Efficiency of the Railroad Industry: A Frontier Production Function Approach

Richard Grabowski
Southern Illinois University, Carbondale

Seyed Mehdian
Temple University

Abstract

This paper uses a production frontier method to measure the revenue efficiency of the railroad industry in the U.S. from 1951 to 1981. The form specified for the frontier is a ray-homothetic production function. It allows returns to scale to vary with output and factor intensity. In addition, it is possible to determine to what extent revenue inefficiency was the result of operating at the inappropriate scale and what part was due to wasting resources. The results indicate that there was little inefficiency due to wasting resources. The main source of inefficiency was operation at decreasing returns to scale.

Introduction

A number of frontier methods for measuring efficiency have been developed. They involve the construction of a best practice frontier and measure inefficiency relative to this frontier. These approaches differ in many ways, but the two main differences involve the method used to determine the shape and placement of the frontier and the interpretation given to deviations from the frontier.[1]

This paper presents a relatively new specification of the frontier for measurement of efficiency. In order to construct the frontier, the corrected ordinary least squares (COLS) method is used. Instead of assuming a Cobb-Douglas specification for the technology, a linear form of the ray-homothetic production function is used. This form is particularly appropriate for analyzing the relationship between size and technical efficiency because returns to scale are allowed to vary with the level of output and relative factor intensity. This approach not only allows measurement of the extent of inefficiency, but also permits inefficiency to be attributed either to operation off the isoquant or operation at an inappropriate scale (nonconstant as opposed to constant returns to scale).

[1]See, for example, Faere, Grosskopf, and Lovell [5], Aigner and Chu [1], and Russell and Young [10].

0747-5535/90/1300-0026 $01.00

The paper applies the ray-homothetic frontier methodology to time series data for U.S. railroads for the 1951-1981 period. The railroad industry for most of this period was a highly regulated industry. The analysis determines the extent to which such a regulated industry operated efficiently. In addition, it is possible to determine to what extent inefficiency was due to operating at an inappropriate scale, i.e., nonconstant returns to scale.[2]

The results indicate that there was little loss of output or inefficiency due to wasting resources. Significant loss of output, however, occurred as a result of operating at nonconstant returns to scale. This was mainly the result of operating at too large a scale, thus experiencing decreasing returns.

Frontier Production Functions

Much of the early work [3, 8] concerned with the efficiency or performance of a specific industry and the role of economies of scale involved the estimation of production and cost functions. These studies generally estimate their cost or production functions using regression analysis; the function fitted had positive as well as negative residuals. Therefore, in the case of a production function, the regression estimated mean output rather than maximal output. If one is interested in questions concerning efficiency, it is maximum output that is important, not mean output. A frontier production function approach avoids this problem by constructing a frontier with all observations either lying on or below the frontier. The maximal output for a particular input combination can be determined.

Before discussing the methodology, a brief discussion of the term *efficiency* is needed. Technical efficiency refers to maximizing the output of various goods produced given any particular input combination. Revenue efficiency refers to maximizing the revenue resulting from the production of various goods. The distinction is illustrated in Figure One. Assume that there are two goods y_1 and y_2 and that the increasing cost situation exists. Thus, pp represents the production possibilities curve, given the technology. Points A, B, and C represent technically efficient points on the frontier. The slope of line ff represents the relative price ratio for the two outputs. Point B is both

[2]The term *inappropriate* as applied to scale needs additional explanation. From an individual firm's perspective, profit maximization in the short-run may lead to operation of the firm at nonconstant returns to scale. Even in the long-run, government regulation or conditions of imperfect competition may lead the firm to operate at nonconstant returns to scale. From society's perspective, however, the optimal situation is one in which firms operate at constant returns to scale. Thus, when the term inappropriate is used, the perspective taken is that of society, not of the individual firm.

technically and allocatively efficient and represents revenue maximization. In this study, revenue efficiency is calculated which allows for the measurement of the extent of both allocative and technical efficiency combined.[3]

