

WIFO ■ BIBLIOTHEK

12. Juni 2003

ÖSTERREICHISCHES INSTITUT FÜR
WIRTSCHAFTSFORSCHUNG

B. Analyses of Costs

MARC NERLOVE

Returns to Scale in Electricity Supply

The study of returns to scale in public-utility enterprises has a long, if not always honorable, history. The question of whether there are increasing or decreasing returns to scale and over what range of output has, as we know, an important bearing on the institutional arrangements necessary to secure an optimal allocation of resources. If, as many writers in the field appear to believe, there are increasing returns to scale over the relevant range of outputs produced by utility undertakings, then these companies must either receive subsidies or resort to price discrimination in order to cover costs at socially optimal outputs.

In addition, as Cheney [2] has pointed out, the extent of returns to scale is a determinant of investment policies in growing industries. If there are increasing returns to scale and a growing demand, firms may find it profitable to add more capacity than they expect to use in the immediate future.

In studying the problem of returns to scale, the first question one must ask is "To what use are the results to be put?" It is inevitable that the Reprinted from *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yngvah Gersbach*, by Carl F. Christ et al., with permission of the author and the publishers, Stanford University Press. © 1963 by the Board of Trustees of the Leland Stanford Junior University.

	Country	L/V	\$w	L/V	\$w	L/V	\$w	L/V	\$w
1. United States (1954)	0.1266	4387	0.1119	4540	0.1332	4314	0.1514	4119	
2. Canada (1954)	0.1410	3769	0.0768	3065	0.1562	3607	0.1570	3296	
3. New Zealand (1953/56)	0.2349	21904	0.2349	3065	0.1562	4314	0.1514	4119	
4. Australia (1953/56)	0.2338	2306	0.2335	3194	0.1562	2190	0.1570	3296	
5. Denmark (1954)	0.2173	1626	0.2331	3194	0.1562	4314	0.1514	4119	
6. Norway (1954)	0.2918	1568	0.2684	1618	0.2800	1526	0.2800	1307	0.1370
7. United Kingdom (1951)	0.4203	1224	0.3870	1208	0.3484	1138	0.3484	1489	0.3266
8. Ireland (1953)	0.4203	1224	0.3870	1208	0.3484	1138	0.3484	1489	0.3266
9. Puerto Rico (1952)	0.2918	1568	0.2684	1618	0.2800	1526	0.2800	1307	0.1370
10. Colombia (1953)	0.4199	1168	0.4199	1168	0.4199	1168	0.4199	1168	0.4199
11. Brazil (1949)	0.5360	572	0.5360	572	0.5360	572	0.5360	572	0.5360
12. Mexico (1951)	0.4738	594	0.7920	591	0.52612	542	0.8747	455	0.5537
13. Argentina (1950)	0.4738	594	0.7920	591	0.52612	542	0.8747	455	0.5537
14. El Salvador (1951)	0.4738	594	0.7920	591	0.52612	542	0.8747	455	0.5537
15. Southern Rhodesia (1952)	0.7920	243	0.7920	243	0.7920	243	0.7920	243	0.7920
16. Iraq (1954)	1.2866	210	1.2866	210	1.2866	210	1.2866	210	1.2866
17. Ceylon (1952)	0.7266	661	0.9860	450	0.5432	621	1.2388	422	0.8034
18. Japan (1953)	0.7266	661	0.9860	450	0.5432	621	1.2388	422	0.8034
19. India (1953)	0.7266	661	0.9860	450	0.5432	621	1.2388	422	0.8034

*Refers to primary iron and steel as well as primary non-ferrous metals.

If the efficiency of firms varies neutrally,¹ as indicated by the error term in (1), and the prices paid for factors vary from firm to firm, then the levels of input are not determined independently but are determined jointly by the firm's efficiency, level of output, and the factor prices it must pay. In short, a fitted relationship between inputs and output is a *confluent* relation that does not describe the production function at all but only the net effects of differences among firms. (For a more general discussion, see [13, 16].)

In such cases, however, it may be possible to fit the *reduced form* of a system of structural relations such as (1) and (3) and to derive estimates of the structural parameters from estimates of the reduced-form parameters. Not only does it turn out to be possible in this case, but an important reduced form turns out to be the cost function:

$$(4) \quad c = k y^{1/r} p_1^{a_1/r} p_2^{a_2/r} p_3^{a_3/r},$$

where

$$k = r(a_0 a_1^{a_1} a_2^{a_2} a_3^{a_3})^{-1/r},$$

$$v = w^{-1/r},$$

and

$$r = a_1 + a_2 + a_3.$$

The parameter r measures the degree of returns to scale. The fundamental duality between cost and production functions, demonstrated by Shephard [17], assures us that the relation between the cost function, obtained empirically, and the underlying production function is unique.² Under the cost minimization assumption, they are simply two different, but equivalent, ways of looking at the same thing.

Note that the cost function must include factor prices if the correspondence is to be unique. The problem of changing (over time) or differing (in a cross section) factor prices is an old one in statistical cost analysis; see [10], pp. 170–76. Most generally, it seems to have been handled by deflating cost figures by an index of factor prices, a procedure that Johnston [10] shows typically leads to bias in the estimation of the cost curve unless correct weights, which depend on (unknown) parameters of the production function, are used. It seems strange that no one has taken the obvious step of *including factor prices directly in the cost function*. If price data are available for the

¹ A model incorporating non-neutral variations in efficiency of the form

$$y = (a_0 n_0)^{a_1 n_1 a_2 n_2 a_3 n_3}$$

was discussed in my paper "On Measurement of Relative Economic Efficiency," abstract, *Econometrica*, 28 (July 1960), 605. It is interesting to note that despite the complex way in which the random elements n_0, n_1, n_2 and n_3 enter, there are circumstances under which it is possible to estimate the parameters in such a production function.

² I owe this point to Hirofumi Uzawa. It is true, of course, only if all firms have the same production function, except perhaps for differences in the constant term, so that aggregation difficulties may be neglected.

construction of an index and prices do not move proportionately, in which case no bias would result from deflation, why not use the extra information afforded?

