Kaye D. Evans
Corroon and Black Corporation

John J. Siegfried and George H. Sweeney
Vanderbilt University

The Economic Cost of Suboptimal Manufacturing Capacity*

The economic cost of suboptimal capacity is estimated by fitting cost curves to late 1960s data describing the cost disadvantage of plants of less than efficient size and multiplying the cost disadvantage at each suboptimal scale by the actual quantity of output produced in plants of that size. Minimum efficient scale (MES) is defined in terms of both production cost and production plus transportation cost minimization. The total estimated economic cost of suboptimal capacity is 2%–4% of the value of shipments for the 15 sample manufacturing industries. Rudimentary empirical tests of alternative interpretations of the cost estimates (e.g., that they reflect technological improvements that have increased MES or that less than MES plants produce more highly valued differentiated products than do larger plants) do not undermine the results.

I. Introduction

If technological constraints generate scale economies, the cost of achieving a more competitive market structure may be the sacrifice of cost advantages arising from large scale production at a single site. This trade-off has been described and explored theoretically by Williamson (1968, 1977). Using a plausible hypothetical model, he demonstrated that the social cost of forgone scale economies could easily outweigh the social gains arising from a reduction in the deadweight loss generated by enhanced competition.

The seriousness of the trade-off between competition and cost efficiency has not been well documented.[1] Bain (1956) and Weiss (1976) attempted to estimate the fraction of output which is produced in plants of less than minimum efficient production scale (MES), that output level at which the average cost curve attains a level close to its minimum. Bain concluded that

* Leonard Weiss graciously provided worksheets from his earlier study of suboptimal capacity that were useful in the present analysis. Helpful comments were received from participants in the Applied Microeconomics Workshop at Vanderbilt University, especially Cliff Huang. Financial support was provided by the Vanderbilt University Research Council.

1. For a review of this literature see Siegfried and Wheeler (1981).

(*Journal of Business,* 1983, vol. 56, no. 1)

0021-9398/83/5601-0001$01.50

the majority of manufactured goods were produced in plants of "reasonably efficient" scale. He further claimed that there was a small "inefficient fringe" of firms that were responsible for 10%–30% of industry output; however, the presence of this fringe did not seem to be influenced by concentration levels or barriers to entry. Rather, its presence seemed to be related to the existence of single plant firms. Bain concluded that the concern over sacrificing economies of scale in order to promote competition was unfounded.

More recently, Weiss (1976) employed late 1960s estimates of minimum efficient scale to update and extend Bain's analysis. He used engineering interview estimates of the minimum efficient scale of plants in 35 industries compiled by Pratten (1971), Scherer et al. (1975), and himself (1976). Combining the MES estimates with census of manufactures employee-size-class information on value of shipments, Weiss concluded: "On the average, about half of total shipments in the industries covered are from suboptimal plants. The majority of plants in most industries are suboptimal in scale, and a very large percentage of output is from suboptimal plants in some unconcentrated industries" (1976, p. 141).

Bain's and Weiss's analyses help to assess the trade-off between scale economies and seller competition, but they are an inadequate guide for public policy. Even though almost half of U.S. output in the 35 industries surveyed by Weiss may be produced in suboptimal plants, it is conceivable that the cost elevation in these plants is relatively small. In order to evaluate the social cost caused by production in suboptimal plants, it is necessary to approximate the value of the additional resources that are devoted to production because of higher than necessary unit costs, which is the goal of this study.

Our estimates provide a basis for evaluating the potential social gains from a policy that discourages suboptimal plant construction. Any such benefits, however, should be weighed against the allocative efficiency losses generated by increased market concentration. Unfortunately, empirical studies of the welfare cost of monopoly (e.g., Harberger 1954; Schwartzman 1960; Kamerschen 1966; and Siegfried and Tiemann 1974) do not provide sufficient information to evaluate the trade-off because those studies all focus on the *remaining* welfare cost congruent with the existing level of monopoly. Thus they provide guidance regarding the social benefits that might be attained from greater encouragement of competition. But what must be traded off against suboptimal capacity cost is just the opposite—the additional welfare loss that would result from a diminution of competition in order to attain minimum efficient scale in all plants. Therefore, comparisons of the estimated social cost of suboptimal capacity with estimates of the current welfare loss due to monopoly do not directly address the traditional dilemma of antitrust.[2] But such comparisons may still be useful for focusing attention on the source of additional efficiency gains, even though the costs of achieving these benefits are not easily measured.

Furthermore, the logic underlying the traditional dilemma of antitrust policy is ambiguous. Although the promotion of more competition might prevent a sufficient agglomeration of sales to support certain firms' plants at the MES level, it also should reduce profit margins to competitive levels, which would require firms with less efficient plants either to expand to the efficient level of capacity or to leave the industry. Thus the promotion of seller competition could yield a reduction in the cost of suboptimal capacity as well as a reduction in the allocative loss caused by pricing above marginal cost.

II. Estimation Methodology

Overview of the Procedure

An exact measure of the cost of suboptimal capacity in an industry requires a comparison between the production costs actually incurred in all plants within the industry and the production costs that could have been achieved had all output been produced in plants of optimal (or larger) scale. Since neither of these cost measures are readily available, we estimate their values indirectly.

The estimates of optimal and actual production costs are derived from three sets of data that (roughly) provide the following information about each industry:

1. (*a*) the size (dollars of sales) of a plant at which all economies of scale are exhausted (MES); (*b*) the cost disadvantage suffered by a plant of size equal to ⅓ MES, measured as the percentage deviation between average cost of a plant of size equal to ⅓ MES and average cost of a plant of size equal to MES;
2. the total sales originating in plants of various size groups; and
3. the industry-wide profit per sales dollar.

Since only the cost disadvantage suffered by plants of size equal to ⅓ MES is known, it is necessary to estimate magnitudes of cost disadvantages suffered by plants of other sizes. We do this by assuming that the cost disadvantage suffered by a plant of a smaller size than MES can be represented by a specific continuous and monotonically declining function of plant scale (S). The parameters of this "cost elevation function" are chosen so that the estimated cost elevation of ⅓ MES equals the reported cost disadvantage of a plant of size equal to ⅓ MES, and

2. This criticism also applies to the use of estimates of the percentage of value of shipments produced in suboptimal size plants to draw inferences concerning the trade-off between allocative and cost efficiency such as those made by Bain (1956) and Weiss (1976).

the estimated cost elevation at MES equals zero. Since this procedure yields results that depend on the assumed nature of the cost elevation function, alternative functional forms are employed, each of which corresponds to a cost function commonly used in empirical work.

Under the assumption that the estimated cost elevation function ($CEP[S]$) accurately represents the percentage elevation of production cost above the production cost of an MES plant, $1 + CEP(S)$ represents an index of average production cost as a function of plant scale. This index is combined with shipments data to derive an estimate of the industry-wide percentage of cost elevation due to production in plants of suboptimal scale. The industry-wide profit data are combined with the cost index and shipments data in order to estimate the actual level of average production cost in plants of optimal scale. This number is then used to calculate the dollar estimate of excess cost due to suboptimal capacity.

