A Further Look at Proper Cost Tests for Natural Monopoly

by Mark A. Jamison
University of Florida
Public Utility Research Center
August 1997

Abstract

This paper extends the analysis of natural monopoly by considering the effects of multilateral rivalry and other-party costs that are not reflected in market transactions. A monopoly firm should be considered a natural monopoly if and only if its production economies dominate those offered by alternative market structures. The monopoly’s production economies are dominant if breaking up the monopoly increases overall production costs for the economy. The natural monopoly must have strict and global cost subadditivity as shown by William J. Baumol (1977), but this is not a sufficient condition. Also, the natural monopoly may not represent the lowest-cost means of producing its products if breaking up the monopoly increases production costs for products that the monopoly does not produce. The other-party costs (such as regulatory costs) or cost savings (such as one-stop shopping) that are not reflected market transactions should be subtracted from or added to the monopoly’s economies of joint production when examining the efficiency of the monopoly market structure.

Mark A. Jamison
Public Utility Research Center
University of Florida
Warrington College of Business Administration
P.O. Box 117142 (205 Matherly Hall)
Gainesville, FL 32606

Phone (352) 392-6148
Fax (352) 392-7796
E-mail: jamisoma@dale.cba.ufl.edu

Home page: http://www/cba.ufl.edu/econ/purc/mjamison.htm
A Further Look at Proper Cost Tests for Natural Monopoly

by Mark A. Jamison
University of Florida

I. Introduction

Early work on the concept of natural monopoly focused on the stability of competition in a market (John Mill, 1926) and on the cost efficiency of a single-product firm being the sole supplier in a market (Alfred Marshall, 1927, as cited by William W. Sharkey, 1982). William J. Baumol (1977) extended the cost efficiency aspect of natural monopoly by considering the case where a multiproduct firm could satisfy the entire market demand for its products and be more cost efficient that all combinations of smaller, more specialized firms.


This paper advances the analysis of natural monopoly in two ways. First, it considers in Section II the case in which firms that produce products that the monopoly does not can produce some of the monopoly’s products and have economies of joint production in doing so. In this situation, the monopoly is a natural monopoly if and only if its production economies are greater than the economies these firms can offer. Second, this paper describes how other-party costs (such as regulatory costs) or cost savings (such as one-stop shopping) that are not reflected market transactions should be subtracted from or added to the monopoly’s economies of joint production when examining the efficiency of the monopoly market structure. This is done in Section III. Section IV describes the implications for proper cost tests for natural monopoly. Section V is the conclusion. As with previous analyses, demand and technology are held constant.

II. Effects of Firms that Produce Other Products

This section examines the situation where some subsets of the monopoly’s products can be jointly produced with products the monopoly does not produce and have economies of scope. In this situation, if the monopoly is to be considered a natural monopoly, its economies of scope must dominate the economies of scope offered by other market structures.
Baumol assumed that the only relevant alternatives to the monopoly are smaller, more specialized firms. Figure 1 illustrates this assumption. The rows in the table in Figure 1 represent firms and the columns depict markets. \( M^A, M^B, \) and \( M^C \) depict smaller and more specialized firms that could also produce the monopoly’s output. (The base character identifies the firm. The superscript identifies the firm’s product mix. \( A, B, \) and \( C \) are all subsets of \( M. \)) The "X"s show points of rivalry -- markets in which the firms could, but not necessarily do, compete. The shaded boxes show which firms actually produce in which markets. The boxes for the smaller, more specialized firms in Figure 1 are not shaded because in Baumol’s framework, the potential rivals need not actually exist. Instead, these firms represent alternative cost structures against which the monopoly’s cost structure should be tested.

Baumol’s analysis of this framework concluded that subadditivity of costs is the critical concept. Example 1 describes this conclusion.

**Example 1.** Assume there are two products -- switched telephone service and non-switched dedicated lines. If a firm produces only switched telephone service, it has a total cost of $10,000,000. If a firm produces only non-switched lines, it has a total cost of $2,000,000. There are large fixed costs in providing these products, so a single firm is the most economical market structure for either product. Finally, if a single firm produces both products, it has economies of scope so that its total cost would be $11,000,000.

