

***CONTESTABILITY: A NEW THEORY OF NATURAL MONOPOLY
AND A VAIN QUEST TO REGULATE EFFICIENCY***

John Brätland

**U.S. Department of the Interior
1849 C Street N.W., Mail Stop 4230
Washington, D.C. 20240**

john.bratland@MMS.gov

(202) 208 3979

WORKING PAPER

SEPTEMBER 2, 2003

ABSTRACT

Under contestability theory, monopolists restrain their pricing behavior by the threat of entry by competitors. Hence, industry structure provides no *a priori* rationale for regulation. However, the analytical elements of contestability theory purport to define and, through regulatory means, attain an 'efficient' industry structure producing at lowest cost without profit. These analytical elements of contestability include *economies of scope*, *subadditivity* and *sustainability*. However, contestability theory ignores the analytical implications of uncertainty, change and entrepreneurship. *Economies of scope* occur if joint production of a slate of products is empirically revealed to be less costly than if each product were to be produced by a separate firm. *Subadditivity* demonstrates that economies of *scale* and *scope* are such that a single firm is the least costly means of satisfying a specific demand for a specific slate of products. However, even with only one firm, competitive pricing is achievable if: (1) prospective entrants are assured of entry free of barriers and the ability to recoup entry costs upon later exit; and (2) the incumbent firm is induced by threat of entry to charge *sustainable prices* such that no profitable entrepreneurial entry is possible. But regulation is thought to necessary if (a) subadditivity holds but no sustainable prices exist to forestall entry or (b) the market is uncontested because of entry barriers and unrecoverable entry costs. However, in both cases, regulatory responses are critically reliant on *opportunity costs being objective and measurable*. But opportunity costs are only objective in a static equilibrium devoid of uncertainty, economic change and genuine entrepreneurial activity. Hence, the policies aimed at increasing economic efficiency are necessarily empty and irrelevant.

CONTESTABILITY: A NEW THEORY OF NATURAL MONOPOLY AND A VAIN QUEST TO REGULATE EFFICIENCY

John Brätland¹

I. Introduction

William Baumol, John Panzar and Robert Willig first presented the theory of contestable markets in their 1982 book, *Contestable Markets and the Theory of Industry Structure* (Baumol et. al. 1988 [1982]). The theory is based on a simple and sensible idea; the threat of entry can induce multi-product incumbent firms in an industry function to moderate pricing behavior even in industries with only a single firm. In this sense, contestability theory offers a new theory of ‘natural monopoly’ and the way in which consumers’ interest are best served by the way in which such a firm can respond to the pressure of threatened entry into the industry. Natural monopolies are traditionally assumed to be characterized by economies scale that make marginal-cost-pricing unworkable. To preserve the benefits of scale economies, conventional thinking has traditionally been guided by the doctrine that such firms must be regulated. But contestability theorists intend to at least partially debunk this notion.

In her foreword to the book by Baumol et. al., Elizabeth Bailey observed that contestability theory is intended to provide “*a new theory of industrial organization...that will transform the field and will render it more applicable to the real world....because of the observability of its properties, the theory of contestable can be extremely helpful in the design of public policy*” (Bailey 1988 [1982], vi, xv) (Emphasis added). In an oddly self-congratulatory tone, these authors go on to describe contestability theory as providing a new standard of welfare-maximizing structure and behavior for industries. This ‘new standard’ of structure and behavior focuses mainly on markets in which multi-product firms have few or no actual incumbent

¹ John Brätland is a Ph.D. economist with the U.S. Department of the Interior. The views expressed in this study are strictly those of the author. The author wishes to thank Robert Bradley for thoughtful comments on an earlier version of this paper.

competitors but which presumably represent the least costly means of satisfying consumer demand for a particular slate of products.² Hence, within the contestability paradigm, efficiency is defined as that industry structure that produces the goods demanded by consumers at the least cost without profits (monopoly rents) being earned by the firms in the industry.

Modern contestability theory challenges 'conventional neoclassical wisdom' on two fronts.

First, the traditional definition of natural monopoly is abandoned and replaced by one combining economies of scale and economies of scope. Natural monopoly exists if a single firm is able to exploit both scale and scope economies and hence produce a particular listing of goods more cheaply than the same combination of goods could be produced by some industry configuration including more than one firm (Train 1991, p.8). The second way in which contestability theory breaks with the traditional neoclassical wisdom is in arguing against presumptive regulation of the natural monopolist. If the market in which the natural monopoly functions is in fact contestable, the pricing behavior of the incumbent firm can be disciplined by the threat of entrepreneurial entry by potential competitors. In other words, the threat of entrepreneurial entry will induce something approaching 'competitive pricing' on the part of the incumbent monopolist.³ Under this deterrence, incumbent firms in the industry earn no profits but only a competitive rate of return on investment. By this standard, contestability theory is squarely in

² On the nature of multi-product production, Baumol and his colleagues note that "this multiplicity can take the form of physically dissimilar offerings, a wide variety of offerings of similar outputs (such as shoes of different sizes) adapted to the demands of individual customers, or just physically similar outputs sold at various places or times. For all of these cases the received theory of demand is well developed, but the theory of production as it relates to industry structure, is not" (Baumol et al. 1988 [1982], 3).

³ In one sense, contestability theory represents a distinct move away from the 'structuralist' or 'concentrationist' approaches to industrial organization theory criticized by Dominick Armentano in his book *Antitrust and Monopoly: Anatomy of Policy Failure* (1990 [1982], 32-40). The essence of this approach to industrial organization is that *existing* market structure or concentration determines pricing behavior; hence, few or one firm in an industry necessarily means an absence of competitive pressure in pricing. However, Armentano is highly critical of the analytical and informational assumptions that seem to underlay antitrust policy and convincingly makes the case that efficiency is not a state of the market that is amenable to empirical measurement. Rather, Armentano defines efficiency in the following way: "If the means employed in the pursuit of selected ends are consistent with those ends, then the means or plans are said to be efficiently employed" (1990 [1982], 29). While Armentano uses the term 'social efficiency' in his discussion, he makes clear that the latter concept can have no meaning outside of the actions of the individual pursuing personally chosen objective. Both Murray Rothbard and Armentano tie the possibility of the individual attaining efficiency to the availability of perfect or complete information which both explicitly acknowledge as an impossibility (Armentano 1990 [1982], 29; Rothbard 1978, 90). Also, Armentano and Rothbard would condemn the notion that any legitimate economic efficiency could ever be achieved through the regulatory actions of government. Rothbard observes that "Efficiency, therefore can never serve as utilitarian

the neoclassical tradition in that zero economic profits are a presumptive indicator of economic efficiency. True contestability is characterized by those situations in which prospective entrants are potentially able to enter the industry and, if necessary, later exit the industry in such a way that they are able to recover all or nearly all of the 'costs' associated with initial entry. Issues related to asset specificity are assumed not to be a problem in recouping costs. Under the precepts of contestability theory, if these conditions are met, market structure, in itself, does not provide a presumptive case for regulatory intervention (Spulber 1989, 142).

But contestability theory presumes to provide a framework in which inefficiency can be detected and in which corrective regulatory sanctions can be imposed. Contestability theory is intended to provide an analytical means to implement an ambitious regulatory agenda designed to deal with either of two sources of 'inefficiency.' *First*, empirically discernable conditions of the market may reveal that a single multi-product firm is the least costly (most efficient) means of satisfying the consumer demand for the slate of products being produced. However, contestability theory claims that regulators are also able to empirically detect situations in which no schedule of prices are available to the incumbent monopolist that will forestall entry of competitors and the loss of production economies. In the name of efficiency and economic welfare, the regulating authority has the normative obligation to prevent the entry of new competitors. *Second*, situations arise in which markets with as few as one firm are found to be *uncontested*. In these circumstances, the new theory purports to provide a coherent framework for determining which prescriptive pricing sanctions are 'welfare enhancing.'

Some of the paradoxes and contradictions of the theory arise from the fact that it was developed during a period that coincided with the deregulation of major industries in the United States. Some of these previously regulated industries include telephony, electric power, airlines, and railroads. One could reasonably assert that under previously existing regulatory regimes, these industries were relatively static. However, deregulation has meant that more dynamic entrepreneurial change will almost certainly characterize these industries. Contestability theory fails to address the fact that in a setting in which entrepreneurial entry is a reality, the implicit

assumptions of a static neoclassical framework no longer pass muster. In this sense, the theory reflects an analytical disconnect; it introduces the prospect of *ex ante* entrepreneurial entry into a market framework that is also assumed to be sufficiently transparent and objective to allow ‘efficiency-enhancing regulatory intervention.’ Clearly, this postulated transparency is necessarily contingent on strong equilibrium assumptions. This presumed transparency is particularly evident in the theory’s reliance upon objective opportunity costs and econometrically-estimated cost functions. Objective information on cost is to be used by regulators in assessing contestability and establishing price ceilings and price floors when contestability is discovered to be ‘imperfect.’ Yet at the same time, contestability theory claims to have regulatory relevance to the inherently entrepreneurial activity of entering and exiting an industry. *It presumes to apply a theory of entrepreneurship in a market setting in which true entrepreneurs do not and cannot exist.*

Some economists have had difficulty in discerning the implications of this paradox. For example, in his book, *Regulation and Markets*, Daniel Spulber has accurately observed the following “The contestable-markets literature has generally focused on a static equilibrium with complete information regarding prices, outputs, demand and costs” (Spulber 1989, 139). Spulber is acknowledging that the entry and exit decisions that are central to contestability are addressed analytically in the assumed absence of market uncertainty. But, oddly and surprisingly, Spulber goes on to note that “the admittedly ideal case presented by the theory of contestable markets is both useful and important for the study of regulation since it provides a standard of comparison against which market equilibria with entry barriers can be judged” (Spulber 1989, 139). But is this ‘ideal case’ really useful in providing standards of comparison? This paper addresses this question and explains why the answer must be ‘no.’

