
CONTESTABLE MARKETS, TRADE, AND DEVELOPMENT

William J. Baumol
Kyu Sik Lee

The design of policies to improve economic efficiency using the private sector as principal agent requires a clear understanding of the role of market structure. Contestability analysis, not ten years old, provides a tool for the purpose. The concept can guide the government that wants to have it both ways: to protect the public and smaller firms from the threat of monopolistic behavior by large firms, at the same time allowing the larger firms enough freedom to meet the requirements of efficiency and to exercise entrepreneurship.

Perfect contestability provides a standard for ideal market behavior. A perfectly contestable market is one in which entry and exit are perfectly costless; in such a market, the mere (perpetual) threat of entry can enforce good conduct by incumbents. So long as sunk costs are zero, a potential entrant can undercut any excessive prices (or unnecessary costs) of incumbent firms yet earn an attractive rate of return. Thus perfect contestability precludes excessive profits and prices as well as waste and inefficiency, and it prevents predatory pricing.

What contestability analysis means for policy in developing countries is that an economy that wishes to take advantage of available economies of scale can use the norms of behavior provided by the theory as a guide for regulation of its larger firms, instead of resorting to nationalization or to stifling restrictions as the means to protect its infant industries and its consumers. Under this standard, the bounds on the firms' behavior set by the regulations replicate those that would be enforced by market pressures in an ideal state of perfect contestability. The article gives a Nigerian example in which regulatory changes applying the theory promise to improve the performance of the electric utility market. Such methods can do more to promote the public interest than privatization, which often results in replacement of a state monopoly by a private monopoly.

Contestability analysis, a recent arrival in the literature of economic theory, has rapidly aroused attention among those involved in determining policy related to economic regulation and antitrust activity. Yet these are not the only promising areas in which the new theory might be applied. It also appears to offer suggestions and ideas for policy related to international trade and developing countries. This article undertakes to suggest some of these applications.

Contestability can be considered a generalization of the concept of perfect competition, interpreted as a (more or less) ideal state of economic affairs. Under contestability, competitive pressures can be supplied by potential entry as well as by the presence of actual current rivals. As a result, unlike perfect competition, perfect contestability is compatible with the presence of economies of scale and the consequent small number of relatively large firms that are disciplined into good behavior by the threat of entry. More industries can thus more easily approximate the circumstances of contestability than those of competition. In other words, contestability may provide a somewhat more realistic description of the behavior and structure of *some* industries. And because economies of scale and oligopoly cannot be wished away or abolished by fiat, contestability constitutes a more attainable goal for government policy toward restraint of monopoly power, and it suggests some rational rules for control of monopolistic industries.

We will show here that freedom of trade is critical in the provision of the beneficent pressures of contestability in the industries of the developed economies. In addition, we will show that in developing countries that are not prepared to expose their infant industries to the unrestrained rivalry of the multinational enterprises or established firms from industrial economies, contestability analysis can offer some guidelines for policies that effectively deny market power to the larger domestic enterprises and to those established under foreign sponsorship, without hampering their ability to compete, their efficiency, or growth in their productivity.

The Theoretical Framework

Before considering possible applications of contestability analysis, we need to define the concept and outline its virtues and limitations. What is contestability, and what can such a state of affairs achieve?

Perfect Contestability as a Standard for the Ideal Market

A perfectly contestable market can be defined as one in which entry and exit are perfectly costless. Freedom of exit in this sense is a critical part of the story because a little consideration confirms that it is merely the obverse of costlessness of entry: if a firm can leave an industry whenever it chooses and can re-

move all the funds it has invested without loss, its decision to enter will obviously have cost it nothing. In the reverse situation, in which exit does entail some loss in value of assets (that is, in which some substantial portion of investment is sunk), the decision to enter clearly incurs, at the very least, some cost in the form of entry risk—that is, risk greater than that incurred by a firm already operating in the arena—and this will inevitably act as a deterrent to entry.

In practice, of course, entry and exit always require some time and this requirement can itself become a source of cost. In particular, delay subjects the entrant to the costly risk of retaliation by incumbents who can use the lag period to reduce prices, mount expensive advertising campaigns, or take other measures to make an invader regret its decision to enter. But even so, a market can remain highly contestable if it is feasible for an entrant to achieve contractual relations with prospective customers. Such contracts can render the entrant immune from retaliation. Where this is true, the incumbent can protect itself from intruders only by behaving well—for instance, by offering customers reasonable prices—all along. Such behavior will indeed foreclose entry opportunities, but only by providing customers with all the benefits that an entrant could be expected to bring. In this the mere (perpetual) threat of entry can enforce good conduct by oligopolists and even by monopolists.