What form of production function is appropriate for electric power? The generalized Cobb-Douglas function presented above is attractive for two reasons: First, it leads to a cost function that is linear in the logarithms of the variables

$$(5) \quad C = K + \frac{1}{r} Y + \frac{a_1}{r} P_1 + \frac{a_2}{r} P_2 + \frac{a_3}{r} P_3 + V,$$

where capital letters denote logarithms of the corresponding lower-case letters. The linearity of (5) makes it especially easy to estimate. Second, a single estimate of returns to scale is possible (it is the reciprocal of the coefficient of the logarithm of output), and returns to scale do not depend on output or factor prices. (The last-mentioned advantage turns out to be a defect as we shall see when we come to examine a few statistical results.) But does such a function accurately characterize the conditions of production in the electric power industry?

A casual examination of trade publications suggests that once a plant is built, fixed proportions are more nearly the rule. Support for this view is given by Konnuya [11], who found that data on inputs and output for individual plants were better approximated by a fixed-proportions model than allowed differences in the proportions due to scale. A simplified version of Konnuya's model is³

$$(6) \quad x_1 = a_1 y^{\rho_1},$$

$$x_2 = a_2 y^{\rho_2},$$

$$x_3 = a_3 y^{\rho_3},$$

At the firm level, however, there are many possibilities for substitution that may go unnoticed at the plant level; for example, labor and fuel may be substituted for capital by using older, less efficient plants more intensively or by using a large number of small plants rather than a few large ones. Given persistent differences in the factor prices paid by different firms, cross-section data should reflect such possibilities of substitution. Certainly, as a provisional hypothesis, a generalized Cobb-Douglas function may be appropriate.

³ Since y is exogenous, this would be inappropriate to estimate the coefficients in (6) by least squares. An objection to this, however, is the fact that, if individual plants are considered, the output allocated to each is not exogenous; see Westfield [19], pp. 15–81. Furthermore, Konnuya does not use output but name-plate net capacity and input levels adjusted to full capacity utilization. It is even more doubtful whether the former can be considered as exogenous in a cross section. My objection here is closely related to the one raised by Hughes (see p. 169); however, while the endogeneity of output at the plant level is clear, its endogeneity at the firm level for a member of a power pool is conjectural.

It would, of course, be preferable to *test* whether significant substitution among factors occurs at the firm level. The use of the generalized Cobb-Douglas unfortunately does not permit us to do so except in a very general way, since its form implies that the elasticity of substitution between any pair of factors is one. A more general form, which has both the Cobb-Douglas and fixed coefficients as limiting cases, has recently been suggested by Arrow, Minhas, Chenery, and Solow [1]. Constant returns to scale are assumed, but the form can be easily generalized; in a more general form it is

$$(7) \quad y = [a_1 x_1^b + a_2 x_2^b + a_3 x_3^b]^{1/b}.$$

In this case returns to scale are given by the ratio b/f and the elasticity of substitution between any pair of factors can be shown to be $1/(1-b)$. In the special case in which $b=f$ it can be shown that the limiting form of (7) as the elasticity of substitution goes to zero is

$$(8) \quad y = \min \left\{ \frac{x_1}{(a_1 + a_2 + a_3)^{1/b} - 1}, \frac{x_2}{(a_1 + a_2 + a_3)^{1/b} - 1}, \frac{x_3}{(a_1 + a_2 + a_3)^{1/b} - 1} \right\}.$$

or fixed coefficients, and the limiting form as the elasticity of substitution goes to one is

$$(9) \quad y = (a_1 + a_2 + a_3)^{1/b} x_1^{a_1/(a_1+a_2+a_3)} x_2^{a_2/(a_1+a_2+a_3)} x_3^{a_3/(a_1+a_2+a_3)},$$

or Cobb-Douglas. Although I have not formally demonstrated the fact, it is possible that the limiting form of the more general case (7) is something like the Komiyama model as the elasticity of substitution tends to zero, and like the generalized Cobb-Douglas as it tends to one.

Unfortunately, in its generalized form (7) is quite difficult to estimate from the data available. Furthermore, although clearly superior to the generalized Cobb-Douglas form, (7) still implies that the elasticity of substitution between any pair of factors (e.g., labor capital and fuel capital) is the same, which hardly seems reasonable. Other generalizations are possible, but none that I have found thus far offers much hope of being amenable to a reasonable estimation procedure.

If the generalized Cobb-Douglas form is adopted, however, relatively simple estimation procedures can be devised for evaluating the parameters of the production function. The reduced form of (1) and (3), that incorporates all but one of the restrictions on the parameters in the derived demand equations (which are the more usual reduced form) is nothing but the cost function. The only restriction not incorporated in (4) or (5) is that the coefficients of the prices must add up to one. It is a simple matter to incorporate this restriction, however, by dividing costs and two of the prices by the remaining price (it doesn't matter either economically or statistically which price we choose). When fuel price is used as the divisor, the result is

$$(10) \quad C - P_3 = K + \frac{1}{r} Y + \frac{a_1}{r} (P_1 - P_3) + \frac{a_2}{r} (P_2 - P_3) + V,$$

which will be called Model A.

Model A assumes that we have relevant data on the "price" of capital and that this price varies significantly from firm to firm. If neither is the case, we are in trouble. Most of the results presented here are based on Model A, but the data used for this price of capital are clearly inadequate. (See Appendix B.) If one supposes, however, that the price of capital is the same for all firms, which is not implausible, one can do without data on capital price and use the restriction on the coefficients of output and prices to estimate the elasticity of output with respect to capital input. The assumption that capital price is the same for all firms implies

$$(11) \quad C = K' + \frac{1}{r} Y + \frac{a_1}{r} P_1 + \frac{a_2}{r} P_3 + V,$$

where $K' = K + (a_2/r) P_2$, since the exponents of the input levels in (1) are assumed to be the same for all firms. Equation (11) is called Model B.

2. SOME STATISTICAL RESULTS AND THEIR INTERPRETATION

Estimation of Model A from a cross section of firms requires that we obtain data on production costs, total physical output, and the prices of labor, capital, and fuel for each firm; for Model B we do not need the price of capital, since it is assumed to be the same for all firms. Details of the construction of these data for a sample of 145 privately owned utilities in 1955 are given in Appendix B and are not discussed here at any length. Suffice it to say that these data are far from adequate for the purpose, and I now believe that a better job could have been done with other sources.