Determination of Minimum Efficient Scale and Production in Suboptimal Plants

Estimates of minimum efficient scale as a percentage of 1967 value of shipments and the percentage cost penalty suffered by a plant of a size equal to ⅓ MES are reported for 12 industries by Scherer et al. (1975). These estimates were based on in-depth interviews of industry engineers and managers. Similar estimates of minimum efficient scale are provided for other industries by Pratten (1971) and by Weiss (1976). The Pratten and Weiss MES estimates are accompanied by estimates of the percentage cost penalty suffered by a plant of ½ MES (rather than ⅓ MES). Complete data requirements for estimating the cost of suboptimal capacity as outlined in the previous section were available for nine of Scherer's industries, three of Pratten's industries, and three of Weiss's industries. These industries' MES levels and cost penalties at ⅓ or ½ MES are reported in table 1.

An estimated tabulation of shipments by plant scale is derived from data in the 1967 census of manufactures (U.S. Department of Commerce 1971).[3] These data report the total value of shipments originating from plants in various size classes, and the number of plants observed within each size class. (The census measures "size" by number of employees.) We estimate the average production scale of plants within one census size class by the average value of shipments per plant originating from plants of that size class.[4] Following the procedure of

3. Although 1977 census of manufactures data are available, the latest estimates of minimum efficient scale are for the mid- to late 1960s (U.S. Department of Commerce 1967, 1971). In our judgment consistency of the various data sources takes precedence over timeliness.

4. A sufficient (but not necessary) condition that would insure that the average production scale is proportional to average shipments is that the value of shipments originat-

TABLE 1 **Economies of Scale and the Extent of Suboptimal Capacity in 15 Manufacturing Industries, 1967**

Industry	SIC Code	Minimum Efficient Scale as % of U.S. 1967 Output	Increase in Average Cost (%)		Value of 1967 Shipments ($ × 10^3)	Value of Shipments Produced in Plants Less than MES (%)	
			At ⅓ MES	At ½ MES		Weiss	Evans, Siegfried, Sweeney
Flour mills	2041	.7	...	3.0	2,454.6	50.6	57.7
Beer	2082	3.4	5.0	...	2,929.7	67.5	67.3
Soybean mills	2092	2.4	...	2.0	2,148.3	87.6	85.8
Cigarettes	2111	6.6	2.2	...	3,044.6	44.7	37.9
Tufted rugs	2272	.7	...	10.0	1,426.9	40.9	32.8
Synthetic fibers	2824	2.8	...	7.0	2,033.2	18.3	15.1
Detergents	2841	2.4	...	2.5	2,593.4	55.6	60.4
Paints and varnishes	2851	1.4	4.4	...	2,911.4	89.8	96.2
Petroleum refining	2911	1.9	4.8	...	20,293.9	82.3	79.5
Glass containers	3221	1.5	11.0	...	1,352.4	69.8	69.9
Cement	3241	17.0	26.0	...	1,246.5	92.2	92.3
Wide strip steel	3312	2.6	11.0	...	19,620.6	51.6	51.6
Iron foundries	3321	.3	...	10.0	2,657.8	61.0	46.2
Refrigerators and freezers	3632	14.1	6.5	...	1,791.9	32.8	66.8
Storage batteries	3691	1.9	4.6	...	577.5	59.7	62.2
Total or weighted average		...	...	...	67,062.7	64.0	63.6

SOURCES.—The figures for SIC codes 2082, 2111, 2851, 2911, 3221, 3241, 3312, 3632, 3691 are from Scherer et al. (1975); 2824, 2841, 3321 are from Pratten (1971); 2041, 2092, 2272 are from Weiss (1976). Value of 1967 shipments from U.S. Department of Commerce, Bureau of the Census (1971).

Weiss (1976), minimum efficient scale is estimated in terms of value of shipments by multiplying percentage of sales required to achieve MES by value of 1967 shipments.

An estimate of the value of shipments originating in suboptimal plants may be derived by summing the shipments originating in size classes smaller than the minimum efficient scale. This procedure encounters an ambiguity in counting shipments from plants in the size class which includes MES. Since one size class may encompass a substantial range of plant sizes, we divide sales originating from plants in the size class containing MES into: (1) sales originating in plants of smaller scale than MES and (2) sales originating in plants of larger (or equal) scale than MES. Our method of allocation is based upon employment data provided with the census of manufactures shipment data. This procedure is described in Appendix A.

Estimation of the Cost Elevation of a Suboptimal Plant

Let S represent the scale of a plant, measured in dollars of shipments per year, and let $C(S)$ represent the cost of producing S dollars of shipments in a plant of scale S. That is, $C(S)$ represents long-run total production cost as a function of shipments. Thus, long-run average production cost (i.e., cost per dollar of shipments) equals $C(S)/S$. It is assumed that average production cost decreases with increased scale for all S less than the minimum efficient scale and that average cost remains constant for all S greater than MES. Let S^* designate the minimum efficient scale, and let c^* designate the average production cost (per dollar of shipment) of a plant of minimum efficient scale.

Define the "production cost elevation" of a plant of suboptimal scale as the elevation of average cost suffered by that plant, expressed as a percentage of average cost of a plant of efficient scale. That is,

$$CEP(S) \equiv \frac{C(S)/S - C(S^*)/S^*}{C(S^*)/S^*}. \quad (1)$$

By the assumed nature of the cost function,

$$CEP(S) > 0 \qquad S < S^*$$

$$CEP(S) = 0 \qquad S \geqslant S^*.$$

The estimation procedure requires the assumption that, for values of S less than S^*, $CEP(S)$ can be approximated by a specific function

ing in each plant is proportional to the scale of the plant. This condition in turn would be satisfied under the following sufficient conditions: (i) all plants have equal capacity utilization; that is, the ratios of actual production to the optimal scale of a plant are equal for all plants; (ii) inventory fluctuations are uniform across all plants; that is, the ratios of the change in inventory to the level of production are equal for all plants; and (iii) output from each plant is sold at the same price.

