Baumol, Panzar, and Willig's Theory of Contestable Markets and Industry Structure: A Summary of Reactions

Voxi Heinrich Amavilah

16. September 2012

((((REEPS)))) Resource & Engineering Economics Publications Services
The Shameless Knowledge Peddlers ©

Baumol, Panzar, and Willig's Theory of Contestable Markets and Industry Structure: A Summary of Reactions[1]

Voxi Heinrich Amavilah[2]
REEPS
PO Box 38061
Phoenix, AZ 85069-8061, USA

This Version: September 16, 2012; *Original Version: Fall 1989*

Abstract: This paper summarizes reactions to the theory of contestable markets and industry structure. The reactions came immediately after the theory was published. The summary finds that the proposed theory stands on sound grounds. However, empirically the theory leaves much to be desired especially for practical policy in developing countries.

Keywords: Baumol-Panzar-Willig, perfectly contestable, perfectly sustainable, Ramsey welfare criteria, constestable markets, industry structure. **JEL Code**: D21, D43, D63, L11, L16, L22, L25

1. Introduction

Professors William J. Baumol, John C. Panzar, and Robert D. Willig [1a, b, c], hereafter the Baumolists, have led a rebellious research effort, culminating in "contestable markets and the theory of industry structure." The theory expands upon the free entry/exit assumption of perfect competition, and permits researchers to look anew at, and derive optimal solutions from, the behavior of markets and industry structures traditionally believed to yield inefficient and non-optimal outcomes. In fact, the new theory holds that potential competition leads to more efficient outcomes in imperfectly competitive settings than it was previously thought.

---

[1]I wrote this paper for Advanced Mineral Economics Theory (MnEc 650A) taught by Michael Rieber in Fall 1989 in the Mineral Economics Program of the Department of Mining and Geological Engineering at the University of Arizona during my first year of a PhD program. I benefitted, without attributing any responsibility for errors, from Dr Rieber's comments and seminar participants. I like the summary and am hereby making it available to the public without alteration.

[2]Amavilah is an independent researcher on economic growth and technological change of developing economies, and an Adjunct Professor of Economics at Glendale College, Glendale, Arizona. Email: vhsamavilah@gmail.com. He is grateful to Tamara Bishop-Amavilah for editorial assistance. *All errors are mine alone.*

This brief paper integrates the views expressed by William Brock [2], Michael Spence [7], Martin Weitzman [9], and M. Schwartz and R. Reynolds [6]on the new theory immediately after it was published [1a, c]. It describes the assumptions underlying market contestability and pins down propositions that determine how industry structures develop, and how contestable markets function. Section 2 is about basic relationships between the new theory and perfect competition. Section 3 characterizes contestable markets and defines related concepts so as to show how contestable markets approach their equilibria. The third section also points out the assumptions which must hold for industry "configurations" to be both feasible and sustainable. Section 4 focuses on the welfare implications of contestable markets and the theory of industry structure, and on the problems and potentials of the theory. The last section concludes the discussion with tentative remarks on the usefulness of the theory to understanding markets and industry structures in developing countries.

2. Theory of Contestable Markets and Industry Structure Versus Perfect Competition

In perfect competition the assumption of free entry/exit implies zero long-run equilibrium profit. Because of this assumption, firms produce at the minimum point on their long-run average cost curves. Traditional theory then treats the results of such behavior as both consistent and efficient, because individual firms' expectations and reaction functions are critical determinants of industry structures. In other words, if firms' expectations and reaction functions change, industry structures change as well. However, how often are firms' expectations accurately realized? And if some expectations are not realized even sometimes, then what can be said about the sustainability of market equilibria?

An important contribution to the debate is the recognition of the need for a comprehensive theory, which can "serve as a standard welfare maximizing structure in industries in which efficiency calls for a very limited number of firms" [1a, b]. Put differently, the assumption of zero-cost entry/exit in perfect competition leads to a zero-profit long-run equilibrium, because in this case firms are price-takers obeying the conventional profit-maximizing decision rule. In imperfectly competitive markets, whether or not firms produce at the minimum points on their long-run average total cost curves (LTATCs) depends upon the nature of the demand functions they face. Since imperfectly competitive firms have control over their average P due to product differentiation, for example, long-run profit tends to zero, but profit-maximizing production (output) may not take place where P = MC = LRATC, even though firms may still face the old negatively sloping demand curve. They often run excess capacity, which allows large price markups, but also large efficiency costs (deadweight losses).

