalensee has observed,

> To the extent that utility regulators in the United States have been concerned with rate structures, they have tended to focus on prices paid by different classes of users. But this focus has typically been motivated and informed by considerations of equity or fairness rather than efficiency.[22]

Hence, regulatory agencies often try to prohibit undue discrimination across customer groups. They require the firm to allocate its total costs to customer groups and then adjust their prices if the revenues by groups do not correspond to the groups' "fully distributed costs."

There is a serious problem implicit in this procedure, however, because a large proportion of a firm's costs are usually common costs. For example, high-voltage power lines are used in common by all customer groups. And although arbitrary accounting rules can be made up to apportion these costs among groups (for instance, in proportion to their respective annual purchases of the product), none are meaningful in an economic sense as a basis for setting prices.

In summary, an important solution to natural monopoly in the United States is regulation. The regulatory solution is not an attempt to implement the ideal pricing schemes discussed earlier. Regulators do not see as their primary objective achieving economic efficiency. Rather, they appear to seek a set of prices that are not unduly discriminatory but that permit total revenues to cover total costs. However, regulatory agencies have become more interested in pricing schemes that promote economic efficiency. For example, peak pricing—charging more when demand presses on capacity, and, therefore, marginal cost is higher—is being implemented by electric utilities in various parts of the country.

Public Enterprise

The second actual solution to natural monopoly is public enterprise, or government ownership and operation of the monopoly. This is not as common in the United States as it is in other countries. The Postal Service is an example in the United States. Other examples include various government-owned electric utilities (for instance, the Tennessee Valley Authority) and Amtrak, the government-owned passenger service railroad.

---

22. Richard Schmalensee, *The Control of Natural Monopolies* (Lexington, Mass.: Lexington Books, 1979).

In principle, public enterprise would appear to be a sensible alternative. Managers would be directed to maximize economic surplus—there would be no need for regulators to try to channel the decisions of profit-maximizing firms closer to the public interest. The efficacy of public enterprise as compared to regulation, however, is a complex issue and will be examined further in Chapter 14.

Summary

This chapter has been an introduction to natural monopoly. Theoretical issues have been introduced and discussed. First, the definition of natural monopoly was developed in both the single-product and the multiple-product cases. Second, alternative policy solutions and their difficulties were discussed. The solutions included "doing nothing," various efficient pricing solutions, competition among bidders for the right to the monopoly franchise, actual regulation, and public enterprise.

In the next chapter we will elaborate extensively on the regulation alternative. Chapter 15 will examine further issues in natural monopoly regulation, with an emphasis on telecommunications.

Questions and Problems

1. Consider a single-product natural monopoly situation with the usual U-shaped long-run average cost curve. Is the range of output over which natural monopoly holds from zero to the output corresponding to minimum average cost? If not, explain how to determine the appropriate range. Use the total cost function $C(q) = 1 + q^2$ to answer this question.
2. Assume a natural monopoly with total costs $C = 500 + 20Q$. Market demand is $Q = 100 - P$.

   a. If price is set at marginal cost, what is the monopolist's profit?

   b. The answer to part a implies that linear (or uniform) marginal cost pricing has a serious problem in natural monopoly situations. Suppose that average cost pricing is employed. Find price, output, and the deadweight loss compared to part a.

   c. Now consider two-part pricing—a type of nonlinear (or nonuniform) pricing. Each consumer must pay a fixed fee regardless of consumption level plus a price per unit. Assume that the market consists of ten consumers with identical demand curves for the product. If the price is set equal to marginal cost, what is the largest fixed fee that a consumer would pay for the right to buy at that price? What fixed fee would permit the monopolist to break even? What is the deadweight loss in this case?
3. Assume the same facts as in question 2 but that now there are six "rich" consumers with each having inverse demands: $p = 100 - 6.3q$; also, there are four "poor" consumers each with demands: $p = 100 - 80q$.

a. What is the largest fixed fee that a poor consumer would pay for the right to buy at marginal cost?

b. Because the poor consumers would not be willing to pay the uniform fixed fee of \$50 necessary for the monopolist to break even, the rich consumers would have to pay a fixed fee of \$83.33. What is the deadweight loss in this case?

c. Third-degree price discrimination could be a solution. That is, if it is legal, resales are not feasible, and consumers could be identified by the monopolist as being rich or poor, the monopolist could charge different fixed fees to the two consumer types. If the price per unit is still equal to marginal cost, what are two fixed fees that are feasible? In this case, what is the deadweight loss?

