Rate-of-Return Regulation: Protecting Whom from What?

Nina W. Cornell

WHEN SHOULD AN industry be subject to classical public utility regulation—that is, the setting of an allowable rate of return, with controls on price and entry? The answer is, "Never." And the reason is plain: this form of regulation, widely viewed as protecting the public from abuse of monopoly power, in fact never has done so, never could, and never will.

Yet the perception lags far behind the reality of failure. Interstate telephone, natural gas, some electrical power, and some railroad shipments still are subject to federal public utility regulation. In some states or at the local level, these same industries plus such others as water, cable television, and taxi companies also are subject to public utility regulation. And even as some industries—railroads, interstate telephone, natural gas, for example—are being partly or wholly freed from such regulation, there are calls for its wider application to new spheres, particularly to all or part of the oil industry and to cable television. So it continues to be relevant to examine why public utility regulation not only fails to protect consumers from abuse of monopoly power but has quite the opposite effect: it tends simply to protect monopolies from competition.

Consider, for example, the case of AT&T, which is generally perceived to be a *natural monopoly.* (A natural monopoly is an industry in which one firm is able to produce all the output demanded—either a single output or a bundle of related outputs—at less cost than if several firms produced the same quantity of output.) AT&T has long been subject to public utility regulation: a rate base—the value of its plant and equipment—has been determined and an allowable rate of return set by the Federal Communications Commission (FCC). AT&T is required to file tariffs and receive certification of public interest, convenience, and necessity for new investments. These or similar requirements are found in the Communications, Interstate Commerce, and Civil Aeronautics acts, among others.

Monopoly exists and is likely to be abused, and . . . society is right in trying to curb that abuse. But public utility regulation is not the way.

Now, obviously, I am not the first to point out the failings of public utility regulation. If

Nina W. Cornell is chief, Office of Plans and Policy, Federal Communications Commission. The views expressed are hers and do not necessarily reflect those of the commission.

ever there is to be an end to it, however, more must be said than simply that its costs outweigh its benefits. Monopoly exists and is likely to be abused, and I believe that society is right in trying to curb that abuse. But public utility regulation is not the way.

Regulatory Setting and Role

Why does public utility regulation fail? Let us look more closely at how the model is supposed to work. If a firm is a natural monopoly, it has no competitors and thus faces no pressure to charge customers lowest cost for its output. Indeed, the railroads' practice in the late nineteenth century of charging higher per-mile rates for short hauls than for long hauls (where there were competing railroad routes) led to the first major federal public utility regulation. Then, too, the more that customers view the output as indispensable—something they must have even if it is very expensive, such as a telephone—the higher the price the monopolist can charge. These characteristics are stated here in overly simple form and not as they arise in individual cases; but they fit the general notion most people have of the type of industry that ought to be regulated.

Rate-of-return regulation is supposed to ensure that public utility prices do not exceed costs. As noted already, the process involves setting an allowable rate of return on a predetermined investment base. Additionally, regulatory agencies impose entry and price controls.

Entry controls are supposed to serve two purposes. One is to prevent the monopoly from overinvesting and then charging the ratepayers for that overage. Thus, for example, interstate telephone companies must get permission before adding to their transmission facilities, and the additional facilities have to be justified by service and demand projections. Entry controls also are justified as a way to ensure that the monopolist (or just a few firms) provide services to everyone equally, if necessary by shielding high profits from one service to provide funds for the more costly ones. Much is made, for example, of the need to block entry into trucking in order to maintain service to small towns by existing firms. More recently, regulatory bodies have been urged to limit entry into the telephone industry so that the monopolist can take advantage of economies-of-scope cost savings that may result from joint production of a number of different outputs.

Like entry controls, price controls also are supposed to serve two functions. One function is to ensure that the price is such that the firm earns no more than the allowable rate of return. The second is to prevent price discrimination.

Problems with the Model

Static Assumptions. The public utility regulatory model can work only if two key assumptions are made—both based on the notion that the world is static. The first is that demand never changes—that consumers do not change in number, income levels, or tastes and preferences, that they demand precisely the same amount of output year after year. The second assumption is that the method of providing that static output—the technology—also never changes. Under those circumstances, it is *possible* that an industry will be and remain a natural monopoly. But this is not all. The regulator first must determine an appropriate rate of return: should AT&T earn 9, 10, or 11 percent on its investments? Then the regulator must determine how much output will be demanded —how many long distance calls will be made or how many kilowatt hours will be used in the upcoming year or years—and the precise costs of producing that amount of output. With these three pieces of information, the regulator can determine the revenue that the firm should be allowed to earn. And from this revenue requirement the regulator can crosscheck individual prices.