"Contestable markets and the theory of industry structure" contends that perfect competition focuses only on the behavior of individual firms already in the market, and hence it overlooks the effects of the behavior of potential entrants on market equilibrium. Thus, the new theory reintroduces the distinction between *competition in the market* and *competition for the market*. The distinction, apparently due to Joe Bain and Harold Demsetz [see 1a], is significant because it enables "... casting (sic) a new light on the relationships between structure, conduct and performance," especially on "... that a particular market does not necessarily equal a particular type of performance" [4], implying that form does not always determine content. For instance, if both entrants and incumbents have identical cost structures, and if the cost of investment is reversible, then can 0-profit obtain. Thus, potential competition enforces cost-minimization regardless of the nature of the industry structure, such that contestability leads to an efficient result.

According to this theory imperfectly competitive structures and behavior are independent of the conjectural variations of incumbents; they are determined instead by potential competition. By contrast, conventional theory treats industry structure as a function of expectations and reaction functions of strategic incumbents behaving much like Stackelberg firms that live in a world in which perfect competition and monopoly are two extremes, marked by "efficient and desirable" on the side of the former and "inefficient and undesirable" on the side of the latter. The implication of all this is that perfect competition defines efficiency as dependent on the number of firms in the industry or market. Baumolists, however, say that such a definition of efficiency is correct only in the absence of externalities, government interventions, public goods and common resources, and the presence of constant returns to scale and symmetric (full) information. Given economies of scale, and therefore a downward-sloping average cost function for the industry, potential competition poses a threat to the least cost producers – the most efficient firms. Thus, rather than being exogenous, the nature of the industry structure is determined endogenously and jointly with price and output decisions – the number of firms in the industry is necessary but insufficient for the optimality of the structure of the industry.

The effect of potential competition causes incumbents to price competitively and to reduce welfare losses unless when there are significant sunk costs, which may exhibit entry and exit. In other words, irreversible costs may lead to natural monopolies, but whether natural or artificial, many monopolies are partial. As an illustration, Paul R. Ferguson [4] shows that utilities in the USA meet the natural monopoly criterion in general. However, in the case of electricity the national grid does not meet the criterion for a natural monopoly, the reason being that monopolies understand that they are partial monopolies and therefore respond to potential competition by behaving "as if" they were competitive firms.

3. Characteristics of Contestable Markets

The hallmark of contestable markets is free entry/exit. Both entrants and incumbents are assumed to have identical production/cost functions. They have equal access to the same technology or comparable technologies. A perfectly contestable market exists only in the presence of potential competitors who constantly seek to enter (exit) the market to take advantage of available profit opportunities (avoid economic loss), suggesting that potential competition is a crucial feature of perfect contestability.

Perfect contestability further assumes competitive behavior among incumbents themselves, not just with respect to potential entrants. Within the market, strategic differences persist and keep incumbents from colluding, while potential competition constrains their behavior from without. The outcome: a determinate contestable equilibrium. This outcome has obvious Walras, Edgeworth, and Keynes aspects described by John S. Chipman [3], although his analysis required Lyapunov stability condition rather than potential competition.

Another feature of contestable markets is that potential entrants can enter the market instantaneously at any scale of production, underprice incumbents and exit without loss. (I have seen this phenomenon referred to as "hit-and-run", but I am unable to locate the specific citation right now.) In other words, incumbents' speed of adjustment to price changes is slower than that of potential entrants. This asymmetric response is one source of the criticisms leveled against the Baumolists [6, 9].

