Economies of Scale in U.S. Electric Power Generation

Laurits R. Christensen and William H. Greene

University of Wisconsin

We estimate economies of scale for U.S. firms producing electric power. Cross-section data for 1955 and 1970 are analyzed using the translog cost function. We find that in 1955 there were significant scale economies available to nearly all firms. By 1970, however, the bulk of U.S. electricity generation was by firms operating in the essentially flat area of the average cost curve. We conclude that a small number of extremely large firms are not required for efficient production and that policies designed to promote competition in electric power generation cannot be faulted in terms of sacrificing economies of scale.

Rates charged for electric power in the United States have recently risen at historically unprecedented rates. This has led to renewed scrutiny of proposals for reorganization of the U.S. electric power industry. One of the most interesting recent proposals has been made by Weiss (1975). He has suggested that firms in the industry be vertically disintegrated by divorcing the generation of electricity from its transmission and distribution. The motivation for this proposal is the belief that additional competition at the generation stage would result in downward pressure on rates. Competition, of course, requires that there be several potential suppliers for any particular load center. The crucial question pertaining to such a development is whether substantial scale economies would be sacrificed. If so, the resulting inefficiency might offset the potential benefits of increased competition.

The regulated entity in the U.S. electric power industry is the firm. Thus, assessment of the effects of reorganization requires information on

Associate professor of economics and graduate student, University of Wisconsin. We are grateful to Ernst Berndt, Dianne Cummings, Arthur Goldberger, Frank Gollop, and Leonard Weiss for helpful discussions and to the University of Wisconsin Graduate School for research support.

[*Journal of Political Economy*, 1976, vol. 84, no. 4, pt. 1]

economies of scale for firms—as opposed to plants or individual generating units. The purpose of this study is to provide such information. Our methodology is to model the structure of production in the industry using the neoclassical cost function approach. Pioneering work with this approach was done by Nerlove (1963) in his study of scale economies in electricity generation. Recent developments in duality theory and functional form specification permit us to specify a much more general model than those available to Nerlove. In particular, our model imposes neither homotheticity nor unitary elasticities of substitution.

In estimation, it is essential to distinguish economies of scale from decreases in cost resulting from technical change. The most convenient way to do so is to use cross-section data on firms which have had access to the same choices in plant design. It is also desirable, however, to gain some perspective on how scale economies have varied over time. To accomplish this, we have chosen to utilize two cross sections, the first for 1955 and the second for 1970. The 1955 sample facilitates comparison with the results of Nerlove, and the 1970 sample provides contemporary estimates.

Analysis of the two cross sections leads us to some important conclusions. First, we find that in 1955 there were significant scale economies available to nearly all firms. By 1970, however, a large share of total electric power was generated by firms which had exhausted scale economies. The explanation for this finding is that the firm size required to exhaust economies of scale increased by only 60 percent between 1955 and 1970, while actual output per firm approximately tripled.

Second, most firms which generated electric power experienced substantial declines in average costs of production between 1955 and 1970. A standard explanation for this phenomenon is made in terms of the exploitation of economies of scale, principally the utilization of very large generator units. Our findings do not support such an explanation. In fact, there was very little relationship between the degree of cost decline and the degree of expansion. Many firms which had very low growth rates achieved cost reductions as impressive as those of the fastest-growing firms.

Our primary finding, then, is that the U.S. electric power industry can be characterized by substantial scale economies at low levels of output. But the implied decreases in average cost diminish in importance for larger firms, resulting in an average cost curve which is very flat for a broad range of output. In 1955 most firms were producing on the low side of this range. By 1970, however, the bulk of U.S. electricity generation was by firms operating in the essentially flat area of the average cost curve. We conclude that a small number of extremely large firms are not required for efficient production and that policies to promote competition in electric power generation cannot be faulted in terms of sacrificing economies of scale.

I. The U.S. Electric Power Industry

The dominant mode of electricity generation in the United States is steam-driven turbines where the steam is produced by fossil-fuel-fired boilers. Water power has traditionally been an important mode as well, but a paucity of attractive dam sites and the tremendous growth of demand for electricity have combined to shrink its relative importance. Nuclear reactors may play a large future role, but their present contribution is still quite small. Internal-combustion engines are also employed but generally only in periods of peak demand. Table 1 indicates the mix of generator units in 1950 and 1970.

The technology and economics of the various forms of electricity generation are quite different. Given the dominant role of steam power, it is the logical starting point for any study which aspires to be relevant to policy decisions affecting the industry. Thus, we limit our attention to conventional steam-powered generators by analyzing only the costs of steam generation for each firm examined. Steam generation accounts for more than 90 percent of total generation for nearly all of the firms in the sample. We also limit our attention to investor-owned utilities with more than $1 million in annual revenues.[1] These firms accounted for approximately 77 percent of the total power produced in the United States in 1970. The rest was generated by several federal projects and several hundred small firms, cooperatives, municipal districts, and state systems. It would clearly be unwise to lump all of these diverse economic entities together in this study.

Most observers agree that there exist economies of scale in electricity generation.[2] There is disagreement, however, about the range over which economies persist. Some engineering estimates predict that economies of scale persist almost indefinitely. For example, Hulbert (1969) claims that there are economies to be achieved for systems up to at least 25,000 megawatts.[3] In sharp contrast are the results of Johnston (1960) and Nerlove (1963), who found that scale economies were exhausted at relatively modest firm size. Of course, many years have passed since their econometric estimates were made, so they are not necessarily inconsistent with contemporary engineering estimates.

In 1955 sales to ultimate consumers by privately owned electric power companies totaled 369 billion kilowatt-hours (kwh). By 1970 the total was 1,083 billion. Since the number of firms declined slightly, the output per firm increased approximately threefold. It is possible that changes in technology between 1955 and 1970 were great enough to have considerably expanded the firm size necessary to exploit scale economies. On the other

[1] Firms having revenues in excess of $2.5 and $1.0 million are designated by the Federal Power Commission as Class A and Class B, respectively.

[2] This paragraph follows closely the discussion of Weiss (1975, p. 148).

[3] This is almost twice the 1970 capacity of the nation's largest privately owned power system, the American Electric Power Company.

TABLE 1

PERCENTAGE OF TOTAL PRODUCTION BY TYPE OF PLANT*

Year	Steam	Hydroelectric	Internal Combustion	Nuclear
1950	69.8	29.1	1.1	0
1970	90.7	6.2	1.5	1.6

* The 1950 figures are from Caywood (1956) and refer to total U.S. production. The 1970 figures are from the Federal Power Commission (1971*a*). These figures are for investor-owned utilities only and neglect the large federal projects. Inclusion of these would increase the percentage of hydroelectric power somewhat. For example, TVA produces 6 percent of the nation's power, and 20 percent of its capacity is hydroelectric (*Forbes Magazine* 1975, p. 25).

hand, if the previous econometric estimates are valid and if technological change has not dramatically changed the shape of the firm cost curve, the rapid growth of firms may have resulted in the exhaustion of economies of scale.

