7

# Returns to Scale in Electricity Supply

MARC NERLOVE, *Stanford University*

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 Chenery [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 purpose of an analysis should affect its form. In particular, the reason for obtaining an estimate of returns to scale will affect the *level* of the analysis: industry, firm, or plant. For many questions of pricing policy, for example, the plant is the relevant entity. On the other hand, when questions of taxation are at issue, the industry may be the appropriate unit of analysis. But if we are concerned primarily with the general question of public regulation and with investment decisions and the like, it would seem that the economically relevant entity is the firm. Firms, not plants are regulated, and it is at the level of the firm that investment decisions are made.

The U.S. electric power industry is a regulated public utility. Privately

I am indebted for a great deal of helpful advice to I. Adelman, K. J. Arrow, A. R. Ferguson, W. R. Hughes, S. H. Nerlove, P. A. Samuelson, and H. Uzawa. Had I been able to take all the advice I received, perhaps I could lay a part of the blame for the deficiencies of this paper on these people. The situation, however, is otherwise.

Support, in part, for the research on which this paper is based has been received from the Department of Economics, University of Minnesota, and under a grant from the Rockefeller Foundation at Stanford University. Stenographic assistance was received from the Office of Naval Research under Contract Nonr-225(50).

I wish also to acknowledge the help of M.S. Arora, G. Fishman, J. Johnston, H. Kanemitsu, and N. K. Rao, who performed the computations on which this paper is based.



Reprinted from MEASUREMENT IN ECONOMICS: STUDIES IN MATHEMATICAL ECONOMICS AND ECONOMETRICS IN MEMORY OF YEHUDA GRUNFELD published by Stanford University Press. 

owned firms, with which I am exclusively concerned in this study, account for nearly 80 per cent of all power produced. The technological and institutional characteristics of the electric power industry that are important for the model I shall develop are as follows:

1. Power cannot be economically stored in large quantities and, with few exceptions, must be supplied on demand.
2. Revenues from the sale of power by private companies depend primarily on rates set by utility commissions and other regulatory bodies.
3. Much of the fuel used in power production is purchased under long-term contracts at set prices. The level of prices is determined in competition with other uses.
4. The industry is heavily unionized, and wage rates are also set by contracts that extend over a period of time. Over long periods, wages appear to be determined competitively.
5. The capital market in which utilities seek funds for expansion is highly competitive and the rates at which individual utilities can borrow funds are little affected by individual actions over a wide range. Construction costs vary geographically and also appear to be unaffected by an individual utility's actions.

From these characteristics we may draw two conclusions, which lead to the model presented below. First, it is plausible to regard the output of a firm and the prices it pays for factors of production as exogenous, despite the fact that the industry does not operate in perfectly competitive markets. Second, the problem of the individual firm in the industry would appear to be that of minimizing the total costs of production of a given output, subject to the production function and the prices it must pay for factors of production. I shall adopt this last conclusion in what follows, although it is subject to some qualifications.

There are two basic objections to the cost-minimization hypothesis. First, rates in the industry are governed by a "cost plus" principle designed to secure investors "a fair return on fair value" (whatever that may mean). Although the application of this principle is a complicated matter in practice, it is clear that if a utility minimizes costs too much, i.e., decreases its costs to such an extent that, under the current rate structure, it obtains a substantial increment in earnings, the regulatory body may initiate an investigation and wipe out the increment through a decrease in rates. My impression, however, is that most utilities operate at a considerable distance from this "danger point."

A second objection to the cost-minimization hypothesis is that it is implicitly static; i.e., it does not reflect the fact that utilities are less concerned with cost minimization at a *point in time* than they are with minimization *over time*. In a dynamic formulation capital costs may be particularly affected. However, two contrary tendencies seem to exist: On the one hand, a steady rate of technological improvement has been experienced and may be expected to continue in this industry; thus, it is advantageous to postpone investment commitments. On the other hand, if there are increasing returns to scale, the steady growth in demand might be expected, *à la* Chenery [2], to lead to capital expenditures in excess of current needs. This tendency to over-capitalization may be aided and abetted by rate commissions, which are often inclined to support it after the fact through an increase in rates.

A related objection has been raised by William Hughes. He pointed out, in effect, that the existence of several power pools among companies treated separately in my analysis means that the outputs of such companies may not be truly exogenous as I have assumed.