The following propositions drive contestability: (1) profit for all firms in the industry is zero – positive profit attracts actual competition; (2) there is no inefficiency of any kind – inefficiency is eliminated because it is associated with positive short-run profit; (3) no output can be sold at a price lower than the marginal cost (MC) of producing it; (4) no predatory pricing is possible – any price greater than MC will attract entrants into the market; and (5) price always equals MC. These propositions are described briefly next below, keeping as close to Brock [2] and Spence [7] as possible.

Suppose we have an industry with $v \geq 1$ firms facing a multiproduct demand function, D(P), for P the vector of industry-determined prices. Assume all firms have identical cost functions, $C(q_i)$, where q_i is output for the ith firm. Define revenue for the ith firm as $\rho_i = Pq_i$ and its cost as $\gamma_i = Cq_i$. An industry structure (κ) is

$$\kappa = f(v, q_i, P) \Leftrightarrow P = f(\kappa). \tag{1}$$

Eq. (1) is said to be feasible if: (a) industry profits are positive; (b) demand facing the industry exhausts production; and (c) industry output is nonnegative. In short,

$$\rho_i - \gamma_i = \pi_i \geq 0;\ Q(v) - D(P) = 0,\ \textit{for all } i \tag{2}$$

where $Q = \sum_i^v q_i = vq_i$ if v firms are of equal size, i.e., $q_i = q_1 = q_2 = q_3 = \ldots = q_k > 0$. However, feasibility is a timeless performance criterion; all it says is that demand depletes "material balances". Hence, a feasible solution may not be necessarily a sustainable solution.

Question: What then is a sustainable industry structure? Eq. (1) is sustainable if (2) holds and entrants' revenue (ρ_E) is at least equal to entrants' cost (γ_E), and entrants' output (q_E) exceeds the demand ($D(P_E)$) facing the industry at its own price (P_E) equal to the industry-given price (P), i.e.,

$$\rho_E \geq \gamma_E;\ q_E \geq D(P_E),\ \textit{for all } P_E \geq P. \tag{3}$$

This means there exists no Bertrand-like possibility; potential competitors cannot profitably underprice incumbents, but incumbents strongly believe entrants can. Hence, a perfectly contestable market yields a long-run equilibrium iff (1) is both feasible (2) and sustainable (3). In that case neither the Cournot-type nor Stackelberg-type trap is possible. Thus, by setting aside strategic interactions among incumbents, the new theory rules out the possibility of collusion, i.e., no acquisitions, mergers, and/or takeovers can happen.

$$\rho_i - \gamma_i = 0;\ \rho_{i,S} > \gamma_i - \gamma_{i,S'};\ P_h > \partial\gamma_i/\partial q_h. \tag{4}$$

From (1), (2), and (3), properties of sustainable structures can be derived. For example, in (1) κ is sustainable iff:

Eq. (4) says that the ith firm's profit is zero, so that for any F set of products for which S is a subset, profits from S will be greater than the difference $S'=F-S$ in cost between producing F and producing S. Spence interprets this as meaning that the incremental cost of producing the subset of products is lower than the corresponding revenue. The third condition in (4) says that the set of prices of the hth product is greater than the associated cost such that (3) is unaltered.

Multiproduct cost structures determine efficient q-vectors as expressed above. Having said that, now the problem turns to what multiproduct cost structures themselves look like. Michael Spence [7], more so than William Brock [2], has illustrated the structures beautifully as follows [cf. 5, 7]: Assume $B=[B_1,B_2,B_3,...,B_k]$ is a nonempty subset set A. Then there are economies of scope at q-vector B of A if the sum of the cost of producing that output vector by more than one firm is greater than the cost of producing q_i by one firm, which is the measure of economies of scope.

Unlike perfect competition the Baumolists provide a better accounting benchmark for nonconstant returns to scale. From Spence (7, p. 985), for example, let us assume a translog cost function $\log Z(q) = C(qx)/x$, where x is a constant. Differentiating $d\log Z(q)/d\log(x) = q\nabla C(q)C^{-1}$. Hence, the degree of the economies of scale, $S^{\circ} = C/[q \nabla C(q)C^{-1}] = 1/[q \nabla C(q)]$, where grad C(q) equals partials of the cost with respect to the whole vector of the set of products produced by the whatever-th firm under consideration. Conventional theory then concludes that $S^{\circ} = C/[q \nabla C(q)C^{-1}] = 1/[q \nabla C(q)] = 1$. For $S^{\circ} = C/[q \nabla C(q)C^{-1}] = 1/[q \nabla C(q)] > 1$ output increases faster than the cost, and for $S^{\circ} = C/[q \nabla C(q)C^{-1}] = 1/[q \nabla C(q)] < 1$, cost rises faster than output [1c, pp. 15-95).