In view of the disparate engineering and econometric estimates of the range of scale economies and the rapid growth of the industry since the econometric studies were performed, there is a clear need for further econometric analysis of the structure of production in the U.S. electric power industry.

II. Modeling the Structure of Production with a Neoclassical Cost Function

The recent application of duality theory to problems in economics has resulted in many useful results for the study of production and cost relationships.[4] A fundamental result is that, given certain regularity conditions, there exist cost and production functions which are dual to each other. The specification of a production function implies a particular cost function, and vice versa. Thus, the structure of production can be studied empirically using either a production function or a cost function; the choice should be made on statistical grounds. Direct estimation of the production function is attractive when the level of output is endogenous. Estimation of the cost function is more attractive, however, if the level of output is exogenous.

With minor exceptions, electric utilities are not allowed to choose the level of production to maximize profits. They are required to supply all the electric power which is demanded at regulated prices. Entrepreneurial decisions are made with respect to the determination of the optimal levels of inputs.[5] Thus, the specification that input levels are endogenous

[4] See Diewert (1974) for an excellent survey of the literature.

[5] Whether or not regulated utilities actually do try to minimize costs has been the subject of considerable debate. Proponents of the Averch and Johnson (1962) hypothesis suggest, instead, that a utility faced with rate-of-return regulation will tend to over-

variables and that the output level is an exogenous variable is plausible. Furthermore, since electric utilities compete with other industries for factors of production, the specification that input prices are exogenous variables is also reasonable. The cost function has as its arguments the level of output and the factor prices. Thus, estimation of the cost function is more attractive than direct estimation of the production function for studying the structure of production in the U.S. electric power industry. This was recognized by Nerlove (1963), and his study was one of the earliest applications of the principles of duality in modeling the structure of production.

Recent developments in duality theory have enhanced the appeal of the cost function approach. Every cost function implies a set of derived demand equations. Functional forms for cost functions have been developed which have two attractive features: they imply derived demand equations which are linear in the parameters, and at the same time they represent very general production structures, even though they cannot be derived from explicit production functions.

We have chosen the translog cost function (Christensen, Jorgenson, and Lau 1971, 1973) for this study. It places no a priori restrictions on substitution possibilities among the factors of production. Equally important, it allows scale economies to vary with the level of output. This feature is essential to enable the unit cost curve to attain the classical U shape. The translog form is also attractive for this application because it contains all the forms estimated by Nerlove (1963) as special cases. This allows us to perform statistical tests of the restrictions implied by Nerlove's models and to assess the impact that these restrictions have on estimates of scale economies.

III. The Translog Cost Function

The translog cost function can be written

$$\left.\begin{aligned}\ln C = \alpha_0 &+ \alpha_Y \ln Y + \tfrac{1}{2}\gamma_{YY}(\ln Y)^2 \\ &+ \sum_i \alpha_i \ln P_i + \frac{1}{2}\sum_i\sum_j \gamma_{ij}\ln P_i \ln P_j \\ &+ \sum_i \gamma_{Yi}\ln Y \ln P_i,\end{aligned}\right\}\ \text{Model A}$$

capitalize. Regarding electric utilities, some tentative support for the hypothesis is offered in the recent studies by Courville (1974), Spann (1974), and Petersen (1975). Breyer and MacAvoy (1974, p. 108) point out, however, that the activities of the electric utilities in the 1960s were actually contrary to predictions of the Averch-Johnson hypothesis. Moreover, rapid increases in capital costs and the existence of considerable regulatory lag suggest that there was scant use of unnecessary capital in 1970—the year of chief interest for this study.

where $\gamma_{ij} = \gamma_{ji}$, C is total cost, Y is output, and the P_i's are the prices of the factor inputs. In order to correspond to a well-behaved production function, a cost function must be homogeneous of degree one in prices; that is, for a fixed level of output, total cost must increase proportionally when all prices increase proportionally. This implies the following relationships among the parameters

$$\sum_i \alpha_i = 1, \tag{1}$$

$$\sum_i \gamma_{Yi} = 0, \tag{2}$$

$$\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = \sum_i \sum_j \gamma_{ij} = 0. \tag{3}$$

A convenient feature of the cost function approach is that the derived demand functions for the factors of production can be easily computed by partially differentiating the cost function with respect to the factor prices, that is,

$$\partial C/\partial P_i = X_i. \tag{4}$$

This result, known as Shephard's lemma (Shephard 1953), is conveniently expressed in logarithmic form for the translog cost function as

$$\partial \ln C/\partial \ln P_i = \frac{P_i X_i}{C} = S_i, \tag{5}$$

where S_i indicates the cost share of the ith-factor input. The translog cost function yields the cost share equations

$$S_i = \alpha_i + \gamma_{Yi} \ln Y + \sum_j \gamma_{ij} \ln P_j. \tag{6}$$

Uzawa (1962) has shown that Allen partial elasticities of substitution (Allen 1938) can be computed from the cost function by the formula

$$\sigma_{ij} = CC_{ij}/C_i C_j, \tag{7}$$

where subscripts on C indicate partial differentiation of the cost function with respect to factor prices. For the translog cost function we have

$$\begin{aligned} \sigma_{ij} &= (\gamma_{ij} + S_i S_j)/S_i S_j, \\ \sigma_{ii} &= [\gamma_{ii} + S_i(S_i - 1)]/S_i^2, \qquad i \neq j. \end{aligned} \tag{8}$$

The own-price elasticity of demand for the ith factor of production is

$$\eta_i = \sigma_{ii} S_i. \tag{9}$$

Economies of scale are usually defined in terms of the relative increase in output resulting from a proportional increase in all inputs. Hanoch

(1975) has pointed out that it is more appropriate to represent scale economies by the relationship between total cost and output along the expansion path—where input prices are constant and costs are minimized at every level of output. A natural way to express the extent of scale economies is as the proportional increase in cost resulting from a small proportional increase in the level of output, or the elasticity of total cost with respect to output. We will define scale economies (SCE) as unity minus this elasticity:

$$\mathrm{SCE} = 1 - \partial \ln C/\partial \ln Y. \tag{10}$$

This results in positive numbers for positive scale economies and negative numbers for scale diseconomies. Furthermore, SCE has a natural interpretation in percentage terms.