None of this should be taken to imply that perfect, or near perfect, contestability is easy to achieve, or that it is pervasive throughout the economy. In some markets heavy sunk costs are the norm, and contractual arrangements between buyers and sellers are rare and difficult to achieve. Such markets will be far from contestable. How widespread contestability is in practice is an empirical issue about which evidence is very limited. Moreover, the contestability of particular industries must be judged case by case; of course, no industry will satisfy fully the theoretical requirements of perfect contestability, and no one is sure how many can be taken to approximate that norm.

The Virtues of Perfect Contestability

It is convenient to classify under four headings the benefits that perfect contestability offers to society: (1) absence of excessive (monopolistic) prices or profits, (2) absence of inefficiency or waste, (3) absence of cross-subsidy or predatory pricing, and (4) (with the possible exception of the single-firm case) Pareto optimality in pricing—that is, pricing consistent with efficiency in the allocation of economic resources to serve the desires of consumers. Another virtue of the contestable markets model, primarily of interest to economic theorists rather than to those concerned with application, is that contestability facilitates the construction and analysis of economic models in many of the ways (and for the same reason) that perfect competition does. The reason is that contestable markets impose prices rather than demand curves on the firm. This fact permits the analytic simplification of dealing in parametric prices,

without injection of the complications that arise in taking demand functions into account in an analysis of the behavior of the firm.

The most obvious benefit of perfect contestability is its complete preclusion of excessive profits and excessive prices. Even a pure monopolist in a perfectly contestable market can earn no higher rate of return on its capital in the long run than is attainable by a small firm under perfect competition. The reason for this startling result is that any profit in excess of the competitive rate of return inevitably constitutes a profit opportunity for a potential entrant so long as sunk costs are zero. Excessive profits permit the entrant to undercut the incumbent's excessive prices to a degree that still leaves an attractive return to the new enterprise. Even if the opportunity proves extremely transitory, costless exit means that the entrant will have benefited from its incursion and will be prepared to repeat it should the incumbent ever again repeat its folly.

The same reasons underlie the second virtue of perfect contestability, its total preclusion of waste and inefficiency. For if any firm incurs costs greater than the lowest permitted by current knowledge and the current state of technology, then a more efficient entrant can benefit in exactly the same way that excessive profits permit. The point is that excessive prices are in themselves always an entry opportunity, whether those prices are attributable to excessive profits, to unnecessary costs, or to anything else.

The absence of waste under perfect contestability encompasses more than is apparent at first glance. Obviously, it precludes allocative inefficiency—the sort of waste on which economic theory has tended to focus. That is, it precludes the use of input x by firm A and input y by firm B when x will yield a more valuable output if used by B and y will produce more if employed by A. In addition, the no-waste attribute of perfect contestability rules out X-inefficiency, the portmanteau class of inefficiencies called to our attention by Harvey Leibenstein, which encompasses all forms of inefficiency aside from the allocative inefficiencies most frequently discussed in standard economic theory. It includes waste attributable to managerial ignorance, wrongdoing, or lack of initiative, among other sources. Perfect contestability is also incompatible with inefficient delays in the introduction of technological improvements because such delays, too, entail unnecessary costs. That is, contestability requires just the efficient rate of introduction of new productive techniques as they become available. (Contestability does not, however, eliminate the free-rider problem that accompanies the acquisition of knowledge or the production of inventions, so that it does not guarantee a socially optimal level of expenditure on research or an optimal flow of innovations.)

The absence of inefficiency guaranteed by perfect contestability extends to industry structure. For example, if the cost structure of an industry is such as to make it most efficient, say, for fifteen firms to produce the industry's current combination of outputs, then in perfectly contestable equilibrium the industry will not be made up of fourteen firms or sixteen or some other number different from the cost-minimizing fifteen enterprises. It is this efficiency property

that has permitted contestability theory to analyze what determines the structure of an industry—that is, what makes one industry an oligopoly, another a monopoly, and a third highly competitive. Since contestability is compatible with any of these market forms, its theory can treat the choice among them as an endogenous selection. Moreover, under perfect contestability the choice among them will not be made on the basis of strategic considerations, but it must instead always follow the cost-minimization requirements just described. Of course, since in reality contestability may often be absent, strategic and other considerations can be important determinants of the structure of industry in practice.

The third desirable attribute of perfect contestability, its prevention of predatory and cross-subsidy pricing, is a crucial consideration for control of monopoly and market power, and prevention of their extension. Predatory pricing occurs when an incumbent firm adopts prices that entail a deliberate sacrifice of profit, with the object of either driving a competitor from the market or preventing the entry of a new rival. This accomplished, the predator can raise prices again to yield excessive profits large enough to make up for the initial profit sacrifice. Since the prospect of excessive future profits is crucial to predation, and perfect contestability always prevents excessive profits by means of the zero cost of entry, with excessive profits impossible, predation must become unattractive.