From this Spence [7] has identified the following cost structures. The first he called the incremental cost of producing an additional bundle of output. The second, associated with both economies of scale and scope, is his concept of "transray convexity" – a very difficult concept to grasp, especially with limited mathematical tuition. However, intuitively, it means that the average cost of producing a pair of products (Baumol, Panzar, and Willig's example is shoes and boots) is smaller than the cost of producing each product separately. The reason is that the effects of complementary production processes exceed those due to scale. And the last types of cost structures are sunk and fixed costs. Investment costs are fully and completely recoverable because capital has multiple uses. In addition to this a market for used capital exists. So demand for investment is instantaneous [7, p. 987ff].

4. Welfare Implications of the Theory

The theory of contestable markets and industry structure offer three normative welfare criteria for industries characterized by economies of scale and scope. These norms derive from the Ramsey surplus maximizing price theory upon the binding condition that industry profit is nonnegative [1c, p. 333ff]. If one assumes a multiproduct surplus value (S(P)) to be equal to a Ramsey optimum, then the optimization problem is to

$$\begin{aligned} &\max\ S(P) + \sum_{i}^{v} (\rho_i - \gamma_i), \\ &s.t.\ D(P) - Q(v) = 0,\ \sum_{i}^{v} (\rho_i - \gamma_i) \geq 0,\ v \geq k. \end{aligned} \tag{5}$$

Equivalently for the industry as a whole the problem in (5) becomes

$$\max\ S(P) + \rho_I - \gamma_I,\ s.t.\ D(P) - Q_I,\ \rho_I - \gamma_I \geq 0, \tag{5'}$$

where I is for industry. Differentiating (5) with respect to each decision variable gives the "viable Ramsey optimum (VIRO)" as a solution [7]. An even stronger criterion is the "viable firm Ramsey optimum - VFRO", which is a solution to

$$\begin{aligned} &\max\ S(P) + \sum_{i}^{v} (\rho_i - \gamma_i) \\ &s.t.\ D(P) - Q((v) = 0,\ \rho_i - \gamma_i \geq 0. \end{aligned} \tag{6}$$

In short, if every firm in the industry is making positive profits, the whole is making profits as well. So, VFRO = VIRO less no-profiting firms in the industry, and it is more restrictive than VIRO in that sense.

The last norm is called the "autarkic Ramsey optimum – ARO" [7]. A solution is ARO if no firm can search and discover some P_z and q_z, the product of which increases S(P). For example, define $P_z q_z = \mathrm{B}_z$ and $C_z q_z = \Gamma_z$. Then $S(P) + \mathrm{B}_z - \Gamma_z = ARO$. Again, ARO is more restrictive than VFRO; it is

VFRO, plus the inability for any firm to search and find a P_z-q_z combination, that increases the sum of producer and consumer surpluses (see 1c, pp. 333-345).

The discussion so far has focused on the integrity of the theory of contestable markets and industry structure. Next below I reflect on the criticisms of the theory. Specifically I note that both Spence and Brock have doubted the realism of the assumptions of theory. They question why: (1) all firms have access to the same technology, (2) sunk costs are zero for entrants and positive for incumbents, and (3) incumbents respond to price changes with longer time lags than entrants. In reality technology is more heterogenous than homogenous even among firms producing identical products. Often technology is also product-specific. Hence, one would expect the costs of producing a Ford and a similar Dodge to be approximately the same, but each family of the two cars would have a specific and unique production technology. Obviously there are instances of multiproduct plants. A GM-Toyota plant, Nummi, outside Fremont (California) that assembles Toyota Corolla, Toyota Tacoma, and Pontiac Vibe, all from the same facility (technology). Here technology appears the same only because it is embodied in product components.