The translog cost function does not constrain the structure of production to be homothetic, nor does it impose restrictions on the elasticities of substitution. However, these restrictions can be tested statistically. If any of the restrictions are valid, it is preferable to adopt the simplified model. If not, it is of interest to investigate the impact of their imposition on the shape of the estimated cost curves.

A cost function corresponds to a homothetic production structure if and only if the cost function can be written as a separable function in output and factor prices.[6] A homothetic production structure is further restricted to be homogeneous if and only if the elasticity of cost with respect to output is constant. For the translog cost function, the homotheticity and homogeneity restrictions are, respectively,

$$\gamma_{Yi} = 0 \tag{11}$$

and

$$\gamma_{Yi} = 0, \qquad \gamma_{YY} = 0. \tag{12}$$

The elasticities of substitution can all be restricted to unity by eliminating the second-order terms in the prices from the translog cost function. Thus, the unitary elasticity restrictions are

$$\gamma_{ij} = 0. \tag{13}$$

We consider five models in addition to the translog Model A. Model B imposes homotheticity. Model C imposes homogeneity. Models D, E, and F correspond to Models A, B, and C, respectively, with unitary elasticities of substitution imposed in each case. The formulas for SCE for Models A–F are presented in table 2.

[6] See Diewert (1974) for formal statements and derivations of the restrictions for homotheticity and homogeneity.

TABLE 2

Scale Economies for Models A–F

SCE (A) $= 1 - (\alpha_Y + \gamma_{YY} \ln Y + \sum_i \gamma_{Yi} \ln P_i)$
SCE (B) $= 1 - (\alpha_Y + \gamma_{YY} \ln Y)$
SCE (C) $= 1 - \alpha_Y$
SCE (D) $= 1 - (\alpha_Y + \gamma_{YY} \ln Y + \sum_i \gamma_{Yi} \ln P_i)$
SCE (E) $= 1 - (\alpha_Y + \gamma_{YY} \ln Y)$
SCE (F) $= 1 - \alpha_Y$

IV. Estimation Procedures

It is feasible to estimate the parameters of each cost function using ordinary least squares. This technique, used by Nerlove (1963), is certainly attractive from the point of view of simplicity. However, it neglects the additional information contained in the cost share equations (6), which are also easily estimable. Furthermore, even for a modest number of factors of production, the translog cost function has a large number of regressors which, aside from the terms involving output, do not vary greatly across firms. Hence, multicollinearity may be a problem, resulting in imprecise parameter estimates.

An alternative estimation procedure, followed by Berndt and Wood (1975), is to estimate the cost shares (6) as a multivariate regression system. This procedure was satisfactory for their purposes, since they treated constant returns to scale as a maintained hypothesis ($\gamma_{Yi} = 0$, $\gamma_{YY} = 0$, $\alpha_Y = 1$). Thus, their cost share equations included all the parameters of the cost function except for the neutral shift parameter α_0, and no economic information was lost by not including the cost function in the estimation procedure. This approach is not satisfactory, however, when constant returns are not imposed, since the crucial parameters in the SCE formulas, α_Y and γ_{YY}, appear only in the cost function.

We conclude that the optimal procedure is to jointly estimate the cost function and the cost share equations as a multivariate regression system. Including the cost share equations in the estimation procedure has the effect of adding many additional degrees of freedom without adding any unrestricted regression coefficients. This will result in more efficient parameter estimates than would be obtained by applying ordinary least squares to the cost function alone.

We specify additive disturbances for each of the share equations and for the cost function. Since the cost share equations are derived by differentiation, they do not contain the disturbance term from the cost function. We assume that the disturbances have a joint normal distribution. Following Zellner (1962), we allow nonzero correlations for a particular firm but impose zero correlations across firms. However, his proposed estimation procedure is not operational for our model. The estimated disturbance covariance matrix required to implement Zellner's procedure is singular because the disturbances on the share equations must sum to zero for each firm. The Zellner procedure can be made operational by deleting one of the share equations from the system. However, the estimates so obtained will not be invariant to which equation is deleted.

Barten (1969) has shown that maximum-likelihood estimates of a system of share equations with one equation deleted are invariant to which equation is dropped. A straightforward extension of his result covers our multivariate system, which includes the cost function with the share equations. Kmenta and Gilbert (1968) and Dhrymes (1970) have shown that iteration of the Zellner estimation procedure until convergence results in maximum-likelihood estimates. Iterating the Zellner procedure is a computationally efficient method for obtaining maximum-likelihood estimates, and is the procedure which we employ here.

Since we obtain maximum-likelihood estimates, we are able to test hypotheses such as homotheticity by using the likelihood ratio test. Denoting the determinants of the unrestricted and restricted estimates of the disturbance covariance matrix by $|\hat{\Omega}_U|$ and $|\hat{\Omega}_R|$, respectively, we can write the likelihood ratio

$$\lambda = (|\hat{\Omega}_R|/|\hat{\Omega}_U|)^{-T/2}, \tag{14}$$

where T is the number of firms. We test hypotheses using the fact that $-2 \ln \lambda$ is distributed asymptotically as chi-squared with degrees of freedom equal to the number of independent restrictions being imposed.

V. Data

Since fuel accounts for a very large share (approximately 65 percent) of the cost of electricity generated by steam power, it is essential to include fuel in the model as a factor of production. Intermediate inputs other than fuel are relatively unimportant in the generation stage of electricity production. Thus, it is realistic to model the structure of production with capital, labor, and fuel as inputs. Hence, we require prices and cost shares for these three inputs in addition to the quantity and cost of total output.

To provide a link with Nerlove's (1963) results, we initially estimated our models using the same data as Nerlove—a 1955 cross section of Class A and B firms. Nerlove did not include the cost share equations in his estimation procedure and therefore did not construct the cost share data. We reconstructed the cost components using Nerlove's primary source (Federal Power Commission 1956).

Weiss (1975, p. 148) has pointed out that Nerlove may have considerably underestimated the extent of scale economies by treating members of holding companies as if they were individual firms. Since the holding company, rather than its subsidiaries, is the proper unit of observation, we revised Nerlove's data by combining these firms where appropriate. This reduces the number of observations in the sample from 145 to 124.