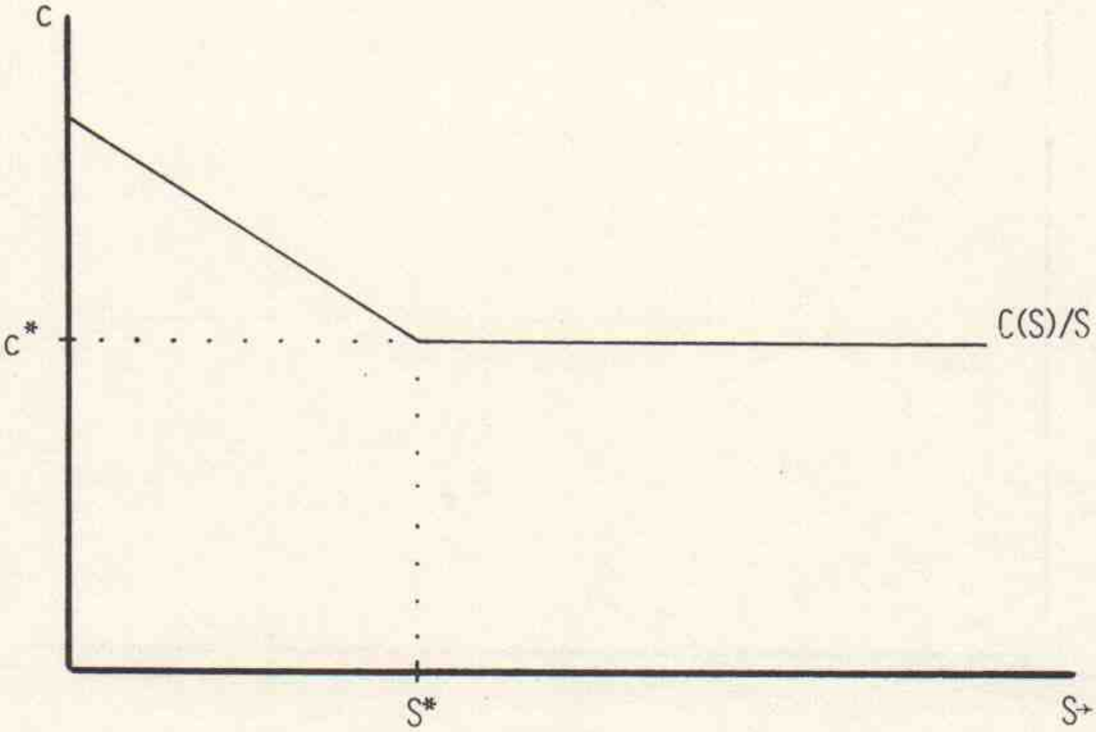

FIG. 1

$g(S)$. The form of the function $g(S)$ must be assumed, with two parameters to be determined by the data. These parameters satisfy the conditions that the estimated cost elevation function, when evaluated at $S = \frac{1}{3} S^*$, equals the reported value of *CEP* at ⅓ MES and that the estimated function equals zero at $S = S^*$.[5] That is, the parameters of the function $g(S)$ must be chosen so that:

$$\begin{aligned} g\left[\left(\frac{1}{3}\right)S^*\right] &= CEP_{\frac{1}{3}\mathrm{MES}}, \\ g(S^*) &= 0. \end{aligned} \tag{2}$$

The estimates of the cost of suboptimal scale were calculated with two alternative forms of the function $g(S)$. The first and simplest form assumes that long-run average cost declines linearly with increased scale, up to the minimum efficient scale. This form of average cost is illustrated in figure 1. In this case, g is a linear function:

$$g_1(S) = a_1 - b_1 S. \tag{3}$$

The second functional form assumes production to be characterized (over the relevant range) by returns to scale of constant order greater than one. The assumed form of the average cost function is illustrated in figure 2. The function, g, that generates this form is a hyperbola:

$$g_2(S) = a_2 S^{b_2 - 1} - 1 \qquad b_2 < 1. \tag{4}$$

5. For Pratten (1971) and Weiss (1976) industries the ⅓ MES is replaced by ½ MES.

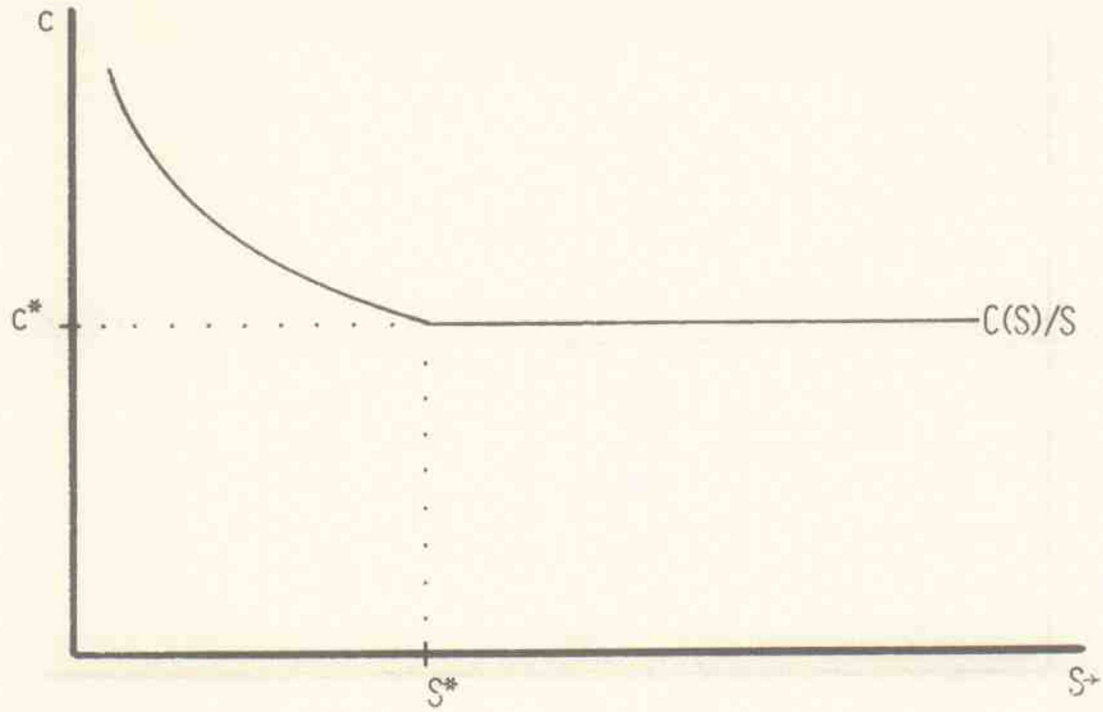

FIG. 2

The constants a_i and b_i are determined so that $g_i(S)$ satisfies both conditions in (2).[6]

Estimation of Total Industry Cost Elevation

Assuming the estimated function $g(S)$ accurately represents the true production cost elevation of a plant of scale S, average production cost in such a plant is given by equation (5):

$$\frac{C(S)}{S} = c^*[1 + g(S)]. \tag{5}$$

The industry-wide cost elevation due to suboptimal scale can be calculated from this equation. Assume the jth plant in an industry is of scale S_j.[7] Then, total production cost incurred in the jth plant

6. Estimates were also performed under the assumption that long-run total cost may be approximated by the sum of a fixed cost, which is independent of plant scale, and a variable cost, which is proportional to plant scale. That is: $g_3(S) = (a_3/S) + b_3$. This cost function generates an extremely high cost elevation for plants with relatively few employees, since the cost function rises asymptotically toward the vertical axis. Several industries have plants with only a few employees, and the extremely high cost elevation estimate therefore generates huge cost estimates. The very large cost elevation estimates at small plant sizes are likely to be misleading, since the estimated equation is based principally on two data points that describe the cost relationship between MES-sized plants and plants of ⅓ (or ½) MES. The very small plants are likely to employ substantially different technologies than even the ⅓ MES size plants and to produce products that are significantly differentiated from the larger plants' products (either geographically or characteristically). Consequently, these plants may be misclassified. Therefore, only estimates based on eq. (3) and (4) are reported.

7. A measure of the true scale of every plant within a size class is not available. Therefore (except in the case of the MES size class), it is assumed that the cost elevation

equals $c^*[1 + g(S_j)]S_j$. Total production cost incurred in the industry is the sum of production costs incurred in each plant in the industry, $\Sigma_j c^*[1 + g(S_j)]S_j$. But if all production had occurred in plants of size greater than or equal to MES, production costs would have equaled $\Sigma_j c^* S_j$.