Actual contestability must be manifested in an entrepreneurial world characterized by genuine-market uncertainty -- not a static equilibrium world as would be required for empirical objectivity. True market contestability necessarily implies entrepreneurial judgement. But in the literature of contestability, there is virtually no mention of market uncertainty and certainly little

mention of entrepreneurship.⁴ The extent to which entry and exit can be profitably accomplished must, in reality, be a totally speculative conjecture on the part of the entrepreneur and not a matter that can be discerned by a regulating authority assessing objective data. There is no empirical content that would provide a framework for the ambitious regulatory agenda envisioned by Baumol and his cohorts. *Ex ante* contestability is fundamentally subjective. But this reality is ignored as evinced by the approach to opportunity cost and cost functions taken in suggested regulatory applications of contestability theory. These and other features of contestability theory are the basis of the following examination.

II. Basic Elements of Contestable Market Theory

While contestability theory is actually intended to provide a framework for regulation of natural monopoly, it also outlines an allocative ideal in which a policy of *laissez faire* is ‘most efficient.’ This allocative ideal is characterized by three principal features. First, potential entrants would be able to respond to profit opportunities within the industry yet at the same time be assured of recouping entry costs should later exit become necessary. This condition would establish *contestability*. Second, the natural monopolist would be able to produce a listing of goods more cheaply than the same listing of goods could be produced by some industry configuration including more than one firm. If this condition is satisfied, the monopolist's cost function is said to be *subadditive*. Third, prices charged by this natural monopolist would be *sustainable* in that they would offer no prospect of profitable entry by a prospective competitor; that is, prices would cover costs and allow only a ‘competitive rate of return’ for the natural monopolist.

However, these ideal conditions may not be fully satisfied in which case regulatory intervention

⁴ In their book, Baumol, Panzar and Willig manage to mention the term ‘entrepreneur’ four times – once on page 25 and three times on page 350. It is noteworthy that, in an earlier publication Baumol himself was a thoughtful critic of the very type of model he advances under the label of ‘contestability.’ “In its simplest form, the theoretical firm must choose among alternative values for a small number of well defined variables.... Explicitly or implicitly the firm is taken to perform a mathematical calculation which yields optimal values for all decision variables....There matters rest, forever until exogenous forces lead to an autonomous change in the environment.... *Obviously the entrepreneur has been read out of the model. There is no room for enterprise or initiative. The management group becomes a passive calculator that reacts mechanically to changes imposed upon it....it does not exert, and does not even attempt to exert, any influence...*” (Baumol 1968, 67) (Emphasis added).

is thought to be in order. But establishing the need for a corrective regulatory agenda is not possible unless *contestability*, *subadditivity* and *sustainability* have objective, measurable content. This content is critically dependent on the ability of regulators to examine the cost structures within firms. But, as the following analysis will attempt to demonstrate, if these cost structures are to be interpreted as reflective of actual opportunity costs, the economic environment must satisfy strictly static equilibrium conditions devoid of uncertainty and possibility of entrepreneurial entry or exit. This paradox is examined in greater depth throughout the remainder of the paper.

A. The contestability criterion: freedom of entry and costless exit

The requirement that exit from the industry be costless has been one of the more controversial aspects of contestability theory (Shepherd 1984, 572-577). Nonetheless, to the extent that this condition holds, the incumbent monopolist would be forced to face the contingency of ‘hit and run’ entry by prospective competitors. This prospect would condition the pricing behavior of the incumbent firm so that no profits would be earned on any of its products over and above a normal, competitive rate of return. In other words, the single incumbent firm would be forced to deal with potential competition rather than actual competition from firms already in the industry. But this general insight is not new; the concept was anticipated by Joe Bain in his 1956 book, *Barriers to Competition* in which he observed “that most analyses of how business competition works and what makes it work have given little emphasis to the force of potential or threatened competition of possible new competitors placing a disproportionate emphasis on competition among firms already established in any industry; [and] that so far as economists have recognized the possible *importance* of [the former] they have no very good idea of how important it actually is” (Bain 1956, 1). The economist Wayne Leeman came to similar conclusions in a paper published in the same year as Bain’s book. In describing Leeman’s research, Murray Rothbard, observes “Leeman concludes, quite correctly, that large rather than small firms dominate many markets, not as a result of victorious cutthroat competition and monopolistic pricing, but by taking advantage of low costs of much large-scale production and keeping prices low in fear of *potential* as well as actual rivals” (Rothbard 1962, 910; Leeman 1956, 333-334).

While the essential idea contained in contestability theory is far from new, in acknowledging this fact one must draw a sharp distinction between the earlier work of Leeman and Bain and the later theory as put forward by William Baumol and his colleagues. Baumol, et. al. make several heroic claims for the theory contestable markets that some earlier economists would not have been prepared to make. For example, (1) a perfectly contestable market permits the monopolist to earn the same profits it can obtain in a perfectly competitive market, and no more. Any earnings that exceed the cost of capital will attract new entrants, who can undercut the incumbent monopolist's prices and take its customers away. (2) Perfectly contestable markets exclude any firm that is inefficient (firms failing to produce at "minimum costs") because inefficiency would enable efficient entrants to take business away. Finally (3) cross-subsidy cannot endure in a perfectly contestable market. That is, in a multi-product firm, each product X must earn revenue at least sufficient to cover the increments in the firm's total cost that result from provision of that product. This is so for the reasons just noted because no other product Y can provide excess profits with which to offset any losses incurred in the production of X . Each product must fully pay its own way if the firm is to avoid insolvency. Hence, in a perfectly contestable market, as in a perfectly competitive one, the prices that prevail will be those required for economic efficiency.⁵ But what is the likelihood that these criteria have any operational content? One notes that such content requires a degree of empirical transparency that could in principle only arise in the most static equilibrium environment free of uncertainty, economic change and entrepreneurial activity.

B. Economies of scope and their source in the multi-product firm

The traditional neoclassical definition of natural monopoly presumes to identify a situation in which economies of scale in the production of a single product are such that the monopolist is the least costly means of satisfying market demand. Economies of scale are present when the average cost of producing a single unit of output decreases as output expands. But economies of

⁵ The three items in this list are those presented by Baumol and Sidak (Baumol and Sidak 1994, 43-44). In his book on contestable markets, William Tye claims another performance criterion; that is the un-profitability of predatory pricing (Tye 1995, 4). This latter issue is addressed below.

scale may be manifested at some ranges of production but not for others. At low levels of output, scale economies may be evident but for expanded levels of production, economies of scale may disappear. Hence, the extent to which a natural monopoly is thought to exist depends upon the degree to which market demand can be satisfied by a volume of output over which scale economies apply. If, however, market demand exceeds this level of output, then the phenomenon of natural monopoly may still hold if market demand can still be met at a lower total cost than if more than one firm were in the industry (Train 1991, 5-6).⁶

The theory of contestable markets is a revamping of the *neoclassical definition of natural monopoly* made necessary by the introduction of what has been labeled ‘economies of scope.’ If the firm is producing more than one product, a situation can arise in which market demand for the affected products will support only one firm. In other words, “economies of scope are said to exist if a given quantity of each of two or more goods can be produced by one firm at a lower total cost than if each good were produced separately by different firms” (Train 1991, p. 8). Suppose three different products X , Y , and Z can be produced separately by three different firms at respective financial outlays of $C(X)$, $C(Y)$ and $C(Z)$. However, the three products can be produced by one firm at a total financial outlay of $C(X, Y, Z)$; if the condition $C(X, Y, Z) < C(X) + C(Y) + C(Z)$ exists, then economies of scope are present. As with economies of scale, economies of scope can be realizable at some levels of output but not at others.

The implicit assumption made in the theory of contestable markets is that the external observer is

⁶ Train emphasizes a sharp distinction between pecuniary and non-pecuniary costs:

A distinction is necessary between “pecuniary” and “non-pecuniary” economies of scale. Often large firms can negotiate with suppliers to obtain lower prices for inputs than would be charged if the firm were smaller. The average cost therefore declines as the size (i.e., output) of the firm increases.. However the reduction in average cost represents simply a transfer of income from suppliers to the firm, such that the total cost to society (including both the firm and suppliers) is unaffected. Reductions in costs that represent transfers only are called pecuniary, while those that represent actual reductions in resources used per unit of output are called non-pecuniary. From a social perspective, only non-pecuniary cost are relevant. We therefore only use the term economies of scale only in reference to non-pecuniary economies (1991, pp. 5-6).