With two exceptions, we used Nerlove's procedures to construct a 1970 cross section of 114 firms and holding companies. First, for the price of fuel Nerlove used the statewide average from the state in which the firm was principally located. The Federal Power Commission (1971*b*) provides a plant-by-plant summary of fuel purchases by type of fuel (coal, oil, and gas), by BTU equivalents, and by prices paid per million BTU for each type of fuel. By simply converting all fuel purchases at all sites to BTU equivalents, we were able to average these prices, using as weights the contribution of each fuel purchase to total BTU consumption. This procedure yields a firm-specific price for fuel. Second, for the price of labor in 29 states Nerlove used statewide averages of wages for utility workers. In the 19 remaining states not even the statewide average wage for utility workers was available. For these states Nerlove estimated utility workers' wages from the regression relationship between utility workers' wages and all production workers' wages in the other 29 states. Unlike Nerlove's source (FPC 1956), FPC (1971*a*) provides enough information to construct a yearly labor cost for each firm. We computed the total labor cost for each firm as the sum of total salaries and wages paid and employee pensions and benefits. To arrive at a yearly price of labor, we divided this by the number of full-time employees plus one-half the number of part-time employees. While this is still not specific to steam generation, it represents a clear improvement over Nerlove's procedure.

VI. Empirical Results

a) *Cost Functions for the U.S. Electric Power Industry*

We present results for cost function Models A–F for three data sets. The data set used by Nerlove is denoted 1955I. Nerlove's data set with firms combined into holding companies is denoted 1955II. Our new data set is denoted 1970. Capital, labor, and fuel inputs are, respectively, denoted by K, L, and F subscripts.

All six models were estimated for all three data sets. The linear homogeneity in factor prices constraints (1, 2, 3) were imposed throughout. Additional regularity conditions which the cost functions must satisfy in order to correspond to well-behaved production structures are monotonicity and convexity in factor prices. Sufficient conditions for these to hold are positive fitted cost shares and negative definiteness of the bordered Hessian of the cost function. These conditions are met at every observation for all of the models which we estimated; hence, we may conclude that the estimated cost functions represent well-behaved production structures.

TABLE 3

Cost Function Parameter Estimates, 1955I Data
(t-Ratios in Parentheses)

Parameters	Model A	B	C	D	E	F
α_0	8.841 (37.21)	8.037 (31.77)	6.761 (31.70)	8.366 (47.29)	7.866 (47.90)	6.542 (62.13)
α_Y	0.252 (4.34)	0.335 (5.90)	0.810 (55.88)	0.223 (3.75)	0.300 (5.14)	0.797 (53.43)
α_K	0.076 (0.91)	0.122 (1.31)	0.122 (1.34)	0.449 (12.98)	0.426 (44.18)	0.426 (44.09)
α_L	0.365 (4.95)	0.117 (1.29)	0.120 (1.36)	0.210 (20.44)	0.106 (27.88)	0.106 (27.88)
α_F	0.559 (9.44)	0.761 (13.93)	0.758 (14.02)	0.341 (9.55)	0.468 (44.78)	0.468 (44.72)
γ_{YY}	0.079 (7.96)	0.079 (7.97)	...	0.083 (8.18)	0.083 (8.24)	...
γ_{YK}	−0.008 (−1.69)	...	...	−0.004 (−0.693)	...	...
γ_{YL}	−0.017 (−11.08)	...	...	−0.016 (−10.49)	...	...
γ_{YF}	0.025 (5.58)	...	...	0.020 (3.69)	...	...
γ_{KK}	0.175 (6.31)	0.155 (5.34)	0.154 (6.20)	...	...	...
γ_{LL}	0.040 (2.34)	0.033 (0.11)	0.004 (0.15)	...	...	...
γ_{FF}	0.167 (6.95)	0.154 (6.07)	0.154 (6.58)	...	...	...
γ_{KL}	−0.024 (−1.51)	−0.022 (−0.054)	−0.002 (−0.073)	...	...	...
γ_{KF}	−0.151 (−6.43)	−0.153 (−6.52)	−0.152 (−6.77)	...	...	...
γ_{LF}	−0.016 (−1.94)	−0.001 (−0.113)	−0.002 (−0.169)	...	...	...
Restrictions .	None	(2)	(3)	(3)	(5)	(6)
Log det $\hat{\Omega}$...	2.261	2.903	3.231	2.550	3.163	3.497

The complete set of parameter estimates for 1955I is presented in table 3. To conserve space, we present in table 4 only the Model A parameter estimates for the 1955II and 1970 data sets. The t-ratios for the nonhomotheticity parameters (γ_{Yi})and the substitution parameters (γ_{ij}) for Model A suggest that neither the homotheticity hypothesis nor the unitary elasticities of substitution hypothesis is consistent with any of the data sets. This suggestion is confirmed by the likelihood ratio statistics given in table 5 for all three data sets.[7] All of the restricted models are

[7] The estimates for Models E and F differ from those given by Nerlove (1963), since he used the OLS estimation procedure. We also computed the OLS estimates of the cost function parameters. They turn out to have generally larger standard errors than the efficient estimates. With these estimates we are not able to reject the homotheticity hypothesis or unitary elasticities of substitution hypothesis for any of the three data sets.

TABLE 4

Cost Function Parameter Estimates for Model A, 1955II and 1970 Data (*t*-Ratios in Parentheses)

Parameter	1955II	1970	Parameter	1955II	1970
α_0	8.412 (31.52)	7.14 (32.45)	γ_{YF}	0.024 (5.14)	0.021 (6.64)
α_Y	0.386 (6.22)	0.587 (20.87)	γ_{KK}	0.175 (5.51)	0.118 (6.17)
α_K	0.094 (0.94)	0.208 (2.95)	γ_{LL}	0.038 (2.03)	0.081 (5.00)
α_L	0.348 (4.21)	−0.151 (−1.85)	γ_{FF}	0.176 (6.83)	0.178 (10.79)
α_F	0.558 (8.57)	0.943 (14.64)	γ_{KL}	−0.018 (−1.01)	−0.011 (−0.749)
γ_{YY}	0.059 (5.76)	0.049 (12.94)	γ_{KF}	−0.156 (−6.05)	−0.107 (−7.48)
γ_{YK}	−0.008 (−1.79)	−0.003 (−1.23)	γ_{LF}	−0.020 (−2.08)	−0.070 (−6.30)
γ_{YL}	−0.016 (−10.10)	−0.018 (−8.25)			

TABLE 5

Test Statistics for Homotheticity, Homogeneity, and Unitary Elasticities of Substitution

	Hypothesis (Restriction on Model A)				
	Homotheticity	Homogeneity	Unitary Elasticities of Substitution	Homotheticity and Unitary Elasticities of Substitution	Homogeneity and Unitary Elasticities of Substitution (Cobb-Douglas)
Number of restrictions ..	2	3	3	5	6
Critical χ^2 (1%)	9.21	11.35	11.35	15.09	16.81
χ^2 1955I	93.09	140.51	41.91	130.79	179.22
χ^2 1955II	78.22	102.27	43.92	112.84	158.89
χ^2 1970	57.91	157.46	94.95	137.77	230.70

rejected at reasonable levels of significance in all three data sets. Thus, it is clear that a model which allows nonhomotheticity and nonunitary elasticities of substitution is required to adequately represent the structure of production for U.S. firms generating electric power.

