The Booth School of Business, University of Chicago

WHY REGULATE UTILITIES?*

HAROLD DEMSETZ
University of Chicago

CURRENT economic doctrine offers to its students a basic relationship between the number of firms that produce for a given market and the degree to which competitive results will prevail. Stated explicitly or suggested implicitly is the doctrine that price and output can be expected to diverge to a greater extent from their competitive levels the fewer the firms that produce the product for the market. This relationship has provided the logic that motivates much of the research devoted to studying industrial concentration, and it has given considerable support to utility regulation.[1]

In this paper, I shall argue that the asserted relationship between market concentration and competition cannot be derived from existing theoretical considerations and that it is based largely on an incorrect understanding of the concept of competition or rivalry. The strongest application of the asserted relationship is in the area of utility regulation since, if we assume scale economies in production, it can be deduced that only one firm will produce the commodity. The logical validity or falsity of the asserted relationship should reveal itself most clearly in this case.

Although public utility regulation recently has been criticized because of its ineffectiveness or because of the undesirable indirect effects it produces,[2] the basic intellectual arguments for believing that truly effective regulation is desirable have not been challenged. Even those who are inclined to reject government regulation or ownership of public utilities because they believe these alternatives are more undesirable than private monopoly, implicitly accept the intellectual arguments that underlie regulation.[3]

[1] Antitrust legislation and judicial decision, to the extent that they have been motivated by a concern for bigness and concentration, *per se*, have also benefited from the asserted relationship between monopoly power and industry structure.

[2] Cf., George J. Stigler and Claire Friedland, What Can Regulators Regulate? The Case of Electricity, 5 J. Law & Econ. 1 (1962); H. Averch and L. Johnson, The Firm under Regulatory Constraint, 52 Am. Econ. Rev. 1052 (1962); Armen Alchian and Reuben Kessel, Competition, Monopoly, and the Pursuit of Pecuniary Gain, in Aspects of Labor Economics 157 (1962).

[3] Thus, Milton Friedman, while stating his preference for private monopoly over public monopoly or public regulation, writes:

55

The economic theory of natural monopoly is exceedingly brief and, we shall see, exceedingly unclear. Current doctrine is reflected in two recent statements of the theory. Samuelson writes:

Under persisting decreasing costs for the firm, one or a few of them will so expand their q's as to become a significant part of the market for the industry's total Q. We would then end up (1) with a single monopolist who dominates the industry; (2) with a few large sellers who together dominate the industry . . . or (3) with some kind of imperfection of competition that, in either a stable way or in connection with a series of intermittent price wars, represents an important departure from the economist's model of "perfect" competition wherein no firm has any control over industry price.[4]

Alchian and Allen view the problem as follows:

If a product is produced under cost conditions such that larger rates . . . [would] mean lower average cost per unit, . . . only one firm could survive; if there were two firms, one could expand to reduce costs and selling price and thereby eliminate the other. In view of the impossibility of more than one firm's being profitable, two is too many. But if there is only one, that incumbent firm may be able to set prices above free-entry costs for a long time. Either resources are wasted because too many are in the industry, or there is just one firm, which will be able to charge monopoly prices.[5]

At this point it will be useful to state explicitly the interpretation of natural monopoly used in this paper. If, because of production scale economies, it is less costly for one firm to produce a commodity in a given market than it is for two or more firms, then one firm will survive; if left unregulated, that firm will set price and output at monopoly levels; the price-output decision of that firm will be determined by profit maximizing behavior constrained only by the market demand for the commodity.

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. To see this most clearly, let us consider the contracting process from its beginning.

Why must rivals share the market? Rival sellers can offer to enter into contracts with buyers. In this bidding competition, the rival who offers buyers the most favorable terms will obtain their patronage; there is no clear or necessary reason for *bidding* rivals to share in the *production* of

---

However, monopoly may also arise because it is technically efficient to have a single producer or enterprise. . . . When technical conditions make a monopoly the natural outcome of competitive market forces, there are only three alternatives that seem available: private monopoly, public monopoly, or public regulation. Capitalism and Freedom 28 (1962).