While Train’s observation is certainly correct, it seems to ignore the fact that the ability of the large firm to bargain for lower input prices would not be possible unless there were actual non-pecuniary economies that could be exploited on the part of suppliers or the supplier’s suppliers.

availed of objective, measurable data on the way in which economies of scope affect the opportunity costs of the multi-product firms. This implausible assumption is made to appear even more questionable in light of the likely source of scope economies. Baumol, et al. make reference to what they refer to as a *public input* that once employed in the manufacture of one good is at the same time available at minimal cost in the production of additional goods (Baumol et. al. 1988 [1982], 76). A variant of this idea would be the production of a set of goods that require a common overhead. Other views hold that scope economies arise from innovative means employed to exploit excess capacity within the firm (Baumol et. al. 1988 [1982], p.76). While the reference to excess capacity is not particularly convincing, this latter observation raises issues that may be more congenial to Austrian perspectives on the firm. Frederic Sautet refers to the capacities of a firm to “generate and explore synergies within its own structure” (Sautet 2000, 92). Sautet acknowledges that these synergies can arise from production processes that share the name assets but adds that “it can also be seen in relation to the stock of knowledge (capabilities, routines, and interconnectedness between assets)..... and coordination of learning processes inside the firm [which] leads to the emergence and exploitation of organizational knowledge” (Sautet 2000, 92). In other words, firms have the capacity to integrate existing dispersed knowledge and to generate new knowledge. But one must note that even though exhaustive case studies may well reveal insights into the ways in which these scope economies may arise in particular firms, it certainly not evident that such information can ever take on an ‘objectified form’ useful or valid in regulating industry structure.

C. Subadditivity: The basis of a new theory of natural monopoly

Subadditivity redefines the concept of ‘natural monopoly.’ It refers to an industry structure in which one multi-product firm is able to produce a combination of goods at a combined cost that is less than the cost that would be born by several firms manufacturing the individual goods or smaller sub groupings of goods. The concept is distinct from and more inclusive than “economies of scope.” Both concepts refer the same listing of goods but *subadditivity* refers to a situation in which a single firm is able to exploit both economies a scale and economies of scope and hence produce any combination of the same listing of goods more cheaply than the same

listing of goods could be produced by *any* alternative industry configuration of more than one firm. (Baumol et. al. 1988 [1982], 170) With presumptive subadditivity, the entry of new firms means that a “loss of efficiency” since both the incumbent and entrants would be supplying the market at a higher cost. As the argument goes, with the market divided between two or more firms, consumers will not be served as efficiently as with one firm.

In most discussions of subadditivity, an effort is made to express the concept mathematically. In the single product case, traditional economies of scale are the principal source of subadditivity. Establishing subadditivity requires a comparison between the presumably objective costs of producing the product by one firm or n separate firms. For the multi-product natural monopolist, both economies of scale and scope must be capable of quantitative examination. A mathematical example posits a vector listing of products expressed as $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m)$ that apparently meet a particular market demand. But this market demand for all of the \mathbf{q}_j can be met by n different planned schedules of production, $(\mathbf{q}^1, \mathbf{q}^2, \dots, \mathbf{q}^n)$, as might be undertaken by different firms. In this formulation, the subscripts represent individual products and the superscripts represent production plans that may arise from different industry configurations (Tirole 1988, 20). An individual production plan \mathbf{q}^j may contain some or all of the total quantity supplied of \mathbf{q}_i and the plan may contain some or all of the m products in the listing. The vector for \mathbf{q}^j could be represented by $\mathbf{q}^j = (\mathbf{q}_1^j, \mathbf{q}_2^j, \dots, \mathbf{q}_m^j)$ but some of the respective components may be zero. Each of these n production plans can be collapsed into a single production plan of one firm. This ‘collapsing’ is represented as the sum of the respective n vectors \mathbf{q}^j .

$$\sum_{j=1}^n \mathbf{q}^j = \left(\sum_{j=1}^n \mathbf{q}_1^j, \sum_{j=1}^n \mathbf{q}_2^j, \dots, \sum_{j=1}^n \mathbf{q}_m^j \right) \quad (1)$$

Hence, the test for the subadditivity condition for the multi-product natural monopoly is expressed as follows:

$$\sum_{j=1}^n C(\mathbf{q}^j) > C\left(\sum_{j=1}^n \mathbf{q}^j\right) \text{ for all vectors } \mathbf{q}^j \text{ such that } \sum_{j=1}^n \mathbf{q}^j \neq \mathbf{0}. \quad (2)$$

But subadditivity is more easily stated than detected and is largely acknowledged as such by its neoclassical champions. Baumol, et. al. note that “unfortunately, the intuitive appeal of the subadditivity concept is counterbalanced by its analytical elusiveness. We often cannot recognize whether a particular function is or is not subadditive simply by looking at its mathematical expression or its graph. Moreover, there apparently exist no straight-forward mechanical criteria that permit us to test whether or not a particular function is subadditive. This is a very demanding task for empirical work because it is likely to require data well outside the range of available observation” (Baumol et. al. 1988 [1982], 171). In a similar vein, William Sharkey calls attention to the fact that “by the very nature of subadditivity, which requires comparison of a single firm's costs with an infinity of alternatives, no regulator can be entirely certain that the firm it regulates is a natural monopoly” (Sharkey 1982, 84).

But is the problem deeper than these acknowledgements would suggest? Can an essentially static concept such as subadditivity *ever* function as an ‘operational’ guide in regulating industry structure?⁷ While the caveats raised by Baumol et. al. and Sharkey are obviously legitimate, their concerns seem to be limited to the possible unavailability of data. But the more troubling issues about subadditivity seem to be tied to the fact opportunity costs do not exist independently of actual decisions and that the nature of the decision affects the degree to which observable data have the desired economic content. For example, in changing and uncertain markets, do ‘data on financial outlays’ measure opportunity costs? If not, no means exist by which to define \mathbf{n} , the number of production plans to be used in making the comparison shown in (2) above. Even if $C(\mathbf{q}^j)$ could be legitimately defined as an objective cost function, with \mathbf{n} as indeterminate, no means are available to quantitatively establish how the vector of products \mathbf{q} is distributed between

⁷ The term “operational” has its own history in the field of economic methodology (Caldwell 1982). The intended meaning here is very practical and concurs with the following observation: “something must be measurable ... which allow either the corroboration or the refutation of central propositions. ... The science consists of efforts to predict the effects on human behavior induced by specific changes in the environment. Operability dictates that the behavioral responses be objectively measurable” (Buchanan 1969 38-39).

hypothetical production plans. Also, what do cost functions actually measure and how does this measurement relate to a firm's valuation of opportunities relinquished by decisions to produce a particular aggregation of goods. Under what circumstance would the content of such a valuation actually be readily available to a regulating authority presuming to monitor and regulate an industry. Assuming that enhancement of efficiency and economic welfare are the ends sought by regulation, can any legitimate prescriptions be made in the absence of firms' actual valuation of opportunities relinquished? Do the opportunity costs for potential entrants have any logical meaning or analytical significance prior to a firm's decision to enter the industry? These questions are explored at greater length below.

D. Sustainability of prices and industry structure

A natural monopoly, as defined here, may function as a competitive market. But is the natural monopoly a *sustainable industry structure*? Three conditions must be satisfied for the prices charged by the natural monopoly in the industry to be sustainable. *First*, the quantities demanded of the goods defining the market must equal the quantities of the outputs produced. *Second*, the prices received by the producing firm must yield revenues no less than the full costs of production. *Third*, there must be no opportunity for profitable entry by potential entrants (Baumol et. al. 1988 [1982], 5). In the event that there is only one firm in the industry, sustainability refers to a set of prices that can be charged by the monopolist that simultaneously allows the firm to avoid losses on any of the products produced yet at the same time creates no inducement for other firms to enter the industry and produce all or some of the products produced by the incumbent firm. The absence of an inducement for potentially rival firms to enter the industry generally means that profits are not earned by the incumbent firm and no cross subsidization occurs in the production and sale of any of the products. The incumbent firm "breaks even and additional entry is unprofitable" (Spulber 1989, 138). "Given free entry and competitive pricing strategies, sustainable market equilibrium prices must be subsidy free. If prices are not subsidy-free, then revenues for some output or set of outputs exceed their stand-alone costs. This would provide an incentive for rival firms to provide those outputs at

competitive prices, thus contradicting their sustainability” (Spulber 1989, 138).⁸

Like subadditivity, the concept of sustainability is highly static and reveals a neoclassical bias in the way profit is viewed in contestability theory. Profit and loss are not seen as ever-present features of the market emanating from the ceaseless changes in economic data that cause new discrepancies and consequent need for new adjustments. Profit and loss are not viewed in terms of success or failure in adjusting the course of production activities to the most urgent demands of consumers. Profits, assuming they are empirically evident, are an unwanted phenomenon and are viewed as an indication the market is functioning ‘inefficiently’ and are interpreted as evidence that the consumers are not being well served by an existing industry structure.