There are a variety of frontier approaches available for measuring revenue efficiency. This paper divides the approaches into four basic types. The first involves the use of programming techniques to construct a nonparametric frontier. The extent of inefficiency of an observation is measured relative to this frontier [7]. This approach has been extended by Faere, Grosskopf, and Lovell to incorporate nonconstant returns to scale and to allow the possibility of input congestion. The major advantage of this approach is that no functional form is imposed on the data. The method is subject, however, to distortions introduced by data outliers, random shocks, measurement errors, etc. The second approach involves using programming techniques to construct a parametric frontier [1]. The only difference between this and the first approach is that a specific functional form is stipulated. The third approach involves the construction of a stochastic frontier.[4] A functional form must be specified, and the frontier is statistically estimated. The main advantage of the third approach is that it allows one to distinguish between economic inefficiency and the effects of random shocks, data outliers, and measurement error. The biggest disadvantage is that a distribution of the error term must be specified, and there are no criteria for doing this.

In this study, the fourth type of frontier, deterministic-statistical, is estimated using corrected ordinary least squares (COLS). A functional form for the production function is assumed, and parameter estimates are obtained using ordinary least squares (OLS) estimation techniques. The intercept is corrected by shifting the value of the intercept until no residual is positive and at least one is zero. This method allows measurement of the revenue inefficiency of each individual observation and analysis of the results.[5]

In other studies using COLS, a Cobb-Douglas production function is often used. In this paper, the ray-homothetic production function is used. The Cobb-

---

[3]In frontier approaches to measuring efficiency, some measure of output is necessary. If one has data on the amount of each type of good or service that is produced, then it is possible to derive separate measures for allocative and technical efficiency by using a multiple output approach. Disaggregated data concerning output were not available to the authors, however; therefore, operating revenue is used as a measure of output.

[4]See the work of Aigner, Lovell, and Schmidt [2], Meeusen and van den Broeck [9], and Schmidt and Lovell [12].

[5]An example of the application of COLS can be found in Russell and Young [10].

Douglas is a homogeneous production function in which returns to scale do not vary with output or input mix. Ray-homogeneous functional forms allow returns to scale to vary with input mix, but not output. Alternatively, homothetic production functions allow returns to scale to vary with output, but not with the input mix. Ray-homothetic production functions are homothetic along each ray of input combinations, but possibly in different ways for different input combinations. For this class of functions, returns to scale can vary both with output and with input mix--thus, the optimal scale (constant returns) can vary with input mix.

Letting x represent an input vector and $\phi(x)$ the maximum output attainable from the input vector x, the ray-homothetic production function can be written as

(1) $$\phi(\lambda x) = F(\lambda^{H(x/\|x\|)} \cdot F^{-1}(\phi(x))),$$

where

$\|x\|$ = the norm of x, $\lambda > 0$, and
F = a monotonically increasing transformation of $\lambda^{H(x/\|x\|)} \cdot F^{-1}(\phi(x))$.[6]

If F is the identity function, then

(2) $$\phi(\lambda x) = \lambda^{H(x/\|x\|)}\phi(x), \text{ where } H(x/\|x\|) > 0.$$

Equation (2) is a ray-homogeneous production function [4]. A ray-homothetic function is a monotonic transformation of a ray-homogeneous function. If the function $H(x/\|x\|)$ is a positive constant for all values of x, it can be seen that equation (2) becomes a homogeneous production function, and equation (1) becomes a homothetic production function [12]. The homothetic, homogeneous, and ray-homogeneous functions are special cases of equation (1).

The returns to scale for a particular value of x are measured by the scale function, which often is referred to as the function coefficient or the elasticity of output. It can be written as

(3) $$u(x) = \lim_{\lambda \to 1} \frac{\lambda}{\phi(\lambda x)} \cdot \frac{\partial \phi(\lambda x)}{\partial \lambda}.$$

For the ray-homothetic function, equation (1), the scale function can be written as

[6] Much of this section draws on the work of Faere, Jansson, and Lovell [6].

(4) $u(x) = u(x/\|x\|), \phi(x))$.