Cross-subsidy, the second *bête noire* of those concerned with monopoly, raises more subtle considerations. A cross-subsidy is defined to be present when one set of customers of a firm obtains its products at a price that entails a loss to the firm and that loss is made up for by excessive prices to other customers. The latter are then said to be providing a cross-subsidy to the former. Usually, this is an accusation brought by a firm complaining that it is subjected to unfair competition by being forced to meet the indefensibly low prices permitted by cross-subsidy, while the allegedly subsidizing firm makes up for its subsidies in markets for its products where it happens to hold a monopoly.

This is not the place to go into the extensive debate over the circumstances that can lead to cross-subsidy or the possibility of rational motives for its adoption. Rather, having noted that it is a standard concern of those who deal with the control of monopolistic practices, we want to show how it is prevented by perfect contestability. For the purpose we observe first that economists today usually propose the incremental cost of a product as the criterion by which to judge whether the price of the product entails the receipt of a cross-subsidy. That is, if the product is priced so as to yield a total revenue at least equal to that product's incremental cost—the increase in the firm's total cost attributable to the supply of that product—it is said that the consumers of the product are receiving no cross-subsidy because they are paying the entire cost incurred by the firm in supplying the product to them. But under perfect contestability, any product price yielding revenue below incremental cost is incompatible with equilibrium, so cross-subsidy cannot occur under that regime. The reason is

most easily described by illustration. Consider a firm that supplies three products, x , y , and z , and assume that the revenue from x is less than its incremental cost. Then an entrant can open for business and offer to sell only y and z at prices slightly below the incumbent's. By dropping x from its product line, the entrant forgoes revenue from x but avoids x 's still greater incremental costs, the result being a net addition to the entrant's profits over and above those formerly enjoyed by the incumbent. Thus cross-subsidy permits the entrant to profit by stealing the incumbent's profitable markets away, sharing with customers of those products part of the supplier's earnings previously going into the cross-subsidy.

We will not take the space required to explain the rather more complex logic behind the fourth desirable attribute of a contestable market: its enforcement of optimal prices in markets containing two or more suppliers. From the viewpoint of such issues as stimulation of growth in a developing country, this is undoubtedly an esoteric matter, not of immediate pertinence. For a full discussion, see Baumol, Panzar, and Willig (1988, chap. 11).

Enough has been said to indicate why perfect contestability can be considered an economic ideal whose approximation by market behavior is to be encouraged when feasible, and whose behavior it may be desirable to replicate by means of policy when markets cannot do the job.

Unsettled Matters in Contestability Analysis

Nothing in the preceding discussion should mislead the reader into believing that everything is settled and definitive in the field of contestability analysis. On the contrary, reservations expressed by several commentators have suggested that the subject was rather controversial, although the authors and most of the commentators seem since to have agreed on positions between the possible extremes.

Perhaps most important, as has several times been suggested here, we do not know yet how many markets or industries constitute reasonably close approximations to contestability and how many are far from meeting its requirements. A number of studies of individual industries have thrown some limited light on this matter. Several such studies have indicated that oceanic freight transportation is highly contestable, or at least prospectively so if some restrictive practices are eliminated. Another piece of research has suggested the conclusion that aluminum production is, in the long run, contestable to a rather surprising degree. (For a review of the literature, see Baumol, Panzar, and Willig 1988, chap. 17, sec. 17E.) In contrast, it is now generally agreed that air passenger transportation is considerably less contestable than had been thought at the time of its deregulation in the United States, although this industry still lies far from the opposite end of the spectrum (see, for example, Morrison and Whinston 1987). A few more studies for particular industries have been carried out, but much remains to be determined.

Note, however, that the analytical power of the theory does not depend on the ubiquity of contestability; indeed, its policy lessons apply primarily to industries that are not contestable, whose regulation can be aided by contestability theory, which provides norms of behavior to which the regulated firms can reasonably be held.

A second important gap in knowledge about the subject in general relates to the behavioral implications of imperfections in the degree of contestability. Of course, real-world instances of perfect contestability, like those of perfect competition, probably do not exist. Schwartz and Reynolds (1983) have argued that in this arena even a little imperfection is apt to be a dangerous thing, producing large departures from the ideal behavior to be expected in a regime of perfect contestability. Opposing arguments have been offered—a series of studies by Kessides (see, for instance, Kessides forthcoming) have provided evidence suggesting that trifling imperfection brings correspondingly trifling deviation from ideal behavior—but all the participants in the discussion seem to agree that the issue can be settled only empirically.

Similarly, some reservations about the theory, such as those expressed by Weitzman (1983), have been pretty well dealt with in the ensuing discussion, yielding an outcome intermediate between the initial views of the discussants. Since the matters raised are fairly technical and are unlikely to be of great interest to readers of this journal, they will not be pursued here.