Under Baumol’s definition, the two-product firm is a natural monopoly because its costs are strictly and globally subadditive.\(^1\)

Focusing the definition of natural monopoly on strict and global subadditivity for the monopoly markets assumes that all subsets of the monopoly’s output could have no economies of joint production if produced with products not produced by the monopoly. That is to say, if \( N \) denotes the set of products in the economy so that \( N = \{1, 2, ..., n\} \) and \( M \) denotes the set of products produced by the monopoly so that \( M \subseteq N \), then by examining only the cost effects of subsets of \( M \), Baumol’s framework implicitly assumes that no elements of \( N \) that are not in \( M \) would have economies of joint production if produced with any \( S \subseteq M \).

Recent trends in technology and markets call into question the appropriateness of this assumption. Sanford Berg and John Tschirhart (1995) and Peter Smith (1996) observed that interindustry rivalry in communications could result from economies of scope that cross industry boundaries. Alexandar Zakem (1991) and Rodney E. Stevenson and David W. Penn (1995) described how energy utilities face rivalry from non-energy companies. Rivalry from diverse firms is called multilateral rivalry. (Mark A. Jamison, 1996) An example of multilateral rivalry exists in telecommunications, in which local exchange telephone companies face rivalry from firms (such as long distance companies and cable television companies) that operate outside the exchange

---

\(1\) Strict subadditivity means that the cost of producing the output jointly is less than the cost of having two or more firms produce the output. Global subadditivity means that the subadditivity condition applies to the entire cost function from the origin up to the entire market demand. (William J. Baumol, 1977)
Figure 2 illustrates multilateral rivalry. As in Figure 1, the rows of the table in Figure 2 represent firms and the columns represent markets. \( M \) depicts the monopoly. \( A, B, C, D, \) and \( E \) represent other firms. Also, as in Figure 1, the "X"s show points of rivalry. The shaded boxes illustrate which firms actually compete in which market. "X"s without shading indicate potential points of competition.

Figure 2 shows two properties of the multilateral framework. The first property is that firms' markets (including potential markets) incompletely overlap or intersect -- i.e., firms have different patterns of interrelations; for example, \( B \) provides rivalry for some of \( M \)'s markets, but not for others, \( C \) provides rivalry for some of \( B \)'s markets, but not for others, and so on. The second property is that it is market boundaries, not industry boundaries, which affect the scope of natural monopoly. Baumol assumed that the relevant industry was equal to the sum of the monopoly’s markets. In the multilateral framework, the sum of the monopoly’s markets may be equal to an industry, less than an industry, more than an industry, or parts of more than one industry. Industry boundaries do not matter.

Multilateral rivalry incorporates some of the features of multipoint or multimarket competition (Michael E. Porter, 1985; and Arjen van Witteloostuijn, 1993), but is more comprehensive in at least two respects. First, multipoint competition exists when firms compete with each other (or could potentially compete with each other) in more than one market. With multilateral rivalry, any two firms that affect the analysis may meet in multiple markets, one market, or no markets. The second difference is that multipoint competition focuses on jointly contested markets – i.e., the competition among firms in the market. Multilateral rivalry explicitly incorporates the effects of multiple firms on each other, including indirect effects such as how firms that are several markets away may affect the competitive positions of direct rivals of a firm being analyzed. The firm \( E \) in Figure 2 illustrates these differences between multipoint competition and multilateral rivalry. \( E \) affects \( M \)'s pricing opportunities and natural monopoly status because it affects the pricing and efficiency of the firms \( A, B, \) and \( D \). \( E \) does not affect \( M \) directly. Mark A. Jamison (1996) explains some of \( E \)'s affects on \( M \)'s pricing options.