Industry-wide percentage production cost elevation is defined as the difference between actual production cost and minimum potential production cost, expressed as a percentage of minimum potential cost:

$$CEP^{\text{ind}} = \frac{\Sigma_j c^*[1 + g(S_j)]S_j - \Sigma_j c^* S_j}{\Sigma_j c^* S_j}. \tag{6}$$

Equation (6) may be rewritten as:

$$CEP^{\text{ind}} = \frac{\Sigma_j [g(S_j)]S_j}{\Sigma_j S_j} = \Sigma_j [g(S_j)]\left[\frac{S_j}{\Sigma_k S_k}\right]. \tag{7}$$

Equation (7) indicates that the industry-wide percentage production cost elevation is simply a weighted average of individual plant cost elevations, with each plant's weight equaling its share of industry sales.[8]

Equation (7) expresses industry-wide excess cost as a percentage of minimum potential production cost. In order to convert this percentage cost estimate into an estimate of actual dollars of excess cost, one needs an estimate of the minimum potential average production cost (c^*), which can be derived from industry-wide profit margin data. The margin data are taken from *Statistics of Income: 1967* (U.S. Department of the Treasury 1971) for the IRS "minor industry" that most closely approximates the respective standard industrial classification (SIC) industry. Although the SIC four-digit industries and the IRS minor industries are considerably different, these margin estimates are probably the best available for the SIC industries. We cannot use margins calculated directly from the census of manufactures because it is not possible to exclude advertising and certain other fixed expenses from the margin estimate. Thus our margins would be grossly overstated and inappropriate for use in estimating the gross cost of suboptimal capacity.

The industry-wide profit per sales dollar equals the ratio of the sum of profits generated in all plants to the sum of sales originating in all plants:

$$\Pi = \frac{\Sigma_j (S_j - \text{Cost}_j)}{\Sigma_j S_j}. \tag{8}$$

of the average size plant within a size class provides a reasonable estimate of the average cost elevation of each plant within that size class.

8. When employing eq. (7) to estimate industry-wide cost elevation, shipments from plants in size class k are treated as originating in n_k plants each of scale equal to VS_k^{total}/n_k.

Assuming that: (i) all costs other than those required for production are insignificant, and (ii) $g(S)$ accurately represents the true production cost elevation function, equation (5) may be rewritten as:

$$\Pi = \frac{\Sigma_j S_j \{1 - c^*[1 + g(S_j)]\}}{\Sigma_j S_j}.$$

This equation may be solved for c^*:

$$c^* = \frac{\Sigma_j S_j (1 - \Pi)}{\Sigma_j S_j [1 + g(S_j)]}. \tag{9}$$

With this estimate of average cost required by a plant of efficient scale, the production cost which would have been required had all plants been efficient may be estimated as the product of total industry value of shipments and minimum average cost. Finally, total industry excess cost due to suboptimal capacity is estimated by the product of the industry-wide percentage of production cost elevation and minimum potential industry cost.

$$\text{Total Excess Production Cost}^{\text{ind}} = CEP^{\text{ind}}\, c^* \sum_j S_j. \tag{10}$$

The excess cost due to suboptimal production capacity is divided by total sales revenue in order to produce a cost estimate that is comparable to the estimates of the allocative efficiency welfare loss due to monopoly pricing.

The Problem of Transportation Costs

The previous analysis assumes that all costs other than production costs are trivial. Although this assumption may be valid for some industries, it is certainly questionable for others (e.g., cement). In industries in which a significant fraction of costs is due to activities other than production (such as research and development and advertising), the assumption results in an exaggerated estimate of average production cost (c^*) and therefore an exaggerated estimate of the dollar amount of excess cost due to suboptimal capacity. As long as these overlooked costs are independent of individual plant scale, however, the estimate of the excess cost expressed as a percentage of minimum potential cost remains unbiased.

If other costs are both large and dependent upon plant scale, a more serious problem arises. In such cases, the plant scale which minimizes average *total* costs may be different from the scale which minimizes average production cost. Therefore, our estimates of optimal scale and excess costs would be distorted.

A likely source of this type of error is transportation costs. Increased output from a plant must be shipped to more distant customers. Average transportation costs thus increase with plant scale. The plant scale that minimizes the sum of average production cost and average transportation cost may, therefore, be considerably smaller than the scale at which production economies are exhausted.[9]

In order to determine the magnitude of the corrections necessary to account for transportation costs, we employ rough estimates of transportation costs to recompute optimal plant scale and the costs of production in nonoptimal plants. Because of the crude nature of our estimates of transportation costs, the results of this exercise should be treated as suggestive rather than definitive numbers.

Specifically, it is assumed that increased output from a plant must be shipped to more distant customers. Transportation costs therefore, increase monotonically with plant scale. An estimate of transportation cost per mile shipped and an assumption about the geographic distribution of demand for each plant's product are employed to derive a relationship between the scale of a plant and average cost required to ship the output of that plant. This estimate is combined with the procedure of the previous section to estimate total (i.e., production plus transportation) cost as a function of plant scale. Details of the assumptions and calculations employed in this analysis are reported in Appendix B.

The assumptions that transportation costs increase monotonically with plant scale and that production costs are constant for outputs exceeding minimum efficient production scale together imply that total costs (both production and transportation) are minimized at a unique output level. That is, costs are assumed to exceed minimum potential costs at output levels both below and above minimum efficient overall scale, generating excess costs due to supraoptimal as well as suboptimal capacity. The main purpose of the transportation cost analysis is to determine what proportion of production cost disadvantages due to suboptimal production scale is offset by corresponding transportation cost savings. Therefore only the excess cost attributable to suboptimal capacity is reported in the empirical results.

III. Empirical Results

Production Costs Alone

Estimates of the percentage of shipments from suboptimal plants are reported in table 1 for the 15 sample industries, which account for 12.0% of the 1967 value of U.S. manufacturing shipments. Our method for converting employment-size-class boundaries to shipments-size-

9. If there exist fixed transportation costs (i.e., transportation costs that are independent of quantity or distance shipped), then average transportation costs may decrease with increased scale. In such cases, the plant scale that minimizes the sum of average production cost plus average transportation cost may be greater than MES.

class boundaries causes some of the differences between Weiss's estimate and our estimate of the percentage of total U.S. output produced in suboptimal plants.[10]

The weighted (by value of shipments) average percentage of suboptimal shipments was 64.0% by the Weiss method and 63.6% after our adjustments. On their surface, these estimates suggest the existence of serious social costs arising from too small plants. However, we now have an explicit (albeit rough) estimate of the importance of these costs.

The estimated aggregate excess production cost due to suboptimal capacity for the 15 industries is approximately 4.7% of minimum efficient scale costs using the linear cost function $[g_1(S)]$, and 6.7% using the nonlinear cost function $[g_2(S)]$. The estimated relative cost penalty resulting from suboptimal production varies considerably across industries. The estimates based on linear cost curves range from a low of .4% for cigarettes to a high of 23.6% MES costs for cement.