Applications

Let us turn now to application of some of these ideas, specifically to international trade and to the efficiency of the public sector in developing countries.

Contestability and International Trade

Some industries are compelled by their technology to consist of a few comparatively large firms within the boundaries of any one country. The reason is that in those industries technology gives rise to economies of scale and scope. Here the term “economies of scope” (which is yet another concept that derives from the literature on contestability) refers to a state of affairs in which a given bundle of several (or many) goods or services can be produced more cheaply by a single multiproduct firm than by a group of more specialized enterprises, each producing only a subset of the industry’s products. Where economies of scale and scope are available, one firm, or at most a small number of firms, will supply consumers’ demands for the industry’s products most cheaply, since a multiplicity of small firms would make those economies impossible.

Under these circumstances, the number of firms that is viable in the industry is limited by the extent of the market. If there is something like a minimum efficient scale of operation for a firm in the industry, and if the quantity of the

industry's product demanded at a viable price is only a small multiple of minimum efficient scale, then only a few enterprises can be expected to survive. If, in addition, entry into the industry requires substantial sunk costs, then neither the presence of a large number of firms nor the pressures of contestability may be able to protect the interests of consumers against prices that exceed competitive levels and other related abuses.

It was Adam Smith who wrote of economies of scale being limited by the extent of the market. The same is obviously true of the number of competitors actually in an industry—the U.S. market faces more competitors if it leaves itself open to Japanese products than if it does not. But the reason is just a bit less obvious. It follows from the relation of the extent of the market to the technologically determined efficient scale of operation of the firm. An oversimplified example will make the point. Suppose that in industry X a firm achieves lowest cost with an annual output of 10 million units of its product, and that costs rise sharply at any other output level. If a price that covers minimum unit costs (including a normal return on capital) will elicit a demand for 30 million units of the product in the U.S. market, it follows that an isolated and sealed-off United States can support at most three firms in the industry. The extent of the market will condemn the industry to the status of a tight oligopoly. But if an open world market is prepared to purchase 100 million units of the product at the same price, the industry will be able to support ten competitors rather than three, and that can be enough to undermine the possibility of collusion that a three-firm industry might have invited.

As a crude rule of thumb, then, the number of firms that an industry can support is given by the ratio of quantity of product demanded by the market at the lowest viable product price to the output of a firm with most efficient size. Because free trade increases the extent of the market, it multiplies that demand without changing the technologically most efficient size of firm. As a result, it increases the ratio of demand to efficient firm output and raises the viable number of participants in the market.

But freedom of trade can also affect the power of potential entry, and that is where contestability analysis comes in. Even an industry whose activities require an entrant to incur substantial sunk costs may have at its disposal a good many firms for which entry and exit are easy. These are firms which already have laid out the sunk expenditures needed for entry but which for some reason are not now present in the market. It will be remembered here that sunk costs are pertinent to an optimal decision only before those sunk costs have been incurred. Once a sunk outlay has been made it becomes irrelevant for a rational decision since that decision, by definition, cannot undo such a cost outlay of the past. Thus, if in Spain's industry X, production by an entrant requires a sunk (and fixed) outlay of \$10 million, a new Spanish firm can be deterred from entering the industry. But it is no impediment to a French firm, already in operation in France, which decides to enter the Spanish market. The French company's sunk cost has already been incurred and is not increased by its decision

to enter the market in another country. In this way, international trade can be a potent source of the sorts of pressures that impart contestability to a market.

Long-run trends have worked to enhance the power of such pressures. With secularly falling transport costs, with the advent of electronic means of transmission that have transformed many services from largely domestic undertakings into routine objects of international trade, and with related developments, each country's markets have become increasingly vulnerable to the threat of entry, potential and actual, from other nations. One set of estimates—the valuable figures provided by Maddison (1982)—suggests that in the 110 years after 1870 the share of exports (and imports) relative to gross national product (GNP) rose perhaps threefold or fourfold. That is, despite the startling rises in gross domestic product (GDP) (and in GDP per capita) over this period, foreign trade has exploded even more sharply, and by a considerable margin. All of this opens markets abroad to a country's domestic producers. But in exchange it makes them increasingly vulnerable to competitive pressures from abroad.

This growth in the extent of the market has undoubtedly transformed some industries that might otherwise have been subjects of concern to the antitrust authorities into highly competitive entities. The fact that the U.S. auto industry for years contained no more than four domestic firms, all of which could have incurred substantial sunk costs, might have been grounds for attention by antitrust agencies, but the competition of cars from France, Germany, Japan, Yugoslavia, and a variety of other places, and the threat that still others would jump in if domestic automakers were to attempt to extract monopoly profits by considerably overpricing their products, have preserved the competitiveness of the U.S. auto market.