Several factors create multilateral rivalry. Globalization and changes in cost structures increase firms' abilities and incentives to enter non-traditional markets. Competitive strategies may lead firms to enter markets as a preemptive strike. Customers and suppliers may vertically integrate to gain strategic advantage. Changes in technology allow firms to exploit new capabilities and cross traditional industry and market boundaries without leaving existing markets. Crossing traditional market boundaries may allow firms to exploit economies of scale. Deregulation may remove arbitrary regulatory distinctions between markets and firms. (Christopher A. Bartlett and Sumantra Ghoshal, 1991; Sanford V. Berg and John Tschirhart, 1995; Stephan Bradley, Jerry Hausman, and Richard Nolan, 1994; Kenchi Ohmae, 1991; and Arjen van Witteloostuijn, 1993) The remainder of this paper focuses on multilateral rivalry that results from economies of scope.

Multilateral rivalry from economies of scope makes it necessary to consider a firm to be a natural monopoly only when it satisfies two conditions: (1) the firm is part of all efficient market
structures for the economy; and (2) the firm satisfies all of the market demand for its products. An efficient market structure minimizes the economy's total cost of production. Example 2 provides an illustration.

**Example 2.** This example is an extension of Example 1. Assume that there are three products: switched telephone service, non-switched dedicated lines, and electricity distribution. Firms can have the following costs:

a. Switched telephone service alone costs $10,000,000.

b. Non-switched lines alone costs $2,000,000.

c. Electricity distribution alone costs $4,000,000.

d. Switched telephone and non-switched lines together cost $11,000,000.

e. Switched telephone and electricity distribution together cost $13,750,000.

f. Non-switched lines and electricity distribution together cost $5,250,000.

g. All three products together cost $15,500,000.

As in Example 1, there are large fixed costs in providing these products, so a single firm is the most economical market structure for any of the three products. The most efficient market structure includes the firm in letter “d” and the firm in letter “c”.

Example 2 illustrates why a monopoly must be part of all efficient market structures before the monopoly can be considered a natural monopoly. In this example, a firm producing both non-switched lines and electricity distribution serves 100% of its markets' demand and is more efficient than all other combination of firms for these products viewed in isolation. However, this firm is an inefficient way to produce these products because non-switched lines can obtain greater economies with switched telephone service than it can with electricity distribution. Indeed, all two-product firms in this economy have subadditive costs, but only a firm producing both switched telephone and non-switched lines is consistent with the efficient market structure. Therefore, only it should be considered a natural monopoly. In Example 2, the firms described in letters “d” and “c” are natural monopolies because each is the only active seller in its respective markets, and each is part of the efficient market structure.

---

2 If the firm satisfies all of the market demand for only some of its products, then the firm is called a partial natural monopoly, implying that it is a natural monopoly for some, but not all, of its products.

3 Because there might be more than one efficient market structure, a firm should be considered a natural monopoly only if it is a member of all efficient market structures. This is consistent with Baumol’s (1977) requirement that costs be strictly subadditive.

4 This arrangement exhibits diseconomies of scope. It could be argued that no firm would ever chose such an arrangement -- that a firm producing all three products would arrange its production to match the cost structure of letter “d” and letter “c”. However, diseconomies of scope could result from common ownership. For example, there might be government restrictions on firm size and diversity that are costly to overcome. Even so, the diseconomies of scope in this example are for illustrative purposes only. They are not critical to the analysis, but make the example easier to follow. The effect of not having these diseconomies of scope is explained later in this section.
A firm that is part of all efficient market structures has dominant cost subadditivity. This means that the economies gained by this firm’s joint production are greater than the economies given up by not having this firm’s products produced by other firms. (Mark A. Jamison, 1996) It also means that a natural monopoly may or may not maximize economies of scope for its products. Electricity distribution in Example 2 illustrates these points. The firm described in letter “c” does not minimize the incremental cost of electricity distribution. The lowest incremental cost for electricity distribution would be the firm in “f” wherein the incremental cost of adding electricity distribution to non-switched lines is $5,250,000 - $2,000,000 = $3,250,000. But even though the firm in “f” can produce electricity distribution at a lower incremental cost than the firm in “c” can, it is inefficient for the economy overall. If the firm in “f” were to produce electricity distribution, the loss in efficiency would be $250,000 -- the $1,000,000 of economies of scope lost by breaking up the firm in “d”, minus the $750,000 of economies of scope gained by jointly producing non-switched lines and electricity distribution.

