CONTESTABLE MARKET THEORY: A REVIEW

Contestability is a concept which relates to the way markets perform. Application of the concept significantly alters perceptions of the impact of monopoly and other market structures. Contestability theory is examined in this article, with a short summary provided in the box.

Abstract

As the main article is relatively technical, this abstract attempts to capture the central ideas in brief.

Governments have a role to play in ensuring that markets perform well — that is, providing goods and services up to the quantities desired by consumers at the cheapest possible price. The economic theory traditionally relied on to analyse how markets perform suggests that there is a relatively simple connection between the number of firms in a market and how well the market performs. Monopoly (one supplier) is consistent with less than optimal performance; many suppliers with optimal performance.

A new approach to analysing markets, going under the name of 'contestable market theory', was developed over the 1970s. Instead of focussing on the number of firms competing *in* a market, it concentrated on what stops firms from successfully contesting *for* the market. Hypothetical economies ('models') in which nothing exists to prevent firms from entering and competing on an equal footing with incumbent firms indicate that a market with one supplier can perform just as well as a market with a large number of suppliers. Such models, therefore, indicate the importance of barriers to free entry (and to free exit, given that exit costs will be taken into account by prospective entrants).

The usefulness of contestable market theory as a basis for designing government policy clearly depends on how common and important are barriers to free entry and exit, in particular where those barriers are a natural part of market structures rather than a product of government regulation. A number of such 'natural' barriers can be identified, some of which are more significant than others. Many of the recent contributions to the economic debate have highlighted such factors, and in particular the relevance of strategic reactions by existing or incumbent firms in face of the threat of a new entrant. Incumbent firms can exploit barriers to entry such as strong customer loyalties, large set-up costs etc by threatening in return to drop prices so as to undercut the newcomer.

Overall, the criticisms of the contestability line have been penetrating. Contestability theory remains a significant advance, however, introducing important qualifications to traditional approaches, highlighting the role of entry and exit barriers, and stressing the need to examine markets in an ongoing dynamic fashion.

Introduction

Economic analysis of markets has traditionally emphasised market structure as a major factor in determining how well a market contributes to welfare. Efficiency is held to be enhanced the greater the number of firms competing within a market. An extreme image of 'perfect competition' has been often used to assist understanding of how economies function. This perfect competition model has a very large number of firms producing the same product, each of which individually has no influence on the price of the good traded in the market.

'Perfect competition' is at one end of a spectrum of market structures focussed on by traditional economic analysis. Three other structures of interest are monopolistic competition, oligopoly, and monopoly. Under perfect competition, each firm can only earn a 'normal' rate of profit — that is a rate of profit no greater than can be earned elsewhere in the economy. Prices and quantities are at levels which maximise economic welfare — in economists' terms, prices are equal to marginal costs, where the marginal cost is the cost of producing one additional unit of the particular good, and at this level prices and quantities are such as to maximise economic welfare. Consumers of the product are thus getting the best deal possible within the constraints of the existing production technology.

In contrast, monopoly at the other end of the spectrum is a situation in which there is only one supplier of a unique product that has no close substitutes. For a variety of reasons the monopolist is able to maintain his position as the only producer. In this world, traditional analysis holds that the monopolist is able to set the price of his output such that he is able to earn greater than normal profits. Economic welfare is severely reduced because consumers are forced to pay higher prices than is possible given the technology available, so that the production of the relevant good is reduced below its optimal level.

Oligopolistic and monopolistic market structures are intermediate cases in which either there is a small number of firms in the market or the firms produce distinctive products. In general the beneficial properties of perfect competition are also held to be absent with these market structures, although they may not be as inefficient as the pure monopoly case.

The welfare implications of this traditional approach to competition theory have been widely accepted and applied in public and competition policy and, in particular, have led to a widespread acceptance of the undesirability of monopoly structures.