A protectionist policy, in addition to the better-known costs that it imposes on the public, is thus likely to undermine the contestability of some industries and weaken that of others. Given the virtues of contestability described earlier, that too can impose a heavy burden on the public.

In an article on export-promoting trade strategy, Bhagwati (1988) supports the argument that economies of scale will be more fully exploited as international trade expands, although he acknowledges that systematic empirical evidence is not yet available on this effect. A study by Harris (1986), however, is indicative: this showed that “a 3.6 percent rise in GNP could follow from the unilateral elimination of Canadian tariffs, if the economies of scale are fully exploited” (p. 39).

In a study on the extent of barriers to industrial countries' imports, Nogués, Olechowski, and Winters (1986, pp. 181–99) show that even though tariff barriers to trade have declined significantly under the General Agreement on Tariffs and Trade—from an average of about 40 percent in the mid-1930s to 4 to 8 percent after the Tokyo Round of 1979—nontariff barriers have become more common. Nontariff barriers include import restrictions, voluntary export restraints, decreed import prices, antidumping and countervailing duties, and others. The study's estimates, based on data from the United Nations

Conference on Trade and Development, indicate that for the sixteen industrial economies examined, 27 percent of imports were subjected to nontariff barriers. The coverage for textiles was as high as 44.8 percent in 1983. The authors estimated the value of imports subjected to nontariff barriers to be some \$231 billion¹ (based on 1981 trade flows)—almost half the total imports of the state-trading East European countries. They also found that nontariff barriers are more prevalent on imports from developing countries than from industrial nations, and they concluded that “the growth of NTBs and their effect on international trade should be taken very seriously” (p. 197).

After a careful review and analysis of recent debates on export promotion policies, Bhagwati concluded: “Export promotion policies emerge with success from the detailed scrutiny offered in this article. Equally important is the fact that their successful adoption will require collaborative and intense efforts to ensure that the protectionist threat, recently escalating, is not allowed to break out into actual protection on a massive scale. The multilateral trade negotiations offer the only reasonable prospect for maintaining a momentum in favor of a freer world trading system. Failure to pursue them successfully, in a spirit of accommodation and mutual understanding of constraints and needs, will only undermine what seems like the best mechanism for containing the protectionist threat” (1988, p. 48).

Contestable Markets and Public Sector Efficiency in Developing Countries

The developing countries have been torn between the desire to take advantage of any available economies of scale and scope that promotion of large firms can offer, and the conviction that in business small is beautiful, whereas large enterprise is inherently dangerous to society. One expedient widely adopted in developing countries has been recourse to nationalization and political control of the activities of the publicly owned enterprises; another has been detailed governmental regulation to a degree not fully appreciated by casual observers from industrial countries. For example, few economists in the wealthier countries can envisage a regulation prohibiting firms producing automobile parts (and hundreds of other industries) from having a total company investment exceeding some \$50,000. Nor can they conceive of a rule that requires each large firm in the economy to obtain from the government every year a license that specifies the maximum amount the company is permitted to produce. Nor are they generally aware of cases in which the insurance industry is prohibited from using computers. Yet such rules have been adopted by some developing countries whose per capita incomes are among the lowest in the world.

It is hard to resist the conclusion that such constricting rules, however good their intentions, significantly impede productive entrepreneurship and system-

atically redirect the energies of the entrepreneurs toward the underground economy. There is good reason to believe that tight regulation and political control of nationalized firms in the developing countries have seriously hampered economic efficiency and substantially depressed outputs and standards of living.

Data for the period since World War II (see, for instance, Summers and Heston 1984) suggest that, measured in both productivity levels and per capita incomes, the industrial countries have been converging on a common standard of living and a common state of efficiency in production. Broadly speaking, the other industrial economies have been approaching (but not so far exceeding) the growth path of the United States, and at the same time, the range of their performance has, with some exceptions, been narrowing sharply.

Matters are very different for the developing countries. They have been converging neither on the standard of industrial economies nor even on a standard among themselves. For example, according to Summers and Heston (1984), who report on a sample of seventy-two countries spanning all income groups over the period 1950–80, real per capita incomes grew at an annual rate of 3.1 percent in the countries they classified as industrial, but at only 1.5 percent a year in their low-income country group—less than half that of the industrial group. The poorer group ended up considerably further behind than they had been at the beginning of the postwar period.

The record for intragroup convergence is similar. For the industrial group the Gini coefficient of real GDP per capita dropped precipitately from 0.302 in 1950 to 0.129 in 1980, meaning that over the period the degree of inequality in the standards of living among those countries fell sharply. But for the poorer countries the Gini coefficient actually rose slightly over the same interval, from 0.103 to 0.112. Other independent calculations confirm these results. Clearly, at least some of the less affluent countries were not outstandingly successful in appropriating and instituting productive techniques either from the more affluent economies or from one another.