[4] Paul A. Samuelson, Economics 461 (6th rev. ed. 1964).

[5] Armen Alchian and William R. Allen, University Economics 412 (1st ed. 1964).

the goods and, therefore, there is no clear reason for competition in bidding to result in an increase in per-unit *production* costs.

Why must the unregulated market outcome be monopoly price? The competitiveness of the bidding process depends very much on such things as the number of bidders, but there is no clear or necessary reason for *production* scale economies to decrease the number of *bidders*. Let prospective buyers call for bids to service their demands. Scale economies in servicing their demands in no way imply that there will be one bidder only. There can be many bidders and the bid that wins will be the lowest. The existence of scale economies in the production of the service is irrelevant to a determination of the number of rival bidders. If the number of bidders is large or if, for other reasons, collusion among them is impractical, the contracted price can be very close to per-unit production cost.[6]

The determinants of competition in market negotiations differ from and should not be confused with the determinants of the number of firms from which production will issue after contractual negotiations have been completed. The theory of natural monopoly is clearly unclear. Economies of scale in production imply that the bids submitted will offer increasing quantities at lower per-unit costs, but production scale economies imply nothing obvious about how competitive these prices will be. If one bidder can do the job at less cost than two or more, because each would then have a smaller output rate, then the bidder with the lowest bid price for the entire job will be awarded the contract, whether the good be cement, electricity, stamp vending machines, or whatever, but the lowest bid price need not be a monopoly price.[7]

The criticism made here of the theory of natural monopoly can be understood best by constructing an example that is free from irrelevant complications, such as durability of distributions systems, uncertainty, and irrational behavior, all of which may or may not justify the use of regulatory commissions but none of which is relevant to the theory of natural monopoly; for this theory depends on one belief only—price and output will be at monopoly levels if, due to scale economies, only one firm succeeds in producing the product.

Assume that owners of automobiles are required to own and display

---

[6] I shall not consider in this paper the problem of marginal cost pricing and the various devices, such as multi-part tariffs, that can be used to approximate marginal cost pricing.

[7] The competitive concept employed here is not new to economics although it has long been neglected. An early statement of the concept, which was known as "competition *for* the field" in distinction to "competition *within* the field" is given by Edwin Chadwick, Results of Different Principles of Legislation and Administration in Europe; of Competition for the Field, as compared with the Competition within the Field of Service, 22 J. Royal Statistical Soc'y. 381 (1859).

new license plates each year. The production of license plates is subject to scale economies.

The theory of natural monopoly asserts that under these conditions the owners of automobiles will purchase plates from one firm only and that firm, in the absence of regulation, will charge a monopoly price, a price that is constrained only by the demand for and the cost of producing license plates. The logic of the example does dictate that license plates will be purchased from one firm because this will allow that firm to offer the plates at a price based on the lowest possible per-unit cost. But why should that price be a monopoly price?

There can be many bidders for the annual contract. Each will submit a bid based on the assumption that if its bid is lowest it will sell to all residents, if it is not lowest it sells to none. Under these conditions there will exist enough independently acting bidders to assure that the winning price will differ insignificantly from the per-unit cost of producing license plates.

If only one firm submits the lowest price, the process ends, but if two or more firms submit the lowest price, one is selected according to some random selection device or one is allowed to sell or give his contracts to the other. There is no monopoly price although there may be rent to some factors if their supply is positively sloped. There is no regulation of firms in the industry. The price is determined in the bidding market. The only role played by the government or by a consumers' buying cooperative is some random device to select the winning bidder if more than one bidder bids the lowest price.

There are only two important assumptions: (1) The inputs required to enter production must be available to many potential bidders at prices determined in open markets. This lends credibility to numerous rival bids. (2) The cost of colluding by bidding rivals must be prohibitively high. The reader will recognize that these requirements are no different than those required to avoid monopoly price in any market, whether production in that market is or is not subject to scale economies.