It has often been suggested that substitution possibilities in electric power generation are scant. For example, Komiya (1962) argued in favor of modeling the structure of production with a fixed-coefficients technology, at least at the plant level. On the other hand, Nerlove (1963, pp. 173–74) argued that there may be significant substitution possibilities at the firm level. To provide some evidence on this question, we averaged

TABLE 6

ESTIMATED ELASTICITIES OF SUBSTITUTION AND OWN-PRICE ELASTICITIES

Elasticities of Substitution	Capital-Labor	Capital-Fuel	Labor-Fuel
1955I	.411	.223	.650
1955II	.547	.195	.574
1970	.639	.218	.165
Own-Price Elasticities	**Capital**	**Labor**	**Fuel**
1955I	−.159	−.499	−.170
1955II	−.163	−.520	−.146
1970	−.238	−.229	−.086

over firms the elasticities of substitution and own-price elasticities implicit in our cost function estimates. The results for all three data sets are presented in table 6.

b) *Economies of Scale*

An estimate of scale economies can be derived for each firm by evaluating the formulas in table 2 at the observed level of output and factor prices. Following Nerlove, we partitioned the sample into five groups of firms according to output; in table 7 we present the 1955I estimates of scale economies for the firm with the median output in each group. The t-ratios given with the estimates use firm-specific standard errors, conditional on observed output and prices. The estimates of scale economies for the homogeneous models, C and F, are constant for all levels of output. The other four models, which allow scale economies to vary with output, confirm our expectation that scale economies diminish as firm size increases. Although the imposition of unitary elasticities of substitution leads to a statistically significant loss of fit, the impact on estimated scale economies is not large. The estimates for Models D, E, and F are roughly the same as those for A, B, and C, respectively. However, the effect of the homotheticity constraint is much more pronounced. The estimates of scale economies for Models B and E, the homothetic models, are considerably less than those for Models A and D.

The homogeneous Models C and F give the erroneous impression that there are significant economies of scale for firms of any size. The nonhomogeneous models, however, indicate that scale economies are exhausted well within the sample output range. The median firm in group 5, which had output less than one-third that of the largest firm, shows no significant scale economies.

TABLE 7

ESTIMATED SCALE ECONOMIES UNDER VARIOUS SPECIFICATIONS OF TECHNOLOGY (t-RATIOS IN PARENTHESES)

	Size Group				
	1	2	3	4	5
Output (million kwh)	43	338	1,109	2,226	5,819
	1955I				
Model:					
Homogeneous:					
F	.203 (13.61)	.203 (13.61)	.203 (13.61)	.203 (13.61)	.203 (13.61)
C	.190 (13.11)	.190 (13.11)	.190 (13.11)	.190 (13.11)	.190 (13.11)
Homothetic:					
E	.388 (17.00)	.216 (16.90)	.117 (6.28)	.059 (2.43)	−.020 (−0.62)
B	.369 (16.62)	.208 (16.66)	.113 (6.20)	.059 (2.44)	−.017 (−0.53)
Nonhomothetic:					
D	.418 (18.00)	.258 (18.53)	.153 (7.94)	.096 (3.83)	.026 (0.77)
A	.408 (17.88)	.258 (18.44)	.157 (8.25)	.104 (4.16)	.040 (1.20)
	1955II				
Model A	.351 (13.66)	.243 (15.67)	.167 (8.68)	.27 (5.15)	.076 (2.26)

It is of interest to compare our estimates of scale economies with those obtained by Nerlove using ordinary least squares.[8] For Model F our estimate is .203, versus Nerlove's estimate of .281. Nerlove's estimates for Model E are inconsistent with his data; apparently he used the formula $(\alpha_Y + \frac{1}{2}\gamma_{YY} \log Y)^{-1}$ instead of $(\alpha_Y + \gamma_{YY} \log Y)^{-1}$ as implied by his model, to compute returns to scale. Table 8 presents our estimates along with Nerlove's corrected and erroneous estimate for Model E. The pattern we observe is that OLS leads to an increase in the estimates of returns to scale at low levels of output and a decrease at high levels. The same pattern emerges from the OLS estimates of Models A, B, and D.

A convenient way to summarize scale economies is to present the average cost curve facing a typical firm. The cost curve can be derived by evaluating the average cost function for a range of outputs while holding

[8] Nerlove computed returns to scale as the sum of the output elasticities. This is the inverse of one minus our estimate of scale economies. We used this transformation of Nerlove's estimates for comparability with our own. Also, Nerlove used common logarithms in his calculations while we used natural logarithms in ours.

TABLE 8

ESTIMATED SCALE ECONOMIES FOR MODEL E (1955I DATA)

	OUTPUT (MILLION kwh)				
	43	338	1,109	2,226	5,819
Christensen-Greene...	.388	.216	.117	.059	−.020
Nerlove (corrected)...	.465	.257	.136	.065	−.032
Nerlove (erroneous) ..	.658	.554	.492	.457	.408

the factor prices fixed at the sample means.[9] These curves are plotted for Models A–F in figure 1 (for 1955I). The size distribution of firms is presented beneath the cost curves. Consistent with the figures in table 7, Models C and F have continuously declining average costs, while the others all have the classical U shape. Dotted arrows indicate the minimum average cost corresponding to Models A, B, D, and E. The imposition of unitary elasticities of substitution has some effect on the shape of the curves, and imposition of homotheticity changes the shape of the cost curve substantially.

Figure 1 reveals that the cost curve for Model A is not well approximated by those from the restricted models. Scale economies persist at outputs considerably above those indicated by Models D, B, and E. Even so, a substantial proportion of electric power appears to have been generated by firms operating in the essentially flat area of the cost curve; and scale economies appear to be exhausted by firms well within the size range of the sample.

We found that Models B, C, D, E, and F can be rejected both on statistical grounds and because they do not provide good approximations to the unit cost curve from Model A. We henceforth limit our attention to Model A in studying the 1955II and 1970 data sets.

Weiss (1975) has suggested that Nerlove (1963) may have understated scale economies by treating members of holding companies as individual firms. In our 1955II data set, observations for firms were combined when they were members of holding companies. It turns out that the five firms in table 7 remained as observations, since they were not subsidiary firms in holding companies. It provides a useful comparison to compute the estimates of scale economies for these five firms. The estimates, presented for Model A in the last line of table 7, show that scale economies were less at lower levels of output—but persisted over a wider range of output. This finding is also apparent from the 1955II unit cost curve in figure 2. The difference between the 1955I and 1955II unit cost curves in figure 2 is

[9] The slope of the unit cost curve is sufficient to infer the presence of scale economies, since SCE = 1 − (marginal cost/average cost). Thus, declining unit costs indicate economies of scale and rising unit costs indicate diseconomies.