III. Presumed Objectivity of Cost and Profit: A Framework for Industry Regulation

Contestability theory contrives a set of scenarios in which regulatory intervention is alleged to improve economic efficiency and to increase welfare. Circumstances requiring regulation would include the following: (1) subadditivity holds but no sustainable prices exist to forestall entry and (2) the market is not contestable because of entry barriers or situations in which costs are not recoupable. Since both prospects are thought to be detrimental to efficiency and economic welfare, contestability theory interprets these phenomena as a rationale for regulatory intervention. The theory presumes to provide an operational framework for the actual regulation of industry structure and the pricing of industrial outputs. For contestability theory to play such a role, cost functions estimated from empirical data must be legitimate representations of opportunity costs and economic profits must be empirically measurable.

⁸ Stand alone cost of a product or a subset of products is the cost which would be incurred if a decision were made to produce only some specified set of products ‘standing alone.’

A. Cost functions and opportunity costs

The central concepts of contestability theory and its proposed regulation of prices are all crucially dependant upon the objectivity of opportunity cost. Opportunity costs would need to be sufficiently objective to allow the construction of a cost function with sufficient empirical legitimacy to allow welfare inference. In their book *Contestable Markets and the Theory of Industrial Structure*, Baumol et al. make the following statement on the criticality of an objectively measurable cost function:

Throughout the book we have found it most convenient and natural to conduct our analyses in terms of firms' cost functions and their properties. Indeed the analysis has driven us to this approach, because our results show that it is the properties of the cost function that are directly related to the elements of industry structure and performance with which we have been concerned. The focus on cost functions is fortunate since recent advances in duality theory and in techniques of empirical analysis have guided the profession away from the problem-ridden task of direct estimation of technological relationships (production functions and multi-output production transformation functions) and toward the estimation of cost functions instead. *Thus we find ourselves felicitously aligned with mainstream trends....* (Baumol et. al. 1988 [1982], 445). (Emphasis added)

This quote helps to highlight issues raised in the questions posed above. Can the econometric estimation of a "cost function" reveal legitimate information of opportunity cost. In his book *Cost and Choice*, James Buchanan has emphasized that this approach to opportunity cost implicitly rests on an assumed equilibrium in which the following conditions must hold: 1) reckoning of opportunity costs must free of any uncertainty, (2) no unexploited profit opportunities can exist anywhere in the economy, (3) decisions must be made at the margin, that is, without discrete discontinuities, and (4) decisions must be made on strictly economic or pecuniary grounds (no non-economic considerations can prompt decisions) (Buchanan 1969,

50).⁹ Buchanan explains that “[i]t is essential for each of the qualifying conditions to be satisfied if measured marginal costs [incremental financial outlays] are to be employed as an objective representation of the subjective elements that actually enters the [firm’s] choice calculus” (Buchanan 1969, 50).

The equilibrium conditions discussed by Buchanan would normally end the discussion since few economists, perhaps with the exception of Baumol and his colleagues, would seriously contend that *any* of these stipulations would ever be satisfied in any real market.¹⁰ Eschewing any reliance of equilibrium assumptions in dealing with the nature of costs, Ludwig von Mises explicitly links the concept of cost to the subjective act of choice: “Costs are a phenomenon of valuation. Costs are the value attached to the most valuable want-satisfaction [foregone] because the means required for its satisfaction are employed for the want-satisfaction the cost of which we are dealing with.If costs were a real thing, i.e., a quantity independent of personal value judgements and objectively discernable and measurable, it would be possible for a disinterested arbiter to determine their height...There is no need to dwell further on the absurdity of this idea” (Mises 1998 [1949], 393).

⁹ Buchanan also stipulates that decision-makers must know their own utility functions. It is worth noting that Baumol himself reviewed the Buchanan book, *Cost and Choice* in the 1970 *Journal of Economic Literature*. Baumol essentially rejected the relevance to policy of subjective opportunity cost. In taking this position, Baumol had to concede the subjectivity of opportunity costs without at the same time acknowledging its devastating operational significance for the fields of industrial organization, monopoly theory and regulatory economics. Baumol observes “... economists should learn the lessons offered to us in this little book – and learn them well. It can save them from serious errors...” However his next words were the following: “... *but it should not be used to justify complete inaction on the ground that the relevant information is largely subjective and unknowable even to a reasonable degree of approximation*” (Baumol 1970, 1211) (Emphasis added). Baumol made no attempt to explain or justify his remarks concerning his view that objectively measurable outlays are representative of opportunity cost ‘to a reasonable degree of approximation’ and how the applied economist is to know when such a judgement may not be warranted. One can conclude, without doing Baumol any injustice, that he learned absolutely nothing from Buchanan’s book and has demonstrated such by his contributions to the development of contestability theory.

¹⁰ James Buchanan notes that “...*cost* ... exists in the mind of the decision-maker and nowhere else;...*cost* is based on anticipations: it is necessarily a forward looking or ex ante concept;... *cost* cannot be measured by someone other than the decision-maker because there is no way that subjective experience can be directly observed;...[and] *cost* can be dated at the moment of decision or choice” (Buchanan 1969, 43). Buchanan adds that cost must be borne exclusively by the decision-maker; it is not possible for costs to be shifted or imposed on others; cost can never be realized because of the fact that choice itself means that that which is relinquished in the act of choosing can never be experienced.

B. Capital use over time and the presumptive objectivity of profit

As noted above, in the contestability paradigm, profits are a presumptive indicator of inefficiency. Working from this misguided premise, regulation of efficiency presumes the existence of objective information on economic profit. Unfortunately for regulators, objective information on profit requires the assumption that financial outlays are faithful representations of opportunity costs of acting. Moreover, any attempt to use the financial outlays of an industrial enterprise as a representation of opportunity cost must inevitably face the fact that the firm owns capital assets the use of which in any given time period will involve a user cost that is not reflected in direct financial expenses. Much of opportunity costs is inevitably grounded in the present use of property that foreclose alternative future employment. Hence, the practical realities of company efforts to establish their user cost of capital mean that future relinquished opportunities must be valued in a way that reckons uncertainty and change.

Contestability theory has been examined in the context of various industries including electrical power transmission, local telephony, railway transportation, airline travel and even banking.¹¹ To varying degrees, these industries are capital intensive and critically reliant on prudent intertemporal management of capital assets. As with any asset employed in production, use in any particular time period involves a sacrifice in the form of a user cost or economic depreciation. This user cost is reflected in a reduction in the present value of future rents and, if it were ever be measurable, would be gauged as a reduction in the market value of already-used capital goods.

Baumol and his colleagues purport to address these issues but do so in the most unrealistic equilibrium framework. They note the importance of formulating a theory of user cost or depreciation that is consistent with both ‘accounting usage’ and economic theory. “Economists generally recognize that accounting depreciation rules bear little relationship to underlying economic relationships. What is less clear is a set of depreciation rules that *is* consistent with the

¹¹ See: (Baumol and Sidak 1994), (Baumol and Sidak 1994), (Tye 1995), and (Dickens 1996).

tenets of economic theory. We will show that our analysis does yield specific rules for depreciation, rules that are not only consistent with recovery of invested resources, but which in addition, satisfy the requirements of optimality in intertemporal resource allocation” (Baumol 1988 [1982], 384). In the context of an elegant mathematical presentation, they go on to note:

The value of the capital can be assessed economically as the present value of the stream of future payments to capital, minus the present value of the stream of future costs [user cost] of the net investment that, in part, make those payments possible.... A set of depreciation payments is any stream of charges that add up to the sum of original investment cost. For any stream of payments to capital there is an equivalent stream of depreciation charges [user cost] and vice versa.....Thus if the firm decides on a set of outputs, input purchases and prices over time, it *automatically decides*, implicitly, on the streams of payments to capital and depreciation charges. Consequently, there is a direct relationship between the firm’s pricing decisions and its depreciation decisions. For example, a decision to use accelerated depreciation for a piece of equipment is tantamount to a decision to charge a high price (relative to variable cost) for those products of the equipment which are supplied early in the life of that item and then reduce the products’s prices (relative to variable cost) as the item ages. ... That is, among the infinite number of different streams of depreciation assessments that cover the cost of capital equipment over its life time, we may select that stream of assessments which induces an ..optimal pattern of quantities demanded and supplied over the relevant time interval (Baumol 1988 [1982], 385 -387). (Emphasis added).

Hence, Baumol and his colleagues purport to ‘objectify’ user cost (depreciation); “[t]he optimal set of depreciation assessments, and, the corresponding prices, are those necessary for efficiency in the intertemporal allocation of resources. It is these optimal depreciation decisions that we are discussing here” (Baumol 1988 [1982], 387). While, this statement is theoretically correct, what

is noteworthy in this observation is that Baumol and his cohorts do not see these depreciation decisions as judgements or conjectures to be made by property owners. Rather, user costs are treated as determined magnitudes or objective data that may be available for regulatory scrutiny.