Previous empirical investigations that have a bearing on returns to scale in electricity supply are those of Johnston [10, pp. 44-73], Komiya [11], Lomax [12], and Nordin [16]. All of these are concerned with returns to scale at the level of the plant, not the firm, and present evidence which suggests that there are increasing or constant returns to scale in the production of electricity. It is shown in Appendix A, however, that because of transmission losses and the expenses of maintaining and operating an extensive transmission network, a firm may operate a number of plants at outputs in the range of increasing returns to scale and yet be in the region of decreasing returns when considered as a unit. Although firms as a whole have been treated in this investigation, the problem of transmission and its effects on returns to scale has not been incorporated in the analysis, which relates only to the *production* of electricity. The results of this analysis are in agreement with those of previous investigators and suggest that the bulk of privately owned U.S. utilities operate in the region of increasing returns to scale, as is generally believed. Nevertheless, the results also suggest that the *extent* of returns to scale at the firm level is overestimated by analyses that deal with individual plants.

As indicated in Table 1, the production of electric power is carried out in three main ways:

1. By internal combustion engines. This method accounts for a negligible fraction of the power produced.
2. By hydroelectric installations. This method accounts for about one-third of all U.S. power production.
3. By steam-driven installations. This method accounts for the remaining two-thirds of U.S. power production.

Few firms rely solely on hydroelectric production because of the unreliability of supply. Furthermore, suitable sites for hydroelectric installations are rather limited and, except for those sites requiring an immense capital investment, almost fully exploited. Because of the great qualitative difference between steam and hydraulic production of electricity, this analysis is limited to steam generation. Since the variable costs of hydroelectric production are extremely low and it appears that firms fully exploit these possibilities, neglect of hydraulic generation should little affect the results on returns to scale.

The costs of steam-electric generation consist of (a) energy costs, and (b) capacity costs. The former consist mainly of the costs of fuel, of which coal is the principal one (see Table 2). Energy costs tend to vary with total output, and depend little on the distribution of demand through time. Capacity costs include interest, depreciation, maintenance, and most labor costs; these costs tend to vary, not with total output, but with the maximum anticipated demand for power (i.e., the peak load). Unfortunately, available data do not permit an adequate treatment of the peak-load dimension of output, hence it has been neglected in this study.

Even if the temporal distribution of demand does not differ systematically from one size firm to another, however, the results may be affected. A large firm with many plants and operating over a wide area has a greater

TABLE 1

Per Cent of Total Kilowatt-hours Produced by Type of Plant, 1930–1950, U.S.

| Year | Steam Generating Plants | Hydroelectric Installations | Internal Combustion Engines |
|---|---|---|---|
| 1930 | 65.1 | 34.2 | 0.7 |
| 1940 | 65.6 | 33.4 | 1.0 |
| 1950 | 69.8 | 29.1 | 1.1 |

TABLE 2

Per Cent of Total Steam-Electric Generation (KWH) by Type of Fuel, 1930–1950, U.S.

| Year | Coal | Oil | Gas |
|---|---|---|---|
| 1930 | 84.8 | 4.7 | 10.5 |
| 1940 | 81.9 | 6.6 | 11.5 |
| 1950 | 66.4 | 14.5 | 19.1 |

Source: R. E. Caywood, *Electric Utility Rate Economics.* New York: McGraw-Hill, 1956.

diversity of customers; hence, a large firm is more likely to have a peak load that is a small percentage of output than a small firm. It follows that capacity costs per unit of output tend to be less for larger firms. But this is a real economy of scale, and one reason for looking at firms rather than plants is precisely to take account of such phenomena. Of course, explicit introduction of peak-load characteristics would be better than the implicit account that is taken here.

## 1. The Model Used

As indicated, the characteristics of the electric power industry suggest that a plausible model of behavior is cost minimization, and that output and factor prices may be treated as exogenous. This suggests that traditional estimation of a production function from cross-section data on inputs and output is incorrect; fortunately, it also suggests a correct procedure. Let

$c$ = total production costs,

$y$ = output (measured in kwh),

$x_1$ = labor input,

$x_2$ = capital input,

$x_3$ = fuel input,

$p_1$ = wage rate,

$p_2$ = "price" of capital,

$p_3$ = price of fuel,

$u$ = a residual expressing neutral variations in efficiency among firms.

Suppose that firms have production functions of a generalized Cobb-Douglas type:

$$y = a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} u . \tag{1}$$

Minimization of costs,

$$c = p_1 x_1 + p_2 x_2 + p_3 x_3 , \tag{2}$$

implies the familiar marginal productivity conditions:

$$\frac{p_1 x_1}{a_1} = \frac{p_2 x_2}{a_2} = \frac{p_3 x_3}{a_3} . \tag{3}$$