On the Shape of the Hospital Industry Long Run Average Cost Curve

Steven A. Finkler

Empirical studies of the hospital industry have produced conflicting results with respect to the shape of the industry's long run average cost (LRAC) curve. Some of the studies have found a classical U-shaped curve. Others have produced results indicating that the LRAC curve is much closer to being L-shaped. Some theoretical support exists for both sets of findings. While classical theory predicts that the LRAC curve will be U-shaped, Alchian has presented theoretical arguments explaining why such curves would be L-shaped. This paper reconciles the results of these studies. The basis for the reconciliation is recognition of the failure of individual hospitals to produce all their individual product lines at efficient volumes. Such inefficient production is feasible and perhaps common, given the incentive structure which exists under current cost reimbursement systems. The implication of this paper is that large hospitals may have a greater potential for scale economies than has previously been recognized.

HOSPITAL costs have been increasing for a number of years, and with each passing year there is greater public interest in cost containment. In this environment the various empirical cost studies that have been done take on great importance. A great deal of research has focused on the shape of hospital cost curves. The results of these studies have serious policy implications.

Unfortunately, we do not have the satisfaction of having a number of studies by different researchers all pointing to the same clear-cut conclusion. In fact, the results of cost curve studies have often been contradictory. Berki perhaps sums up the state of the literature on cost functions best when he states that " 'The exact general form of the function is unimportant' . . . , but 'whatever its exact shape' . . . , and depending on the methodologies and definitions used, economies of scale exist, may exist, may not exist, or do not exist, but in any case, according to theory, they ought to exist."[1]

Part of the source of the contradiction was discussed by Berry[2]. He noted that hospital output is not homogeneous, and researchers who fail to recognize the multiproduct nature of the hospital industry will have erroneous results. However, this still does not resolve the question of why even researchers who adjust for case mix have not uncovered the economies of scale that, as Berki pointed out, ought to exist. This problem should be resolved before policy decisions regarding optimal hospital size are made.

Cost Curve Theory

For many years the U-shaped average cost curve held an established place in

Address communications and requests for reprints to Steven A. Finkler, Assistant Professor of Accounting, The Wharton School, W280 Dietrich Hall–CC, University of Pennsylvania, Philadelphia, PA 19104. Professor Finkler is also Senior Fellow at the National Health Care Management Center, Leonard Davis Institute of Health Economics, University of Pennsylvania.

0017–9124/79/1404–0281/$02.50/0

economic theory. In the short run, based on increasing returns at first due to higher utilization of discrete, fixed inputs, followed by decreasing returns due to the law of diminishing returns, the theoretical construct which applies to both plant and firm size seemed secure. It was thought that eventually increasing costs might set in because some inputs (e.g., available land) are fixed, even in the long run. It has never been proven empirically, however, that firms in most industries ever become large enough for these diseconomies to become significant.

Work by Alchian vastly changed the outlook by distinguishing between the rate of production and the volume of production[3]. He contended that a new theory was necessary to explain the fact that most empirical studies found L-shaped cost curves and economies of scale rather than the U-shape predicted by classical theory. Johnston's summary of studies of cost curves indicated that the L-shaped curve is the one predominantly observed in empirical analyses[4].

Alchian showed that by segregating the production process into two components, rate and volume, the observed L-shape results are readily explainable. Marginal costs are a rising function of the rate of output and a falling function of the volume. Thus, if the percentage change in volume of output rises faster than the percentage change in the rate of output, we expect marginal costs to continue their fall and an L-shaped average cost curve with scale economies to be evident. Increases in volume occur over time, not at a moment in time. Simply doubling volume implies that it is produced over twice as long a time span. This puts no pressure on management, such as that usually attributed as being the cause of diseconomies of scale. The responsibility and required level of control by management is unchanged, but merely lasts longer.

Some empirical studies in the hospital industry have observed the more traditional U-shaped curve. Carr and Feldstein[5] and Cohen[6,7] observed U-shaped curves that showed a minimum cost per bed-day at hospitals of bed size 190 and 150–350, respectively. Since this agrees with the classical model, these results were not viewed as being counterintuitive or in need of further explanation. However, other researchers found significantly fewer diseconomies of scale. M. Feldstein[8] determined that there was only a shallow or slight U-shape and could not accept or reject a hypothesis that overall diseconomies of scale exist. Similarly, Lave and Lave[9] found a curve that, while not U-shaped, did not indicate the economies of scale that the Alchian theory would predict. This leaves an unresolved dilemma. Within the hospital industry, the evidence is conflicting as to whether the long run average cost (LRAC) curve is L- or U-shaped.