Now, even in a world with demand and technology constant, this is a mind-boggling process. Just consider any of the industries that is currently regulated and list all the inputs that have to be costed: the system of accounts necessary to give that cost information is (to understate the case) very large indeed.

Thus, even held static, the model suffers a major flaw because it asks a central agency to know an enormous lot! The information is, perhaps, knowable. But it is knowable only from one source, the regulated firm itself, and only *after* that firm has begun operations and made at least a first approximation at getting to

the proper size. Moreover, the cost information has to be developed in a way that enables the agency to prevent wasteful use of inputs.

Dynamic Assumptions. Let us now relax our static assumptions one-by-one, beginning with the more unrealistic of the two—the assumption that the total number of consumers and their incomes, tastes, and preferences are fixed over time.

Changing demand. This means that consumer demand will change—and, as it does, the enormous problem of calculating costs becomes even more so. Each time demand changes, all costs have to be recalculated. For an industry to be a natural monopoly at all, the cost per unit cannot be constant, without regard to supply. Rather, the unit cost *must* decline as the size of plant is increased over some range of output levels.

Once again the regulatory agency must get information on cost per unit, but now it is not estimating a single cost per unit that it can use over and over again to make prescriptions. Instead, it has to get the cost per unit when output is 100 units, when it is 101 units, and 102, and so on. Let us suppose further than the agency can accurately estimate costs at output levels other than the firm's present one. The job is still only half done. In order to fix the proper rates that would result in the allowable rate of return for the next year, the agency has to predict future demand. That is, the agency must be able to predict that demand will be such that if the firm produces x units of output at predicted cost y and sells them at predicted price z, the firm will earn its allowable rate of return and no more.

This kind of prediction is at best extremely difficult: it involves predicting how changes in price will change the demand for output. For example, when airline deregulation expanded competition and lowered air fares in the last few years, passengers increased more than the airlines had expected. Similarly, when the FCC required AT&T to lower interstate evening and weekend telephone rates in the 1960s, demand grew more rapidly than expected. Because both the passenger and telephone call increases were met by making better use of equipment during relatively slack periods, increased usage added more to revenues than to costs—and thus raised the actual rate of return earned.

Once again, the regulator can get the relevant data on costs only from the regulated firm. And because the firm must supply a wide variety of different cost estimates that depend on multiple contingencies, it is much tougher to verify the accuracy of the data. Additionally, as demand grows—assuming no change in technology and no change in costs of inputs—the quantity demanded at some point will outrun the amount a single supplier can produce at least cost. Which is to say, we are no longer in the presence of a natural monopoly.

Changing technology. Yet all these problems are minor compared to those the regulator faces once we relax the assumption that technology never changes. In the real world, firms do not tear out a whole operating plant and replace it with the latest version. Instead, they introduce change incrementally, a little bit here, a little bit there. AT&T, for example, has been installing digital switches, which make possible new services such as call forwarding, in its local exchanges since the mid-1960s and hopes to have all traffic on digital equipment by the end of this century. The type of digital switching equipment being installed, moreover, has changed significantly over these years. This process of gradual substitution results in a very large number of choices of technologies that the firm may employ. And the regulator has to match each of these choices to all possible levels of demand if it is to ensure that the profits earned are equivalent or even close to the outcomes of competitive markets.

With data problems of these magnitudes, just to derive some prices to "play with" is patently impossible—and I have not even begun to consider the problem of ensuring that the costs are efficient ones and not subject to waste. Furthermore, whatever data the agency does get are going to be rapidly out-of-date, almost "on delivery," and thus of little value for predicting the future.