The importance of extending estimates of the percentage of shipments from suboptimal plants into an average cost elevation estimate is revealed by comparing the industry-by-industry estimates reported in table 2. Several of the industries with high percentages of output from suboptimal plants incur relatively low costs from their suboptimal production capacity. This occurs because either (1) the cost disadvantage of suboptimal production is not very large or (2) most of the output from suboptimal capacity is produced in plants only slightly smaller than MES. For example, although 37.9% of the value of shipments in the cigarette industry is produced in suboptimal plants, costs relative to value of shipments are only about 0.3% greater because of this suboptimal production. On the other hand, the cost disadvantage of suboptimal capacity relative to value of shipments for the synthetic fiber industry is 0.9%, and only 15.1% of these shipments come from suboptimal plants. The simple correlation coefficient between our estimate of the percentage of shipments produced in suboptimal plants and the relative excess cost based on the linear cost curve is .44. That this is not significantly greater than zero at the .05 level supports our contention that the simple calculation of the percentage of shipments from suboptimal plants can be a misleading indicator of potential sources of efficiency gains.[11]

10. A major discrepancy occurs only in SIC 3632, refrigerators and freezers. Weiss apparently used only refrigerators to compute the MES value of shipments level and applied this estimate to the total refrigerators and freezers industry in determining the volume of shipments from plants exceeding MES capacity. Our estimate is derived by applying Scherer's MES physical output estimate to the total refrigerators and freezers industry value of shipments to determine the value of shipments of an MES plant. Given the similar production technology of refrigerators and freezers, this seemed to us to be the more reasonable procedure.

11. The corresponding simple correlation coefficient between the percentage of ship-

TABLE 2 **The Cost of Suboptimal Capacity in 15 Manufacturing Industries, 1967**

Industry	SIC Code	Production Cost of Suboptimal Capacity as % of Value of Shipments		Production + Transportation Cost of Suboptimal Capacity as % of Value of Shipments		IRS Industry Code	Siegfried-Tiemann Welfare Loss as % of Value of Shipments
		Linear Cost	Nonlinear Cost	Linear Cost	Nonlinear Cost		
Flour mills	2041	1.55	2.12	1.28	1.84	2040	.004
Beer	2082	3.31	4.31	1.79	2.71	2082	.006
Soybean mills	2092	1.48	1.61	.86	1.00	2040	.004
Cigarettes	2111	.33	.26	.32	.23	2100	.162
Tufted rugs	2272	3.42	4.57	3.39	4.53	2270	.032
Synthetic fibers	2824	.90	1.19	.88	1.17	2812	.338
Detergents	2841	1.49	2.84	1.28	2.73	2841	.073
Paints and varnishes	2851	4.62	7.68	4.19	7.24	2850	.017
Petroleum refining	2911	3.42	4.40	.44	1.31	2911–2912	.072
Glass containers	3221	4.38	3.99	3.67	3.31	3210	.052
Cement	3241	18.67	19.79	8.70	9.40	3240	.012
Wide strip steel	3312	6.23	9.42	4.95	7.91	3310	.020
Iron foundries	3321	5.13	8.05	4.94	7.82	3310	.020
Refrigerators and freezers	3632	2.82	3.36	2.39	2.89	3630	.002
Storage batteries	3691	2.29	2.72	1.98	2.41	3691	.006
Total or weighted average		4.21	5.87	2.59	4.14		.058

Sources.—Last col., Siegfried and Tiemann (1974); all other data Evans, Siegfried, and Sweeney.

The estimated annual economic cost of suboptimal production capacity ranges from \$9.9 million in cigarettes to \$1,223.2 million in integrated wide-strip steel works and totals \$2.8 billion (using the linear cost curve). The total estimated economic cost (using the linear cost curve) is approximately 4.2% of the 1967 value of shipments of the 15 industries in the sample.

Production and Transportation Costs

The estimates of the cost of suboptimal capacity when both production and transportation costs are used to determine minimum efficient scale are also reported in table 2.

The importance of considering costs other than production costs is illustrated vividly by the cement industry. For cement, the cost of suboptimal capacity when transportation costs are included in the analysis reveals that over half of the estimated production cost disadvantage due to plants of less than MES is offset by estimated transportation cost savings, so that the estimated net cost of suboptimal capacity is around 8.7% of the value of shipments for the industry.

Overall, the estimates of the percentage increase in cost due to suboptimal capacity are reduced from 4.2% to 2.6% of the value of shipments for the linear cost curve and from 5.9% to 4.1% for the nonlinear cost curve when transportation costs are included in the analysis.[12] However, the estimates remain substantial, constituting around \$1.7 billion annually from the 15 industries alone (using the linear cost curve).

Costs of Suboptimal Capacity versus the Welfare Loss Due to Monopoly

The economic cost of suboptimal production cannot be considered as the cost of promoting competition, since (1) plant size is not necessarily correlated with firm size and market concentration because of the

ments produced in suboptimal plants and the excess cost based on the nonlinear cost curve is .43, also not significantly different from zero.

12. The cost of nonoptimal capacity (suboptimal plus supraoptimal) when transportation costs are included in the analysis is \$1,889 million using the linear cost curve and \$2,926 million using the nonlinear cost curve. The total estimated cost of supraoptimal capacity is only 8.6% of the suboptimal capacity cost if we use the linear cost curve and 5.5% in the case of the nonlinear cost curve. However, the importance of supraoptimal costs does vary by industry, ranging from 0.1% of suboptimal costs for paints and varnishes to 52.1% of suboptimal capacity for petroleum refining. For nine of the 15 sample industries the increase in costs attributable to supraoptimal capacity is 3.6%–13.6%. The general conclusions of the analysis are unaffected by the consideration of supraoptimal capacity. It is omitted in the text in order to facilitate direct comparisons with previous estimates of suboptimal capacity and to err in the direction of a conservative estimate, there being reasons to believe that the supraoptimal capacity cost estimates may be exaggerated, since such costs could be avoided by constructing multiple plants.

existence of multiplant firms, and (2) the promotion of additional seller competition would not generate the currently existing *level* of costs attributable to suboptimal production capacity, but presumably would generate a marginal increase in such costs. It is that increment in costs attributable to suboptimal capacity that must be compared to the benefits of any proposed improvement in seller competition. Our estimates do not provide direct guidance regarding the magnitude of such an increment. Nevertheless, a comparison of the estimated deadweight welfare loss due to monopoly with the economic cost of suboptimal capacity estimates is instructive.

Siegfried and Tiemann (1974) computed the deadweight welfare loss due to monopoly for the narrowest industry classifications and nearest time period to the present analysis. Their results are reported in table 2. They used Internal Revenue Service "minor industries" rather than SIC four-digit industries. The use of IRS data, which are classified on the basis of the primary product of each firm rather than on the basis of establishments, means that the industry definitions are not comparable. In order to assess industry comparability, we computed the ratio of the value of shipments for the 15 1963 IRS industries that best correspond to the 15 SIC industries to the 1963 value of shipments of the respective SIC industries. Only paints and varnishes and cement had substantially the same value of shipments in the IRS and SIC industry, and even for those industries we cannot determine if the composition of the totals is similar. The ratios ranged from .11 for iron foundries to 1.09 for cement. Nevertheless, there is undoubtedly more comparability between these sets of industries than between the 15 SIC industries and the total manufacturing sector. From these data it appears that there are substantially greater costs associated with production in suboptimal plants than are associated with the existing level of monopoly power in the U.S. economy. Indeed, the suboptimal capacity costs as a fraction of value of shipments are about two orders of magnitude greater than the estimated allocative efficiency losses due to monopoly, with the exception of cigarettes and synthetic fibers. The deadweight welfare loss relative to value of shipments is .06% for the 15 industries together while the lowest corresponding fraction for the 15 industry cost of suboptimal capacity is 2.6%.