4. If third-degree price discrimination is not a feasible alternative in question 3c, consider the optimal two-part tariff. That is, what is desired is the two-part tariff that minimizes deadweight loss—or that maximizes total surplus. One way to think about it is to imagine the case of a zero fixed fee and price equal to marginal cost. This causes a loss of \$500 that must be covered. Imagine raising both the fixed fee and the price simultaneously—both can cause losses: the fee by excluding poor consumers and the price by causing deadweight consumption losses. One possibility is to exclude poor consumers and go to solution 3b. The other possibility is to keep all consumers in the market; this implies that the fixed fee should equal the consumer surplus of a poor consumer. It is optimal to take all of the poor consumers' surpluses as a fee. To see why, consider the opposite case where the poor have some excess of surplus over the fee. Then the price could be lowered, reducing deadweight losses and the surplus could be used to offset the reduction in revenues without excluding the poor from the market.

a. Find the sum of consumer and producer surplus minus the \$500 fixed cost (that is, find total surplus) for case 3b where the poor are excluded.

b. Find total surplus for the case of all consumers retained in the market. Hint: An equation in P can be defined that equates to \$500 the total contributions to fixed cost (10 times the fixed fee, equal to the consumer surplus of a poor consumer, plus the revenues net of variable cost generated by consumption). Hence, what is the optimal two-part tariff where all are retained in the market?

c. Compare the efficiency of the tariffs in parts a and b.

5. A multipart tariff can be superior to the optimal two-part tariff found in question 4. A multipart tariff involves a fixed fee plus multiple prices per unit, which depend upon predefined blocks of consumption.

a. Show that by making an additional two-part tariff available to the consumers that they can use at their option, the "two" two-part tariffs are Pareto superior to the optimal tariff in question 4 (that is, $F = \$38.55$, $P = \$21.50$). Let the optional two-part tariff be $P = \$20.50$ and $F = \$51$. These two two-part tariffs are equivalent to a multipart tariff that has a fixed fee of \$38.55 and a price of \$21.50 for the first 12.4 units and a price of \$20.50 for all units above 12.4. Show this result by plotting the two tariffs on a graph that has total expenditure on the vertical axis and total units on the horizontal axis. The two straight lines representing the tariffs intersect at 12.4 units. Because consumers will always operate on the lowest line that they can attain to minimize expenditure, the multipart tariff is just the lower boundary (that is, the kinked line defined by $F = \$38.55$ and the marginal prices of \$21.50 for the first 12.4 units and \$20.50 thereafter).

b. Demonstrate that the two two-part tariffs are Pareto superior to the optimal two-part tariff in question 4b. Note that the optional tariff will not change the poor consumers' behavior at all. Why?

c. As a result we can focus solely on the rich consumers and the monopoly. If both are made better off by the optional tariff and the poor are kept the same, then the optional tariff results in a Pareto improvement—which is a stronger welfare statement than simply saying one tariff yields a higher total surplus. (That is, if we focus on total surplus comparisons, we ignore the fact that some people may be made worse off even though total surplus is higher.) Find the consumer surplus of a rich consumer under the two-part tariff of question 4b.

d. Find the consumer surplus of a rich consumer under the multipart tariff.

e. Find the change in profit of the monopolist. Hence a movement from two-part tariffs to multipart tariffs clearly has the potential for gains in efficiency. The intuition is that the more the "parts," the better the tariff can be tailored to the differences in willingness to pay across consumers.

6. Assume that a water distribution monopoly serves two consumer types: industrial and residential. The demands by the two classes are as follows. Industrial: $Q_I = 30 - P_I$ and Residential: $Q_R = 24 - P_R$. The company has no costs other than the fixed cost of the pipeline, which is \$328. Find the Ramsey prices. Hint: See note 18.