Hirshleifer[10] attempted to resolve such conflicts by showing that, using Alchian's model, the classical result is simply a special case in which the volume and rate are both increasing proportionately (see Appendix). This, however, is an unlikely circumstance. While Hirshleifer is correct in asserting that the classical model is accurate under this condition, there is no reason to believe that one would observe this condition frequently.

Product Heterogeneity

The measure of output frequently used in hospital industry studies is hospital beds. Thus, in comparing the efficiency of hospitals of differing sizes, one compares a 100-bed hospital versus a 300-bed hospital versus a 500-bed hospital. Several researchers have questioned the use of the number of beds as an output measure. First, beds measure capacity rather than output. Further, McNerney [11] and Berry [2] explain that since we are dealing with multiproduct firms, the quality of the service or the case mix may be changing over the observed range of firm sizes, therefore changing the definition of output along the quantity axis.

In the hospital industry, a larger firm is likely to offer more products and more levels of quality for their products than a smaller firm would. Larger hospitals tend to be centers for referral. They treat a wider variety of ills and must have a substantially larger set of equipment and personnel. Thus, the 600-bed hospital does not sell the same products as the 200-bed hospital, and it is inappropriate to compare them without adjusting for product lines offered. Also, some services are more expensive than others, so it must be realized that product line adjustments which are based simply on the number of product lines offered are also inaccurate.

Empirical studies frequently group many multiproduct firms into one industry for practical considerations, even though the mix and quality of their products may differ substantially. The problems of such aggregation are clear. For example, assume that one auto maker decides to aim for a low price market and therefore produces only a subcompact economy car. If we measure efficiency of production in the auto industry by average cost per car, this firm will appear to be the most efficient, simply because its product is not really the same as that throughout the industry.

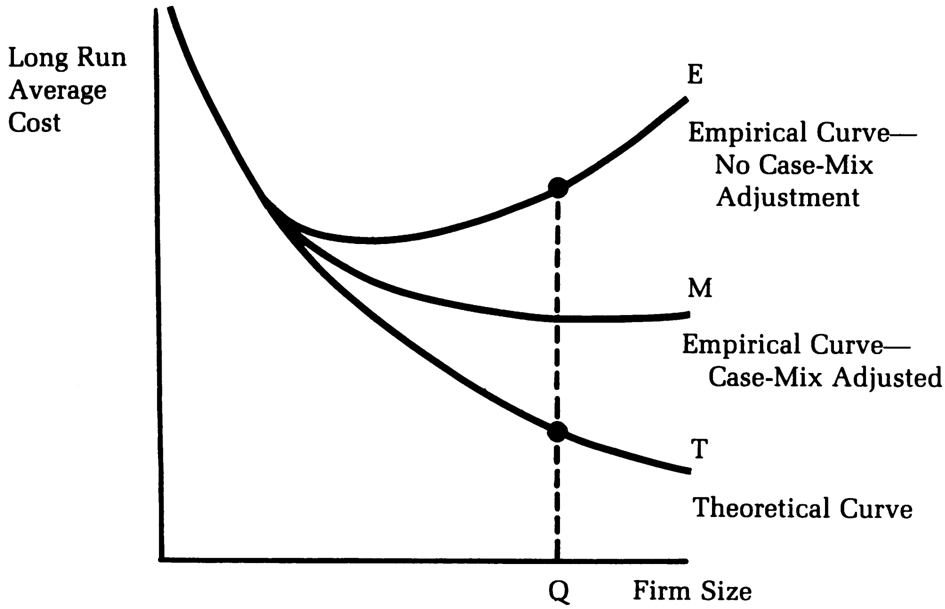
We can get some idea of the magnitude of the heterogeneity problem by looking at which hospitals tend to offer which services. Data collected by the American Hospital Association on 47 services[12] support the contention that larger hospitals offer more services in general and more of the costlier services. For almost every service the percentage of hospitals in any size group offering the product increases monotonically with hospital size (with slight statistical anomalies at the very largest hospitals). We cannot determine an industry's LRAC curve if the industry produces multiproducts unless we account for product mix. Of the studies mentioned earlier, those which found a strong U-shape did not adjust for product mix, while those that did not find strong eventual diseconomies did make some adjustment.