IV. Conclusion

If the 15 industries analyzed in this study are representative of all manufacturing, the existence of suboptimal capacity is likely to represent a theoretical cost to the U.S. economy between \$14 billion and \$23 billion (1967) per year in the manufacturing sector alone (after netting out transportation cost savings). Although scale economies are not likely to be as important in agriculture, construction, services, trade, and finance as they are in manufacturing, there may be some additional costs of suboptimal capacity contributed by these other sectors of the economy. Particularly striking is the huge cost of suboptimal capacity vis-à-vis recent estimates of the welfare cost due to monopoly.

There are many assumptions and limitations that underlie this estimate of the cost of suboptimal capacity. For example, the industries examined were not chosen randomly; we have assumed that capacity utilization and X-inefficiency do not vary with plant size; and the functional form of the estimated cost curves is arbitrary. The plausibility of the assumptions varies. Nevertheless, the finding of a difference between the cost of suboptimal capacity and the deadweight welfare loss estimates attributed to monopoly that exceeds two orders of magnitude in most cases leaves substantial room for compromise on assumptions without tampering with the general impression that one obtains from the numbers.

A more important criticism of the estimated cost of suboptimal capacity concerns the interpretation of the findings. When we observe more than half of all output produced in suboptimal plants, we might want to ask why this occurs. One possibility is that transportation cost savings from smaller, dispersed plants partially compensate for the production cost inefficiencies of the smaller plants. Our results suggest the importance of product shipping costs, and our estimated costs of suboptimal capacity are adjusted for this consideration.

There are other rationales, however, that might equally plausibly explain the "cost" of suboptimal capacity in ways that are not undesirable. For example, if there has been technological progress that has altered the minimum efficient scale, many of the plants that are classified as below MES in 1967 may have been efficient when they were built, and vice-versa.[13] Thus the estimated cost of suboptimal capacity may not necessarily reflect poor management decisions. Indeed, we would want suboptimal plants to be replaced only if the total costs of a replacement plant of MES were less than the variable costs of continuing to operate the suboptimal plant. Consequently the continued operation of suboptimal plants may be efficient. In this case the cost of suboptimal plants is really the savings that society might expect to enjoy once long-run equilibrium is attained congruent with the technological change.

A rough clue to the importance of technological progress in increasing the MES of plants in the 15 sample industries is how firms are adjusting their plant sizes. The survivor technique indicates (in a world of perfect competition) the direction of changes in technological prog-

13. Technological progress can either increase or decrease the level of MES, although it probably increases it.

ress.[14] A comparison of the relative share of value of shipments produced by plants of less than (1967) MES in 1963 with the share in 1972 for each of the 15 industries revealed three cases where the share of suboptimal plants was declining over time and nine cases where it was increasing.[15] This suggests that some of the cost of suboptimal capacity might dissipate as plants are replaced using state of the art technology, but that this is not a serious flaw in our estimates.

Another interpretation of the results is that the suboptimal plants produce different (specialized) products than the larger plants classified in the same industry; in particular the smaller plants might produce higher quality products. Consequently, firms with high cost plants can remain viable, and indeed, might even be more "efficient" than firms with larger plants. Or, alternatively, the small plants might have some other (nonproduction) cost advantage that offsets higher production costs, such as lower wages due to monopsony power in rural locations, or cost advantages derived from a strategic location near raw material sources. If this were the reason for the higher production costs of smaller plants we should expect to observe margins for the smaller plants at least as great as margins for the larger plants.

A comparison of the price-cost margins of plants below MES with the price-cost margins of plants at or above MES in 1967 revealed only one industry where the smaller plants seemed to have a distinct competitive advantage and nine industries in which the smaller plants seemed to have a competitive disadvantage (lower margins).[16] Consequently this qualification to our findings does not appear to be of major significance.

Although there are reasonable explanations for some or much of the cost of suboptimal capacity, it is important to recognize the enormity of this cost and to insure that the other benefits that might be derived from a sacrifice of plant size (e.g., greater seller competition, increased flexibility of assets, diversified decision making, etc.) are worth this cost.

14. A comprehensive evaluation of the survivor technique is in Shepherd (1967).

15. Data are found in U.S. Department of Commerce, Bureau of the Census (1967, table 3, "Selected Statistics by Employment Size of Establishments: 1963"); and U.S. Department of Commerce, Bureau of the Census (1976, table 4, "General Statistics by Employment Size of Establishments: 1972"). The relative share of value of shipments produced in suboptimal plants declined from 1963 to 1972 in cigarettes, carpets, and iron foundries. It was relatively stable in synthetic fibers, paint, and strip steel. We used a 2-percentage-point change in the share of the less than MES plants as the criterion to identify a qualitative change.

16. A detailed evaluation of the advantages and disadvantages of census price-cost margin data is contained in Collins and Preston (1970). The data are from the U.S. Department of Commerce (1971, table 3, "Selected Statistics by Employment Size of Establishment: 1967"). The only industry in which suboptimal plants had price-cost margins at least 1 percentage point higher than MES plants was iron foundries. For five industries, there was no difference exceeding 1 percentage point.

Appendix A

Allocation of Sales within the Size Class Containing Minimum Efficient Scale

Utilization of employment data to determine firm scale requires the assumption of a functional relationship between employment and scale. It is assumed that, for plants within the size class containing MES, the scale of a plant (S) (measured in dollars of shipments) may be approximated by a linear function of the plant's average annual employment (e). That is, for all plants in size class containing MES,

$$S(e) \simeq a + be, \tag{A1}$$

where a and b are constants that must be determined. A consequence of assumption (A1) is that the scale of a plant employing the average number of employees (of the size class) equals the average scale of plants within that size class.