The results from the least-squares regression suggested by equation (10) are given in line I of Table 3; the interpretation of these results in terms of the parameters of the production function is given in line I of Table 4. The R^2 is 0.93, which is somewhat unusual for such a large number of observations; increasing returns to scale are indicated, and the elasticities of output with respect to labor and fuel have the right sign and are of plausible magnitude; however, the elasticity of output with respect to capital price has the wrong sign (fortunately, it is statistically insignificant).

The difficulties with capital may be due in part to the difficulty I encountered in measuring both capital costs and the price of capital. The former were measured as depreciation charges plus the proportion of interest on long-term debt attributable to the production plant; the figure for capital price was compounded of the yield on the firm's long-term debt and an index of construction costs. Depreciation figures reflect past prices and purchases of capital equipment, whereas the price of capital as I constructed

TABLE 3

Results from Regressions Based on Model A for 145 Firms in 1955

Regression No.	Y	Coefficient			R^2	Elasticity of Output with Respect to			
		$P_1 - P_2$	$P_2 - P_3$	π		Scale	Labor	Capital	
I	0.721 (±0.175)	0.562 (±0.198)	-0.003 (±0.192)	—	0.931	I	1.39	0.78	
II	0.496 (±0.173)	0.512 (±0.199)	0.033 (±0.185)	-0.046 (±0.022)	0.932	II	1.44	0.74	
IIIa	0.398 (±0.079)	0.641 (±0.691)	-0.093 (±0.669)	—	0.512	IIIa	2.52	1.61	
IIIb	0.608 (±0.116)	0.105 (±0.275)	0.364 (±0.277)	—	0.635	IIIb	1.50	0.16	
IIIc	0.931 (±0.198)	0.408 (±0.199)	0.249 (±0.189)	—	0.571	IIIc	1.08	0.44	
IID	0.915 (±0.108)	0.472 (±0.174)	0.133 (±0.157)	—	0.871	IId	1.09	0.52	
IIIe	1.045 (±0.065)	0.604 (±0.197)	-0.295 (±0.175)	—	0.920	IIIe	0.96	0.78	
IVa	0.394 (±0.055)	{ }{ }{ }			0.950	IVa	2.52	1.10	
IVb	0.651 (±0.189)	{ }{ }{ }				IVb	1.53	0.65	
IVc	0.877 (±0.376)	0.435 (±0.297)	0.100 (±0.196)	—		IVc	1.14	0.50	
IVd	0.908 (±0.354)	{ }{ }{ }			—	IVd	1.10	0.48	
IVe	1.062 (±0.169)	{ }{ }{ }				IVe	0.94	0.41	

Figures in parentheses are the standard errors of the coefficients.
The dependent variable in all analyses was $C - P_3$.
The variables are defined as follows:

$$\begin{aligned} C &= \log \text{costs} & Y &= \log \text{output} & P_1 &= \log \text{wage rate} & P_2 &= \log \text{capital "price"} \\ P_3 &= \log \text{fuel price} & x &= \left| \frac{\text{output } 1955 - \text{output } 1954}{\text{output } 1954} \right| \end{aligned}$$

it does not; it is perhaps not so surprising then that the price has little effect on costs. Model B is designed to evade this difficulty. Results based on Model B are presented in line V of Table 5 and the implications of this

TABLE 4

Returns to Scale and Elasticities of Output with Respect to Various Inputs Derived from Results Presented in Table 3 for 145 Firms in 1955

Regression No.	Y	Coefficient			R^2	Elasticity of Output with Respect to			
		$P_1 - P_2$	$P_2 - P_3$	π		Scale	Labor	Capital	
I	0.721 (±0.175)	0.562 (±0.198)	-0.003 (±0.192)	—	0.931	I	1.39	0.78	
II	0.496 (±0.173)	0.512 (±0.199)	0.033 (±0.185)	-0.046 (±0.022)	0.932	II	1.44	0.74	
IIIa	0.398 (±0.079)	0.641 (±0.691)	-0.093 (±0.669)	—	0.512	IIIa	2.52	1.61	
IIIb	0.608 (±0.116)	0.105 (±0.275)	0.364 (±0.277)	—	0.635	IIIb	1.50	0.16	
IIIc	0.931 (±0.198)	0.408 (±0.199)	0.249 (±0.189)	—	0.571	IIIc	1.08	0.44	
IID	0.915 (±0.108)	0.472 (±0.174)	0.133 (±0.157)	—	0.871	IId	1.09	0.52	
IIIe	1.045 (±0.065)	0.604 (±0.197)	-0.295 (±0.175)	—	0.920	IIIe	0.96	0.78	
IVa	0.394 (±0.055)	{ }{ }{ }			0.950	IVa	2.52	1.10	
IVb	0.651 (±0.189)	{ }{ }{ }				IVb	1.53	0.65	
IVc	0.877 (±0.376)	0.435 (±0.297)	0.100 (±0.196)	—		IVc	1.14	0.50	
IVd	0.908 (±0.354)	{ }{ }{ }			—	IVd	1.10	0.48	
IVe	1.062 (±0.169)	{ }{ }{ }				IVe	0.94	0.41	

regression for the parameters in the production function are given in line V of Table 6. It is apparent that the estimates of returns to scale and the elasticities of output with respect to labor and fuel are changed very little; the elasticity with respect to capital is of the right sign but still unreasonably low for an industry that is so capital-intensive.⁴

⁴ K. Arrow has pointed out that considerations of plausibility implicitly involve an alternative method of estimating the coefficients in the production function: From the marginal productivity conditions (3), we find that for any pair of inputs i and j ,

$$\frac{p_x i}{p_x j} = \frac{a_i}{a_j},$$

Hence, by constraining some average of the ratios of expenditures on factors, we obtain estimates of the ratios of exponents in the production function. Had the data been arranged in such a manner as to facilitate computation of expenditures on individual factors, a comparison of the ratios a_i/a_j , obtained in this way with those derived from the cost function would have been a useful supplement to the analysis. Arrow also pointed out that one could also verify the results by the fit of the production function derived from them. Unfortunately, it is not feasible to obtain good physical measures of the inputs, and such measures are required for this test.