One concludes that to neoclassical economists such as Baumol and his colleagues, neoclassical equilibrium is a presumptive description of reality. Were it to be such, efficiency would be an empirically measurable economic phenomenon. But all of this mechanistic, deterministic precision dissolves into emptiness and absurdity if the assumption of a static intertemporal equilibrium is dropped, which, of course, it must be. True markets are dynamic, uncertain environments. In an equilibrium world, the reduction in the market value of capital goods should be equivalent to the reduction in the present value of future net receipts arising from current use of the asset. In these static, idealized circumstances, user cost would be precisely equivalent to the incremental reduction in the market value of the capital goods employed in the production process. But in real world markets, characterized by genuine uncertainty, used capital goods, even if they are non-specific to particular employments, are not standardized products with quotable market prices. Their market value is always a matter of conjecture or judgement that will necessarily differ between decision makers depending upon their appraisals and understanding of the market.

Hence, user cost is always reckoned subjectively. In this reckoning, the owner of capital assets attempt to establish, at the margin, a balance between the valuation of current benefit of using capital and his valuation of future productive benefits relinquished because of current use. This user cost is based on the acting entrepreneur's understanding or expectation of the market's future. In a realistic disequilibrium world, no two actors are likely to see the future of the market in exactly the same way. The fact that "tenders submitted by different firms for the same order often quote prices that differ widely is no doubt due to the differences in user cost calculated"(Lachmann 1986, 66-67). These are basic realities of real markets that are essentially ignored in the suggested applications of contestability theory. The user cost imputed to the current use of capital will reflect the highly personal outlook of owner-entrepreneurs who may embrace optimism or pessimism, boldness or timidity at any one moment in time in which they

must make decisions.

In no sense can the user cost ever be presented to a regulating authority as objective data. It cannot be embedded in an econometrically-estimated cost function to be applied to the regulation of industry. This empirical barrier has stark implications for the way in which contestability theory presumes to use the concept of economic profit; “[w]hat is present profit depends no less on expectations of future revenue than our present magnitudes. Firms in the same industry, even if they all had the same physical assets, could earn the same rate of profit on them only if their directors put the same valuations on them and had depreciated them at the same rate” (Lachmann 1986, 77, 71). The reckoning of profits by a firm is really a matter of uncertainty, judgement and discovery meaning that profits cannot be empirically evident even to the accountants employed by the firm. As markets change unexpectedly, historical costs and profits are always subject to review after the fact. As Peter Lewin has observed, “The judgement involved in measuring the latter [user cost] affects the profit calculation and lends it an unavoidable element of arbitrariness. This means that profit, even measured retrospectively, necessarily contains elements of subjective judgement or convention” (Lewin 1999, 164).¹² Hence, the regulatory act of trying to discern profits of a firm cannot simply be a matter of examining accounts; as Ludwig von Mises noted “cost accounting is therefore not an arithmetical process which can be established and examined by an indifferent umpire. It does not operate with uniquely determined magnitudes that can be found out in an objective way. Its essential items are the result of an understanding of future conditions, necessarily always colored by the entrepreneur’s opinion about the future state of the market” (Mises 1998 [1949], 346) (Emphasis added).

¹² In this context, historical reassessment of profits is not irrelevant for future action. If reassessment reveals that user cost associated with past decisions were actually higher than previously thought, capital assets thought to be available for planned production are actually more greatly diminished than earlier valuations may have indicated.

IV. Regulatory Imperatives and Presumed Applications of Contestability Theory

If opportunity costs and economic profits are matters of subjective judgement by the owners of invested property, what remains of the regulatory agenda underlying the theory of contestability? The remainder of this paper will answer this question by examining the scenarios thought to require regulatory intervention. *First*, subadditivity may hold for the natural monopolist but there are no prices that can be charged that will forestall the entrepreneurial entry of competitors. What is thought to be lost by the entry of such firms is the presumed welfare benefits of both scale and scope economies. After entry of competitors, firms then in the industry would be operating at smaller scales and probably narrower scopes of production than were possible when the single firm was supplying the market. Consumers are thought to be hurt by the prospect of the same slate of goods being supplied at higher costs and higher prices by more than one firm. Hence, in the name of increasing welfare, regulatory prohibitions on entry are thought to be necessary. *Second*, subadditivity may hold but the market is not perfectly contestable. What appears to be a lack of perfect contestability may arise from the perception on the part of regulators that there is little threat of entry to discipline the pricing behavior of the incumbent firm. Perhaps entry into the industry is thought by regulators to involve sunk cost for prospective entrants or the entrant may be at a cost disadvantage in having to bear costs that are not borne by the incumbent firm. In the event that the market were somehow found to be uncontested, the regulating authority would be called upon to impose 'operational' pricing regimes that would replicate the pricing that would occur if the market were contestable.

A. Preventing entry when subadditivity holds but sustainable prices do not

One can begin by asking a Hayekian question: to whom is this information given? Of course, the answer is that it is given to no one. But, for the moment, one can set aside the most compelling fact about subadditivity, namely that no regulating authority would ever be in a position to identify it. Assume that a hypothetical circumstance has somehow been identified in which the single firm represents the most efficient (least costly) means of meeting market demand for the

particular slate of goods. *Subadditivity* holds. Recall that *subadditivity* refers to a situation in which a single firm is able to exploit both economies a scale and economies of scope and hence produce any combination of a listing of goods more cheaply than the same listing of goods could be produced by some industry configuration including more than one firm. “The question arises: are there always prices that a natural monopolist can charge that will prevent entry, that is, are sustainable? If not, then allowing entry will prevent the attainment of an optimal equilibrium” (Train 1991, 308). Assume there is no set of *sustainable* prices that will forestall entrepreneurial entry of new competitors. In this situation, contestability is actually thought to be detrimental to economic welfare since, after entry of additional firms, consumers will be paying higher prices for goods that are being produced at higher cost. Is this a regulatory problem? Should the government ever presume to have a policy position on such as situation – assuming that it could be identified?¹³

Consider the following example.¹⁴ Three products can be produced in a market in which demand is fixed at 1,000 units for each product. If produced by separate firms, total cost for each product is \$30,000; total cost for the three products would be \$90,000. Also suppose that economies of scope exist allowing any two of the products to be provided by one firm at a cost of \$48,000. But assume that a subadditive monopoly is possible in which all three products can be produced for a cost of \$75,000 earning total revenue of \$75,000. Hence, with three firms, total cost is \$90,000; with two firms, total cost is \$78,000 and with a monopolist, total cost \$75,000. For the monopolist, profits would be zero but another firm could enter and provide *two* of the products, *X* and *Y*, charging \$24.50 for each and earning \$49,000 in total revenue and a net profit of \$1,000. Assume that an incumbent monopolist already exists and tries to forestall entry of the second firm by lowering the price of products *X* and *Y* to \$24.00 such that a potential entrant earns no profit. Then the monopolist must charge \$27.50 for Product *Z* to break even. However, at these prices, a prospective entrant could charge \$23.50 for product *X* and \$26.50 for product *Z*, earning revenues of \$50,000 and a profit of \$2,000. Under any other price combination a new firm would be able to provide *X* and *Y*, *X* and *Z* or *Y* and *Z* at a lower price and make a profit,

¹⁴ The example given here is derived from one presented by Kenneth Train (1991, 309-310). However, Train credits Edward Zajac with the example (Zajac 1978).

even though the incumbent monopolist would be earning zero economic profit.

Does the situation outlined in this example call for regulatory intervention? Theorists plunge ahead with suggested regulatory proscriptions on entry. “Clearly, in this situation, an equilibrium with one firm, which is optimal from a cost perspective, is not possible if entry is allowed” (Train 1991, 310). In his book *The Theory of Natural Monopoly*, William Sharkey claims that “to the extent that entry results in inefficient production, it may be desirable to place restrictions on the competitive process when an unsustainable natural monopoly is known to exist...Because entry into a natural monopoly market necessarily results in higher costs of production, a possible goal of regulation would be to prevent such entry...” (Sharkey 1982, p. 148, 151-152). Baumol et. al. are only slightly more circumspect in recommending a strict prohibition on entry into the industry:

In an industry characterized by subadditive costs but no sustainable prices, freedom entry threatens the viability of natural monopoly. If entry succeeds because the incumbent has no prices available that are capable of preventing it, then the public will bear the resulting increase in cost. It will be served by a multiplicity of firms whose product is *in toto* more expensive, rather than by a single firm whose relatively low cost reflect subadditivity. Therefore to prevent such waste, it would appear appropriate for the policy maker to remove the threat by prohibiting entry... the argument ...applies with full force only to perfectly contestable markets. ...But that does not preclude the possibility that a natural monopoly can find prices that yield a slightly lower rent and leave invulnerable to entry. In such a case, despite the unsustainability, the threat of entry may still force the monopoly to behave better than it otherwise would have. Yet it is true that our analysis shows the possibility of ground for occasional and limited restrictions upon freedom of entry. Where markets are perfectly contestable and monopoly is natural but unsustainable, limitation upon entry may be needed to ensure that the socially optimal set of products can be produced in the most efficient manner. This theoretical finding must be interpreted as no more than the

first step in the process of identifying the economic characteristics of the special situations that may warrant concern over the effects of free entry. (Baumol et. al. 1988 [1982], 221-233).¹⁵