That a natural monopoly may not maximize economies of joint production for its products implies that strict and global cost subadditivity for a firm is not a sufficient condition for natural monopoly. This is demonstrated by showing that a firm may have strictly and globally subadditive costs, but not be part of an efficient market structure. Assume that the economy consists of four products $N = \{i, j, b, d\}$ and that $M = \{i, j\}$. Further assume that $\mathcal{M}_b$, $\mathcal{L}_b$, and $\mathcal{D}_d$ all have cost structures that exhibit strict and global subadditivity. This subadditivity results in cost savings from joint production for $\mathcal{M}_b$ of $\zeta = C(\mathcal{M}_b) + C(\mathcal{M}_j) - C(\mathcal{M}_j)$, where $\zeta$ represents an amount of cost savings from joint production of the products in arg. Also assume that $\mathcal{M}_b$ is subject to multilateral rivalry from $\mathcal{L}_b$ and $\mathcal{D}_d$ because costs of $\mathcal{L}_b$ and $\mathcal{D}_d$ are strictly and globally subadditive as are the costs of $\mathcal{D}_d$. $\mathcal{M}_b$ is part of the efficient market structure if and only if $\zeta > 0$, strict and global subadditivity is not a sufficient condition for natural monopoly.

Even though strict and global cost subadditivity is not a sufficient condition for natural monopoly, it is a necessary condition. (William J. Baumol, 1977)

### III. Effects of Other-Party Costs

Previous analyses of natural monopoly have not explicitly incorporated costs that customers, suppliers, and the government may incur because of market structure. These other-party costs should be considered because they consume economic resources. Such costs include, for example, transaction costs that affect the efficiency of a firm’s boundaries (Oliver E. Williamson, 1992).
Suppliers’ and customers’ transaction costs may be higher or lower if multiple firms rather than the monopoly, produce the products in question. Examples of suppliers’ costs that may increase with the number of firms include contracting, billing, and developing customer-specific assets. On the other hand, suppliers may find that a single, larger customer more problematic to deal with, which could lead to more costly contract negotiations and enforcement. Examples of customers’ costs that may be affected by market structure include time (Gary S. Becker, 1996), product research, and bill payment -- costs saved by the so-called one-stop shopping. Customers may also find the single firm less responsive and, therefore, more costly to deal with. For example, a utility company may make bill payment more costly for customers by closing bill payment centers. Closing the payment centers lowers the company’s costs, but could increase customers’ bill-paying costs.

Government costs are also affected by market structure. Costs of regulation and antitrust are among the costs most likely to be affected. The government is more likely to engage in the economic regulation of a highly concentrated or monopoly market than of a highly competitive one. Likewise, the existence of a monopoly, or the movement of an industry towards monopoly, raises antitrust issues that the government and the firms spend money to address.

Unless the other-party costs are reflected in market transactions, the amount of these costs that result from a monopoly market structure should be considered an offset to the monopoly’s economies of joint production. Likewise, unless other-party cost savings from a monopoly market structure are reflected in lower market transactions, these cost savings should be considered an addition to the monopoly’s economies of joint production. Other-party costs or cost savings that are reflected in market transactions are already considered in the conditions for natural monopoly described in Section II. Examples include defendant and non-government plaintiff costs of antitrust actions, regulatory and other government costs not funded through taxes or assessments against firms, and consumer purchases of consumer buying guides.
V. Proper Cost Tests

Baumol’s assumption that no subsets of the monopoly’s products could have economies of scope with other products allowed for cost tests of the monopoly’s products to determine if the monopoly qualified as a natural monopoly. For example, William W. Sharkey (1982) explained that strict and global subadditivity exists if the firm’s cost function: (1) exhibits cost complementarity; (2) is convex and ray subadditive; (3) is trans-ray convex and has declining-ray average costs; or (4) is quasiconvex and has declining-ray average costs. David S. Evans and James J. Heckman (1984, 1986), Richard T. Shin and John S. Ying (1992), Thomas O. Armstrong and Karen Leppel (1994), and National Economic Research Associates (1995) applied these cost tests to empirical data.