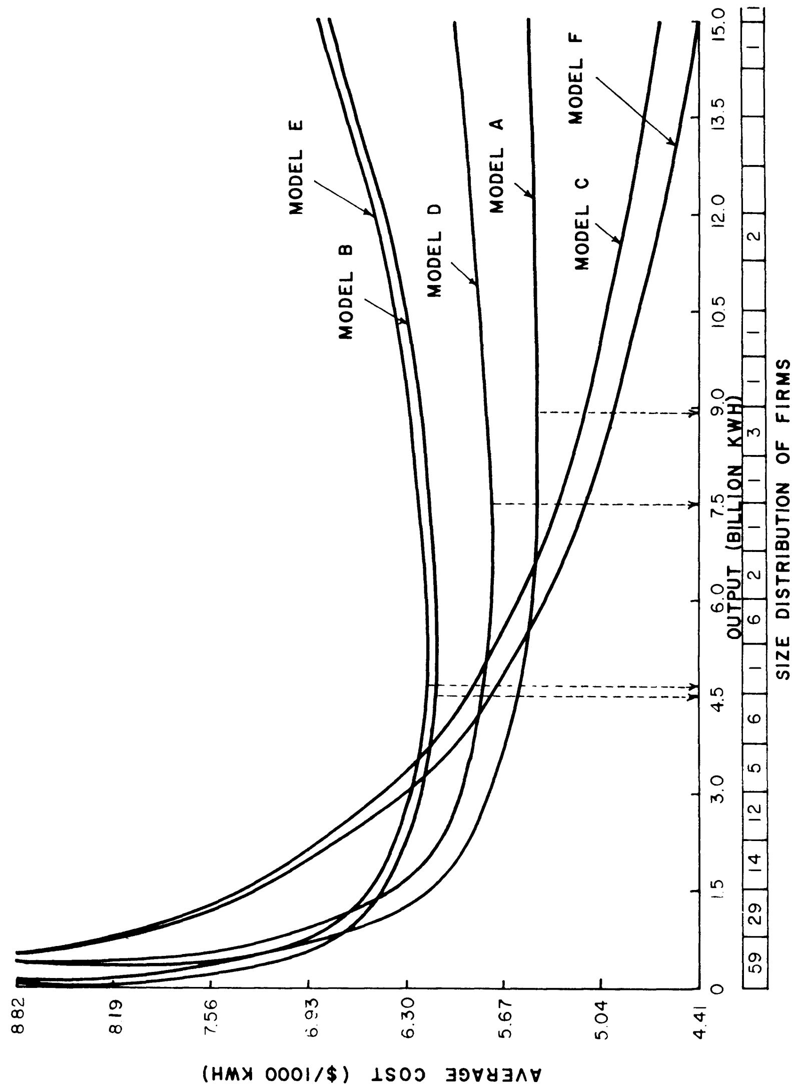
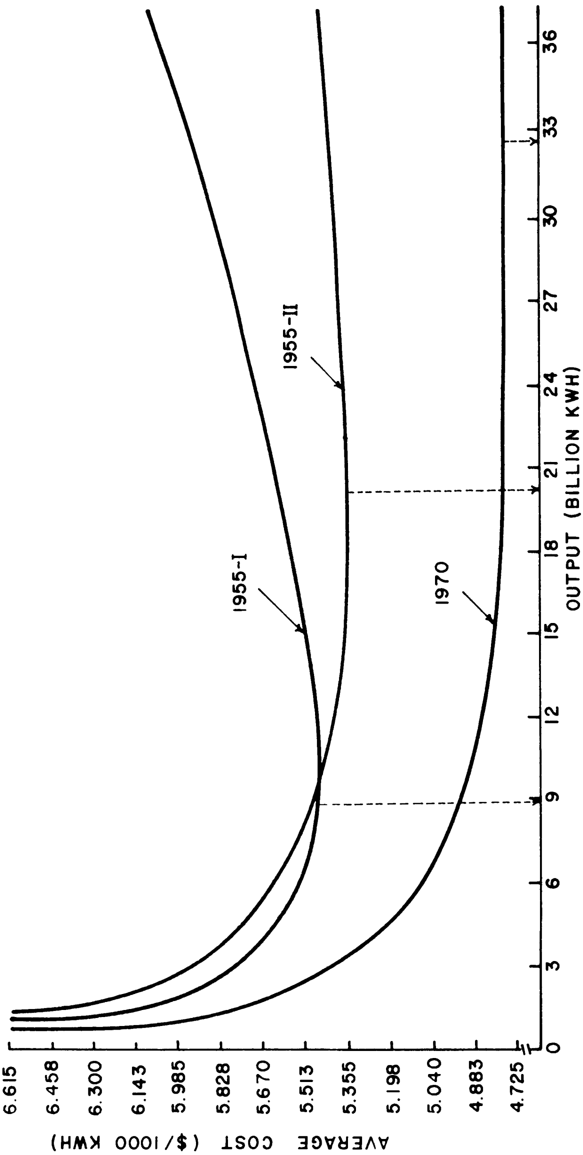


FIG. 1.—Average cost curves, 1955I data (dotted arrow indicates minimum average cost)

striking. The 1955II curve is relatively flat over a much broader range of output than the 1955I curve. In addition, the range of decreasing costs is more than double that of the 1955I curve. The size distribution of firms in figure 2 indicates that no firm had exhausted scale economies in 1955. Thus, Weiss's conjecture that the 1955I data set yields an understatement of scale economies is supported by our results.

Between 1955 and 1970, sales of electricity to ultimate consumers by privately owned producers grew from 369 to 1,083 billion kwh. Table 9



SIZE DISTRIBUTION OF FIRMS

1955-I	88	26	11	7	3	4	2	2		1		1													
1955-II	76	15	8	8	3	7	2	2		2		1													
1970	26	22	7	10	9	1	6	4	3	2	1	4	1	1	1	2	2	1	2	1	1	1	2	1	3

FIG. 2.—Average cost curves, Model A (dotted arrow indicates minimum average cost)

illustrates the pattern of growth for a sample of firms. Results from fitting Model A with our 1970 data set are indicated in table 9 and figure 2. The last two columns of table 9 show that firms typically had lower scale economies available in 1970 than in 1955. The reasons for this are clear from figure 2. The average firm tripled its output level between 1955 and 1970, but the shape of the cost curve changed very little. In particular,

TABLE 9

ESTIMATES OF SCALE ECONOMIES FOR SELECTED FIRMS (FIRMS ORDERED BY 1970 OUTPUT)