Applying this concept to other industries as well as the hospital industry gives an explanation more palatable than Hirshleifer's for why we sometimes observe U-shaped cost curves. If the regression analysis has taken the approach of assuming the firm to be a single-product producer, then in many cases industry diseconomies of scale may incorrectly appear to exist because of failure to control for product mix. The result is a false signal about the shape of the cost curve.

Individual Products

It is suggested here that there is another factor which must be considered in determining the shape of the LRAC curve. We assert that some products are produced by hospitals at volumes insufficient for full economies of scale to be achieved. In addition to the hetereogeneity problem, there also exists the problem of the effect on cost of the volume at which each individual product is produced. This factor, which has not been previously discussed with respect to its effect on the LRAC curve, is consistent with the Alchian theory without requiring Hirshleifer's special case.

Figure 1:
Empirical versus Theoretical Costs

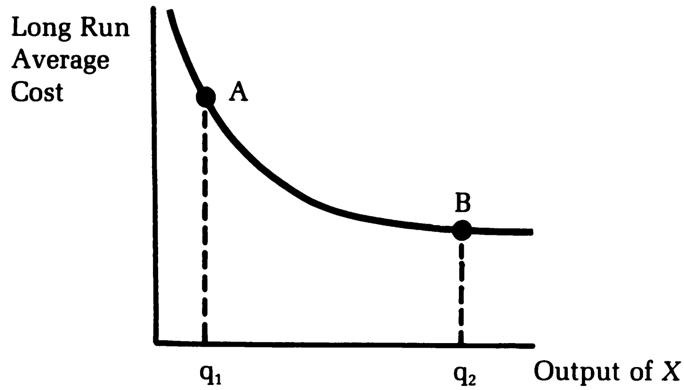


Assume that the hospital industry's LRAC curve is represented by Figure 1. Note that there are three hypothetical curves in the diagram. Curve T is the hypothetical theoretical LRAC curve as predicted by Alchian's theory. Implicit in that theory is efficient production of all products. The actual shape of this curve depends on the rate and volume of production. However, it is generally agreed that, theoretically, economies of scale are expected. Curve E is the hypothetical empirically observed LRAC curve, without adjusting for product or case mix. This curve will have a definite U-shape because it is the larger hospitals which tend to have more products and more of the highly sophisticated, expensive products. Between these curves is hypothetical Curve M, which is adjusted for case mix but not for the efficiency with which each product is produced. This curve will be identical with T if there are no inefficiencies in production. For multiproduct producers with inefficiency in the production of at least one product, M will lie between T and E and may exhibit diseconomies, no economies, or some economies of scale.

The early studies of hospital cost curves tended to look like Curve E because of their failure to adjust for case mix. The more recent studies, adjusting for case mix, have fallen into the M curve category. Our main concern is determining why the M and T curves do not coincide. We propose that this results from failure to achieve full scale economies in the production of individual products. Such production can be sustained in hospitals because of the lack of competitive markets.

Existing payment schemes are not capable of forcing inefficient producers out of business. For example, perhaps under free market conditions there might be 100 producers of open heart

Figure 2:
Individual Product Line Long Run Average Cost Curve



surgery, each performing 500 cases per year. Given the present cost reimbursement system, it is possible for there to be 500 producers, each performing 100 cases per year.

We know that the larger hospitals are more likely to offer the sophisticated, expensive services, such as heart surgery[12]. If they offer it inefficiently, the larger hospitals will appear to suffer from diseconomies of scale when the estimated LRAC curve E or M is observed. In that case, large hospital size would be incorrectly attacked as being inherently inefficient, while the real problem lies in the failure of hospitals to realize full scale economies in production of individual products.