Moreover, if we are willing to consider the possibility that collusion or merger of all potential bidding rivals is a reasonable prospect, then we must examine the other side of the coin. Why should collusion or merger of *buyers* be prohibitively costly if an infinite or large number of bidding rivals can collude successfully? If we allow buyers access to the same technology of collusion, the market will be characterized by bilateral negotiations between organized buyers and organized sellers. While the outcome of such negotiations is somewhat uncertain with respect to wealth distribution, there is no reason to expect inefficiency.

Just what is the supply elasticity of bidders and what are the costs of colluding are questions to be answered empirically since they cannot be

deduced from production scale economies. There exist more than one firm in every public utility industry and many firms exist in some public utility industries. And this is true even though licensing restrictions have been severe; the assertion that the supply of potential *bidders* in any market would be very inelastic if licensing restrictions could be abolished would seem difficult to defend when producing competitors exist in nearby markets. The presence of active rivalry is clearly indicated in public utility history. In fact, producing competitors, not to mention unsuccessful bidders, were so plentiful that one begins to doubt that scale economies characterized the utility industry at the time when regulation replaced market competition. Complaints were common that the streets were too frequently in a state of disrepair for the purpose of accommodating competing companies. Behling writes:

There is scarcely a city in the country that has not experienced competition in one or more of the utility industries. Six electric light companies were organized in the one year of 1887 in New York City. Forty-five electric light enterprises had the legal right to operate in Chicago in 1907. Prior to 1895, Duluth, Minnesota, was served by five electric lighting companies, and Scranton, Pennsylvania, had four in 1906 . . . . During the latter part of the nineteenth century, competition was the usual situation in the gas industry in this country. Before 1884, six competing companies were operating in New York City . . . . Competition was common and especially persistent in the telephone industry. According to a special report of the Census in 1902, out of 1051 incorporated cities in the United States with a population of more than 4,000 persons, 1002 were provided with telephone facilities. The independent companies had a monopoly in 137 of the cities, the Bell interests had exclusive control over communication by telephone in 414 cities, while the remaining 451, almost half, were receiving duplicated service. Baltimore, Chicago, Cleveland, Columbus, Detroit, Kansas City, Minneapolis, Philadelphia, Pittsburgh, and St. Louis, among the larger cities, had at least two telephone services in 1905.[8]

It would seem that the number of potential bidding rivals and the cost of their colluding in the public utility industries are likely to be at least as great as in several other industries for which we find that unregulated markets work tolerably well.

The natural monopoly theory provides no logical basis for monopoly prices. The theory is illogical. Moreover, for the general case of public utility industries, there seems no clear evidence that the cost of colluding is significantly lower than it is for industries for which unregulated market competition seems to work. To the extent that utility regulation is based on the fear of monopoly price, *merely because one firm will serve each market*, it is not based on any deducible economic theorem.

The important point that needs stressing is that *we have no theory that*

[8] Burton N. Behling, Competition and Monopoly in Public Utility Industries 19-20 (1938).

allows us to deduce from the observable degree of concentration in a particular market whether or not price and output are competitive. We have as yet no general theory of collusion and certainly not one that allows us to associate observed concentration in a particular market with successful collusion.[9]

It is possible to make some statements about collusion that reveal the nature of the forces at work. These statements are largely intuitive and cannot be pursued in detail here. But they may be useful in imparting to the reader a notion of what is meant by a theory of collusion. Let us suppose that there are no special costs to competing. That is, we assume that sellers do not need to keep track of the prices or other activities of their competitors. Secondly, assume that there are some costs of colluding that must be borne by members of a bidders' cartel. This condition is approximated least well where the government subsidizes the cost of colluding—for example, the U.S. Department of Agriculture. Finally, assume that there are no legal barriers to entry.