TABLE 5
Results from Regressions Based on Model B for 145 Firms in 1955.
Dependent Variable Was $\log(Y)$ - Log C_{fixed}

Regression No.	Y	P_1	P_2	R^2
V	0.723 (±0.019)	-0.483 (±0.303)	0.496 (±0.106)	0.914
VIA	0.361 (±0.086)	0.212 (±1.250)	0.655 (±0.350)	0.438
VIIB	0.661 (±0.106)	-0.401 (±0.333)	0.490 (±0.134)	0.672
VIC	0.985 (±0.180)	-0.014 (±0.261)	0.330 (±0.138)	0.647
VID	0.327 (±0.106)	0.327 (±0.228)	0.426 (±0.064)	0.884
VIE	1.035 (±0.067)	0.704 (±0.272)	0.643 (±0.132)	0.934

Figures in parentheses are the standard errors of the coefficients.

A second difficulty with these regressions is not apparent from an examination of the coefficients and their standard errors. As part of these analyses, the residuals from the regressions were plotted against the logarithm of output. The result is schematically pictured in Figure 1. It is clear that neither regression relationship is truly linear in logarithms. To test this visual impression the observations were arranged in order of ascending output, and Durbin-Watson statistics were computed; the values of the statistics indicated highly significant positive serial correlation, which confirmed the visual evidence.

Aside from difficulties with the basic data, there appear to be at least two plausible and interesting hypotheses accounting for the result.

1. The first explanation of the result derives from dynamic considerations closely related to those underlying Friedman's Permanent-Income Hypothesis [7]. The important thing to note is that actual costs are underestimated by the regressions at both high and low outputs. Consider the situation pictured in Figure 2. Firms operate not on the long-run cost curve, but at points on the various short-run curves. If firms are evenly distributed about their optimal outputs (i.e., outputs at which long-run marginal cost equals short-run marginal cost), the effect will be to increase the estimate of the extent of increasing returns to scale if they are increasing, or diminish further

TABLE 6

Returns to Scale and Elasticities of Output with Respect to Variable Inputs Presented from Results Presented in Table 5 for 145 Firms in 1955

Regression No.	Returns to Scale:	Elasticity of Output with Respect to
V	1.38	0.67
VIA	2.77	0.59
VIIB	1.61	-0.62
VIC	1.02	-0.01
VIN	1.08	0.35
VIE	0.97	0.68

the estimate of returns to scale if they are decreasing.⁵ But elsewhere Friedman holds that a uniform distribution is not likely to occur; in fact he says, "The firms with the largest output are unlikely to be producing at an unusually low level; on the average they are likely to be producing at an unusually high level; and conversely for those that have the lowest output" ([14], p. 237).

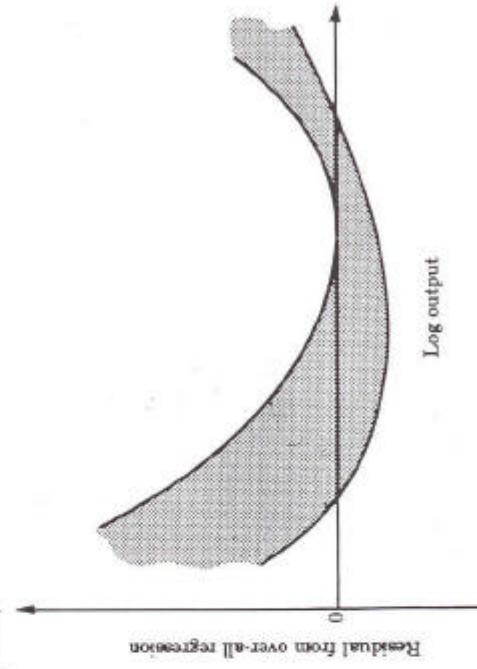


FIGURE 1

⁵This argument rests partly on the form of the function that constrains it to pass through the origin.

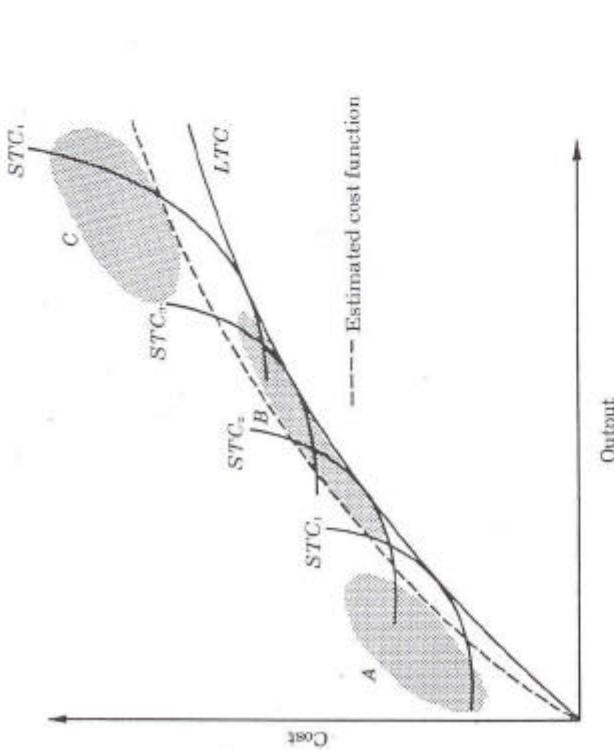


FIGURE 2

The situation described by Friedman is pictured in Figure 2 by the shaded areas A, B, and C, which refer, respectively, to observations on firms with unusually low, usual, and unusually high outputs. The Friedman explanation does produce a residual pattern similar to that observed. Regression II, Table 3, is designed to test this explanation for Model A. A corresponding test for Model B was not made. Since "usual" output cannot be directly observed, the hypothesis was modified slightly by identifying departure from the usual with large changes in output from the previous year, the assumption being that firms with stable output were likely to be near the optimal long-run output.⁸ Thus, the absolute percentage changes in output should be positively related to total costs. Unfortunately, they are negatively related and significantly so.

Part of the explanation for this unexpected result is suggested by a more careful examination of the data. Almost all firms with large changes

⁸ Capacity figures might have been used. However, those available appear to be somewhat unrealistic. These are based on generator name-plate ratings, which refer to the maximum output that can be produced without overheating. According to the Federal Power Commission, however, units of the same size, general design, and actual capability may show as much as a 20 per cent difference in rating ([5], p. xi). Furthermore, in a multiple-plant firm, total generator capacity is not the only factor to be considered. Such deficits in the capacity figures also led to grouping firms by output rather than by capacity in the analyses of covariance presented below.

had positive changes and had been experiencing rapid growth for some time. It is well known, though unfortunately not taken into account in these analyses, that there is a steady rate of technological progress in generating equipment. Since expanding firms purchase new equipment in the process, the average age of a plant in those firms experiencing large changes in output is lower than that of firms with more stable outputs. Hence, the former tend to have lower costs because of the inadequacy of the capital-cost data to reflect obsolescence.⁹ Thus, while one would not want to reject the Friedman hypothesis on the basis of this evidence, it clearly does not explain the residual pattern.