7. Assume a natural monopoly with total cost $500 + 20Q$ facing a demand of $Q = 100 - P$.

a. Find the price that enables the monopolist to break even. (This is the same problem as 2b.) Call this price P^*.

b. Loeb and Magat show that if the monopolist is allowed to choose its own price and to have the regulatory agency subsidize the firm by an amount equal to consumer surplus at the selected price, the monopoly will select price equal to marginal cost. What is the price and amount of government subsidy?

c. Loeb and Magat also note that a bidding process for the monopoly franchise would enable the government to recover some of the subsidy. What is the amount recovered and what is the net subsidy after bidding?

d. An alternate proposal would make use of two-part tariffs. For example, assume that the current regulated price is P^*. Now assume that the regulatory agency offers the firm the right to select any two-part tariff that it wishes as long as the consumer continues to have the option of buying at P^*. (For simplicity, assume a single consumer.) What is the two-part tariff that the monopolist will choose and what is its profit? What is the deadweight loss?

e. Assume that the government uses a bidding process to eliminate the monopoly profit in part d. The bid is in the form of a single price, like P^*, that the consumer will always have as an option to the two-part tariff. That is, the same rules are in effect as in part d except that now the bidding is for the right to offer a two-part tariff optional to some P^* that the bidding will determine. What is the low bid?

f. Compare the Loeb and Magat proposal in part c with the proposal in part e. Do both proposals give efficient prices? Are there any substantive differences?

Board Staff Interrogatory # 3

Ref: #1, p. 5

Has AMPCO received any indication from Hydro One that the highest hour of each of the 5 highest peak days of demand in Ontario is the most appropriate number of hours and days to reflect cost causation, for either the whole network or Hydro One's predominant share of the network? Conversely, has AMPCO received any indication that some other number of hours, days, or another combination would be more appropriate for that purpose?

Response

AMPCO has received no constructive information from Hydro One regarding any change in the network charge determinant. Our advocacy of the 5CP was initially informed by precedent in the Pennsylvania-New Jersey-Maryland definition of capacity obligations for direct customers and load-serving entities and has been reinforced through discussions with our members regarding practical strategies for demand response in relation to observed inflection points (or regions) in price duration curves for Ontario.

Board Staff Interrogatory # 4

Ref: #1, p. 5, and Exhibit H1 / Tab 3 / Schedule 1 / p. 5

a. Please confirm that the numerator in equation 1 should read June 2012. Alternatively, please explain the relevance of June 2010.

b. Is the purpose of Equation 1 to clarify Hydro One's formula with respect to an example with dates, or is it to correct the formula by removing one of the terms in the equation?

Response

a) June 2010 in the numerator is a typographical error. The term should read June 2011.

b) We are unclear what formula is referred to as "Hydro One's formula". Our intent is simply to show how the network charge determinant might be calculated if the Board were to decide to change the design to that proposed by AMPCO, implemented for a 12 month period commencing in July of each year, based on a customer's coincident critical peak demand in the prior 12 month period.

Board Staff Interrogatory # 5

Ref: # 1, p. 9

AMPCO's evidence shows that industrial customers in at least some market sectors have shifted their loads toward off-peak periods in response to the hourly price structure of the electricity commodity.

a. Does AMPCO have information on the load-carrying capability of the Ontario network as it existed prior to the introduction of the commodity market, and as a result does AMPCO have information on what proportion of the existing network capability was planned or placed in service prior to the load shifting that has been done by industrial customers?

b. Does AMPCO consider that its members should have some responsibility for the Network revenue requirement associated with capacity that may be under-utilized as a result of load shifting by those customers?

Response

a) No, although the Power Advisory report cites research conducted for Ontario Hydro and published in 1993 that found evidence of load-shifting by industrial customers then.

b) AMPCO's members are a sub-set of industrial customers. All customers and classes of customers are able, indeed encouraged by policy and regulation, to shift consumption from peak to off-peak periods. While this policy is not new, the implementation of time-of-use rates for regulated price plan customers is an obvious and recent example. We see no possible justification for levying additional charges on our members to recover costs the causation of which they make no contribution.

Board Staff Interrogatory # 6

Ref: #1, p. 13

Does AMPCO shows in Figure 2 that line losses are a non-linear function of total load, and makes the additional point that the higher cost of energy when load is high augments the cost of losses at peak times.

a. AMPCO recommend that Hydro One implement a loss factor for transmission that would be a separate component of the tariff?

b. If so, does AMPCO recommend that the loss factor should vary by time-of-use, or in a real time manner responsive to total load, in order to reflect the non-linear function shown in Figure 2?

Response

a) No. The IESO collects hourly uplifts from transmission customers (wholesale market participants) to recover the costs associated with losses. This is described by the IESO:

   "Wholesale Market Services Charge – This charge includes the cost to operate the wholesale electricity system, administer the electricity market, and maintain the reliability of the provincial grid. These rates are set by the OEB and include the Wholesale Market Service Charges.