Under these conditions, new bidding rivals will be paid to join the collusion. In return for joining they will receive a pro rata share of monopoly profits. As more rivals appear the pro rata share must fall. The cartel will continue paying new rivals to join until the pro rata share falls to the cost of colluding. That is, until the cartel members receive a competitive rate of return for remaining in the cartel. The next rival bidder can refuse to join the cartel; instead he can enter the market at a price below the cartel price (as can any present member of the cartel who chooses to break away). If there is some friction in the system, this rival will choose this course of action in preference to joining the cartel, for if he joins the cartel he receives a competitive rate of return; whereas if he competes outside the cartel by selling at a price below that of the cartel he receives an above-competitive rate of return for some short-run period. Under the assumed conditions the cartel must eventually fail and price and output can be competitive even though only a few firms actually produce the product. Moreover, the essential ingredient to its eventual failure is only that the private per-firm cost of colluding exceeds the private per-firm cost of competing.

Under what conditions will the cost of colluding exceed the cost of competing? How will these costs be affected by allowing coercive tactics? What about buyer cartels? What factors affect how long is "eventually"? Such questions remain to be answered by a theory of collusion. Until such questions are answered, public policy prescriptions must be suspect. A market in which many firms produce may be competitive or it may be collusive; the large number of firms merely reflects production scale diseconomies; large numbers do not necessarily reflect high or low collusion costs. A market in

---

[9] However, see George J. Stigler, A Theory of Oligopoly, 72 J. Pol. Econ. 44 (1964).

which few firms produce may be competitive or it may be collusive; the small number of firms merely relects production scale economies; fewness does not necessarily reflect high or low collusion costs. Thus, an economist may view the many retailers who sell on "fair trade" terms with suspicion and he may marvel at the ability of large numbers of workers to form effective unions, and, yet, he may look with admiration at the performance of the few firms who sell airplanes, cameras, or automobiles.

The subject of monopoly price is necessarily permeated with the subject of negotiating or contracting costs. A world in which negotiating costs are zero is a world in which no monopolistic inefficiencies will be present, simply because buyers and sellers both can profit from negotiations that result in a reduction and elimination of inefficiencies. In such a world it will be bargaining skills and not market structures that determine the distribution of wealth. If a monopolistic structure exists on one side of the market, the other side of the market will be organized to offset any power implied by the monopolistic structure. The organization of the other side of the market can be undertaken by members of that side or by rivals of the monopolistic structure that prevails on the first side. The co-existence of monopoly *power* and monopoly *structure* is possible only if the costs of negotiating are differentially positive, being lower for one set of sellers (or buyers) than it is for rival sellers (or buyers). If one set of sellers (or buyers) can organize those on the other side of the market more cheaply than can rivals, then price may be raised (or lowered) to the extent of the existing differential advantage in negotiating costs; this extent generally will be less than the simple monopoly price. In some cases the differential advantage in negotiating costs may be so great that price will settle at the monopoly (monopsony) level. This surely cannot be the general case, but the likelihood of it surely increases as the costs imposed on potential rivals increase; legally restricting entry is one way of raising the differential disadvantages to rivals; the economic meaning of restricting entry *is* increasing the cost of potential rivals of negotiating with and organizing buyers (or sellers).

The public policy question is which groups of market participants, *if any*, are to receive governmentally sponsored advantages and disadvantages, not only in the subsidization or taxation of production but, also, in the creation of advantages or disadvantages in conducting negotiations.

At this juncture, it should be emphasized that I have argued, not that regulatory commissions are undesirable, but that economic theory does not, at present, provide a justification for commissions insofar as they are based on the belief that observed concentration and monopoly price bear any necessary relationship.

Indeed, in utility industries, regulation has often been sought because of the inconvenience of competition. The history of regulation is often written

in terms of the desire to prohibit "excessive" duplication of utility distribution systems and the desire to prohibit the capture of *windfall* gains by utility companies. Neither of these aspects of the utility business are necessarily related to scale economies. Let us first consider the problem of excessive duplication of facilities.

Duplication of Facilities. Communities and not individuals own or control most of the ground and air rights-of-way used by public utility distribution systems. The problem of excessive duplication of distribution systems is attributable to the failure of communities to set a proper price on the use of these scarce resources. The right to use publicly owned thoroughfares is the right to use a scarce resource. The absence of a price for the use of these resources, a price high enough to reflect the opportunity costs of such alternative uses as the servicing of uninterrupted traffic and unmarred views, will lead to their overutilization. The setting of an appropriate fee for the use of these resources would reduce the degree of duplication to optimal levels.