For example, if we select any firm size to place in Figure 1, for instance Q, we are selecting an aggregation of costs for all of the products produced by firms of that size. Each of the products making up the aggregation has its own LRAC curve. Figure 2 presents a LRAC curve for product X. What if large firms produce a number of specialized products at a high cost point on the LRAC curves of the specific products, such as point A in Figure 2? If such a hypothetical situation were the actual case, then Curve M in Figure 1 would be the expected empirical observation (after adjusting for case mix), even though Curve T is the correct theoretical function.

Such production is not necessarily bad. If the specialized product offered at high cost on its LRAC curve is an emergency product that requires the speed of close proximity, then high cost may be a socially efficient outcome. On the other hand, for elective services it is less likely that society's resources are being used efficiently. Rather than large size being the culprit, policy makers should confront hospital product scope, allowing some hospitals to expand production and causing some hospitals to cease production of certain product lines.

The next step should be to empirically test the assertions that have been made here. Essentially, this requires estimation of the LRAC curve for each product at a number of hospitals. The aggregation of the individual product line cost curves should produce an industry LRAC curve similar to T in Figure 1. (This assumes the use of efficient costs such as point B in Figure 2). This could be compared to an empirically observed curve based on aggregate information. Such a curve would correspond to E or M in Figure 1.

This test would achieve two ends. First, it would more accurately display the true shape of the hospital industry LRAC curve. Second, in theory, the potential social inefficiency from failure to achieve economies of scale in the production of individual products can be measured by the difference between the area under the observed curve E or M and the estimated Curve T which is based on disaggregated product line cost curves.

However, to make that measurement, we must meet the following conditions: 1) we must compare only firms offering the same product line mix, or must devise a way to adjust for product line mix (note that for one group of firms with a more costly product mix, the cost curves will be higher than for a group with a less costly product mix); 2) there must be adjustment for travel, inconvenience, and other costs that change as average firm size changes; and 3) there must be adjustment for other potential reasons for cost differences such as different levels of quality. It is not likely that we will be able to make all of these adjustments satisfactorily.

Of greater significance, however, is a lack of data regarding product line costs. Hospitals collect cost information based on services such as surgery or radiology rather than product lines. This would be analogous to an auto manufacturer's knowing the total cost and average cost of bumpers for all car models combined, but not knowing the bumper costs associated with any particular model. As a result, the LRAC curve for specific product lines is not generally known, and it would require a research effort of enormous magnitude to estimate it for all product lines at a number of hospitals. Without the individual product line LRAC curves, we cannot estimate the true shape of the industry curve.

However, there is some available empirical evidence. Two separate case studies on the elective product line of open heart surgery both found significant economies of scale in production of that elective service[13,14]. In each study, costs per unit of output (measured by annual open heart surgery) fell sharply for at least the first 100 cases. One of the two studies found significant scale economies at production levels of up to 400 cases[13]. The state of California collects data on open heart surgery production[15]. During the most recent year for which data are available, a total of 91 hospitals in California offered open heart surgery. Of those, 48 produced 100 or fewer cases per year, and all but 7 produced 400 or fewer cases per year.

The implications of this data are that a significant number of hospitals offering open heart surgery are doing so at volumes insufficient to achieve potential scale economies. Attempts to measure a LRAC curve for a group of hospitals including these 91 hospitals, without adjusting for the inefficiently low volume at which many of them produce open heart surgery, would fail to give accurate information, even if allowance is made for case mix (i.e., the fact that these hospitals do offer open heart surgery). Such an attempt would incorrectly estimate diseconomies of scale.

Summary and Conclusions

Empirical studies in the hospital industry have produced conflicting results with respect to the shape of the industry's long run average cost curve. Some of the studies have found a classical U-shaped curve. Others have produced results indicating that the LRAC curve is much closer to being L-shaped.

Some of the studies failed to account for the heterogeneous nature of hospital production. They found a U-shaped LRAC curve and concluded that large hospitals are inefficient. Studies acknowledging and adjusting for case mix find significantly fewer or no diseconomies of scale in hospital size, but do not find strong cost justification for large hospitals.

While classical theory predicts that the LRAC curve will be U-shaped, Alchian has presented theoretical arguments for such curves to be L-shaped. Despite attempts by Hirshleifer to salvage the classical theory, Alchian's arguments are quite strong, and in general, economies of scale are expected to occur for firms in equilibrium.