Secondly, fixed and sunk costs, while theoretically separable, are empirically inseparable. Often where fixed costs exist, sunk costs are present as well. To say one type exists and other does not is, obviously different from saying the two are coexisting but difficult to separate.

Hence, thirdly, the difference in the response lags to price changes assumed in the theory of contestable market and industry structure is tantamount to game theoretical models in which one player is given the advantage of the first move. If symmetry is assumed and incumbents instead of entrants are given the first move, a similar outcome may still obtain – raising questions about whether or not that too would be a sustainable outcome. This observation suggests that the proposed solutions depend both on the game rules and on potential competition. In a recent paper Richard Gilbert [4] examines the role of potential competition in industrial organization, focusing on whether markets approach equilibria according to the "classic limit pricing model", "dynamic limit pricing model", perfect contestability model, or "efficiency difference model". I highly recommend the paper, although the comparison it sought turned out difficult to make mainly because Baumolists do not define what they mean by markets, industry, and even potential competition. What is clear from Gilbert is that "potential competition is important, but maybe not as important as the theory of contestable markets implies" (p. 123). He also concludes that it is unlikely that established firms set aside their strategic behavior in response to potential competition. In fact, Gilbert insists, incumbents sometimes respond aggressively to potential competition.

For Brock [2] perfect competition is a more robust model in that solutions to market games converge to one unique solution as the number of players tends to infinity. Players take prices as given parameters. Therefore, the speed of convergence is rapid, consistent, and subject to some describable limit theorem. The theory of contestable markets and industry structure lacks such robustness, because the number of firms alone does not influence the speed at which the market converges to an optimal solution. Incumbents are treated as passive actors. The question is why that should be the case, since, if demand turns out to be sluggish, imperfectly competitive firms do not only have control over the price and output, they also have market power, and market power can serve as a barrier to entry [1c,

p. 347ff, cf. 8].

The new theory deals well with economies of scale. However, the existence of economies of scale implies the presence of huge fixed and sunk costs; the two are interrelated as may be reflected in the constraints on the production technology [1c, pp. 389-394] and the intertemporal effect of learning-by-doing on the shape of that technology [1c, pp. 429-434]. The last part of the preceding statement is a familiar story from A.A. Young [10] and his many students. Because of the effect of learning-by-doing on the shape of the production function, one would find lower and lower degrees of economies of scale the shorter the run.

Thus, the theory sounds great, but only as a theory; although its potential applications span the gamut from international competition to cost determination at the industry level. However, its greatest use may be in perishable goods markets and industries. A direct irony of that is that not all goods are perishable, and many do carry huge storage costs and for long periods of time.

The three Ramsey welfare criteria the theory describes serve as benchmarks for policy. Obviously when all assumptions of perfect contestability are satisfied, Ramsey solutions approximate Pareto optimality, although they conceal a lot of information. For example, VIRO, due to its nonnegative profit constraints, rules out the possibility of some firms in the industry incurring losses large enough to reduce average industry profit to negative. This contradicts the conventional shut-down rule.

Despite these problems, the new theory is a substantial piece of work. Whether or not it will stand the test of time and gain acceptance, and how quickly, depends mainly on its ability to guide policy and further empirical research. The empirical hurdle may be its most difficult obstacle to jump over since some of its assumptions and propositions appear untestable.

5. Concluding Remarks

This paper attempts to integrate the comments made by Professors Brock, Spence, and Weitzman on the theory of contestable markets and industry structure. First, it outlines the basic relationships between it and perfect competition. Second, it characterizes market contestability and related concepts of feasibility, sustainability, and market equilibrium. It turns out that vulnerability to potential competition defines perfect contestability. Potential competition depends upon the assumption of costless entry/exit.

Also sketched is how the theory describes economies of scale and scope. From these descriptions three types of cost structures are identifiable: incremental costs, sunk costs, and fixed costs. Related to these costs is “ray average cost” associated with transray convexity – a difficult concept left to the original sources.