2. Fortunately, the observed result can be explained by a much simpler hypothesis, namely, that the degree of returns to scale is not independent of output, but varies inversely with it. Figure 3 illustrates this explanation: The solid line gives the traditional form of the total cost function, which shows increasing returns at low outputs and decreasing returns at high outputs. If we try to fit a function for which returns to scale are independent of the level of output, e.g., one linear in logarithms, a curve such as the dashed one will be obtained. The shaded areas A and B show the output ranges, high and low, for which total costs are underestimated.

If the true cost function is not linear in logarithms, we can either fit an over-all function that reflects this fact or attempt to approximate the actual function by a series of segments of functions linear in logarithms. Because of fitting difficulties and the problem of determining the form in which factor prices enter the cost function, I initially chose the latter course. Firms, arrayed in order of ascending output, were divided into 5 groups containing 29 observations each. A list of the firms used in the analysis appears in Appendix C. The results of fitting five separate regressions of the form indicated by Model A are given in lines IIIA through IIIE of Table 3 and the corresponding implications for the parameters in the production function in lines IIIA through IIIE of Table 4. Similar results for regressions of the form indicated by Model B are presented in lines VIa through VIe of Tables 5 and 6.

The results of these regressions with respect to returns to scale are appealing: Except for statistically insignificant reversal between groups C and D, returns to scale diminish steadily, falling from a high of better than 2.5 to a low of slightly less than 1, which indicates increasing returns at a diminishing rate for all except the largest firms in the sample. However, in the case of regressions III, the elasticity of output with respect to capital

⁹ Treatment of capital costs is the source of one of the most serious shortcomings of the present study, as indeed capital measurement is in most studies of production. Solow's recent contribution to the study of the aggregate production function [18] offers considerable promise of an appropriate measure of capital used in the production of electric power. I hope, in future work, to make use of a model of production that involves fixed coefficients *ex post* at the plant level, but that permits substitution of inputs and that changes over time *ex ante*.

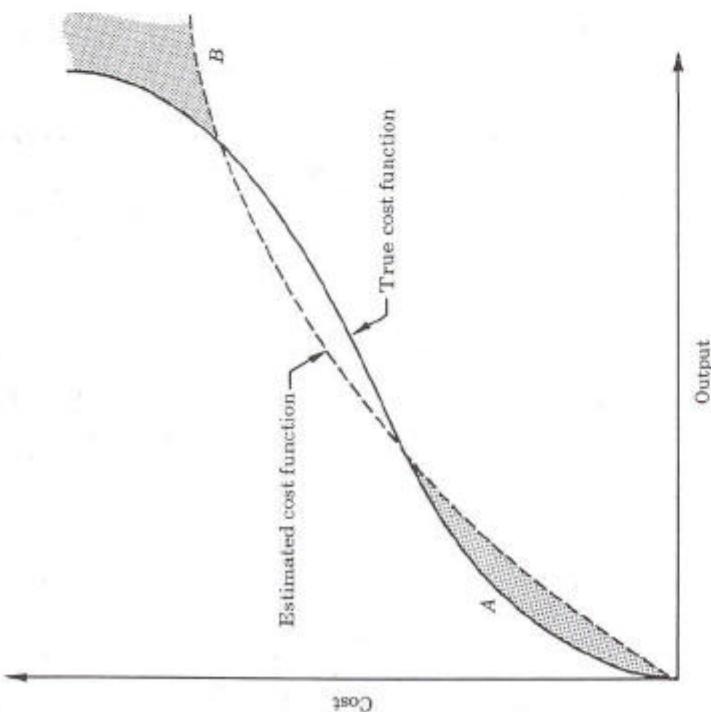


FIGURE 3

is the hypothesis of *neutral variations in returns to scale*. A general test of this hypothesis is equivalent to testing the hypothesis that the coefficients for the various prices in the individual group regressions are the same for all groups while allowing the constant terms and the coefficients of output to differ.⁸ The hypothesis of neutral variations in returns to scale is tested in this way only in the context of Model A. The regression results are presented in lines IVA through IVF of Table 3 and their implications for the production function in Table 4. An analysis of covariance comparing regressions III and IV gives an *F*-ratio of 1.576. With 133 and 225 degrees of freedom, a ratio this high is significant at better than the 99 per cent level; hence, we cannot confidently reject the hypothesis of non-neutral variations in returns to scale on statistical grounds alone with this test. Examining the results derived from regressions IV, however, we find that the degree of returns to scale steadily declines with output until, for the group consisting of firms with the largest outputs, we find some evidence of diminishing returns to scale.⁹ Furthermore, the elasticities of output with respect to the various input levels are all of the correct sign and of reasonable magnitude, although I still feel that the elasticity with respect to capital is implausibly low.¹⁰ Thus, on economic grounds, one might tentatively accept the hypothesis of neutral variations in returns to scale.

If one accepts the hypothesis of neutral variations in returns to scale, a somewhat more refined analysis is possible; since we may then treat the degree of returns to scale as a continuous function of output. That is, instead of grouping the firms as we did previously, we estimate a cost function of the form

$$(12) \quad C = K + \frac{1}{r(Y)} Y + \frac{a_1}{r} P_1 + \frac{a_2}{r} P_2 + \frac{a_3}{r} P_3,$$

where $r(Y)$, the degree of returns to scale, is a function of the output level. Since neutral variations in returns to scale are assumed, the coefficients of the prices are unaffected. A preliminary graphical analysis indicated that returns to scale as a continuous function of output might be approximated by a function of the form

$$(13) \quad r(y) = \frac{1}{\alpha + \beta \log y}.$$

⁸ For a generalized (which) includes the marginal rate of substitution between x_i and x_j is

$$\frac{\partial y / \partial x_i}{\partial y / \partial x_j} = \frac{a_i / a_j}{x_j / x_i},$$

Hence, if the ratio of a_i to returns to scale, r , is restricted to be the same for each output group, the marginal rates of substitution will be invariant with respect to output level at each given factor ratio.