In other words, the ray-homothetic production function allows returns to scale to vary with relative input intensity, $(x/\|x\|)$, and output, $\phi(x)$. Optimal scale, the level of output for which there are constant returns to scale, occurs where $u(x) = 1$. Decreasing and increasing returns are indicated if $u(x)$ is less than or greater than unity.

In this paper, labor (N), capital (K), and fuel (F) are used as inputs while revenue (y) is used as output. The ray-homothetic function can be written as

$$(5) \qquad y = \ln\theta + a_N\left(\frac{N}{N+K+F}\right)\ln N + a_K\left(\frac{K}{N+F+K}\right)\ln K + a_F\left(\frac{F}{N+F+K}\right)\ln F,$$

where θ, a_N, a_K, and a_F are all parameters.[7]

The returns to scale function specified by equation (5) can be written as

$$(6) \qquad u = \frac{a_N\left(\frac{N}{N+K+F}\right)}{y} + \frac{a_F\left(\frac{F}{N+K+F}\right)}{y} + \frac{a_K\left(\frac{K}{N+K+F}\right)}{y}$$

or more simply

$$(7) \qquad u = \frac{a_N N' + a_K K' + a_F F'}{y},$$

where

$$N' = \frac{N}{N+K+F}$$

$$K' = \frac{K}{N+K+F}$$

$$F' = \frac{F}{N+K+F}$$

The optimal scale of output is obtained where constant returns to scale prevail. This occurs where $u = 1$. From equation (6), the optimal scale of output (y^*) can be written as:

[7]During this time period, it is likely that technical innovation occurred. Thus, equation (5) was modified to incorporate a time variable. The coefficients for this variable were not statistically significant, however, and it was dropped from the analysis.

(8) $y^* = a_N N' + a_K K' + a_F F'$.

In order to determine revenue efficiency and the relative importance of scale inefficiency utilizing the ray-homothetic production function, the following procedure must be used. First, equation (5) is estimated using ordinary least squares (OLS). The intercept is corrected by shifting the value of the intercept until no residual is positive and at least one is zero. Potential output is calculated by substituting the quantity of each input actually used by the firm into the estimated ray-homothetic production function with the intercept corrected. Total output lost due to inefficiency is the difference between actual output and potential output. This lost output is characterized as pure revenue inefficiency.

This analysis is illustrated in Figure Two. The x_i axis measures the vector of inputs (held in constant proportions) and the y axis measures output (revenue). Production function A represents the estimated ray-homothetic function with a corrected intercept. Suppose observation one uses x_1 of the inputs and produces y_1. Because potential output is y_1', the ratio of actual to potential output is y_1/y_1'.

The ray-homothetic function also allows for scale inefficiency. An intuitive explanation involving Figure Two shows the output lost as a result of scale inefficiency (see appendix). Production function B represents a constant returns to scale function that is tangent to function A at point C, the optimal scale for function A. If observation one used input vector x_1 and achieved constant returns to scale, output would be y_1''. If pure technical inefficiency were zero (so that actual output were y_1'), scale inefficiency would be y_1'/y_1''. The total output lost due to scale inefficiency would be $(y_1'' - y_1')$.

To summarize, for input combination x_1 and output y_1, the total output lost from revenue inefficiency is $(y_1'' - y_1)$. The lost output due to pure revenue inefficiency is $(y_1' - y_1)$. The output lost due to scale inefficiency is $(y_1'' - y_1')$.

Pure revenue inefficiency occurs when, with the existing technology and input combination, a firm could produce more revenue with the inputs it employs (or the same level of revenue with fewer inputs). A firm is scale inefficient when it operates at nonconstant returns to scale. This may not represent any inability to optimize on the part of the firm. The market may not be competitive, or price distortions may occur. Consequently, it may be optimal (in terms of maximizing profit, revenue, or some other activity) for the firm to operate at nonconstant returns to scale. Alternatively, the firm may be adjusting toward an equilibrium position that is constantly changing as a result of risk, technical change, etc. Operating at nonconstant returns to scale may be socially inefficient, but not necessarily inefficient from the individual firm's point of view.