COMPANY	OUTPUT (MILLION kwh) 1955	OUTPUT (MILLION kwh) 1970	AVERAGE COST ($/1,000 kwh) 1955	AVERAGE COST ($/1,000 kwh) 1970	SCE 1955I	SCE 1955II	SCE 1970
Newport Electric	68	50	11.69	10.75	.379	.332	.324
Community Public Service	63	183	8.73	7.03	.413	.363	.247
United Gas Improvement	235	467	10.39	8.44	.442	.267	.216
St. Joseph Light & Power	253	938	8.03	5.45	.284	.263	.181
Iowa Southern Utilities	299	1,328	11.36	6.07	.265	.248	.160
Missouri Public Service	209	1,886	9.26	5.47	.299	.274	.143
Rochester Gas & Electric	1,156	2,020	7.66	8.89	.158	.167	.136
Iowa Electric Light & Power	1,166	2,445	7.37	5.37	.158	.167	.133
Central La. Gas & Electric	353	2,689	7.53	5.54	.253	.263	.127
Wisconsin Public Service	1,122	3,571	6.90	6.02	.162	.170	.103
Atlantic City Electric	1,291	4,187	7.96	7.00	.150	.161	.094
Central Illinois Public Service	2,304	5,316	5.48	4.43	.110	.133	.097
Kansas Gas & Electric	1,668	5,785	5.08	3.36	.136	.153	.094
Northern Indiana Public Service	1,137	6,837	7.01	4.96	.167	.174	.079
Indianapolis Power and Light	2,341	7,484	5.51	3.94	.112	.135	.080
Oklahoma Gas & Electric	2,353	10,149	4.94	3.01	.119	.142	.066
Niagara Mohawk Power	8,787	11,667	5.47	6.40	−.003	.047	.049
Potomac Electric Power	3,538	13,846	6.79	6.95	.070	.102	.037
Gulf States Utilities	2,507	17,875	3.85	3.27	.121	.146	.036
Virginia Electric Power	5,277	23,217	6.12	4.85	.037	.077	.015
Consolidated Edison	14,359	29,613	9.71	8.43	−.041	.019	−.003
Detroit Edison	11,796	30,958	6.19	6.05	−.026	.030	−.004
Duke Power	9,956	34,212	4.51	4.84	−.013	.038	−.012
Commonwealth Edison	19,170	46,871	7.10	5.43	−.035	.024	−.014
Southern	13,702	53,918	4.74	4.30	−.056	.008	−.028

the bottom of the cost curve shifted out by a relatively modest 60 percent (from 20 to 32 billion kwh). Thus, most firms were operating in a portion of the 1970 cost curve which was flatter than their point of operation on the 1955II cost curve. In effect, unexploited economies of scale were much smaller in 1970 than in 1955.

Up to this point, we have spoken rather loosely about "flat" or "relatively flat" portions of the average cost curves. We now give some precision to this notion by defining the "flat" portion of the curve as that part which exhibits no statistically significant economies or diseconomies of scale. We choose the 5 percent significance level and allocate $2\frac{1}{2}$ percent to each tail for SCE. Thus, any point on the average cost curve with a corresponding SCE which is less than 1.96 times its standard error is considered to be in the "flat" region of the curve. When a firm reaches a level of output in the "flat" region, we say that it has no statistically significant unexploited scale economies. The "flat" regions of the cost curves in figure 2 are indicated in table 10 along with the number of

TABLE 10

RANGES OF SIGNIFICANT SCALE ECONOMIES

	1955I	1955II	1970
Bounds of the region with no significant economies or diseconomies (billion kwh):			
Lower	4,400	9,700	19,800
Upper	34,000	57,000	67,100
Significant scale economies:			
Number of firms	119	118	97
Proportion of total output	.377	.741	.487
No significant economies or diseconomies:			
Number of firms	26	6	16
Proportion of total output	.623	.259	.446
Significant scale diseconomies:			
Number of firms	0	0	1
Proportion of total output	0	0	.067

firms and proportion of output produced in each of three regions: (1) statistically significant scale economies; (2) no significant scale economies or diseconomies; and (3) statistically significant scale diseconomies. The changes in industry production are dramatic. In 1955, 74.1 percent of steam-generated electric power was produced by firms with significant unexploited scale economies. By 1970, the figure had fallen to 48.7 percent. Not only was a large portion of all power produced in the flat region of the cost curve, but a nontrivial amount (6.7 percent) was produced by one firm (the American Electric Power Company) which showed significant diseconomies of scale.

VII. Concluding Remarks

In figure 3 we display the 1970 average cost curve over the full range of output observations. In addition, we include on the figure solid vertical arrows to indicate the bounds of the "flat" portion of the curve. This demonstrates the large range of output which can be interpreted as exhibiting constant returns to scale. Identification of this range provides an important tool for analyzing proposals to restructure the electric power industry. The aggregate cost of producing electric power is clearly related to the proportion produced by firms operating in the "flat" area of the cost curve.

Table 9 and figure 2 indicate that the cost of producing electric power fell absolutely during the 1955–70 period. This phenomenon is often attributed to the exploitation of scale economies or to scale-related technical change. Such explanations are contradicted by figure 2. To a