⁹ Note, however, that the estimated value is insignificantly different from one, so that we cannot reject the hypothesis of constant returns to scale for this group of firms.

¹⁰ See p. 179.

price behaves very erratically from group to group and has the wrong sign in groups A and E; in regressions VI the elasticity of output behaves erratically, both with respect to labor and with respect to capital, having the wrong sign in groups B and C for the former and in group D for the latter. Analyses of covariance for regressions III and VI, compared with the over-all regressions I and V, respectively, gave *F*-ratios of 1.569 and 1.791 in that order. With 141 and 125 degrees of freedom, these ratios are significant at better than the 99 per cent level. Thus, breaking the sample into five groups significantly reduces the residual variance. However, because of the erratic behavior of the coefficients of independent variables other than output, it appears that we may have gone too far. Regressions III and VI are based on the assumption that all coefficients differ from group to group. Economically, this may be interpreted as the hypothesis of *non-neutral variations in returns to scale*; i.e., scale affects not only returns to scale but also marginal rates of substitution.

A halfway house between the hypothesis of no variation in returns to scale with output level and the hypothesis of non-neutral variations in scale

Thus, instead of regressions of the form suggested by (10) or (11), we fit

$$(14) \quad C - P_3 = K + \alpha Y + \beta Y^2 + \frac{a_1}{r} [P_1 - P_3] + \frac{a_2}{r} [P_2 - P_3] + V \quad (\text{Model C})$$

and

$$(15) \quad C = K' + \alpha Y + \beta Y^2 + \frac{a_1}{r} P_1 + \frac{a_3}{r} P_3 + V$$

(Model D).

The results obtained for regressions based on Model C and Model D are reported in Table 7 for regressions VII and VIII, respectively. The implications of these results for the production function are given in Table 8. Note that returns to scale and the other parameters have been computed at five output levels only, so that the results in Table 8 may be readily compared with those in Tables 4 and 6.

Perhaps the most striking result of the assumption of continuously and neutrally variable returns to scale of the form suggested in (13) is the substantial increase in our estimate of the degree of returns to scale for firms in the three largest size groups. Whereas before, we found nearly constant returns to scale, it now appears that they are increasing.¹¹ In addition, all

TABLE 7
Results from Regressions Based on Models C and D for 145 Firms in 1955;
Continuous Neutral Variations in Returns to Scale

Regression No.	Model C: Dependent Variable Was $C - P_3$				R^2	
	VII	Y	Y^2	$P_1 - P_3$	$P_1 - P_3$	
	0.151 (± 0.002)	0.117 (± 0.012)	0.498 (± 0.101)	0.062 (± 0.151)	0.958	

Model D: Dependent Variable Was C

Regression No.	Coefficient				R^2
	Y	Y^2	P_1	$P_1 - P_3$	
VIII	0.137 (± 0.064)	0.118 (± 0.013)	0.279 (± 0.224)	0.255 (± 0.054)	0.952

Figures in parentheses are the standard errors of the coefficients.

¹¹ Using the variance-covariance matrix for the coefficients in (14) or (15), one could easily compute, for a given y , a conditional standard error for $1/r$, which could then be

TABLE 8

Returns to Scale and Elasticities of Output with Respect to Various Inputs Derived from Results Presented in Table 7 for 145 Firms in 1955

Group	Regression VII (Model C)				R^2
	Return to Scale ^a	Labor	Capital	Fuel	
A	2.92	1.45	0.18	1.29	
B	2.24	1.12	0.14	0.98	
C	1.97	0.98	0.12	0.87	
D	1.84	0.92	0.11	0.81	
E	1.69	0.84	0.10	0.75	

Group	Regression VIII (Model D)				R^2
	Return to Scale ^a	Labor	Capital	Fuel	
A	3.03	0.85	1.41	0.77	
B	2.30	0.64	1.07	0.59	
C	2.01	0.66	0.94	0.61	
D	1.88	0.52	0.88	0.48	
E	1.72	0.48	0.80	0.44	

^a Evaluated at the medium output for each group.

the coefficients in both analyses are of the right sign, and the results based on Model D yield results of plausible magnitude for the elasticity of output with respect to capital as compared with the elasticities with respect to labor and fuel. Analyses of covariance, comparing regressions VII and I

used to test whether $1/r$ were significantly less than one (i.e., whether the finding of increasing returns was statistically significant). Unfortunately, the regression program used did not print out the inverse of the moment matrix, so this test could not be made. But there is little doubt, in view of the extremely small standard errors of the estimated α and β , that such a test would have shown the increasing returns found to be statistically significant.

with regressions VIII and V, yield F -ratios of 1.631 and 9.457, respectively; both are highly significant, with 141 and 140 degrees of freedom. A comparison of regression VII with regression III yields an F -ratio of 1.032, which, though not significant, does suggest that neutral variations in returns to scale of the form used are indistinguishable from non-neutral. Hence the hypothesis of neutral variations in returns to scale may be accepted both on economic grounds and on grounds of simplicity.

3. CONCLUSIONS AND PROSPECTS

The major substantive conclusions of this paper are that

1. There is evidence of a marked degree of increasing returns to scale at the firm level; but the degree of returns to scale varies inversely with output and is considerably less, especially for large firms, than that previously estimated for individual plants.
2. Variation in returns to scale may well be neutral in character; i.e., although the scale of operation affects the degree of returns to scale, it may not affect the marginal rates of substitution between different factors of production for given factor ratios.

These substantive conclusions derive from two conclusions of methodological interest:

1. The appropriate model at the firm level is a statistical cost function which includes factor prices and which is uniquely related to the underlying production function.
 2. At the firm level it is appropriate to assume a production function that allows substitution among factors of production. When a statistical cost function based on a generalized Cobb-Douglas production function is fitted to cross-section data on individual firms, there is evidence of such substitution possibilities.
- Inadequacies in the estimation of capital costs and prices and in the treatment of transmission suggest, however, that a less aggregative approach is called for. On a less aggregative level, it may be possible to produce more adequate measures of capital and to introduce transmission explicitly. A simple model of optimal behavior on the part of the firm may then allow us to combine this information in a way that will yield more meaningful results on returns to scale at the firm level.