Consider that portion of the ground controlled by an individual and under which a *utility's* distribution system runs. Confront that individual with the option of service at a lower price from a company that is a rival to the present seller. The individual will take into consideration the cost to him of running a trench through his garden and the benefit to him of receiving the service at lower cost. There is no need for excessive duplication. Indeed, there is no need for any duplication of facilities if he selects the new service, provided that one of two conditions holds. If the *individual* owns that part of the distribution system running under his ground he could tie it in to whatever trunk line serves him best; alternatively, once the new company wins his patronage, a rational solution to the use of that part of the distribution system would be for the utility company owning it to sell it to the utility company now serving the buyer.

There may be good reasons for using community property rather than private property to house the main trunk lines of some utility distribution systems. The placement of such systems under or over streets, alleyways, and sidewalks, resources already publicly owned (a fact taken as datum here), may be less costly than routing them through private property. The failure of communities to charge fees for the use of public property, fees that tend to prevent excessive use of this property, can be explained in three ways.

(1) There was a failure to understand the prerequisities for efficient resource use. Some public officer must be given the incentives to act as a rational conservator of resources when these resources are scarce.

(2) The disruption of thoroughfares was not, in fact, costly enough to bother about.

(3) The setting of fees to curtail excessive use of thoroughfares by utility companies was too costly to be practical.

The first two explanations, if true, give no support to an argument for regulating utility companies. The third explanation may give support to some sort of regulation, for it asserts that the economic effects that are produced by the placing of distribution systems are such that it is too costly to economize through the use of a price system. The costs of taking account of these effects through some regulatory process must be compared with the benefits of realigning resource use, and if the benefits are worth the costs some regulation may be desirable. Note clearly: scale economies in serving a market are not at issue. To see this, imagine that electrical distribution systems are thin lines of a special conducting paint. The placing of such systems causes no difficulties. They are sprayed over either public or private property. Nonetheless, suppose that the use of each system is subject to scale economies. Clearly, the desire to regulate cannot now be justified by such problems as traffic disruption, even though scale economies are present. "Excess" duplication is a problem of externalities and not of scale economies.

Let us suppose that it is desirable to employ some sort of regulation because it is too costly to use the price system to take account of the disruptive effects of placing distribution systems. Regulation comes in all sizes and shapes, and it is by no means clear what type of regulation would be most desirable.

A franchise system that allows only a limited number of utility companies to serve a market area was employed frequently. A franchise system that awarded the franchise to that company which seemed to offer the best price-quality package would be one that allowed market competition between bidding rivals to determine that package. The restraint of the market would be substituted for that of the regulatory commission.

An alternative arrangement would be public ownership of the distribution system. This would involve the collection of competing bids for installing the distribution system. The system could then be installed by the bidder offering to do the specified job at the lowest price. This is the same process used by communities to build highways and it employs rival bidding and not commissions to determine that price. The community could then allow its distribution system to be used by that utility company offering to provide specified utility services at lowest cost to residents. Again the market is substituted for the regulatory commission. Public ownership of streets may make public ownership of distribution systems seem desirable, but this does not mean that the use of regulatory commissions is desirable.

The Problem of Windfalls. We must now consider a last difficulty that has sometimes been marshalled to support the regulation of utilities. This argument is based on the fact that events in life are uncertain. The application of this observation to the utility business goes like this. After a buyer enters into an agreement with a utility company for supplying utility service, there may be changes in technology and prices that make the agreed upon price

obsolete. In such cases, it is asserted, the price should be changed to reflect the current cost of providing utility services. The regulation by commission of prices on the basis of current costs is needed in the utilities industries because of the durability of original investments in plant and distribution systems. This durability prohibits the use of recontracting in the market place as a method for bringing about appropriate changes in price.