Policy proscriptions on entry of additional firms largely ignore the analytical and policy implications of an entrepreneurial environment. *First*, no empirical means would provide a regulator, the incumbent monopolist or firms contemplating entry with objective information on sustainable prices. The most critical point to note in this example is that the stated 'facts' are never revealed as objective information either to regulators or to prospective competitors. What are listed as 'facts' in the example can never be more than conjecture or judgement that will almost always differ between the various affected parties and observers. Hence, sustainability can never serve as a quantitatively determined guideline in making regulatory decisions on allowing or prohibiting entry in to the industry. *Second*, no demand schedule is rigidly stable. The schedule of demand around which the concept of subadditivity rests is in fact not static. In the example given above, demand is assumed to be fixed at 1000 for each the three services provided by the natural monopolist. In reality, the quantities demanded of the respective products in the list would be subject to dynamic change. Demand may become smaller or larger depending upon changes in the tastes or goals of customers. The volume of each good demanded by consumers may be growing to a degree that would eventually allow entrants and the incumbent to fully exploit the economies and scale and scope to the degree initially enjoyed by the incumbent prior to the entrance of new firms. *Third*, another form of market change affecting the product list is that its actual composition may undergo significant change over time. Entrants into the industry are not necessarily restricted to production from within the initial slate of goods produced by the incumbent natural monopolist. Entrants may offer differentiated slates of goods that may intersect but not replicate the slate offered by the incumbent monopolist. After

¹⁵ A portion of this quote from Baumol et. al. highlights the fact that the example given above is not totally clear on the pricing issues facing a subadditive natural monopoly. Baumol and his colleagues suggest that the natural monopolist may reduce prices somewhat and 'reduce rent.' The latter suggestion implicitly takes note of the fact that a subadditive monopoly cannot survive in a situation in which its prices are equivalent to marginal cost. To allow for economies of scale and scope, prices must exceed marginal costs by an amount sufficient to allow the natural monopoly to break even. However, the example given in the text seems to suggest that the natural monopolist is able to earn no profit by setting its prices for the three products equal to 'cost' of the products. But the latter pricing policy is not one that would allow the natural monopoly to survive even if there were no threat of entry. The remainder of the paper will discuss the pricing issues facing natural monopolists.

the entry of new firms into the previously one-firm industry, competitors may well begin to offer new products as part of their entrepreneurial efforts to better satisfy consumers. Assuming that new slates of goods are conducive to cost complementarities in production, more firms will be able to more fully exploit economies of scale and scope as the market grows. In other words, subadditivity may be eventually destroyed by competition with *no loss to consumers*. *Fourth*, economies of scale and scope as manifested in production costs do not remain constant over time. Improvements in old production technologies and introduction of new technologies will drive down costs as firms seek higher profits in their efforts to better satisfy consumer demand. These types of market change suggest that no natural monopoly should ever be protected from entry by potential rivals. Subadditivity and sustainability as they may apply to a particular firm at a particular moment time are always likely to be only transitional and temporary.

B. Uncontested monopoly: regulation through Ramsey pricing

Ramsey-optimal prices are designed to achieve optimal deviations from marginal cost in regulating a natural monopoly for which marginal cost pricing would mean losses. An optimal upward deviation from marginal cost is sought with intent of providing the regulated firm with sufficient revenue to remain in business. The deviation from marginal cost pricing takes into account the elasticities of demand of the respective products produced by the regulated natural monopoly. If one assumes that upward deviation from marginal cost can be chosen ‘optimally,’ the regulated firm is assured a flow of revenue sufficient to cover the costs of all goods produced. In their application of contestability theory to electric-power transmission, Baumol and Sidak note the following in explaining Ramsey pricing: “If this set of prices yields revenues insufficient to cover the suppliers total cost, however, the prices must be modified [from marginal cost pricing] for the goods to continue to be supplied by private enterprise. But, under the neoclassical paradigm, every deviation from marginal costs creates some inefficiency--first because it provides an incentive for consumers to switch to those goods whose prices are raised only modestly relative to their true marginal costs, and second, because every rise in price restricts demand by cutting into consumer purchasing power” (Baumol and Sidak 1995, 31-32).

Ramsey prices are achieved when the deviation from marginal cost multiplied by the *net* elasticity of demand is the same for all products produced by then firm. Application of the Ramsey rule yields the following relationship (Train 1991, 126-127):

$$\left[\frac{p_i - MC_i}{p_i} \right] \left(E_i - \sum_{j=1}^{n-1} E_{ji} \right) = \left[\frac{p_j - MC_j}{p_j} \right] \left(E_j - \sum_{i=1}^{n-1} E_{ij} \right), \quad i \neq j, \quad E_{ji} \geq 0, \quad E_{ij} \geq 0. \quad (3)$$

The following definitions apply to (3): p_i and p_j are the prices of products i and j respectively ($i, j = 1, \dots, n-1$); MC_i and MC_j are respectively the marginal costs of products i and j ; E_i and E_j are the own price elasticities of demand for products i and j respectively; E_{ji} is the cross elasticity measuring the percentage change in the quantity of good j demanded in response to a one percent change in the price of good i ; and E_{ij} is the cross elasticity measuring the percentage change in the quantity of good i demanded in response to a one percent change in the price of good j .

The formula in (3) addresses instances in which the cross elasticities are not necessarily zero for all combinations of products produced by the regulated monopolist. The difference between the demand elasticities and the summation of cross elasticities yields a net elasticity of demand to be used as a weighting factor in deriving the Ramsey prices (Train 1991, 139).¹⁶ For each product, the largest revenue contribution is achieved by the applying the “inverse net elasticity” relationship in such a way that larger deviations from marginal cost are imposed for those products with the lowest net elasticity of demand. For example, if a particular good q_i has a large net elasticity of demand, an increase of one percent in the price of q_i will add little to the revenue of the regulated natural monopoly. The increase in price results in a significant reduction in the quantity demanded. However, for a good q_j with a small price net elasticity of demand, a one percent increase in price will be accompanied by more than a one percent increase

¹⁶ The net elasticity takes account of the possibility that some of the cross elasticities may be positive and others negative. In the version given above, subtracting the summation of cross elasticities (which can be positive or negative) from the own elasticity of demand, , yields a net elasticity of demand in the Ramsey formula. While the Ramsey formula is virtually useless for its intended purpose, the logic behind the subtraction of cross elasticities to derive net elasticities of demand is obvious. Not subtracting the sum of cross elasticities could involve a situation in which a Ramsey increase in the price of one good with a low elasticity of demand could have an indirect effect on the demand of other products produced by the firm and may have unintended consequences with respect to the firms

in the revenues of the firm. Hence, Ramsey prices are thought to represent second-best optima because they selectively establish deviations from marginal cost pricing which assure the firm sufficient revenue to continue to produce the products in question. Moreover, Ramsey-optimal prices purport to accomplish the above while minimizing distortions in consumer choices and ‘loss of economic efficiency.’

To Baumol and his colleagues, a critical shortcoming of Ramsey-optimal pricing is that it imposes daunting data requirements with respect to product demand. But they fail recognize or acknowledge the fact that marginal opportunity costs are subjective and that the marginal cost information required to apply the formula will be unavailable -- or more accurately ‘non-existent.’ Moreover, one notes that contestability theory exhibits and perpetuates a fundamental error with respect to the meaning of the relationship between price and marginal cost. For example, Baumol and Sidak assert that “[e]conomic efficiency requires that the price of every product be set equal to its marginal cost, provided that doing is consistent with the economic viability of the firm (which will be true in the absence of scale economies, as must be true in a world of perfect competition)... A more direct way of viewing the matter is to interpret marginal cost as the true cost the consumer imposes upon the economy in buying an additional unit of product... As a result, if the consumer makes purchase decisions rationally, achieving a given level of satisfaction at minimum cost to himself, he automatically does so at minimum cost to society, as economic efficiency requires” (Baumol and Sidak 1995, 27-28). However, the presumed normative significance of any observable equivalence between price and observable outlays is debunked by James Buchanan: “...equalities between prices and marginal costs, as objectively observed quantities in fully competitive equilibrium, are inferred predictions which depend on the behavioral assumptions upon which the whole theory is constructed. *These equalities have no normative significance and they have no direct relationship to allocational efficiency*” (Buchanan 1969, 40). While Ramsey prices are thought to be “optimal deviations” from marginal costs, the same types of mistaken inferences are drawn with respect to the efficiency of their use as have traditionally been imputed to marginal cost pricing.

C. Uncontested monopoly: regulatory imposition of price ceilings

The same analytical barriers arise with respect to the alternatives recommended for use instead of Ramsey prices. Constrained-market prices have been offered as a more viable regulatory pricing procedure and is based on the idea that a regulating authority can, on the basis of available cost data, derive and impose price ceilings and price floors between which the regulated firm free to pursue its own pricing strategy for all of the products produced (Baumol and Sidak 1995, pp. 37-38). Of interest is the fact that between the price ceiling and price floor, the regulated firm would be able to design its own set of discriminatory prices in maximizing the income allowable under the regulatory constraint. In applying constrained-market pricing to uncontested markets, the objective is to (a) provide assurance that the incumbent firm earns only a competitive return on investment, and (b) prevent cross subsidization since deliberate “losses” for one product must be compensated by economic profits from other products (Baumol et. al. 1988 [1982], 24-29). Hence, the rationale of the price constraint is to duplicate these results of contestability even in situations in which the market is not contestable. The remaining analysis will concentrate on the operational validity of the proposed price ceilings and price floors. In the case of price ceilings, opportunity costs present an even more severe analytical barrier since this latter regulatory framework is dependent upon the quantitative reckoning of opportunity costs that can be only hypothetical. Since the market is not in fact contestable, the regulator must be able to infer of costs or glean information on costs independently of actual entry decisions by entrepreneurs.