Actual Effects of Rate-of-Return Regulation

As the more unrealistic assumptions are relaxed and we close in on the real world, it begins to be apparent why public utility regulation never could have held monopoly profits to or even near the competitive level. The amazing thing is that anyone ever expected it to. If

the only outcome of applying an unworkable model to the control of monopoly were that monopoly power went uncurbed, however, little would be lost except illusions. Unfortunately the actual effects are not so benign. To begin with, applying the model imposes substantial costs on society, including the direct administrative costs of the process itself. More important, however, rate-of-return regulation, by impeding or blocking innovation, most likely has significantly maintained some industries as monopolies and prevented the search for truly effective alternative means to control monopoly power. These costs are the hardest to measure because they involve the valuation of *what did not happen*—goods and services not made available to consumers.

[Rate-of-return regulation] imposes substantial costs on society.... More important, however... by impeding or blocking innovation, [it] most likely has significantly maintained some industries as monopolies and prevented the search for truly effective alternative means to control monopoly power.

Slowing of innovation has important consequences for the limiting of monopoly power. Chief among the historical concerns about monopoly power is that one customer will be asked to pay more, sometimes a great deal more, than another for essentially the same output. That concern is reflected in the concentrated political attention on higher railroad rates per ton/mile on short hauls than on longer hauls, on the possibility that telephone rates within rural areas and between rural and urban areas will be higher than rates within and between urban areas, and so forth. From that political attention comes the requirement to provide service to all comers at nondiscriminatory rates.

What gives a monopoly such power? First of all, as already noted, the output being provided must be seen by the customer as being very important, rather than a luxury. Therefore demand for it is quite inelastic—that is, the quantity demanded by any individual is not likely to vary a great deal with price. Demand tends to be inelastic in the absence of close substitutes that can serve the same or almost the same function, which provides the user with a choice. Gasoline provides a good example. As long as all of us have cars, we must have a liquid fuel that does not destroy the internal combustion or diesel engines. There are few substitutes for gasoline or diesel fuel. Moreover, because alternative modes of transportation are so poorly developed, particularly for short-haul movement of people, not only is there no alternative choice of fuel but also no alternative mode of transportation for many people. It is not surprising therefore that demand for gasoline is *relatively* inelastic. (The conventional wisdom, now proven erroneous, had been that demand for gasoline was *totally* inelastic. Recently, however, very high gasoline prices have in fact reduced consumption.)

The most certain way to reduce the potential for abuse of monopoly power in the face of inelastic demand is actively to encourage the search for substitutes. This means encouraging innovation. And innovation is most likely to be encouraged if there are alternative sources of supply—that is, if there is lively competition. Compare, for example, the degree of change since World War II in the telephone instrument, on the one hand, and the desk calculator on the other. The contrast is all the more striking in that roughly the same technology could have been used to upgrade both products.

Impact on Innovation Further Considered

Rate-of-return regulation with price and entry controls has the effect of slowing product innovation and technological change by regulated firms; by firms that might want to enter the market, using a better idea to make the very same output; and by firms that might develop new products to serve the same basic functions.

Impact on Regulated Firms' Innovative Activity. In the absence of competition, neither regulated firms nor the regulatory agency have any reason to depreciate any faster than physically necessary the plant and equipment that in a competitive industry would be economically obsolete long before it wears out. In many regulated industries, the life of much equipment is very long, often between twenty and fifty years or more. The failure to replace it before the end

of its physical life is beneficial to society only if no cheaper means of accomplishing the same purpose has been found—that is, some new technique that can perform at sufficiently lower cost to pay back the cost of removing the old equipment before its physical life is over. The cost that has to be paid back, however, depends on the rate of depreciation the firm has chosen. If it chooses a fast rate, the firm is gambling that better equipment will be available soon; long depreciation periods reflect the opposite view.

The effects of long depreciation periods can be illustrated in telephone switching equipment. Central office telephone panel switches were the first automatic switches developed to replace operators in larger cities in the 1920s. A few panel switches were still in use in the mid-1970s, fully fifty years later. In the same period, at least three new generations of switching equipment, including all digital switches, had been developed. (Also in these years, computers were invented; they have gone through five generations already.) Each of these new generations of technology permitted more rapid connections among telephones connected to the same switch or an increase in the number of connections a single switch could handle, at much lower cost and with much simplified maintenance. Yet, from the 1920s to the 1970s, telephones connected to panel switches never utilized any of these newer technologies.

If a firm uses a very long depreciation period and new, improved equipment becomes available before the old wears out, its customers may then start to pay more than the minimum possible costs of providing the output. If a firm in these circumstances faces no pressure to install the new equipment anyway and write off the undepreciated part of the old, it simply will not do so. And the customers will end up paying for the technological lag.