This paper discusses the role of case mix in causing the different reported observations of the shape of the hospital industry LRAC curve. It takes a further step in attempting to reconcile the studies which show slight diseconomies with Alchian's predictions of scale economies. The basis of this reconciliation is recognition of the failure of individual hospitals to produce individual products at efficient volumes. Such inefficient production is feasible, and possibly common, in hospitals, given the incentive structure which exists under current cost reimbursement payment systems.

Large hospitals may have a greater potential for scale economies than has been recognized, but such economies have been ignored because they are, at least in part, offset by diseconomies in the production of individual products by individual hospitals. Cost reimbursement makes it financially possible for hospitals to offer some products at volumes so low that they would not be offered by some of the hospitals in a competitive environment.

Econometric techniques are unable to distinguish between diseconomies due to size and those due to inefficient production of specific products. This is a result of the type of cost data available. Such data are highly aggregate in nature at the level of the firm. As a result, inefficiencies in the production of specific products are hidden, and study results may lead to potentially incorrect policy recommendations about appropriate hospital size.

Further research into the efficiency of production of specific product lines is needed. Until the data are available to permit such studies to be undertaken, we must at least remain aware of the potential impact of such inefficiencies on the shape of the hospital industry LRAC curve.

ACKNOWLEDGEMENTS

The author is indebted to the helpful comments of Alain Enthoven, Hal Luft, J. Michael Cummins, and David Teece.

APPENDIX

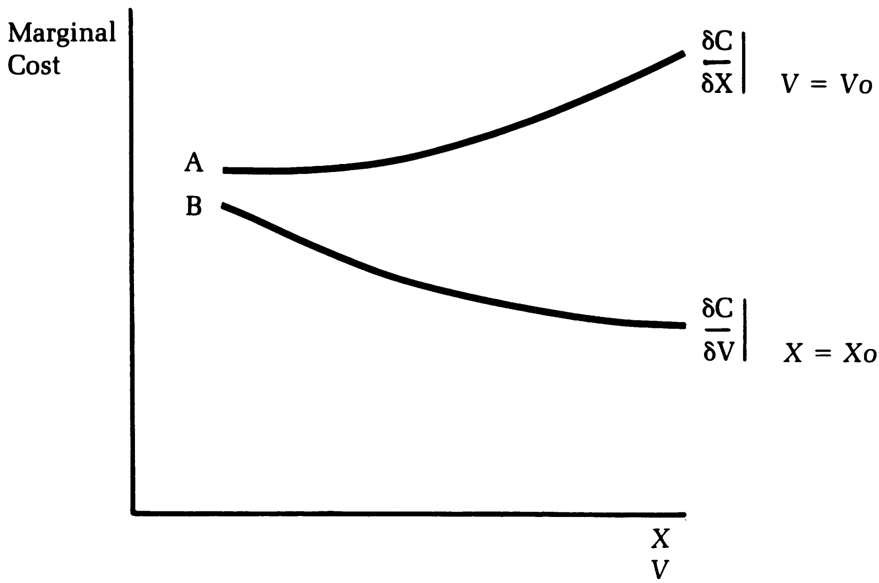
Alchian's model starts with the assumption that $C = f(V,X,T,m)$, where C is total cost in dollars to produce V, the scheduled volume of output; X is the rate of output; T is the time at which production begins, and m is the length of period during which the production process takes place. Further, $C \geq 0$. Then, by identity,

$$V = \int_{T}^{T+m} X(t)dt \tag{1}$$

Now, by assuming that X is constant over the production period m, and assuming that we cannot control the commencement of production, T, then the above identity reduces to

$$V = mX \tag{2}$$

Figure 3: Cost Curve—Rate and Volume



Looking at the partial derivatives of cost with respect to the independent variables X and V, we note Alchian's assertion (discussed below) that

$$C_X>0 \qquad C_V>0$$
$$C_{XX}>0 \qquad C_{VV}<0 \qquad (3)$$

Comparing Curves A and B in Figure 3, we can readily see the implication of Alchian's assertions. In Curve A we note that the marginal cost with respect to changes in the rate of production, C_X, is constantly increasing while we hold volume constant. Alchian contends that producing at a higher rate is always more costly than at a lower rate. His justification is that producing a set volume over two periods would require less fixed capital than producing that same volume in only one period. This assertion is of questionable validity because production at a higher rate might allow for achievement of economies of scale. Only at volumes above the point at which diseconomies of scale set in would Alchian's assertion apply.