Development of contestable market theory in the late 1970s and early 1980s, however, calls into question the universal application of the 'bad per se' rule to market structures other than those described by the 'perfect competition' model. Instead of placing emphasis on competition 'in' the market, contestable market theory focuses attention on competition 'for' the market.[1] The theory asserts that in the absence of barriers to costless entry or exit of the market by firms, the threat of potential entrants willing to enter a market to exploit any transient profit opportunity constrains the behaviour of firms already in the market. The threat of new entrants undercutting price and displacing an incumbent's market share induces optimal pricing across all market

[1] Baumol, W.J., Panzar J.C., and Willig R.D. (1982) in 'Contestable Markets and the Theory of Industry Structure'. Harcourt-Brace, Jovenovich draw together the undercurrents of the contestability literature of the late 1970s and early 1980s to provide an explicit theoretical statement of the mechanics and implications of the theory.

structures, including monopoly. Free movement of resources into and out of markets thus provides the check upon sub-optimal behaviour within industries and not the number of firms actively participating in the market. Markets with few suppliers may therefore exhibit perfectly competitive outcomes. The conclusions of contestable market theory are therefore significantly at variance with the traditional interpretation of market structures and implies a radically different approach to competition policy.

The remainder of this article provides a brief description of contestable market theory, drawing together the main threads of the theory, and then discussing its applicability to real world markets.[2]

Contestable Markets

Assumptions

To arrive at the fundamental conclusion of economic welfare-maximising prices and quantities in a perfectly contestable world the theory makes a number of simplifying assumptions. First, where there are sunk costs (e.g. costs of capital equipment which cannot be resold) associated with entry into the market by a new firm, it is assumed that those costs can be recovered (through profits) in less time than it takes for incumbent firms to make price changes. Prospective entrants can thus act on the assumption of no price change by the incumbent, and the attractiveness of entry can be evaluated at the pre-entry price level of the incumbent firm. Both incumbents and potential entrants are constrained by consumer demand and a market will not withstand increased quantity without downward pressure on price.

Secondly, it is assumed that consumers respond instantaneously to price differentials between competing suppliers allowing full return on investment to the new entrant immediately.

The third assumption of free entry and exit provides the foundation upon which the theory is based. For an entrant to face no significant lags or costs in the entry process, it must be able to instantly replicate all dimensions of size, technology, costs, product array, brand loyality, perceived product quality and other attributes or advantages of the existing firm. There also need to be no other special costs borne by the new entrant that are not faced by the incumbent — the case of sunk costs having already been dealt with. Legal restrictions on size, numbers, structure and such like are precluded from inhibiting entry or exit. Capital has to be highly mobile from market to market, otherwise there has to exist a well developed lease or resale market allowing easy liquidation of investment with no unusual loss of value after its use in a particular market. Furthermore, it is assumed that 'human capital' is not industry specific and, like physical capital, can be easily deployed in alternative markets without large personnel costs such as retraining. Only if these conditions prevail will there be no lags. It is important that both entry and exit are free from impediment, as the decision to enter will take account of expected exit conditions.

The Model

Given the assumptions above an incumbent firm must price optimally at the competitive level otherwise a new entrant may enter with zero or minimal risk, and profitably undercut the market. The theory rests crucially upon this dichotomy between incumbents and a pool of potential entrants. Incumbents' behaviour is disciplined by the existence of aggressive firms ever ready to enter at the slightest sign of profitability or inefficiency.

The concept is therefore very simple — in the absence of entry and exit barriers a firm must price its product at the minimum level consistent with making only normal profits, otherwise above normal returns induce new entrants to enter, undercut price and eliminate both supernormal profits and the existing firm. Market performance and the degree of 'contestability' follow directly from the presence or absence of barriers to entry and exit. Contestability, and likewise welfare, may therefore be interpreted as a function of barriers to entry and exit, rather than as a function of the number of incumbent firms. Indeed, if all the assumptions of the theory hold, even a market with a single producing firm will produce an outcome of equivalent benefit to society as that which would be produced by 'perfect competition'.

Entry Barriers, Sunk Costs and Contestability

There are, of course, degrees of contestability. Perfect contestability represents the limiting case in which no barriers to entry or exit exist. The imperfect case in which barriers do exist can be introduced. In the presence of barriers the theory asserts that incumbents can earn economic profits above zero, but only up to the level of economic 'rent' derived solely from the entry barrier. The potential threat of entry constrains profit to that level.

Entry barriers are defined as anything imposing an expenditure or cost upon the entrant but not upon the incumbent. The theory tends to focus on the presence of sunk costs, rather than of economies of scale, as the major source of production-related entry barrier. This is because the problem of scale economies is assumed away in the theory by the assumption that existing capital equipment, including fixed (but not sunk) costs, affects neither entry nor exit costs. The large fixed costs associated with economies of scale do not therefore constitute a barrier to entry.