Assumption (A1) and its implication are illustrated in figure A1. Suppose that MES is contained in the ith size class. Let m_i and M_i represent the minimum and maximum number of employees per plant in size class i. Furthermore, let E_i^{tot} represent the total employment in all plants in size class i, and let VS_i^{tot} represent total value of shipments originating in all plants of size class i. Finally, let n_i equal the number of plants observed in size class i. Assumption (A1) implies that the function $S(e)$ can be represented by a straight line that passes through the point (E_i^{tot}/n_i, VS_i^{tot}/n_i). That is,

$$S\left(\frac{E_i^{tot}}{n_i}\right) = \frac{VS_i^{tot}}{n_i}. \tag{A2}$$

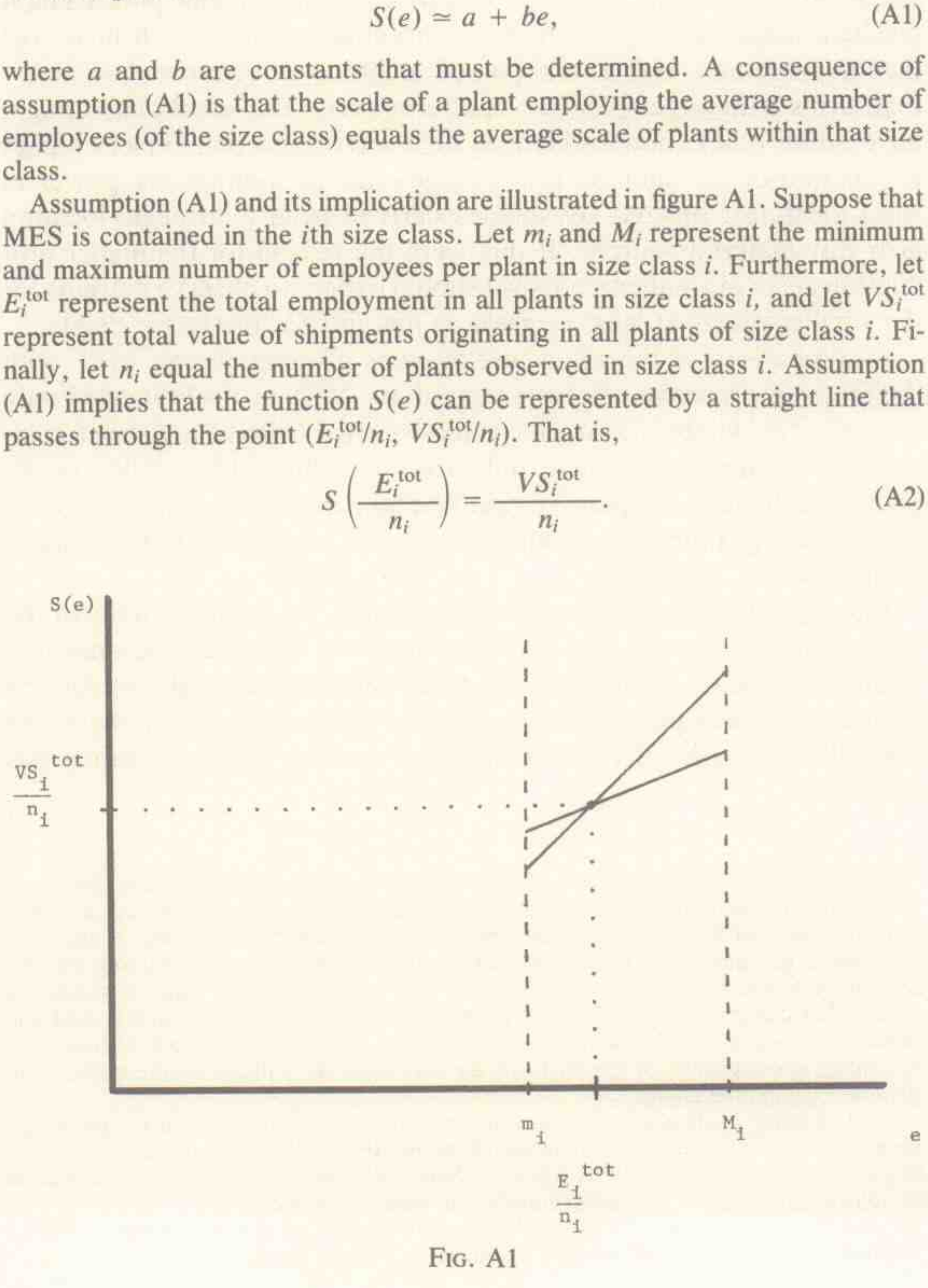

FIG. A1

Two of the infinite number of possible S functions are illustrated in figure A1.

Determination of the slope of the function $S(e)$ requires a second assumption, concerning the manner in which output per employee varies with plant size. It is sufficient to set the rate of change of output per employee as plant size (in this case measured by employment) varies between m_i and M_i employees. It is assumed that the average rate of change of output per employee throughout size class i equals a weighted average of the rate of change of output per employee between size class $i - 1$ and size class i, with the rate of change of output per employee between size class i and size class $i + 1$. That is;

$$\frac{[S(M_i)/M_i] - [S(m_i)/m_i]}{M_i - m_i} = w_1 \frac{v_i - v_{i-1}}{\epsilon_i - \epsilon_{i-1}} + w_2 \frac{v_{i+1} - v_i}{\epsilon_{i+1} - \epsilon_i}, \qquad (A3)$$

where:

$$v_j \equiv \frac{VS_j^{\text{tot}}}{E_j^{\text{tot}}},$$

$$\epsilon_j \equiv \frac{E_j^{\text{tot}}}{n_j},$$

$$w_1 \equiv \frac{1}{\epsilon_i - \epsilon_{i-1}} \left[\frac{1}{\epsilon_i - \epsilon_{i-1}} + \frac{1}{\epsilon_{i+1} - \epsilon_i} \right]^{-1},$$

$$w_2 \equiv \frac{1}{\epsilon_{i+1} - \epsilon_i} \left[\frac{1}{\epsilon_i - \epsilon_{i-1}} + \frac{1}{\epsilon_{i+1} - \epsilon_i} \right]^{-1}.$$

Assumption (A3) and equation (A2) provide sufficient information to calculate the constants a and b in (A1). Therefore (A1) and (A3) together provide a means of determining plant size (measured in dollars of shipments per year) from plant employment (measured in employees per year).

Finally, in order to divide shipments from plants of size class i into shipments from plants of suboptimal scale and shipments from plants of optimal scale, it is necessary only to assume a distribution of actual plants according to employment size. Plants are segregated into two subclasses: those with fewer than the mean number of employees ($= \epsilon_i$) and those with more than the mean number of employees. It is assumed that within each subclass, plants are evenly distributed by employment size. Given this assumption, the proportion of plants within each subclass may be calculated from the requirement that the implied average number of employees per plant approximately equal ϵ_i. (One can insure that the implied average number of employees per firm precisely equal ϵ_i only by relaxing the requirement that the number of plants within each subclass be an integer.) A typical such distribution of plants by employment size is illustrated in figure A2. Assumed levels of employment in actual plants are denoted by X.

Once the distribution of plant employment sizes and the relationship between employment and shipments have been determined, shipments from plants smaller than MES and shipments from plants larger than MES may be readily estimated.

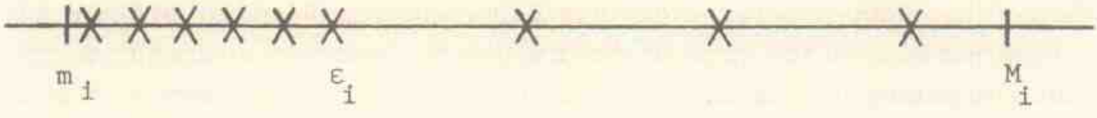

FIG. A2

Appendix B

Estimating the Relationship between Scale and Transportation Costs

To estimate transportation costs in each industry, the following assumptions are made:

1. Within any industry, the average transport cost per dollar of shipments incurred by a plant depends only on the distance over which its output is shipped.