If new equipment also makes possible new services that the old equipment cannot provide, the loss to customers becomes much larger (and harder to measure besides). Now, not only are customers potentially paying more than the minimum cost for the output they do receive but also they cannot receive outputs that are technically possible and for which they might be willing to pay.

Telephone switching equipment also illustrates this cost. Before the advent of switches using digital technology, each new generation of switches offered mainly speedier connections or much simplified maintenance—either of which lower the costs of interconnecting telephones. With the use of digital technology, however, more than just speed-and-cost economies are possible. Now the switch can be made "smart"—that is, it can offer forwarding services, call waiting, and international direct dialing, among other services. AT&T's plan to convert all local exchanges to digital switching equipment only by about the turn of the century means that, for some customers, the first time they will be able to buy such services will be more than twenty-five years after the relevant technology was developed.

Nor is delay in the deployment of new technologies the only barrier to innovation imposed by rate-of-return regulation. Insofar as a regulatory agency does succeed in holding down the rate of return to "normal" levels, it takes away the incentive for the regulated firm to engage in high-risk research and development. Such activities pay off only if a high rate of return can be earned on the successful inventions.

Barriers to Innovation by Nonregulated Firms. Rate-of-return regulation also slows innovation by firms not in the industry but that might have discovered a better means of providing the regulated output. Any firm wanting to offer such service legally usually must apply to the regulatory agency for permission to enter. In applying, the firm has to reveal much of the detail about its new idea, thus providing valuable information to those already in the industry. Frequently the existing firms can block the would-be entrant by adopting the proposed innovation.

In the 1950s, for example, the telephone companies planned to expand the long distance network mainly by the use of cable, which would have meant substantially higher long distance rates. In response, the television networks considered building their own interconnections using much less expensive microwave technology (which had been developed by the military during World War II). The telephone companies promptly decided that they should use microwave technology and asked the FCC to bar anyone other than telephone companies from building such networks. The commission initially agreed with the telephone companies;

subsequently it allowed private microwave systems as well.

Thus potential new firms may never get into the industry at all and almost certainly cannot inject ideas into the market before existing firms learn of them and even adopt them. It is scarcely surprising, therefore, that few firms are willing to innovate in regulated areas unless these innovations are likely to be so massively profitable that they still would pay off even if the original innovating firm were not first into the market.

Barriers to Intermodal Competition. A third problem with rate-of-return regulation is that it encourages regulatory creep. As similar services are discovered, or instituted, they too become subject to regulation. And so not only has innovation in ways to supply the product been curbed; so also has innovation to find close substitute products that might have curbed the monopoly power that led to concern in the first place.

Furthermore, regulated firms have positive incentives to expand their activities into previously unregulated areas. Because all of the firm's various activities use in common at least some of the same inputs—at a minimum, the firm's overhead—it may be possible for the firm to cover common costs in the price of the regulated output and to gain immediate advantage in the previously unregulated sphere. So more regulation is needed to block unwanted cross subsidies. The development of "smart" data processing terminals that also can be used as telephones is an example of this spread. The movement of AT&T into the manufacture of these terminals first resulted in extending tariff regulations to such offerings. Subsequently it has led to a more than twelve-year effort to establish rules for telephone companies to prevent cross subsidies. No end to this regulatory proliferation is yet in sight.

Alternative Approaches to Curbing Monopoly Power

So, rate-of-return regulation does not work, creates distinctly bad side-effects, and takes on the status of a self-fulfilling prophecy. There is no possibility of keeping profits at the competitive level and, what is worse, two of monopolies' major counteracting forces, competition and innovation, have in effect been nullified. Indeed, one of the strangest things about the choice of rate-of-return regulation to cope with monopoly power is that it is equivalent to saying, "Because in an unfettered market you might exercise or you have in fact exercised your monopoly power, we're going to give it to you in perpetuity!"

Assuming still that monopoly power can and probably will be abused, ways must be found to produce the desired output competitively and to produce alternative outputs that meet the same basic need. In brief, what is called for are government actions that end the monopoly *but do no more than that*, leaving market forces free to act in all other normal ways. There may not be any one universally available solution. But a few examples suggest routes to follow in at least some cases.