Frontier Efficiency

The data on inputs and output used in this paper were drawn from the *Yearbook of Railroad Facts* published by the economics and finance department of the Association of American Railroads in Washington, D.C. The time period used was from 1951 to 1981.[8] Output was measured as total operating revenue, labor by labor compensation measured in dollars, capital by the total investment in road and equipment in dollars, and fuel by total fuel expenditures. All were measured in thousands of dollars and deflated, using a GNP deflator with 1970 as base year, to represent real expenditures.

The first step in applying the corrected ordinary least squares approach is to estimate equation (5) using ordinary least squares. The Durbin-Watson statistic derived from the results indicates the existence of serial correlation. The equation was estimated using an error term that can be written as

(9) $\varepsilon_t = p\varepsilon_{t-1} + u_t,$

where

u_t is assumed to be homoscedastic and nonautoregressive. The results are presented in Table 1. All of the coefficients are highly significant.

The next step is to adjust the intercept of the ray-homothetic function upward until no residual is positive and at least one is zero. Using the procedure outlined in the previous section, the overall revenue efficiency measure ((y_1/y_1'') in Figure 2), the pure revenue efficiency measure ((y_1/y_1') in Figure 2), and the scale efficiency measure ((y_1'/y_1'') in Figure 2) can be calculated. These measures are presented in Table 2.

Throughout the time period under consideration, there was little output lost as the result of pure revenue inefficiency. The main source of revenue inefficiency seems to have been the result of scale inefficiency. To determine whether this was the result of operating at decreasing or increasing returns to scale, the optimal (constant returns to scale) level of output was calculated for each year. This was done by taking the values for a_N, a_K, and a_F from Table 2 and substituting them into equation (8). Substituting the values for N', K', and

[8]Data availability determined the particular time period chosen.

F' for each year into the equation allowed calculation of y* (optimal) for each year. This is presented in Table 3 with the actual output levels (y).

Throughout the time period, the actual revenue produced exceeded the optimum. A t-test was performed in order to determine whether the means of y and y* were statistically different; this indicated that the difference was significant at the 1 percent level. This indicates that decreasing returns to scale prevailed. In other words, on average, the railroad industry's main source of inefficiency stemmed from operating at too large a scale. Table 3 further indicates that scale inefficiency declined markedly during the 1950s and early 1960s, stabilized throughout the late 1960s and the mid 1970s, and worsened in the late 1970s and early 1980s. The particular problems experienced in the early 1980s were most likely the result of the economic slowdown and recession that occurred during that time period.

Before closing, several additional points need to be made. It should be remembered that the data used in this study were highly aggregated time series data. Thus, the results should be interpreted as indicating possibilities for further research rather than firm conclusions. For example, an extension would involve using disaggregated data allowing for multiple outputs and a more detailed breakdown of inputs. The ray-homothetic functional form still could be used. In this case, however, the methodology would have to be different. One could use either a parametric programming approach or a programming cost function approach. In the first procedure, a linear ray-homothetic, multiple output production frontier would be calculated using linear programming techniques. In the second, a ray-homothetic cost frontier would be created using linear programming techniques. In either case, one would be able to determine the extent of inefficiency and the source of the inefficiency.

Summary

In this paper, a production frontier method was used to evaluate the revenue efficiency of the railroad industry from 1951 to 1981. The functional form used to construct the frontier was a ray-homothetic production function. This form allowed returns to scale to vary with output and relative factor intensity. Corrected ordinary least squares was the methodology used to construct the frontier.

The results indicated that the main source of revenue inefficiency was operating at a nonoptimal scale. Throughout this time period, the railroad industry was subject to decreasing returns to scale. On average, the scale of operations was too large. This may have been due to the fact that during much of this time period the industry was regulated heavily. The industry was required to provide some minimum level of services and, thus, was operating at too large a scale.

This study is different from almost all previous work on the railroad industry. Prior work involved the estimation of either average production or cost functions, with the goal of determining appropriate scale. The implicit assumption is that production is technically efficient. In this paper, the assumption of technical efficiency is relaxed.