   These costs include:

   Physical Limitations and Losses: These are losses that occur as electricity flows across transmission lines. The IESO also collects other costs incurred in operating the power grid, such as when it must take actions to avoid overloads on the transmission system in cases of surges in demand."

More information is available at:
http://www.ieso.ca/imoweb/siteshared/electricity_charges.asp?sid=bi.

b) While it is not necessary for Hydro One to implement charges to recover the costs of transmission losses from transmission customers (as we point out in answer (a) above), the question raises a valid issue with respect to the recovery of transmission losses and distribution losses from distribution customers. The current practice, to recover these costs via an annual average loss factor for all customers within each distribution franchise area, is an obvious opportunity for significant improvement in rate design, to promote efficiency and efficient demand management, to apportion costs in a way that is more just and reasonable among customers and customer classes, and to create opportunities for distribution companies to be accountable for, and have incentives to reduce, the prevalence and cost of losses attributable to their customers (both within the distribution franchise and on the transmission system). AMPCO looks forward to an opportunity to take this issue up with the Board and other parties at the Board's earliest convenience.

Board Staff Interrogatory # 7

Ref # 2, p. 9

The IESO provided hourly demand data by industry sector, and Hourly Ontario Energy Price (“HOEP”) is publicly available.

a. Please confirm that adequate data was available to enable an analysis of load shifting from a daily peak period of say, four or six hours daily, into an off-peak period or a shoulder period elsewhere in the day.

b. Please confirm that the analysis of elasticities in this paper is of demand during a twelve-hour peak period and a twelve-hour off-peak period.

c. If the previous statements are confirmed, and since this paper is apparently submitted in support of the AMPCO’s High 5 proposal, why is the analysis not designed to estimate load shifting out of a shorter peak period?

Response

a) Yes. Adequate data are available to enable analysis of shifting between 4-6 hour periods.

b) Correct.

c) The analysis could be done for shorter time periods. However, we think that conducting such analysis would only result in even larger price elasticities as it takes into account not only shifting between peak and off peak periods but within such time periods. Previous studies such as Boisvert et al. (2004) find relatively large elasticities of substitution when peak hours are of short duration. This is consistent with the idea that firms then have more hours and therefore flexibility, in order to compensate for reductions in output during peak hours. We view our approach to be more conservative as we are only taking into account shifting across periods.

Board Staff Interrogatory # 8

Ref: # 2, pp. 7-8

Please state whether any of the other analyses of demand elasticities cited in the paper provide information on load shifting from a short peak period. (Please include only those that concern industrial customers' demand. Include all studies in which the peak periods are shorter than the Ontario uniform network charge.) If possible, please state whether the hours of the peak period were fixed, or alternatively were determined in a responsive manner, for example based on system cost or load.

a. Please provide a copy of the unpublished document "Industrial and Commercial Customer Response to Real Time Electricity Price", Boisvert et al, 2004, if possible.

b. Please describe the extent of response by industrial customers to the highest prices amongst the real time prices in the study by Boisvert et al. In particular, please state whether the response found by Boisvert et al is greater than found in Dr. Sen's study which is based on twelve-hour fixed time intervals.

Response:

a) Boisvert et al. (2004) find elasticities of substitution from 0.10 to 0.27. They also find that price responses are the highest for high prices of short duration high, ranging from 0.20 to 0.27. Although not directly comparable, they do seem larger in magnitude than the results in Sen (2010). Further, elasticities fall significantly as peak hours become longer in duration. Boisvert et al. (2007) is another good example of recent research on the elasticity of substitution by large customers. The highest elasticity of substitution is for firms for the manufacturing sector with a value of 0.16. However, this is for a short time period (2-5 pm). Specifically, they define off peak and on peak periods empirically through the actual load. They calculate elasticities of substitution for the 12 pm – 5 pm, 1pm – 5 pm, and 2pm – 5 pm time periods and obtain the highest elasticities for the 2 pm – 5 pm period.

b) A copy of Boisvert et al. (2004) is available at http://eetd.lbl.gov/ea/EMS/reports/Boisvert.pdf.

c) As noted above, the high end elasticities of substitution found by Boisvert et al. (2004) seem larger than the elasticities obtained by Sen (2010). However, it should be noted that the elasticities are not directly comparable as Dr. Sen did not calculate the actual elasticity of substitution because of the lack of individual customer data.