Long-term fixed costs are costs that can only be eliminated by the cess-

[2] A more extensive critique is contained in Lloyd, K.A. (1986) 'Contestability in Context', Reserve Bank of New Zealand Discussion paper G86/5, Wellington.

ation of production. Sunk costs, on the other hand, cannot be eliminated by closing down production. Expenditure on legal and other fees necessary to create the firm structure, advertisisng expenses needed to make consumers aware of the presence of a new entrant, the loss of liquidity involved in the transfer of liquid assets into fixed capital etc. are all costs which tend to a greater or lesser degree to be irreversible. The presence of such costs creates a difference between the prosective new entrant and the incumbent. The incumbent has already spent money on sunk costs, and can consequently ignore those costs in assessing the best response to the threat of entry. The prospective entrant, however, faces the choice of whether or not to spend money on sunk costs. It appears, therefore, that incumbent firms have a direct incentive to increase barriers to entry through advertising, research and development, etc., because the building of such barriers allows increased profits — at the expense of decreased welfare for consumers and society in general.

The discussion so far has centered on production-related barriers to entry i.e. those that might be a natural part of the marketplace, or of the technology being used. Such barriers have, as already noted, been the main focus of the theory, probably in large part because they involve key questions about the real world applicability of the theory. There are of course other forms of barriers to free entry and exit which arise from regulatory restrictions, legal requirements etc. Some of these regulatory barriers can readily be removed. Other types are, however, more problematic for the contestability theorists. For example, the establishment of property rights as in the development of patent rules is seen by most to be a necessary part of providing incentives for research and development and innovation. Patent rules, however, prevent the ready duplication of the incumbent firm's technology and so may facilitate monopoly pricing in the absence of further safeguards.

Policy Implications of Contestable Market Theory

Contestable market theory has potentially important implications for the interpretation of market performance and for the design of economic policy (in particular 'competition' or 'antitrust' policy). The welfare maximising results of the theory logically apply to all market structures that meet the theory's assumptions, and in particular to monopolies and oligopolies which have traditionally been seen as output restricting and welfare diminishing. According to the theory, the number of firms in an industry is simply the minimum number necessary to supply total demand and therefore does not reflect market power but is instead the outcome of efficient production and pricing. The existence of monopoly or oligopoly does not automatically imply inefficiency or welfare loss and should not necessarily draw the attention of regulators or encourage antitrust action.

Instead, entry and exit barriers are seen as the appropriate targets of policy. Sunk costs are identified as the major source of entry barrier and potential competition does not become an effective force until these large, irretrievable entry costs decline. This implies that competition policy should concentrate on eliminating the power associated with controlling sunk facilities. This may be achieved in some cases by government ownership of sunk facilities (for example, airport runways), or in other cases, by encouragement of consortium or multiple ownership (for example, satellites). Contractual or other arrangements must be non-discriminatory and allow, for example, easy transfer of leases or shared use of sunk facilities. Detachment of sunk facilities from the firm eliminates (within the model, that is) much of the need for government intervention and prevents excess profits being extracted from sunk costs. The theory therefore suggests policymakers should concentrate on removing certain types of regulatory restrictions, ranging from those on competitive pricing, to those on market size, numbers, etc.

Criticisms of Contestable Market Theory

Contestable market theory provides a framework in which entry barriers are represented solely as a return over and above normal returns. Profit above the imputed value of the entry barriers induces entry and provides a check upon excessive profit-taking. While the theory is conceptually appealing it has been strongly criticised on a number of grounds. The strongest criticism, perhaps, is the failure of the theory to include analysis of the strategic interactions between potential entrants and incumbents. Traditional theory has analysed behaviour in imperfect market structures from a strategic perspective using a wide array of 'reaction' models. Contestable market theory seems to go only part way in examining these factors. While explicitly recognising the power of potential entrants the theory does not examine any retaliatory strategic or tactical responses by the incumbent. This follows in part from the asumptions made about the feasible speed of pricing changes made by the incumbent. In reality it is likely that incumbents would take action to preserve their market position in the face of entry. A whole range of strategic moves can be taken by the incumbent to strengthen their position and build up real or perceived entry barriers. Similar measures can be taken to pre-determine the outcome of any post-entry price war. The assumed absence of a counter-reaction by incumbents is clearly important for the validity of the policy prescriptions of contestability theory.