Finally, one should view this particular study as a tentative first step. Future research should be directed toward utilizing frontier methodologies that allow multiple outputs. In addition, disaggregated cross-sectional data would allow the construction of separate measures of technical and allocative efficiency for each observation. Also, one could attempt to determine whether efficiency is related to various structural characteristics of particular firms.

References

1. Aigner, D.J. and S.F. Chu, "On Estimating the Industry Production Function," *American Economic Review*, 58 (September 1968), pp. 826-839.

2. Aigner, D.J., C.A.K. Lovell, and P.J. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6 (July 1977), pp. 21-37.

3. Borts, George H., "The Estimation of Rail Cost Functions," *Econometrica*, 28 (January 1960), pp. 108-131.

4. Eichorn, W., "A Generalization of the Concept of Homogeneous Production Functions," *Unternehmensforschung* (Now known as the *Journal of Operations Research*) 13 (1969), pp. 99-109.

5. Faere, R., S. Grosskopf, and C.A.K. Lovell, *Measurement of Efficiency of Production* (Boston: Kluwer Nijhoff, 1985).

6. Faere, R., L. Jansson, and C.A.K. Lovell, "Modeling Scale Economies with Ray Homothetic Production Functions," *The Review of Economics and Statistics*, 67 (November 1985), pp. 624-629.

7. Farrell, M.J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Association* 120 Series A - General (1957), pp. 253-281.

8. Keeler, Theodore, "Railroad Costs, Returns to Scale, and Excess Capacity," *The Review of Economics and Statistics*, 56 (May 1974), pp. 201-208.

9. Meeusen, W.J. and J. van den Broeck, "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review*, 18 (June 1977), pp. 435-444.

10. Russell, N.P. and T. Young, "Frontier Production Functions and the Measurement of Technical Efficiency, *Journal of Agricultural Economics*, 34 (May 1983), pp. 139-150.

11. Shephard, R.W., *Cost and Production Functions* (Princeton: Princeton University Press, 1953).

12. Schmidt, P. and C.A.K. Lovell, "Estimating Stochastic Production and Cost Frontiers When Technical and Allocative Inefficiency are Correlated," *Journal of Econometrics*, 13 (May 1980), pp. 83-100.

Table 1
Ordinary Least Squares Estimates
(Inputs and Outputs Measured in Thousands of Dollars/Deflated)

Variable	Coefficient	t-statistic
Intercept	-115,715,620	-8.81
Labor: N' (lnN)	15,195,048	9.22
Capital: K' (lnK)	6,782,089	9.72
Fuel: F (lnF)	6,988,427	9.01

$R^2 = .92$
Durbin-Watson = 1.80

Table 2
Efficiency Measures

Year	Overall Revenue Efficiency	Pure Revenue Efficiency	Scale Efficiency
1951	.50	.91	.55
1952	.50	.91	.55
1953	.60	1.00	.60
1954	.62	.95	.65
1955	.63	.96	.66
1956	.64	.98	.65
1957	.61	.95	.65
1958	.61	.88	.69
1959	.62	.90	.69
1960	.61	.87	.70
1961	.63	.86	.73
1962	.64	.87	.73
1963	.65	.88	.74
1964	.66	.89	.74
1965	.68	.90	.75
1966	.70	.92	.75
1967	.67	.88	.77
1968	.68	.88	.77
1969	.68	.88	.77
1970	.66	.87	.77
1971	.66	.87	.76
1972	.65	.86	.75
1973	.66	.87	.75
1974	.68	.92	.73
1975	.66	.87	.75
1976	.65	.90	.73
1977	.64	.90	.71
1978	.65	.90	.72
1979	.60	.91	.66
1980	.52	.86	.61
1981	.49	.86	.57

Table 3
Optimal and Actual Output
(Measured in Thousands of Dollars/Deflated)