The ceiling price of constrained-market pricing is intended to be the upper-bound price which the regulated natural monopoly is able to charge for any particular product. The ceiling is based on a regulator's presumed quantitative knowledge of the price at which a hypothetical entrant may attempt to enter a hypothetically contestable market; of course, in this case, the natural monopoly is not threatened by entry of competitors. The price ceiling for an individual product derives from ‘stand alone cost costs’ of individual products or sub-groupings of products that would presumably be associated with production by a more specialized firm. Stand alone cost (**SAC**) of a product or a subset of products “is the cost which would be incurred by an efficient entrant to the industry in question if it were to decide to produce only some specified set of

[products]...That is, it is the cost to produce just those items `standing alone’ (Baumol and Sidak 1995, 52).

Assume that a multi-product regulated firm were producing products (goods or services) q_1, q_2, \dots, q_n . The cost to the firm of producing these products is supposedly given by the function $C = C(q_1, q_2, \dots, q_n)$. The cost of producing each of these products separately by individual firms is supposedly given by $SAC_1 = C(q_1), SAC_2 = C(q_2), \dots, SAC_n = C(q_n)$. But in this case, the SAC 's would be hypothetical costs imputed by the regulating authority. Assuming that costs were measurable or imputable magnitudes, economies of scope would reveal the following inequality:

$$C(q_1, q_2, q_3, \dots, q_n) < \sum_{i=1}^n SAC(q_i). \quad (4)$$

The relationship in (4) would apply for the sum of the stand-alone costs for any combinations of subsets of the original listing of goods produced q_1, q_2, \dots, q_n . The price ceiling or maximum price for the product q_i is established by its stand-alone cost $SAC_i = C(q_i)$. Why? The stand-alone cost is intended to represent the most costly potential alternative source for the product in question. This cost is equivalent to a price for q_i that is imputed to be just sufficient to attract new firms prepared to produce only the one product. Baumol and Sidak outline the responsibilities of the regulator: “such a price [in excess of SAC_i] should be rejected by the regulator in the non-contestable regulated market, where the regulators task is to impose the attributes of competitive behavior wherever the powers of competition are inadequate to automatically to prevent the price from being adopted...” (Baumol and Sidak 1995, 81-82). In other words, the price ceiling is based upon the regulator’s alleged ability to impute the costs that would be incurred by a hypothetical new competitor in a hypothetically contestable market. Undaunted by any analytical concerns, Baumol and Sidak refer to this ceiling as the “combinatorial cost test” and assert that it prevents the regulated incumbent firm from using its economies of scope to “achieve earnings that would not be permitted to it by the forces of an effectively [hypothetical and non-existent] competitive market...the combinatorial stand-alone

price ceiling means that the prices of every combination of the firm's products must yield combined revenues not exceeding the corresponding stand-alone cost of the combination of products in question” (Baumol and Sidak 1994, 78).

However, this analytical device is empty. Stand alone costs do not exist in any meaningful sense independently of actual entrepreneurial decisions by actual entrants to produce specific subsets of the monopolist’s product slate. Again, to reiterate a central point, entrepreneurs in the process of weighing the costs and benefits of entry subjectively experience opportunity cost but this experience is never revealed as objective information to external observers such as a regulating authority. But opportunity costs of hypothetical entrants cannot be imputed by a regulating authority in the way suggested in the application of contestability theory. If the regulated natural monopoly were functioning in an environment characterized by uncertainty and economic change, the price ceiling based on stand-alone hypothetical costs is nothing more than the most fanciful theoretical speculation.

D. Uncontested markets and the non-operational regulation of price floors

Baumol asserts “that an excessively low price can be as damaging to economic efficiency as an excessively high price. Prices that are too low may come at the expense of adequacy of investment and may prevent expansion of capacity, modernization of facilities, and ultimately, continued operation of supplier firms.” But from the regulatory perspective, the principal stated concern is the prevention of “cross subsidization” and “predatory pricing.” Predatory pricing is a strategy thought to be pursued by a firm with the intent of ultimately driving competitor from the market. As the theory goes, once competition has been driven out of the market, the monopolist would be free to raise prices to ‘monopolistic levels.’ Another possible strategy would one in which the multi-product firm would price a product “below cost” and to subsidize this “loss” by charging a price for another product that exceeds its ‘cost.’ The regulation of a price floor would be intended to prevent the incumbent regulated firm from pursuing price strategies capable of driving out competitors or of preventing entry by potential competitors.¹⁷

¹⁷ For an incisive analysis demonstrating that predatory pricing behavior cannot be logically be explained within

One of the measures of cost suggested as a price floor for the single product q_i is *average incremental cost*. In the regulatory framework described by Baumol and Sidak, the average incremental cost AIC is not the equivalent of marginal cost. *Incremental cost*, in this case, would be it the addition to total cost borne by the firm through the act of adding product q_i to the slate of products already produced. Hence, AIC_i is the cost per unit of q_i added to the regulated firm total 'cost' as a result its production of the product. Moreover, AIC_i as shown in (5) differs from average variable cost in that it includes those "product-specific" fixed cost that are necessary in the production of q_i :

$$AIC_1 = \{C(q_1, q_2, q_3, \dots, q_n) - C(0, q_2, q_3, \dots, q_n)\} / q_1. \quad (5)$$

Hence, for the individual product q_i , the price floor constraint assures that $p_i q_i \geq AIC_i$ holds when the product is considered individually.¹⁸ To the extent that this latter relationship were ever to have any legitimate empirical content, it would represent presumptive evidence that the firm is not engaging in predatory price behavior. The price floor idea advanced by Baumol and his colleagues also applies to particular subsets of products sharing common costs. Assume that the regulated firm produces a set k of 4 products which share common costs $S_k(q_1, q_2, q_3, q_4)$. The price floor applied to this subset of products would be given by the following:

$$\sum_{i=1}^4 p_i q_i \geq \sum_{i=1}^4 AIC_i + S_k. \quad (6)$$

Finally, the price-floor condition must satisfy the following condition for all product sets

the theoretical boundaries defining the neoclassical theory of the firm, see William Anderson (2003, pp.23-40).

¹⁸ Baumol and Sidak note that average-incremental costs or marginal costs can be the price floor depending upon the presence or absence of economies of scale. They observe that "If there are economies of scale in the sense that average-incremental cost falls lower and lower the larger the volume of output, then marginal cost must be below average-incremental cost, so that the latter must constitute the effective floor on prices. If average incremental cost neither declines nor increases when output grows, marginal cost and average-incremental cost must be equal and either constitutes the effective price floor. Finally, if average-incremental costs rise when output increases, then marginal cost will always be the higher of the two, and it must therefore serve as the effective price floor" (Baumol and Sidak 1994, 67).

produced by the regulated firm:

$$\sum_{i=1}^n \mathbf{p}_i \mathbf{q}_i \geq \sum_{i=1}^n \mathbf{AIC}_i + \sum_{k=1}^K \mathbf{S}_k. \quad (7)$$

In (7), the \mathbf{S}_k are the subsets of products which share common costs. The sum over \mathbf{K} subsets represents all the combinations of product subsets sharing costs within the regulated firm.

Several criticisms of these price floors become apparent. First, and most obviously, these constructions are operationally empty; a regulatory agenda designed to prevent such pricing activity is premised on costs being sufficiently transparent so that cross subsidization and predatory pricing strategy can be detected. In essence, the regulatory strategy is operational only if costs can be seen as objective and hence measurable. As emphasized throughout this paper, opportunity costs are subjective which means that these procedures provide no legitimate basis for regulatory intervention aimed at promoting ‘economic welfare’ or efficiency. No legitimate procedures exist by which one can determine if a predatory price is being charged nor is there any means by which one can discern predatory motives in the pricing behavior of a firm. “The cutting of price may just as well be due to inability to dispose of a stock at any higher price as to ‘cutthroat’ competition, and it is impossible for an observer to separate the two elements” (Rothbard 1962, 604). In general one cannot detect any legitimate distinction between aggressive pricing and pricing practices that are allegedly ‘predatory.’

Second, the concern over predatory pricing appears odd and disjointed in the case of uncontested natural monopoly. Recent research has shown that predatory pricing cannot be rationally explained within the context of the neoclassical theory of the firm (Anderson, 2003). Moreover, what could possibly be the rationale for so called ‘predation’ if the market is in fact ‘uncontested?’ Is the case being made that low prices in themselves are a barrier to entry that should be precluded because such prices offer no prospects of profitable entry? Is this not the type of behavior that one would hope to see in a market characterized by the presence of a single

firm within an industry?’ In fact, “the process of attempting to eliminate rivals is inherently competitive. Large firms bent on eliminating competitors would presumably seek to reduce final prices, increase their own efficiency and productivity, and offer additional services to potential buyers, all in order to secure business from rivals. But such competitive activity is delightful. From a consumer perspective and ought not to be discouraged in any way” (Armentano 1990 [1982], 42). Much of this absurdity regarding the issue of predatory pricing arises from the notion that contestability theory provides insight into optimal industry performance, optimal pricing, optimal size of firms and the optimal number of firms in an industry. But as Murray Rothbard has noted: “We do not know, and economists cannot tell us, the optimum size of a firm in any given industry” (Rothbard 1962, 573).