The general definition of natural monopoly provided in this paper complicates the use of cost tests for natural monopoly. All cost tests to date have focused on comparing the cost or production functions of monopoly production to those of smaller, more specialized firms. This paper’s definition of natural monopoly requires that any cost test compare the monopoly’s economies of joint production to economies that could be gained by dividing the monopoly’s production among other firms, including firms that do not currently produce any of the monopoly’s products. For example, a cost complementarity test would need to incorporate both the incremental costs of having the monopoly produce the products and the economies lost by not having the monopoly’s products produced by other firms. Likewise, economies measured in the other tests would have to be netted against economies that other firms could have by producing some subset of the monopoly’s products.

An optimization algorithm of the economy could be used to test whether a monopoly firm was a member of all efficient market structures. This would require data and knowledge of all existing and potential cost functions based on current technology. A less demanding test could be used if it could be determined that only a certain subset of firms and markets were relevant to the analysis. The first step in such an analysis would be to determine the limits of the multilateral rivalry from economies of scope. Consider the case of Example 2, which had an economy consisting of two telecommunications services and electricity distribution service. Assume that these services were only a subset of the services provided in the economy. If it could be determined that the telecommunications services could have economies of scope only with electricity distribution, and that electricity distribution could have no economies of scope with anything but the telecommunications services, then a test for natural monopoly could be limited to only these three services.

VI. Conclusion

This paper extends the analysis of natural monopoly to consider the effects of multilateral rivalry and other-party costs. It demonstrates that defining natural monopoly under conditions of multilateral rivalry is more restrictive than under William J. Baumol’s (1977) framework in that the monopoly must have dominant cost subadditivity in addition to strict and global cost subadditivity. This paper also shows that considering other-party costs may increase or decrease the incidence of natural monopoly, depending on how the monopoly market structure affects the costs of consumers and government, and whether these costs are reflected in market transactions. Conducting tests for
natural monopoly may be impractical if not impossible where there is multilateral rivalry because of the potentially large number of scenarios that must be tested.

This analysis of natural monopoly is primarily applicable to markets with stable technologies. Incorporating technology change requires comparisons of various technology and market structure paths that an economy could follow. In this case, a natural monopoly would be a firm that served 100 percent of the market demand for its products and was a member of all efficient technology and market structure paths over a specified period of time. Further research is needed on to develop a definition for path efficiency, to explain when a firm should be considered a member of an efficient path, and to quantify the costs of erroneously either allowing or denying a monopoly market structure.
Appendix

This appendix provides a more formal analysis.

Let $M$ denote a set of products such that $M = \{1, 2, ..., m\}$, and $q_1, ..., q_m$ are the quantities supplied. For all sets $S \subseteq M$,

$$q_i^S = \begin{cases} q_i & i \in S \\ 0 & i \in M \setminus S \end{cases}$$

so that $q^M = (q_1, q_2, ..., q_m)$ and $q^{[1]} = (q_1, 0, ..., 0)$. $\mathcal{N}^M$ denotes the firm producing $M$ and $\mathcal{N}_S^M$ represents the quantities of products $S$ produced by the firm $\mathcal{N}^M$. (The notation $MS$ represents all of the products in $M$ that are not also contained in $S$.) Also, let $C(\cdot)$ represent the cost function, so that the cost of a firm $\mathcal{N}$ providing $S \subseteq M$ is $C(\mathcal{N}^S)$. Let $T(\cdot)$ represent a partitioning of production into subsets to be produced by separate firms. Then $\mathcal{N}^M$ is a natural monopoly under William J. Baumol’s (1977) definition if

$$C(\mathcal{N}^M) < C(T(M)) \quad \forall T(M)$$

and $\mathcal{N}_i^M$ is equal to the market demand for $i \quad \forall i \in M$. Equation (1) states that $\mathcal{N}^M$’s cost function is globally and strictly subadditive.