Acknowledgments

APPENDIX A

A Relation Between Returns to Scale at the Plant Level and at the Firm Level from a Statistical Utility

Consider a firm that produces x_i units in each of n identical plants. If plants and demand are uniformly distributed, all plants will produce identical outputs, so that the total output produced will be nx , where x is the common value. Under these circumstances, a general formula that has been developed by electrical engineers to express transmission losses [8] reduces to

$$(A.1) \quad y = bn^2x^2,$$

where y is the aggregate loss of power. That is, with uniformly distributed demand and identical plants, transmission losses are proportional to the square of total output.

If z is delivered power, we have

$$(A.2) \quad z = nx - y = nx - bn^2x^2.$$

Let $c(x)$ be the cost of producing x units in one plant. Production costs of the nx units are thus $nc(x)$. And let $t = T(n, x)$ be the cost of maintaining a network with n plants, each of which produces x units. We may expect that t increases with x , $\partial T/\partial x > 0$, since larger outputs require more and heavier wires and more and larger transformers. However, t may or may not increase with n . It is likely to decrease with n if the expense of operating and maintaining long transmission lines is large relative to the cost of a number of short lines, and likely to increase if the converse is true.

The total cost of delivering an amount z of power $T(z)$ is the sum of production costs of a larger amount of power and transmission costs:

$$(A.3) \quad T(z) = nc(x) + T(n, x).$$

Suppose that the firm chooses the number and size of its plants in order to minimize $T(z)$ for any given z . The values of n and x that minimize $T(z)$ subject to (A.2) are given by solving

$$(A.4) \quad c(x) + \frac{\partial T}{\partial n} - x\dot{c}\mu = 0,$$

$$(A.5) \quad nc(x) + \frac{\partial T}{\partial x} - n\dot{x}\mu = 0,$$

$$(A.6) \quad z - (nx - bn^2x^2) = 0,$$

where

$$(A.7) \quad \begin{aligned} \mu &= 1 - 2bx \\ &\quad \frac{z-y}{nx}. \end{aligned}$$

The degree of returns to scale at the plant level, $p(x)$, may be defined as the reciprocal of the elasticity of production costs with respect to output:

$$(A.8) \quad p(x) = \frac{c(x)}{x c'(x)}.$$

It follows from (A.4), (A.5), and (A.8) that

$$(A.9) \quad p(x) = 1 + \frac{t}{(nx)c'(x)} (e_x - e_n),$$

where

$$e_x = \frac{x \partial T}{t \partial x}, \quad e_n = \frac{n \partial T}{t \partial n}.$$

Since nx , t and $c'(x)$ are positive, it follows that returns to scale are greater or less than one, according to whether the elasticity of transmission cost with respect to output exceeds or falls short of the elasticity with respect to number of plants. If transmission costs decrease with a larger number of plants, then under the particular assumptions made here, the firm will operate plants in the region of increasing returns to scale. It may nonetheless operate as a whole in the region of decreasing returns to scale.

Let $P(z)$ be the degree of returns to scale for the firm as a whole when it delivers a supply of z units to its customers:

$$(A.10) \quad P(z) = \frac{\Gamma(z)}{z \Gamma'(z)}.$$

It is well known that the Lagrangian multiplier λ is equal to marginal cost; hence, from (A.5),

$$(A.11) \quad \Gamma'(z) = \lambda = \frac{1}{\mu} \left[nc'(x) + \frac{\partial T}{\partial x} \right].$$

Substituting for $\Gamma'(z)$ from (A.11), μ from (A.7), and $\Gamma(z)$ from (A.3), we obtain the following expression for $P(z)$:

$$(A.12) \quad P(z) = \frac{\Gamma(z)}{z} \cdot \frac{n(z-y)}{nx[n c'(x) + \partial T / \partial x]} \\ = \left(1 - \frac{y}{z} \right) \frac{ne(x) + t}{n[x c'(x)] + x(\partial T / \partial x)}.$$

By definition,

$$p(x) = \frac{c(x)}{x c'(x)},$$

hence

$$(A.13) \quad P(z) = p(x) \left(1 - \frac{y}{z} \right) \frac{nc(x) + t}{n c(x) + [p(x)e_x]t}.$$

Neglecting the last term in the product on the right-hand side of (A.13) for the moment, we see that returns to scale at the firm level will typically be less than at the plant level, solely because of transmission losses; how much less depends on the ratio of losses to the quantity of power actually delivered. The final term in the product is a more complicated matter: If there are increasing returns to scale and if the costs of transmission increase rapidly with the average load (i.e., $e_x > 1$), then it is clear that the tendency toward diminishing returns at the level of the individual firm will be reinforced. It is perfectly possible under these circumstances that firms will operate individual plants in the range of increasing returns to scale and yet, considered as a unit, be well within the range of decreasing returns to scale.

Although this argument rests on a number of extreme simplifying assumptions, it nonetheless may provide an explanation for the divergent views and findings concerning the nature of returns to scale in electricity supply. Davidson [3] and Houthakker [9], for example, hold that there are diminishing returns to scale, while much of the empirical evidence and many other writers support the contrary view. The existing empirical evidence, however, refers to individual plants, not firms, and many writers in the public utility field may have plants rather than firms in mind.

APPENDIX B

THE DATA USED IN THE STATISTICAL ANALYSES

Estimation of Equation (7) from cross-section data on individual firms in the electric power industry requires that we obtain data on production costs, total physical output, and the prices paid for fuel, capital, and labor. Data on various categories of cost are relatively easy to come by, although there are difficulties in deriving an appropriate measure of capital costs. Price data are more difficult to come by, in general, and conceptual as well as practical difficulties are involved in formulating an appropriate measure of the "price" of capital. Such problems are, in fact, the *raisons d'être* for Model B, which permits us to ignore capital prices altogether.

A cross section of 145 firms in 44 states in the year 1955 was used in the analyses. The firms used in the analysis are listed in Appendix C. Selection of firms was made primarily on the basis of data availability. The various series used in the analyses were derived as follows.