(1) *Franchise bidding.* When the monopoly produces a single output that is not subject to rapid shifts in technology, franchise bidding may be superior to unfettered markets. Bidding for the right to supply the output would transfer most or all of the potential monopoly profits from the firm to the governmental unit (and to its tax-paying citizens). For example, retail distribution of water or electricity within a locality could be awarded to the company willing to pay the most for the privilege. Currently cable television franchises are being awarded in some localities on the basis of the socially beneficial goods and services (free governmental coverage, studios, and the like) the company is prepared to offer.

(2) *Mandatory interconnection.* When society deals with firms engaged in providing network services—railroads, telephones, electric utilities, and the like—structuring the market with a few carefully drawn rules could be much more effective than applying rate-of-return regulation. One such rule should be mandatory interconnection of all those locations in the area where multiple lines come together—that is, requiring all railroads entering a city to transfer cars to any other railroad at all other railway terminals in the city, or requiring all local telephone exchange switches to pass message traffic to all other local exchange switches.

(3) *Encouraging intermodal competition.* As noted already, not only competition within

(Continues on page 49)

straints is added to the phasing out of deposit rate ceilings, will fundamentally alter the competitive environment, the pricing of bank services, and the structure of the banking industry. The number of competitors in local markets now characterized as monopolies or oligopolies will increase. At the same time, the number of banks will decline sharply over the next twenty years, perhaps by a third or a half. Accordingly, average bank size will rise sharply, with most of the change involving consolidations among small and medium-sized banks. The trend toward putting explicit prices on all bank services will spread. Large depositors will no longer subsidize small ones, nor will depositors as a group subsidize borrowers. Idle account balances will yield a rate of return reflecting market conditions less the costs of maintaining those accounts (and handling transactions). Savings at regulated depository institutions will be more remunerative and may, in consequence, increase as a percentage of disposable personal income.

The result would seem to be that net savers will benefit, but not as much as they might have in the absence of greater performance regulation; and net borrowers will pay higher costs, though not as high as they would have if banks had been denied the opportunity for more equitable competition with non-deposit institutions. Competition will be shown to work. Unfortunately, the changes that accompany its working—structural shifts, higher credit costs, and a significant redistribution of the burden of costs among users of financial services—will lead to criticisms that may undermine its support. Ideally, competitive deregulation should reduce the need to control performance directly. The danger is that the critical voices will succeed in undoing the deregulation or in so expanding performance controls as to offset the benefits greater competition can bring. ■

Ideally, competitive deregulation should reduce the need to control performance directly. The danger is that the critical voices will succeed in undoing the deregulation or in so expanding performance controls as to offset the benefits greater competition can bring.

Rate-of-Return Regulation
(Continued from page 41)

the existing industry but also competition among similar industries offers consumers better protection from monopoly power than rate-of-return regulation. Thus, for example, in the early days of railroads, more public attention to the state of the roads might have generated earlier and better alternatives to the railroads for very short hauls. Ultimately, of course, the advent of trucks provided precisely that kind of competition (which then got regulated also!).

Similar competitive possibilities are arising today in communications. Competition to traditional local telephone companies that interconnect telephones by wire could come from wireless systems—two-way radios that operate on new frequencies—or systems that combine radio and wire links. A few relatively modest changes in regulatory restrictions on the use of existing radio systems could open the way to such competition. Similarly, new video technologies such as cassettes and discs, as well as more relaxed rules on subscription and low-power television stations, could be more effective anti-monopoly techniques than state or local rate regulation of cable television systems.

(4) *Antitrust restrictions.* When a monopoly offers a multiplicity of closely related services, its power to abuse can be held in check by antitrust restrictions against tie-ins and refusals to deal. For example, government could require that, insofar as a firm is a monopoly, its various services be offered individually and be subject to resale. These two techniques—unbundling (requiring, for example, that the telephone service and the telephone set be offered separately) and resale (allowing the customer to share his purchase with other users and to charge for that sharing)—would impose pressures to keep charges close to costs and induce competitive offerings of at least some services.

So SOCIETY DOES have techniques available for reducing monopoly power. These alternatives all involve governmental intervention in markets. Some even involve regulation—but not rate-of-return regulation. Going that route, no consumer has yet been protected from abuses of monopoly power, nor ever will be. It is both a snare and a delusion—and an unacceptable fraud on the public. ■