Year	y* (Optimal Output)	y (Actual Output)
1951	7,787,714	17,945,788
1952	7,803,123	17,994,493
1953	7,612,805	17,922,973
1954	7,563,057	16,004,648
1955	7,524,549	16,092,882
1956	7,359,038	16,257,231
1957	7,505,435	15,896,045
1958	7,518,406	14,148,769
1959	7,507,930	14,301,397
1960	7,518,819	13,729,140
1961	7,478,178	13,015,776
1962	7,498,691	13,165,823
1963	7,48,453	13,131,212
1964	7,498,485	13,248,020
1965	7,494,791	13,291,480
1966	7,470,357	13,469,869
1967	7,452,706	12,564,898
1968	7,459,056	12,505,389
1969	7,468,068	12,527,708
1970	7,499,263	12,491,310
1971	7,523,264	12,689,016
1972	7,538,420	12,674,683
1973	7,501,646	12,832,391
1974	7,245,016	13,452,179
1975	7,243,798	12,413,728
1976	7,245,608	13,240,344
1977	7,226,306	13,358,033
1978	7,242,720	13,293,349
1979	7,148,087	14,397,613
1980	7,059,211	14,374,908
1981	7,016,301	14,835,869

Figure 1
Technical and Allocative Efficiency

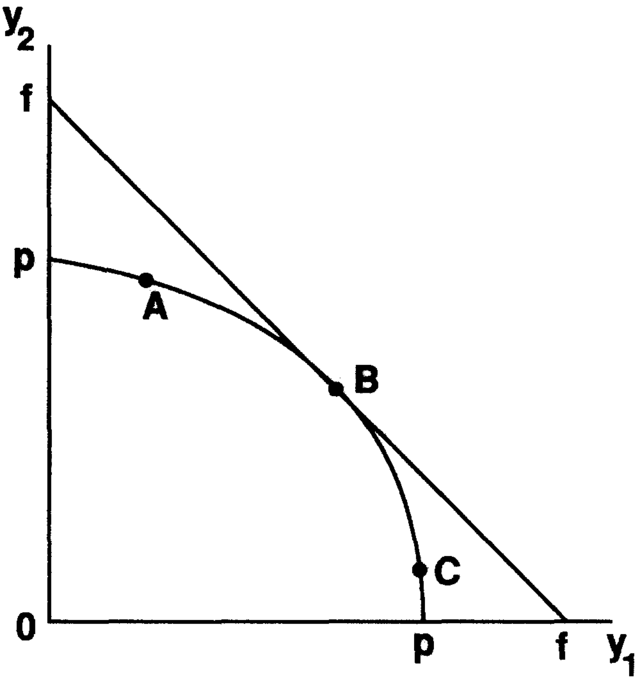


Figure 2
Revenue Efficiency

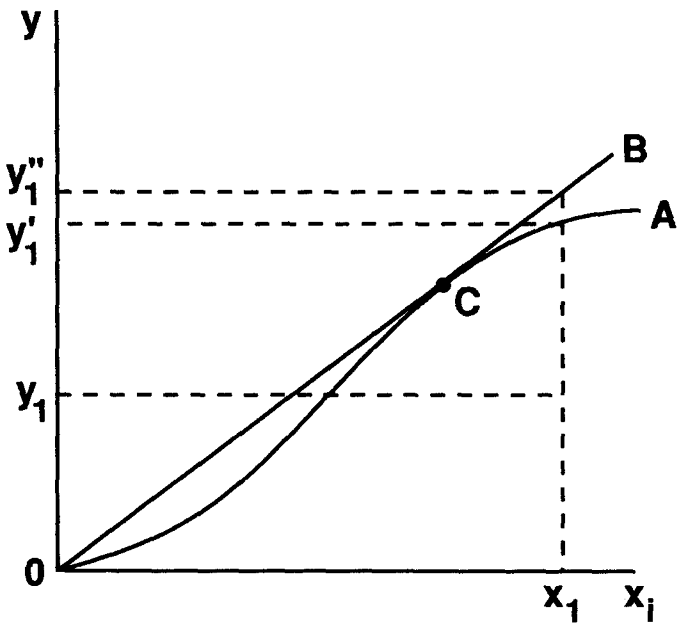
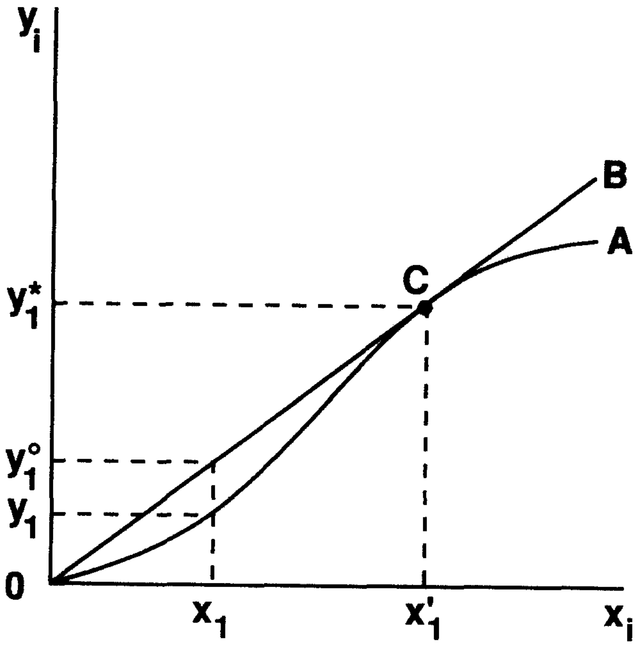