2. Within any industry, average transport cost per dollar of shipments per mile shipped is a constant.

It is further assumed that each plant ships its output to a specific "market area" that has the following characteristics:

3. The center of each plant's market area is located at the plant.

4. Demand for each plant's output is uniformly distributed over its market area.

Following Mohring and Williamson (1969), we assume that each market area is square, and that products are shipped from each plant in north-south and east-west directions.

Let ρ represent the density of demand in a plant's market area (measured in dollars of sales per square mile), let t represent the average transportation cost per dollar of shipment per mile shipped, and let S represent the value of shipments from a plant. Then assumptions 1–4 imply that average transportation cost per dollar of shipments $[T(S)/S]$ equals:[17]

$$\frac{T(S)}{S} = \theta\, S^{1/2}, \tag{B1}$$

where

$$\theta = (\sqrt{2}/3)\, \rho^{-1/2}\, t \simeq .4714\rho^{-1/2}\, t. \tag{B2}$$

For each industry, Scherer et al. (1975) provide an estimate of transportation cost per dollar of shipment shipped over a standardized distance. These figures are employed to derive estimates of average transportation cost per dollar of shipments per mile shipped (t).

Demand density ρ is estimated by the average density of shipments throughout the continental United States in 1967 (i.e., total value of shipments divided

17. Alternative estimates could be derived by assuming that all markets are circular, and that products are shipped radially from each plant. Equation (B1) would still apply in this case. However, eq. (B2) must be replaced by:

$$\theta = (2/3)\pi^{-1/2}\rho^{-1/2}t \simeq .3761\rho^{-1/2}t.$$

That is, average transportation cost would be proportional to the square root of plant scale. The assumption of circular markets implies an average transportation cost approximately 20% smaller than that implied by the assumption of square markets.

by the area of the continental United States). This estimate would yield an accurate measure of the density of demand for an individual plant's output if all demand were in fact uniformly distributed throughout the continental United States and if all plants were located at equidistant points composing a uniform lattice. Obviously, we cannot claim such an assumption to be realistic. However, one cannot predict a clear direction of magnitude of the bias resulting from this assumption.[18]

Under the assumption that all costs are composed of either production costs or transportation costs, equation (8) implies:

$$\Pi = \frac{\Sigma_j S_j \left\{1 - c^*[1 + g(S_j)] - \dfrac{T(S_j)}{S_j}\right\}}{\Sigma_j S_j}, \tag{B3}$$

which may be solved for c^* as follows:

$$c^* = \frac{\Sigma_j S_j \left[1 - \dfrac{T(S_j)}{S_j} - \Pi\right]}{\Sigma_j [1 + g(S_j)] S_j}. \tag{B4}$$

Employing this revised estimate of c^*, average total cost may be calculated as the sum of average production cost plus average transportation cost. The economically optimal plant scale, $\bar{S}$, may be estimated as that scale which minimizes average total cost. The total percentage cost elevation suffered by a plant of scale S_j, $CET(S_j)$, is then estimated as:

$$CET(S_j) = \frac{[C(S_j)/S_j] + [T(S_j)/S_j] - \left[\dfrac{C(\bar{S})}{\bar{S}} + \dfrac{T(\bar{S})}{\bar{S}}\right]}{\dfrac{C(\bar{S})}{\bar{S}} + \dfrac{T(\bar{S})}{\bar{S}}}. \tag{B5}$$

The total percentage excess cost elevation suffered throughout the industry is estimated as the weighted average of plant cost elevations:

$$CET^{\text{ind}} = \Sigma_j \, CET(S_j) \left(\frac{S_j}{\Sigma_k S_k}\right). \tag{B6}$$

Finally, the total dollar value of excess cost due to nonoptimal capacity may be estimated as the product of the industry-wide cost elevation and the total industry minimum potential cost:

$$\text{Total Excess Cost}^{\text{ind}} = [CET^{\text{ind}}]\left[\frac{C(S)}{\bar{S}} + \frac{T(S)}{\bar{S}}\right]\left[\sum_j S_j\right]. \tag{B7}$$

18. True market demand obviously exceeds average density in some geographic areas and is smaller than average density in others. Areas with large density of demand, however, are likely to be served by competing plants. In such cases, the density of demand for an individual plant's output must be smaller than the density of demand for the product. Therefore, the relationship between the density of demand for a plant's output and the average U.S. density of demand for the product is ambiguous. This bias is a major factor contributing to the tenuity of our transportation cost estimates.

References

Bain, J. S. 1956. *Barriers to New Competition*. Cambridge, Mass.: Harvard University Press.

Collins, N., and Preston, L. 1970. *Concentration and Price-Cost Margins in Manufacturing Industries*. Berkeley and Los Angeles: University of California Press.

Harberger, A. 1954. Monopoly and resource allocation. *American Economic Review* 44 (May): 77–87.

Kamerschen, D. R. 1966. An estimation of the "welfare loss" from monopoly in the American economy. *Western Economic Journal* 4 (Summer): 221–36.

Mohring, H., and Williamson, H. F., Jr. 1969. Scale and "industrial reorganisation" economies of transport improvements. *Journal of Transport Economics and Policy* 3 (September): 251–71.

Pratten, C. F. 1971. *Economies of Scale in Manufacturing Industry*. Cambridge: Cambridge University Press.

Scherer, F. M.; Beckenstein, A.; Kauffer, E.; and Murphy, R. D. 1975. *The Economics of Multi-Plant Operations*. Cambridge, Mass.: Harvard University Press.

Schwartzman, D. 1960. The burden of monopoly. *Journal of Political Economy* 68 (December): 627–30.

Shepherd, W. G. 1967. What does the survivor technique show about economies of scale? *Southern Economic Journal* 34 (July): 113–22.

Siegfried, J. J., and Tiemann, T. K. 1974. The welfare cost of monopoly: an inter-industry analysis. *Economic Inquiry* 12 (June): 190–202.

Siegfried, J. J., and Wheeler, E. H. 1981. Cost efficiency and monopoly power: a survey. *Quarterly Review of Economics and Business* 21 (Spring): 25–46.

U.S. Department of Commerce, Bureau of the Census. 1967. *Census of Manufactures: 1963*. Industry Statistics Series. Washington, D.C.: Government Printing Office.

U.S. Department of Commerce, Bureau of the Census. 1971. *Census of Manufactures: 1967*. Industry Statistics Series. Washington, D.C.: Government Printing Office.

U.S. Department of Commerce, Bureau of the Census. 1976. *Census of Manufactures: 1972*. Industry Statistics Series. Washington, D.C.: Government Printing Office.

U.S. Department of the Treasury. Internal Revenue Service. 1971. *Statistics of Income: 1967. Corporation Income Tax Returns*. Washington, D.C.: Government Printing Office.

Weiss, L. 1976. Optimal plant size and the extent of suboptimal capacity. In R. T. Masson and P. D. Qualls (eds.), *Essays in Industrial Organization in Honor of Joe S. Bain*. Cambridge, Mass.: Ballinger.

Williamson, O. E. 1968. Economies as an antitrust defense. *American Economic Review* 58 (March): 18–36.

Williamson, O. E. 1977. Economies as an antitrust defense revisited. *University of Pennsylvania Law Review* 125 (April): 699–736.