Many of the other assumptions made also turn out to be important. The freedom of exit assumption is required as it enables both incumbents and entrants to liquidate their investment in a market at little cost. It appears almost without exception that production requires specific assets which cannot be transferred or sold costlessly. This applies to physical equipment, advertising, research and development, expert skills and other commitments needed to establish entry. Human capital may

not be instantaneously transferable between markets as training, relocation, etc. may be required. Friction of this sort represents a sunk cost and is therefore an entry barrier allowing the extraction of above normal profits.

The assumption that consumer demand will shift instantaneously to lower priced products is also very stringent. It has important implications as it ensures an immediate return on investment to the new entrant, so reducing risk. Furthermore, sluggishness on the demand side creates market power and, in some instances, represents an entry barrier. Strong brand loyalty or consumer preference for existing products has been found in a number of empirical studies and was regarded as an obstacle facing new entrants in a study of the deregulated United States aviation industry. There are also early indications that established banker-customer relationships have been important barriers to costless entry of new banks following the recent Australian banking system deregulation.

It is further assumed that information is freely available in perfectly contestable markets allowing entrants to readily replicate existing production techniques. While this somewhat unrealistic assumption is used in a wide range of economic models it is doubtful whether potential entrants could immediately duplicate the expertise and knowledge of incumbents. Efficiency gains derived from experience or 'learning by doing' are clearly inconsistent with this assumption.

An inter-related criticism is the 'total entry' aspect of new entry into contestable markets: entry is supposed to be absolute, totally displacing the incumbent. In reality entry occurs as a continuous process, beginning at a preferred scale of output and expanding as and if conditions are favourable. Typically, initial entry is partial in order to minimise risk and permit learning and growth as circumstances allow. Observed entry behaviour therefore appears inconsistent with that postulated by contestable market theory. If entry is typically small scale then high fixed costs present a barrier as fixed cost per unit is higher for the entrant producing at a smaller scale of output. This would therefore allow the incumbent to extract economic profits without inducing entry.

In the light of these criticisms where does this leave contestable market theory?

Contestability in Perspective

Proponents of contestability claim the theory provides a more 'general' and applicable view of markets than traditional analysis and in the extreme, some suggest traditional theory is redundant. In the light of the discussion above one must, however, treat the more extreme claims with some sceptism. The theory is, however, intuitively appealing and provides significant insights into the effect of barriers on competition. Furthermore, it shows that under certain conditions all market structures could generate the level of welfare commonly associated only with perfect competition.

Without the way of thinking about markets and competition provided by contestability theory, quite improper conclusions can be drawn from traditional economic analysis and used as the basis for government regulation. For example, a single dominant producer of a particular product line has in the past been typically regarded as undesirable. Contestability theory requires one to consider the reasons for the market dominance, thereby more readily allowing for the possibility that the market is in fact quite contestable and that dominance has occurred because of the particular efficiencies of the single producer.

Development of the theory represents an important synthesis, focussing attention on barriers as significant market determinants. The effect of barriers on internal market behaviour has traditionally been analysed in a negative sense — i.e. the negative effect of barriers allowing incumbents to earn economic rent. Contestability turns this interpretation around and focusses attention on the positive effect of removing barriers. The theory also points towards a dynamic interpretation of markets. The ability of firms to enter on an ongoing basis contrains market behaviour of incumbents. The degree of contestability of a market may be changing over time with technology, regulatory breakdown, or changes in other barriers altering the entry and exit conditions. An incumbent pricing optimally can protect himself against new entrants using the same technology, but he cannot protect against innovation or technological advancements.

Contestability has also highlighted the fact that competition for a market is not necessarily restricted to those firms falling strictly within the traditional industry boundaries. Instead 'contestability' forces one to define industries, markets, etc. broadly to take account of substitute products and international as well as domestic potential competitors.

It is these wider notions of emphasis on the importance of barriers, a dynamic interpretation of markets, and the focus on broader industry definitions which are the major contributions of contestable market theory to our understanding of how real markets work. Application of the theory in a policy environment requires careful consideration of the type of market under examination, and should not be utilised without regard to traditional competition concepts. The intuitive concepts and wider notions of the theory are very useful but should not be used in isolation of the considerable knowledge accumulated on the behaviour of firms and market structures within traditional theoretical frameworks. Contestability is a useful concept which adds another perspective to policymakers' interpretation of markets.