There is no definitive and undisputed list of explanations for this lackluster performance. But it seems widely agreed that a likely culprit is the frequency with which the poorer countries have restricted the range of activities left open to entrepreneurs. And, at least in part, this must be ascribed to fear of market power and of the income inequalities to which it leads. This implies that the national income pie has inadvertently been kept small by measures intended to prevent entrepreneurial success from slicing the pie unevenly. The statistics suggest that the price of pursuing this goal has been high, with whatever domestic equality it has been able to achieve being purchased at the cost of exacerbated inequality among nations.

What has contestability analysis to do with any of this? The answer is that it may offer a way out of the dilemma that allows entrepreneurs and firms the freedom to innovate, grow, and take advantage of economies of scale and scope and yet protects the public from exploitation by them. For contestability

analysis provides some principles for controlling enterprises suspected of possession of market power that offer the public all the protection from monopolistic exploitation that they would have received if the enterprises had been effectively constrained by the forces of market competition, yet do not impede economic efficiency, thwart initiative, or handicap the exercise of entrepreneurship.

To understand what contestability theory can contribute to the analysis of policy in this area, we must first inquire a bit further into the roots of the problem. The role of economies of scale and scope is crucial—without the availability of such economies no one would seriously advocate promoting or perhaps even tolerating bigness. It is only because the production techniques in some industries make production by large enterprises less costly to society than production by small firms that developing countries have been willing to live with larger enterprises. In any case, in industries with diseconomies of scale the problem is not even likely to arise, because their larger firms with their necessarily higher costs will be apt to succumb to the competition of more efficient smaller enterprises. Thus the problems that justify government control of industry are peculiar to the presence of economies of scale.

The presence of economies of scale, however, poses several special problems for pricing policy. First, the marginal-cost-pricing rule that guides pricing under perfect competition—that is, the setting of the price of each product at a level equal to marginal cost, as market forces require in a regime of perfect competition—becomes a recipe for bankruptcy where economies of scale are present. (Marginal cost is defined as the increment in the supplier's total cost that results from a one-unit increment in that firm's output. In other words, the marginal cost of product X is the first partial derivative of the firm's total cost with respect to the quantity of X it supplies.) The easiest way to see this is to recall that in the single-product firm, economies of scale mean that the average cost curve must decline as output increases and that when average cost is falling marginal cost must be lower than average cost. In that case, a price equal to marginal cost cannot possibly cover the firm's average cost and so total revenue must be less than total cost. With economies of scale, then, marginal-cost pricing is incompatible with viable operation. This, and the fact that survival of multiple firms is neither possible nor desirable where economies of scale occur, rules out the concept of perfect competition as a guide for policy in the arena we are considering.

But whereas perfect competition is incompatible with economies of scale, perfect contestability is not. Thus the theoretical concept of perfect contestability, with the demonstrated benefits it offers to the public, can serve as a proper guide to a government that wants to have it both ways—to permit relatively large firms to operate where they are appropriate, and to operate with a considerable degree of freedom, while extending effective government protection of the interests of the public.

Just what government action does contestability analysis call for in these circumstances? We have seen that contestability prevents prices from going too high because beyond some level (that is, beyond the cost of operation of an entrant) they will bring an influx of new competitors. Prices in a perfectly contestable market can approach this level, but they can never exceed it. Thus the forces of perfect contestability impose a ceiling on prices at their entry-inducing levels.

Similarly, perfect contestability imposes a floor on prices—one that protects the legitimate interests of competitors, just as the ceiling protects the interests of consumers. We have seen that this floor is set at the level of incremental cost, that is, while price in a perfectly contestable market can be nearly as low as incremental cost, it can go no lower.

Between these price bounds, the contestable firm is free to set its price at any level or rather to adapt the price of each of its products to the current state of demand for that item. And prices constrained in this way can yield no excessive profits; for under contestability excessive profits will always induce entry, so that excessive profits mean that at least some of the firm's prices must exceed the proper ceilings—the boundary between profitability and unprofitability of entry.

What all this means for policy in a developing country is that such an economy using contestability as its guide can protect both consumers and smaller rivals of large firms without recourse to drastic and economically costly measures such as nationalization or highly restrictive constraints on enterprise. If competitive behavior is what the society wishes to achieve, it can attain this with the aid of some fairly straightforward rules, notably the adoption of upper and lower bounds for the prices of firms considered to possess market power, bounds calculated with care to replicate the bounds that would be enforced by market pressures under the ideal state of perfect contestability.