B.1. PRODUCTION COSTS

Data on expenditures for labor and fuel used in steam plants for electric power generation are available by firm in [6], but the capital costs of produc-

tion had to be estimated. This was done by taking interest and depreciation charges on the firm's entire production plant and multiplying by the ratio of the value of steam plant to total plant as carried on the firm's books. Among the shortcomings of this approach, three are worthy of special note:

(a) For many well-known reasons, depreciation and interest charges do not reflect capital costs as defined in some economically meaningful way. Furthermore, depreciation practices vary from firm to firm (there are about four basic methods in use by electric utilities), and such variation introduces a non-comparability of unknown extent.

(b) The method of allocation used to derive our series assumes that steam and hydraulic plants depreciate at the same rate, which is clearly not the case.

(c) Because of their dependence on past prices of utility plant, the use of depreciation and interest charges raises serious questions about the relevant measure of the price of capital. The use of a current figure is clearly inappropriate, but unless we are prepared to introduce the same magnitude on both sides of the equation, it is difficult to see how else the problem can be handled.

B.2. OUTPUT

Total output produced by steam plant in kilowatt hours during the entire year 1955 may be obtained from [6]. This was the series used, despite the fact that the peak load aspect of output is thereby neglected. Since the distribution of output among residential, commercial, and industrial users varies from firm to firm, characteristics of the peak will also vary and this in turn will affect our estimate of returns to scale if correlated with the level of output.

B.3. WAGE RATES

At the time this study was undertaken, I was unaware of the existence of data on payroll and employment by plant contained in [5]; hence, inferior information was used to obtain this series. Average hourly earnings of utility workers (including gas and transportation) were available for 19 states from Bureau of Labor Statistics files. A mail survey was made of the State Unemployment Compensation Commissions in the remaining 20 states. All replied, but only ten were able to supply data. A regression of the average hourly earnings of utility workers on those for all manufacturing was used to estimate the former for states for which it was unavailable. The resulting state figures were then associated with utilities having the bulk of their operations in each state. In only one case, Northern States Power, were operations so evenly divided among several states that the procedure could

not be applied. In this case an average of the Minnesota and Wisconsin rates was employed.

B.4. PRICE OF CAPITAL.

As indicated, many practical and conceptual difficulties were associated with this series. Be that as it may, what was done was as follows: First, an estimate of the current long-term rate at which the firm could borrow was obtained by taking the current yield on the firm's most recently issued long-term bonds (obtained from Moody's Investment Manual). These were mainly 30-year obligations, and in all cases had 20 or more years to maturity. This rate was in turn multiplied by the Handy-Whitman Index of Electric Utility Construction Costs for the region in which the firm had the bulk of its operations ([4], p. 69). Two shortcomings worth special mention are:

- (a) The neglect of the possibility of equity financing by the method.
- (b) The fact that the Handy-Whitman Index includes the construction costs of hydroelectric installations.

B.5. PRICE OF FUEL

Since coal, oil, or natural gas may be burned to produce the steam required for steam electric generation, and since many plants are set up to use more than one type of fuel, prices were taken on a per-Btu basis. These were available by state from ([4], p. 49), and the state figures were assigned to individual utilities in the same manner as wage rates.

APPENDIX C

Names of Prices and Corresponding Costs, Output, Wage Rate, Fuel, Price, and Capital, Prior to 1955

Firms used in the analysis are listed here in order of ascending output (measured in billions of kilowatt-hours). They are divided into 5 groups containing 29 observations each. These appear on pp. 434-438 following.

(References appear on p. 439, following this Appendix.)

REFERENCES

- [1] Arrow, K. J., B. Minhas, H. Chenery, and R. Solow, Capital-Labor Substitution and Economic Efficiency, Memo. C-11 (Revised) Project for Quant. Res. in Econ. Dev., Stanford Univ. (August 1960).
- [2] Chenery, H. B., Overcapacity and the Acceleration Principle, *Econometrica*, 20 (January 1952), 1-28.
- [3] Davidson, R. K., *Price Discrimination in Selling Gas and Electricity*, Baltimore: Johns Hopkins Press, 1955.
- [4] Edison Electric Institute, *Statistical Bulletin*, 1957, New York, 1958.
- [5] Federal Power Commission, *Steam Electric Plants: Construction Cost and Annual Production Expenses*, 1956, Washington, D.C., 1957.
- [6] Federal Power Commission, *Statistics of Electric Utilities in the United States, 1955, Classes A and B Privately Owned Companies*, Washington, D.C., 1956.
- [7] Friedman, M., *A Theory of the Consumption Function* (Gen. Ser. 63). New York: National Bureau of Economic Research (distr. Princeton Univ. Press), 1957.
- [8] George, E. E., Intra-system Transmission Losses, *Electric Eng. (Trans. A.I.E.E.)*, 62 (March 1943), 153-58.
- [9] Houthakker, H. S., Electricity Tariffs in Theory and Practice, *Econ. J.*, 61 (1951), 1-25.
- [10] Johnston, J., *Statistical Cost Analysis*, New York: McGraw-Hill, 1960.
- [11] Korniyev, R., "Technological Progress and the Production Function in the United States Steam Power Industry" (September 1959), duplicated.
- [12] Lomax, K. S., Cost Curves for Electricity Generation, *Economica*, 19, new series (1952), 193-97.
- [13] Marschak, J., and W. H. Andrews, Random Simultaneous Equations and the Theory of Production, *Econometrica*, 12 (July-October 1944), 143-205.
- [14] *Business Concentration and Price Policy* (Span. Conf. Ser. n). New York: National Bureau of Economic Research (distr. Princeton Univ. Press), 1955.
- [15] Nerkow, M., "Notes on the Identification and Estimation of Cobb-Douglas Production Functions" (1958-59), duplicated.
- [16] Nordin, J. A., Note on a Light Plant's Cost Curves, *Econometrica*, 15 (July 1947), 231-36.
- [17] Shephard, R. W., *Cost and Production Functions*, Princeton, N.J.: Princeton Univ. Press, 1953.
- [18] Solow, R. W., Investment and Technical Progress, in K. J. Arrow, S. Karlin, and P. Suppes (eds.), *Mathematical Methods in the Social Sciences*, 1959. Stanford: Stanford Univ. Press, 1960, pp. 93-104.
- [19] Westfield, F. M., "Static and Dynamic Optimization Problems in the Multi-Plant Firm with Particular Reference to the Electric Power Industry," unpublished Ph.D. dissertation, M.I.T., 1957.