The welfare criteria the theory puts forward are restrictive even though under certain conditions some approximate Pareto optimality. By restricting industry profit to positive levels, however, the VIRO criterion, for instance, hides that some firms within the industry may be losing money. This gives the impression of efficient performance by the industry, which may obviously not be the case, especially if one or more products of one or more multiproduct firms is incurring economic losses. The theory also conceals technical inefficiency, the same thing it claims to eliminate through potential competition. Even so, the theoretical basis of the theory appears well-established.

The theory of contestable markets and industry structure appears to have far-reaching policy implications for developing economies. In fact, the theory does seem to suggest that government created monopolies and oligopolies in these countries may be more efficient than previously thought. At one point it goes as far as stating that it may be desirable on efficiency grounds to restrict actual entry/exit and encourage potential competition. But is it possible to regulate entry/exit and still expect potential competition to thrive? Would incumbents soon learn that potential competition that stops short of actual competition is just an empty bluff?

Suppose the proposition of potential competition does indeed force incumbents to behave "as-if" they were perfectly competitive firms. To eliminate inefficiencies would it be a feasible policy for government to set up real or pseudo potential competitors to police the behavior of incumbents? In most developing nations state-owned enterprises, justified by seemingly reasonable propositions like the "infant industry hypothesis", abound, but they are mainly sources of distortions in factor markets, which invariably over- or under-state the marginal value of the factors of production, which then spillover into product markets. The new theory is not nearly as clear about the contestability of factor markets as it is of product markets.

But even in product markets, if potential competitors seek to underprice incumbents, then potential competition does either not mean anything, or else it fosters development of black markets. It seems potential competition is primarily a feature of advanced free markets.

The theory also assumes physical capital to be fungible and technology to be homogenous and given. This is not such a terrible assumption in industrialized countries with well functioning capital and technology markets. However, it is easy to conclude that, at least for now, the theory of contestable markets and industry structure does not seem useful for either policy decisions or empirical research in developing countries.

References

[1a] Baumol, William J. 1982. "Contestable markets: An uprising in the theory of industry structure," *American Economic Review,* Vol. 72, No. 1, pp. 1-15 (March).

[1b] Baumol, William J. 1983. "Contestable markets: An uprising in the theory of industry structure: Reply," *American Economic Review*, Vol. 73, No. 3, pp. 491-96 (June).

[1c] Baumol, William J., Panzar, John C., and Willig, Robert D. 1982. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich, Inc.

[2] Brock, William A. 1983. "Contestable markets and the theory of industry: Review article," *Journal of Political Economy*, Vol. 91, No. 6, pp. 1055-1066.

[3] Chipman. John S. 1965. “The nature and meaning of equilibrium in economic theory.” In *Functionalism and Its Limits in the Social Sciences* (edited by Don Martindale), American Academy of political and Social Science, Monogrpah No. 5 (February), pp. 35-64.

[4] Ferguson, Paul R. 1988. *Industrial Economics: Issues and Perspectives*. New York: Macmillan Education, Ltd, Chapters 1 and 5.

[5] Gilbert, Richard J. 1989. “The role of potential competition in industrial organization,” *Journal of Economic Perspectives,* Vol. 3, No. 3, pp. 107-127 (Summer).

[6] Schwartz, Marius, and Robert J. Reynolds (1983) ‘Contestable Markets: An Uprising in the Theory of Industry Structure: Comment’, *American Economic Review*, 73, No. 3, pp. 488–490 (June).

[7] Spence, Michael. 1983. “Contestable markets and the theory of industry structure: A review article,” Journal of Economic Literature, Vol. XX, pp. 981-990 (September).

[8] Sweezy, Paul M. 1970. *The Theory of Capitalist Development*. New York: Modern Reader Paperbacks.

[9] Weitzman, Martin. 1983. “Contestable markets: An uprising in the theory of industry structure: Comment,” American Economic Review, Vol. 73, No. 3, pp. 486-87 (June).

[10] Young, Allyn A. 1928. “Increasing returns and economic progress,” *Economic Journal*, Vol. 38, pp. 527-42.