An Illustration from the Developing World

We conclude with a discussion of a specific situation to illustrate the applicability of contestable market analysis to developing countries. In Nigeria, as in many other developing countries in Africa, Asia, and Latin America, state monopoly enterprises with substantial capital investment already in place fail to deliver infrastructural services such as electric power, water, and telecommunications sufficient to meet demand. A World Bank research project that studied the economic effects of infrastructural deficiencies in Nigeria (Lee and Anas 1989) confirmed that the costs of poor and unreliable public infrastructural services fall heavily on manufacturing firms, which have to go to considerable expense to overcome these deficiencies. According to the establishment survey conducted for the study, all but 14 of 179 firms have standby generators for electric power. The firms without their own generators are mostly small enterprises; they do not have the equipment not because the burden of poor

electric supply is less per unit of output for them, but rather because of economies of scale in the generation of electricity. At the same time, the survey indicates that small firms are the ones that cannot afford expensive capital investments for boreholes for self-provision of water, for vehicles for shipment of products, for motorcycles for couriers, and for radio equipment to be used in lieu of telephones. Hence, the burden of inadequate infrastructural services may be much higher for small firms than large ones.

Nigeria, therefore, affords two extreme manifestations of inefficiency in the provision of infrastructural services: first, a public sector with heavy capital investment that fails to provide the services; second, the costly provision of those services by the firms. These infrastructural failures have at least two sources. The first is relatively well understood and relates to shortcomings in the procedures and technology used by the public sector, including problems in day-to-day management, operation, and maintenance of the facilities. The second, more complex and intransigent, relates to general deficiencies of administration, bureaucracy, planning, billing and revenue collection, and personnel training, and the lack of appropriate incentives for efficient management. This second set of influences that lead to X-inefficiency, as described earlier, has remained a stubborn obstacle to progress over the years because investment in additional facilities is futile if the necessary institutional organization and logistical support systems are lacking.

For example, the Nigerian Electric Power Authority (NEPA), heavily subsidized by the government, had kept the price of electricity at 7 kobos (a little less than 1 U.S. cent) a kilowatt-hour for the ten years preceding 1989. An indication of how much firms are willing to pay for reliable services is the fact that, despite this low rate (compared with 7 cents in industrial countries), individual firms have maintained expensive standby generators even though, on average, 75 percent of the generating capacity of the firms remained idle, and the average cost of self-generation was 4.6 naira (60 cents) a kilowatt-hour. In July 1989 the NEPA price was raised to 30 kobos a kilowatt-hour, a fourfold increase. How quickly NEPA will improve the quality of services with the aid of this higher rate remains to be seen.

At best, the public sector's performance is likely to improve very gradually. Thus the most promising way to correct these basic inefficiencies, for Nigeria as for many developing countries, is to combine efforts to improve public sector performance with incentives for private sector participation in the provision of services. In Nigeria, some minor regulatory changes can yield significant benefits to individual firms and households and to the economy. For example, Nigerian manufacturers are not allowed to sell any excess power they may produce to the public agency or to other firms. The cost saving attainable by allowing such transactions can be large. More important, however, such regulatory changes can make the electric utility market more nearly contestable. Large manufacturing firms with unused generating capacity can produce more electricity and sell their excess power to other firms. Because of the economies

of scale in electricity generation, the firms will be able to lower their price as they increase their production of electric power. According to the survey results, at full utilization of their generating capacity, the average cost would decline from 1.75 naira a kilowatt-hour for firms with 20 to 49 employees to 40 kobos for those having 1,000 or more employees. The price could decline until all excessive profits are eliminated. This would put strong pressure on NEPA to perform and compete with the manufacturing firms engaged in joint production (of electricity and their own products). Suppose such forces of contestability set a ceiling price at 40 kobos, the average generating cost of the firms in the largest category. The customers would still buy electric power from these firms even though NEPA's price is lower at 30 kobos as long as NEPA's service remains unreliable. If NEPA should deliver reliable power service at 30 kobos, which is unlikely, the manufacturers would exit from the electric utility market.

Contestability would attract third-party vendors (that is, private utility companies) into the market if they could compete with the existing manufacturing firms selling electricity. Moreover, it is quite possible that these entrepreneurs specializing in the generation of electric power will produce electricity at an average cost lower than that of the large manufacturing firms, perhaps at 30 kobos a kilowatt-hour, either because of economies of scale that can be exploited further or by use of better technology. If so, the manufacturing firms will no longer find it profitable to produce and sell electric power as joint production. Thus the utility companies in this example would set the ceiling price. Indeed, a private entrepreneur began producing and selling electricity in the Maroco low-income area in Lagos. NEPA allowed this individual to operate there because this slum area was excluded from NEPA's network. The private operator charged 30 kobos a kilowatt-hour even in this low-income neighborhood and was thriving and doing good business even though NEPA's price was only 7 kobos—a strong indication of users' willingness to pay for reliable services.

Another example of a minor regulatory change that would benefit Nigerian manufacturers is permission for the management board of an industrial area to operate and manage "utility pools," which would include a wide range of services such as electric power generation, water supply, garbage collection, and telecommunications. Such an arrangement would make it possible to exploit whatever economies of scope are available, along with economies of scale.