In Curve B, the marginal cost with respect to changes in the volume, C_V, is constantly decreasing, although always positive, while we hold the rate constant. Thus, when volume increases and the production rate is constant, the marginal cost falls. Hirshleifer explains that the larger volume allows better scheduling and therefore matching of inputs and outputs, and the use of more durable equipment. Also learning-curve effects exist as management and labor both add a number of improvements to the production process. Thus, the decrease in cost for each additional unit is logical.

Hirshleifer attempts to show that the U-shaped average cost curve is a special case of the Alchian model. By plotting C on the vertical axis with X and V in the horizontal plane, he draws attention to the shape of the marginal cost function over the ray of slope $V/X=m$. He then asserts that

$$\left.\frac{dC}{dX}\right|_{V=mx} = \left.\frac{\delta C}{\delta X}\right|_{V=Vo} + \left.\frac{\delta C}{\delta V}\right|_{X=Xo} + \left.\frac{\delta V}{\delta X}\right|_{V=mX} \quad (4)$$

or simplifying the notation

$$\frac{dC}{dX} = C_X + mC_V \quad (5)$$

$$\frac{d^2C}{dX^2} = C_{XX} + 2mC_{XV} + m^2C_{VV} \quad (6)$$

Proof of the desired point requires that d^2C/dX^2 becomes positive beyond some point as X and V increase proportionately and indefinitely. This is proved because the positiveness of C_{XX} must sometime dominate the negativeness of C_{VV} and possible negativeness of C_{XV} since they are limited by the constraint that $C_V>0$. For further details of this proof, see Hirshleifer[9].

While Hirshleifer has salvaged the classical theory from complete obliteration by Alchian's new theory, his results are very limited, holding only for a special case. This author rejects the idea that Hirshleifer's explanation is adequate to cover all observed cases of U-shaped average cost curves.

REFERENCES

1. Berki, S. *Hospital Economics*, p. 115. Lexington, MA: Lexington Books, 1972.
2. Berry, R. Product heterogeneity and hospital cost analysis. *Inquiry* 7:67, Mar. 1970.
3. Alchian, A. Costs and outputs. M. Abramovitz (ed.), *The Allocation of Economic Resources: Essays in Honor of B. F. Haley*, pp. 23–40. Stanford: Stanford University Press, 1959.
4. Johnston, J. *Statistical Cost Analysis*. New York: McGraw-Hill Book Co., 1960.
5. Carr, W. and P. Feldstein. The relationship of cost to hospital size. *Inquiry* 4:45, June 1967.
6. Cohen, H. A. Variations in cost among hospitals of different sizes. *Southern Economic Journal* 33:355, Jan. 1967.
7. Cohen, H. A. Hospital cost curves. In H. Klarman (ed.), *Empirical Studies in Health Economics*, pp. 279–93. Baltimore: Johns Hopkins Press, 1970.
8. Feldstein, M. *Economic Analysis of Health Service Efficiency*. Chicago: Markham Publishing Co., 1968.
9. Lave, J.R. and L.B. Lave. Hospital cost functions. *American Economic Review* 60:379, June 1970.
10. Hirshleifer, J. The firm's cost function: A successful reconstruction? *Journal of Business* 35:235, July 1962.
11. McNerney, W.J. et al. *Hospital and Medical Economics*. Chicago: Hospital Research and Educational Trust, 1962.
12. American Hospital Association. *Hospital Statistics*, Table 12A, p. 188. Chicago: American Hospital Association, 1976.
13. Finkler, S. Nonprofit hospital product-line cost assessment and interfirm allocation of production. Doctoral dissertation, Stanford University, 1978.
14. McGregor, M. and G. Pelletier. Planning of specialized health facilities. *New England Journal of Medicine* 299:179, July 1978.
15. The State of California, Department of Health, Office of Statewide Planning. Numbers of Cardiovascular Surgical Services Provided in 1975 by Facility. In 1975 *Annual Report of Hospitals*. Sacramento: The State of California, Department of Health, 1976.