V. Conclusions

Contestability theory demonstrates that the pricing behavior of a multi-product natural monopolist is disciplined by the threat of entrepreneurial entry. The contestability model introduces several new concepts into an essentially static neoclassical theory of the firm. These include (1) *economies of scope* in which joint production of a slate of products is less costly than if each product were to be produced by separate firms; (2) *subadditivity*, a formal demonstration that the joint cost function of a firm exhibits both economies of *scale* and *scope* to such a degree that the single firm is the least costly means of satisfying a specific demand for a specific slate of products. But even in such a case, ‘benefits of competitive pricing’ are thought to be achievable if: (a) prospective entrants are assured of entry free of barriers and the ability to recoup entry costs; and (b) the incumbent firm is induced by threat of entry to charge *sustainable prices* such that no profitable entry is possible. (3) *Sustainability* is attained if market demand is being fully satisfied and the monopolist is able to charge prices that fully cover the cost of production.

Regulatory intervention is purportedly warranted if the above conditions are not met. However, establishing such is critically dependent upon opportunity costs and profits being revealed as objective data. But opportunity costs and economic profits are subjective. This analytical barrier becomes particularly apparent in dealing with the issue of user cost or opportunities relinquished

in use of capital over time. While contestability theory claims to deal with these issues, techniques offered for user-cost measurement are hopelessly empty under uncertainty, change and entrepreneurship.

Two scenarios present an alleged rationale for regulatory intervention: (1) subadditivity holds but no sustainable prices exist to forestall entry or (2) the market is not contestable because entry barriers or situations in which costs are not recoupable. In those instances in which natural monopoly exists and no sustainable prices are possible, entry of competitors would raise the cost of production. In this case contestability theory would proscribe entry. But, quite aside from the fact that the information on cost necessary to detect subadditivity does not exist, such a regulatory agenda ignores the fact that the slate of products defining the subadditivity condition is subject to change as are production technologies. Hence, a proscriptive policy with respect to entry is absolutely without foundation and is likely to be detrimental to economic welfare.

In the event that there is thought to be no threat of entry by competitors, contestability theory would prescribe prices to be charged by the incumbent monopolist. One of these strategies is Ramsey pricing in which prices represent 'optimal deviations' from marginal cost. Ramsey prices employ a weighting factor equal to the net elasticity of demand for each product produced by the natural monopolist. This approach to pricing is advanced as a means of covering production costs while minimizing inefficiencies arising from marginal cost pricing. While this approach to regulated pricing is ultimately dismissed for reasons related to calculation of net demand elasticities, the more formidable barrier is the non-existence of objective information on marginal opportunity cost. The same problems beset the offered alternatives to Ramsey prices. Price ceilings are proposed to ward off any possibility of monopoly pricing. But in practice, there is no way of identifying a monopoly price; moreover, the requisite information on opportunity cost required to determine the price ceiling does not exist as objective information. The same problems beset any effort to establish price floors prescribed to prevent predatory pricing and cross subsidization.

The preceding discussion should establish in the reader's mind an understanding of the fact that

economies of scope or scale, the condition of subadditivity and sustainability or its absence are largely beyond empirical investigations. While empirically available data may suggest something about the presence or absence of these phenomena within an industry, no information would ever be available to a regulating authority that would justify intervention or the presumption that economic welfare would be enhanced by a particular regulatory sanction. To this extent, the theory of contestable markets does not apply to real markets since it is too significantly detached from reality to have any operational significance in regulatory policy.

REFERENCES

Anderson, William. 2003. "Pounding Square Pegs into Round Holes: Another Look at the Neoclassical Theory of Predatory Pricing." *The Quarterly Journal of Austrian Economics* 6(1): 23-40.

Armentano, Dominick T. 1990 [1982]. *Antitrust and Monopoly: Anatomy of Policy Failure*. New York, N.Y.: Holmes and Meier Publishers, Inc.

Bailey, Elizabeth. 1988 [1982]. "Foreword." In Baumol, William, John C. Panzar and Robert D. Willig. 1988 [1982]. *Contestable Markets and the Theory of Industry Structure*. New York, NY: Harcourt Brace Jovanovich.

Bain, Joe. 1956. *Barriers to Competition*. Cambridge Massachusetts: Harvard University Press.

Baumol, William. 1968. "Entrepreneurship in Economic Theory." *American Economic Review* 58(2): 64-71.

Baumol, William. 1970. "Review of 'Cost and Choice: An Inquiry in Economic Theory'." *Journal of Economic Literature* 8(2): 1210-1211.

Baumol, William and J. Gregory Sidak. 1994. *Toward Competition in Local Telephony*. Cambridge, Massachusetts and Washington DC: MIT Press and American Enterprise Institute (AEI) Press.

Baumol, William and J. Gregory Sidak. 1995. *Transmission Pricing and Stranded Costs in the Electric Power Industry*. Washington, DC: American Enterprise Institute (AEI) Press.

Baumol, William, John C. Panzar and Robert D. Willig. 1988 [1982]. *Contestable Markets and the Theory of Industry Structure*. New York, NY: Harcourt Brace Jovanovich.

Buchanan, James. 1969. book, *Cost and Choice: An Inquiry in Economic Theory*. Chicago, Illinois: Markam Publishing Company.

- Buchanan, James. 1979. "General Implications of Subjectivism in Economics." In *What Should Economists Do?* Indianapolis, Indiana: Liberty Press.
- Caldwell, Bruce. 1982. *Beyond Positivism*. Boston, MA: Allen Unwin.
- Coase, Ronald. 1980. "The Business Organization and the Accountant." In James Buchanan and G.F. Thirlby, ed. *LSE Essays on Cost*. New York, NY: New York University Press.
- Dickens, Ross N.. 1996. *Contestable Market theory, Competition and the United States Commercial Banking Industry*. New York, NY: Garland Publishing Company.
- Dreze, J. 1964. "Some Postwar Contributions of French Economists to the Theory of Public Policy, with Special Emphasis on Problems of Resource Allocation." *American Economic Review* 54(4).
- Lachmann, Ludwig. 1986. *The Market as Economic Process*. New York, NY: Basil Blackwell, Inc.
- Leeman, Wayne. 1956. "The Limitations of Local Price Cutting as a Barrier to Entry." *Journal of Political Economy* 64(August): 329-332.
- Lewin, Peter. 1999. *Capital in Disequilibrium: The Role of Capital in a Changing World*. New York, NY: Routledge.
- Mises, Ludwig. 1952. "Profit and Loss." In *Planning for Freedom and Other Essays and Addresses*. South Holland, Illinois: Libertarian Press.
- Mises, Ludwig von. [1949] 1998. *Human Action: A Treatise on Economics*, The Scholar's Edition. Auburn, Alabama: Ludwig von Mises Institute.
- Rizzo, Mario. 1978. "Uncertainty, Subjectivity and Economic Analysis." In *Time, Uncertainty and Disequilibrium*. Mario Rizzo, ed. New York, NY: Lexington Books.
- Rothbard, Murray. 1962. *Man, Economy and State: A Treatise on Economic Principles*. Princeton, N.J.: Van Nostrand Company.
- Rothbard, Murray. 1978. "Comment: The Myth of Efficiency." In *Time, Uncertainty and Disequilibrium*. Mario Rizzo, ed. New York, NY: Lexington Books.
- Sautet, Frederic. 2000. *An Entrepreneurial Theory of the Firm*. New York, NY: Routledge.
- Sharkey, William. 1982. *The Theory of Natural Monopoly*. New York, NY: Cambridge University Press.

Shepherd, William G. 1984. "'Contestability' vs. Competition." *American Economic Review* 74(September): 572-587.

Spulber, Daniel. 1989. *Regulation and Markets*. Cambridge, Massachusetts: MIT Press.

Thirlby, G.F. 1980 [1946]. "The subjective theory of value and accounting 'cost'." In James Buchanan and G.F. Thirlby ed., *L.S.E. Essays on Cost*. New York, NY: New York University Press.

Thirlby, G.F.. 1980 [1952]. "The economist's description of business behavior." In James Buchanan and G.F. Thirlby ed., *LSE Essays on Cost* New York, NY: New York University Press.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, Massachusetts: MIT Press.

Train, Kenneth.. 1991. *Optimal Regulation: The Economic Theory of Natural Monopoly*. Cambridge, Massachusetts: MIT Press.

Tye, William B. 1995. *The Theory of Contestable Markets: Applications to Regulatory and Antitrust Problems in the Rail Industry*. New York, NY: Greenwood Press.

Zajac, Edward. 1978. *Fairness or Efficiency*. Cambridge, MA: Ballinger.