Focusing the definition of natural monopoly on strict and global subadditivity for the monopoly’s products assumes that all subsets of $M$ could have no economies of joint production if produced with products not in $M$. That is to say, if $N$ denotes the set of products in the economy so that $N = \{1, 2, ..., n\}$ and $M \subseteq N$, then examining only the cost effects of subsets of $M$ implicitly assumes that $\exists B \subset N$ such that $B \cap M \subseteq B$ and $B \cap M = S$, and $C(\mathcal{N}^B) < C(\mathcal{N}^{B \setminus S}) + C(\mathcal{N}^S)$.

In practice, $\mathcal{N}^M$ is but one of many firms in an economy. Multilateral rivalry exists for $\mathcal{N}^M$ when it faces rivalry from $\mathcal{N}^B$ for products $S$ if $\mathcal{N}^B$ is an actual competitor (in which case $\mathcal{N}^M$ is not a monopoly for $S$), or such that $\mathcal{N}^{B \setminus S}$ is a potential competitor. Multilateral rivalry for $S$ in $\mathcal{N}^M$ results from economies of scope if $C(\mathcal{N}^S) < C(\mathcal{N}^{B \setminus S}) + C(\mathcal{N}^B)$. $\mathcal{N}^M$ faces multilateral rivalry from economies of joint production for all of its products if $\forall i \in M, \exists S \subseteq M$ such that $i \in S$ and $C(\mathcal{N}^B) < C(\mathcal{N}^{B \setminus S}) + C(\mathcal{N}^S)$.

The efficiency of market structure can be explained as follows. Let $\mathcal{S}(\cdot)$ denote the cost minimizing partition of production among firms such that $C(\mathcal{S}(N)) \leq C(T(N)) \forall T(N)$. There may be more than one $\mathcal{S}(N)$ for an economy, so the generating collection of possible cost-minimizing market structures is denoted as $GC(P^N)$. $\mathcal{N}^M \in \mathcal{S}(N) \in GC(P^N)$ states that $\mathcal{N}^M$ is part of an efficient market structure.

If there are no other-party costs, a natural monopoly can be defined as follows:
Definition 1. The firm $\mathcal{M}^M$ is a natural monopoly if:

Condition 1. $\mathcal{M}^M \in \mathfrak{Z}(N) \forall \mathfrak{Z}(N) \in GC(P^N)$.

Condition 2. $q_i^M$ is equal to the market demand for $i \forall i \in M$.

$\mathcal{M}^M$ may be consistent with efficiency but not be a natural monopoly if $\exists \mathfrak{Z}(N) \in GC(P^N)$ for which $\mathcal{M}^M \notin \mathfrak{Z}(N)$.

Other-party costs are incorporated as follows. Let $C_0(T(N))$ represent other-party costs that are not included in firms’ cost functions and that are created by market structure. $\mathfrak{Z}(\cdot)$ is now the cost minimizing partition of production among firms such that $C(\mathfrak{Z}(N)) + C_0(\mathfrak{Z}(N)) \leq C(T(N)) + C_0(T(N)) \forall T(N)$. The generating collection of possible cost-minimizing market structures, including other-party costs, is denoted as $GC(P_0^N)$.

Considering other-party costs, a natural monopoly can be defined as follows:

Definition 2. The firm $\mathcal{M}^M$ is a natural monopoly if:

Condition 1. $\mathcal{M}^M \in \mathfrak{Z}(N) \forall \mathfrak{Z}(N) \in GC(P_0^N)$.

Condition 2. $q_i^M$ is equal to the market demand for $i \forall i \in M$. 


Bibliography


Porter, Michael E. *Competitive Advantage: Creating and Sustaining Superior Performance*. New