Problems of uncertainty create a potential for positive or negative windfalls. If market negotiations have misjudged the development of a better technology and if there is some cost to rewarding contracts to other producers once they are agreed upon, then an unexpected improvement in the technology used by those who are awarded the contracts may generate a price that is higher than per-unit cost, but higher by an amount no greater than the cost of rewarding contracts. In such cases, the firms now holding the contracts may collect a positive windfall for a short-run period. Or, if input prices increase by more than is expected, these same firms may suffer from a negative windfall. But the same thing is true of all markets. If a customer buys eggs today for consumption tomorrow, he will enjoy a positive windfall if the price of eggs is higher tomorrow and a negative windfall if the price is lower. The difference in the two cases is that, where long-term contracts are desirable, the windfalls may continue for longer periods. In such cases it *may* be desirable to employ a cost-plus regulatory scheme or to enter a clause that reserves the right, for some fee, to renegotiate the contract.

The problem faced here is what is the best way to cope with uncertainty. Long-term contracts for the supply of commodities are concluded satisfactorily in the market place without the aid of regulation. These contracts may be between retailers and appliance producers, or between the air lines and aircraft companies, all of whom may use durable production facilities. The rental of office space for ninety-nine years is fraught with uncertainty. I presume that the parties to a contract hire experts to provide relevant guesses on these matters and that the contract concluded resolves these issues in a way that is satisfactory to both parties. Penalties for reopening negotiations at a later date can be included in the contract. I presume that buyers and sellers who agree to contract with each other have handled the problem of uncertainty in a mutually satisfactory way. The correct way to view the problem is one of selecting the best type of contract. A producer may say, "if you agree to buy from me for twenty-five years, I can use facilities that are expected to produce the service at lower costs; if you contract five years, I will not invest much in tooling-up, and, hence, I will need a higher price to cover higher per-unit costs; of course, the longer-run contract allows more time for the unexpected, so let us include an escape clause of some kind." The buyer and seller must then agree on a suitable contract; durability of equipment

and longer-term commitments can be sacrificed at the cost of higher per-unit costs, but there is no reason to expect that the concluded contract will be biased as to outcome or nonoptimal in other respects.

Cost-plus rate regulation is one way of coping with these problems, but it has great uncertainties of its own. Will the commission be effective? Does a well defined cost-plus arrangement create an inappropriate system of incentives to guide the firm in its investment and operating policies? Do the continual uncertainties associated with the meaning of cost-plus lead to otherwise avoidable difficulties in formulating investment plans? Rate regulation by commissions rather than by market rivalry may be more appropriate for utility industries than for other industries, but the truth of this assertion cannot be established deductively from existing economic theory. We do not know whether regulation handles the uncertainty-rent problem better or worse than the market.

The problem of coping with windfalls must be distinguished from the problem of *forecastable* rents. Suppose that it is known that buyers will incur considerable recontracting cost if they decide to change sellers after they are part way through an awarded contract. It would appear that the seller who wins the initial contract will be able to collect a rent as large as this recontracting cost. But this is not true if this recontracting cost is forecastable, that is, if it is not a windfall. The bidding for the initial contract will take account of the forecastable rent, so that if the bidding is competitive the rent will be forfeited by the lower bid prices to which it gives rise.

To what degree should legislation and regulation replace the market in the utilities or in other industries and what forms should such legislation take? It is not the objective of this paper to provide answers to such questions. My purpose has been to question the conventional economic arguments for the existing legislation and regulation. An expanded role for government can be defended on the empirical grounds of a documented general superiority of public administration in these industries or by a philosophical preference for mild socialism. But I do not see how a defense can be based on the formal arguments considered here; these arguments do not allow us to deduce from their assumptions either the monopoly problem or the administrative superiority of regulation.

In the case of utility industries, resort to the rivalry of the market place would relieve companies of the discomforts of commission regulation. But it would also relieve them of the comfort of legally protected market areas. It is my belief that the rivalry of the open market place disciplines more effectively than do the regulatory processes of the commission. If the managements of utility companies doubt this belief, I suggest that they re-examine the history of their industry to discover just who it was that provided most of the force behind the regulatory movement.