Figure 3
Measuring Scale Inefficiency



Appendix

The analysis discussed in this section is based on a simple two input case. It can be generalized easily to the situation of n inputs. The ray-homothetic production function in this case can be written

$$(10) \qquad y_i = \ln\theta + a_K \frac{K_i}{K_i + L_i} \ln K_i + a_L \frac{L_i}{K_i + L_i} \ln L_i \,,$$

where

y_i, K_i, and L_i = revenue, capital, and labor used by firm i respectively.
θ, a_L, and a_K = parameters to be estimated.

The actual level of output and quantities of inputs used are known. The optimal level of output for each individual firm, y_i^*, can be determined by solving

$$(11) \qquad y_i^* = a_K\left(\frac{K_i}{K_i + L_i}\right) + a_L\left(\frac{K_i}{K_i + L_i}\right) .$$

Multiplying K_i and L_i in equation (10) by u and setting equation (10) equal to the optimal level of output for firm i, y_i^*, gives

$$(12) \qquad y_i^* = \ln\theta + a_K\left(\frac{K_i}{K_i + L_i}\right)\ln(K_i \cdot u) + a_L\left(\frac{L_i}{K_i + L_i}\right)\ln(L_i \cdot u) .$$

Thus, u would be the number by which L_i and K_i would be multiplied if the optimal level of output were to be produced by the ith firm. Note that the ratios

$$\frac{K_i}{K_i + L_i} \text{ and } \frac{L_i}{K_i + L_i}$$

would not change. Solving for lnu gives

$$(13) \qquad \frac{y_i^* - y_i}{y_i^*} = \ln u \text{ or}$$

$$1 - \frac{y_i}{y_i^*} = \ln u.$$

Taking the antilog will give the actual value for u. Finally, u will be positive and less than one for firms experiencing decreasing returns to scale and positive and greater than one for firms with increasing returns to scale.

Examining Figure 3, the x_i axis measures the vector of inputs, with movements to the right representing an equiproportional increase in all inputs, and the y_i axis output. Production function A represents the variable returns to scale production function estimated in this paper. Firm one uses x_1 of the inputs and produces y_1 actual output. Production function B represents the constant returns to scale function. (u-1) represents the percent by which all input usage would have to increase to allow the firm to produce at the optimal scale represented by point c, using x_1 inputs and producing y_1^* output. In addition, $(\frac{u-1}{u})$ can be calculated; y_1^o could be derived as

$$(14) \qquad y_1^o = y_1^* - (\frac{u-1}{u})\, y_1^* .$$

For firms operating at decreasing returns to scale the same sort of logic prevails. The output that could be produced by the firm if it were operating at constant returns to scale would be

$$(15) \qquad y_1^o = y_1^* + (\frac{1-u}{u})\, y_1^* .$$