On the basis of what we can observe in Nigeria and from similar situations in many other developing countries, a sensible way of ending the inertia that permits continued X-inefficiency in the public sector is to encourage private sector participation in various infrastructural functions and activities, ranging from production of services to distribution, operation and maintenance, and billing and revenue collection. The extent of private sector participation in a market for the supply of a particular type of activity will depend on the contestability of such an activity. For example, production and distribution of electricity will incur larger sunk costs than operations and maintenance activities. The contestability framework will make it easier to understand variations in

the emerging market structures of particular types of activity and will offer policy choices. This seems a more promising strategy than the now popular privatization schemes, which often end up by replacing a state monopoly with a private monopoly with no regard for the underlying market structure in the economy.

The challenge is to provide an environment with appropriate incentive systems for private entrepreneurs to engage in the supply of the pertinent infrastructural services, thus creating the market mechanisms needed for effective conduct of activities such as production, distribution, and maintenance. The opportunity to create and expand such markets is opened by the failure of the public sector to perform in these countries and by the willingness of users to pay for reliable services when they become available. To the extent that the markets for particular infrastructural services are contestable because the sunk costs involved in capital investment are low, it should be possible to liberalize restrictions against the entry and exit of private entrepreneurs. In this new environment the role of the government can be redefined as that of carrying out appropriate monitoring and supervision of market operations, with little or no provision of such services by the public sector.

Conclusion

The contestable market framework, not yet ten years old, has already made its mark on the design of policy. In the United States, it has explicitly been used as the basis for the design of new approaches to policy and new relations between government and industry. Moreover, it regularly enters the deliberations of regulatory agencies, the courts, and legislatures. This article has described the key elements of the theoretical framework and attempted to extend some of the lessons learned to the areas of international trade and economic development. Our illustrative discussion of public-sector inefficiency in Nigeria has shown applicability of this concept to developing countries. The issue is particularly pertinent at this juncture, when many developing countries are pursuing development of the private sector and making use of privatization, and when the international trade regime is increasingly affected by the protectionist threat. One of the main messages of this article is that the design of policies to improve economic efficiency through private sector participation requires a clear understanding of the emerging market structure. The contestability analysis provides a tool for the construction of such policies.

Note

William J. Baumol is a professor of economics at New York and Princeton universities. Kyu Sik Lee is a principal economist in the Infrastructure and Urban Development Department of the World Bank.

1. One billion equals 1,000 million.

References

The word "processed" describes works that are reproduced from typescript by mimeograph, xerography, or similar means; such works may not be cataloged or commonly available through libraries, or they may be subject to restricted circulation.

- Baumol, W. J., J. C. Panzar, and R. D. Willig. 1988. *Contestable Markets and the Theory of Industry Structure*. Rev. ed. San Diego, Calif.: Harcourt Brace Jovanovich.
- Bhagwati, Jagdish N. 1988. "Export-Promoting Trade Strategy: Issues and Evidence." *World Bank Research Observer* 3, no. 1: 27-57.
- Harris, R. 1986. "Market Structure and Trade Liberalization: A General Equilibrium Assessment." In T. N. Srinivasan and J. Whalley, eds., *General Equilibrium Trade Policy Modelling*. Cambridge, Mass.: MIT Press.
- Kessides, I. N. Forthcoming. "Market Concentration and Sunk Costs." University of Maryland. Processed.
- Lee, Kyu Sik, and Alex Anas. 1989. "Manufacturers' Responses to Infrastructure Deficiencies in Nigeria." Policy, Planning, and Research Working Paper 325. World Bank, Policy, Research, and External Affairs Office, Washington, D.C. Processed.
- Maddison, Angus. 1982. *Phases of Capitalist Development*. New York: Oxford University Press.
- Morrison, S. A., and Clifford Whinston. 1987. "Empirical Implications and Tests of Contestability Hypothesis." *Journal of Law and Economics* 30: 53-66.
- Nogués, J. J., A. Olechowski, and L. A. Winters. 1986. "The Extent of Nontariff Barriers to Industrial Countries' Imports." *World Bank Economic Review* 1, no. 1: 181-99.
- Schwartz, Marius, and R. J. Reynolds. 1983. "Contestable Markets: An Uprising in the Theory of Industry Structure: Comments." *American Economic Review* 73 (June): 488-90.
- Summers, Robert, and Alan Heston. 1984. "Improved International Comparisons of Real Product and Its Composition, 1950-1980." *Review of Income and Wealth* 30 (June): 207-62.
- Weitzman, M. L. 1983. "Contestable Markets: An Uprising in the Theory of Industry Structure: Comments." *American Economic Review* 73 (June): 486-87.