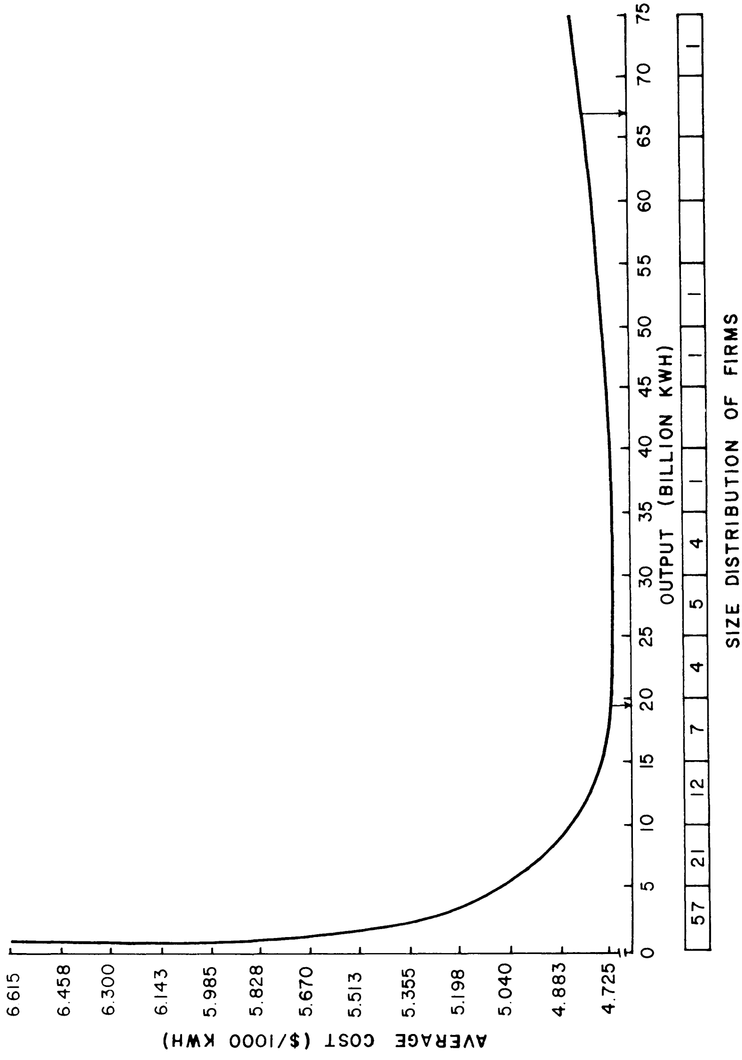


FIG. 3.—Average cost curves, Model A, 1970 (arrows indicate region of no significant economies or diseconomies of scale).

close approximation, the 1970 cost curve is simply a downward vertical displacement of the 1955II curve. This implies that technical change unrelated to increases in scale deserves the primary attribution for declines in the cost of production. This is corroborated by the fact that there is little correlation between the degree of cost reduction and the rate of growth of individual firms. It is true that many firms did exploit scale economies between 1955 and 1970. However, for the typical firm, movements along the cost curve resulted in cost declines which were small relative to the downward displacement of the entire curve.

Hughes (1971) and others have discussed the desirability of rationalizing the U.S. electric power industry—reducing the number of firms or planning units. Our cost functions yield estimates of the potential cost savings from rationalization. In 1970 if the combined output of all the firms in our sample had been produced at the minimum point on the average cost curve (0.473¢ per kilowatt hour), the total cost of production would have been $175.1 million less.[10] The estimated cost decline is 3.2 percent of observed costs. Of the estimated savings, $83.8 million (47.9 percent) is for labor ,$49.6 million (28.3 percent) for capital, and $41.7 million (23.8 percent) for fuel. Production of the observed output by the 114 firms in our sample would have required 33 firms of optimal size. The composition of the cost decline differs considerably from the average cost shares (14 percent for labor, 23 percent for capital, and 63 percent for fuel). The difference results from the nonhomotheticity of the production process.

References

Allen, R. G. D. *Mathematical Analysis for Economists*. London: Macmillan, 1938.

Averch, H., and Johnson, L. L. "Behavior of the Firm under Regulatory Constraint." *A.E.R.* 52 (December 1962): 1052–69.

Barten, A. P. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Econ. Rev.* 1 (Fall 1969): 7–73.

Berndt, E. R., and Wood, D. O. "Technology, Prices, and the Derived Demand for Energy." *Rev. Econ. and Statis.* 57, no. 3 (August 1975): 259–68.

Breyer, S. G., and MacAvoy, P. *Energy Regulation by the Federal Power Commission*. Washington: Brookings Inst., 1974.

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

Christensen, L. R.; Jorgenson, D. W.; and Lau, L. J. "Conjugate Duality and the Transcendental Logarithmic Production Function" (abstract). *Econometrica* 39 (July 1971): 255–56.

———. "Transcendental Logarithmic Production Frontiers." *Rev. Econ. and Statis.* 55 (February 1973): 28–45.

Courville, L. "Regulation and Efficiency in the Electric Utility Industry." *Bell J. Econ. and Management Sci.* 5 (Spring 1974): 53–74.

Dhrymes, P. J. "Equivalence of Iterative Aitken and Maximum Likelihood Estimates for a System of Regression Equations." Unpublished. Univ. Pennsylvania, 1970.

Diewert, W. E. "Applications of Duality Theory." In *Frontiers of Quantitative Economics*, edited by M. D. Intriligator and D. A. Kendrick. Amsterdam: North-Holland, 1974.

Federal Power Commission. *Statistics of Privately Owned Electric Utilities in the United States, Classes A and B, 1955 Supplement*. Washington: Government Printing Office, 1956.

———. *Statistics of Privately Owned Electric Utilities in the United States, Classes A and B, 1970 Supplement*. Washington: Government Printing Office, 1971. (*a*)

[10] This figure is arrived at by moving all production along the fitted average cost curve to the minimum point.

———. *Steam Electric Plant Construction Costs and Annual Production Expenses, 1970 Supplement*. Washington: Government Printing Office, 1971. (*b*)

Forbes Magazine. "Tennessee Valley Authority: The Yardstick with Less than 36 Inches" (April 1, 1975), pp. 24–28.

Hanoch, G. "The Elasticity of Scale and the Shape of Average Costs." *A.E.R.* 65, 3 (June 1975): 492–97.

Hughes, W. R. "Scale Frontiers in Electric Power." In *Technological Change in Regulated Industries*, edited by William M. Capron. Washington: Brookings Inst., 1971.

Hulbert, G. C. "Power Generation in the 1970's." In *1969 Future Power Forum: Perspective for the 70's*. New York: Westinghouse, 1969.

Johnston, J. J. *Statistical Cost Analysis*. New York: McGraw-Hill, 1960.

Kmenta, J., and Gilbert, R. F. "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions." *J. American Statis. Assoc.* 63 (December 1968): 1180–1200.

Komiya, R. "Technical Progress and the Production Function in the United States Steam Power Industry." *Rev. Econ. and Statis.* 44, no. 2 (May 1962): 156–67.

Nerlove, M. "Returns to Scale in Electricity Supply." In *Measurement in Economics—Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, edited by Carl F. Christ. Stanford, Calif.: Stanford Univ. Press, 1963.

Petersen, H. C. "An Empirical Test of Regulatory Effects." *Bell J. Econ.* 6, no. 1 (Spring 1975): 111–26.

Shephard, R. W. *Cost and Production Functions*. Princeton, N.J.: Princeton Univ. Press, 1953.

Spann, R. M. "Rate of Return Regulation and Efficiency in Production: An Empirical Test of the Averch-Johnson Thesis." *Bell J. Econ. and Management Sci.* 5, no. 1 (Spring 1974): 38–52.

Uzawa, H. "Production Functions with Constant Elasticities of Substitution." *Rev. Econ. and Statis.* 44, no. 4 (October 1962): 291–99.

Weiss, L. W. "Antitrust in the Electric Power Industry." In *Promoting Competition in Regulated Markets*, edited by Almarin Phillips. Washington: Brookings Inst., 1975.

Zellner, A. "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *J. American Statis. Assoc.* 57 (June 1962): 585–612.