Board Staff Interrogatory # 9

Ref: #2, and Exhibit I / Tab 4 / Schedule 67

Power Advisory stated, in response to VECC interrogatory # 67(c), that it is reasonable to expect that the elasticity of substitution between peak and off-peak is greater with a shorter definition of the peak period.

a. Does Dr. Sen agree with Power Advisory's statement?

b. Does Dr. Sen agree that the elasticities derived in ref # 2, and/or the other studies cited, are likely lower than the elasticities that would be found if the peak period were defined as a narrower period?

Response

a) Yes. As stated in our response above (in 7(c)), conducting research based on shorter time periods should result in larger price elasticities as it takes into account not only shifting between peak and off peak periods but within such time periods.

b) Yes. Please see response above.

Board Staff Interrogatory # 10

Ref: # 2, pp.10-14, and Ref # 3, pp. 39-40

Commenting on Dr. Sen's previous analysis, cited in this study as Sen (2009), Power Advisory stated that the "estimated coefficients are not robust under different estimation time frames" (p. 39), and go on to summarize results using two definitions of the off-peak price that differ from each other (p. 40).

a. Does the model specification in this study include any modifications to improve the robustness of the coefficients, in particular with respect to the time frame of peak and off-peak definitions, relative to the results of the earlier study that Power Advisory was commenting on?

b. If so, please describe the modification(s) that have been made.

Response

a) Yes.

b) We have used data for all hours of the day, which we did not do before. Further, as noted in our submission, we employ additional data from 2008 as well as new information on total industrial demand and demand by electricity generators, distributors, and transmitters. The empirical estimates have also been redone using Feasible Generalized Least Squares (FGLS) which account for first order autocorrelation and unknown heteroskedasticity. We also evaluate the sensitivity of our findings through the use of Instrumental Variables intended at correcting for measurement error and pooling the data across all years of our sample. Finally, more right-hand side controls are added (monthly unemployment rates, the daily exchange rate, and dummy variables for weekends and holidays) to capture the effects of other potential determinants of industrial electricity consumption.

Board Staff Interrogatory # 11

Ref: #2, p. 43

One of the industry sectors analyzed by Dr. Sen is electric power generation, transmission, and distribution excluding LDCs. The coefficients for the price variables are found in column G, and are larger for this sector than for any of the other six sectors.

a. Please provide a description of what the electricity is used for in this sector, if available

b. If the demand of the sector includes use within the generating stations, pumped storage, line losses, and use within transformer and distribution stations, is it reasonable to expect that the electricity demand for any of these uses would be responsive to peak and off-peak prices in a pattern similar to the other sectors? Are there other uses in this sector that would be expected to be sensitive to peak and off-peak prices?

Response

a) We have no detailed information on the components of demand by electric power generation, transmission and distribution, excluding LDCs, although we understand that generation station service is a significant component.

b) Line losses are not included as discussed in our response in 6(a) above.

   Pumped storage would be included. We would expect that the operation of pumped storage facilities should be highly sensitive to changes in peak and off-peak prices and rates, since the value of pumped storage depends directly on the difference between them.

   While we have no detailed information, we understand that the nature and extent of demand for generation station service varies from station to station and is a function both of how the generator operates in the market (i.e., whether it provides base-load, mid-peak or peaking services) and of how the generator connection to the grid is configured. For example, a generator connection could be configured so that station service is provided directly by the generator itself, i.e., before power reaches the grid, or, alternatively, so that all output is transmitted to the grid and station service demand is subsequently withdrawn. In the latter case, changes in the design of rates should be expected to influence the economics of possible changes in the configuration of the generator grid connection. A further issue, however, confounds such a simple analysis, and that is the extent to which generator contracts or regulations provide any incentive to a generator to reduce its demand for station service.

Board Staff Interrogatory # 12

Ref: # 2, p. 15

a. Please explain the rationale for including the Herfindahl-Hirschman Index as a variable in the regression analysis. What is the expected sign of the coefficient?

b. How frequently is the Herfindahl-Hirschman Index recalculated for use in this analysis: hourly, monthly, annually, other?

Response

a) The rationale for including the Herfindahl Hirschman Index is to capture the effects of market power among large wholesale suppliers.

b) The Index is calculated at the hourly level.