This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: New Developments in Productivity Measurement

Volume Author/Editor: John W. Kendrick and Beatrice N. Vaccara, eds.

Volume Publisher: UMI

Volume ISBN: 0-226-43080-4

Volume URL: http://www.nber.org/books/kend80-1

Publication Date: 1980

Chapter Title: A Study of High and Low "Labor Productivity" Establishments in U.S.Manufacturing

Chapter Author: Benjamin Klotz, Rey Madoo, Reed Hansen

Chapter URL: http://www.nber.org/chapters/c3914

Chapter pages in book: (p. 239 - 292)

3 A Study of High and Low "Labor Productivity" Establishments in U.S. Manufacturing

Benjamin Klotz, Rey Madoo, and Reed Hansen

3.1 Introduction

There is a consensus that policy makers concerned with both the macro problems of inflationary growth, incomes, and employment policies, and the micro problems of production and exchange at the industry level, need to know more about the productivity process at work in the economy. Productivity has often been studied by using a production function to link the growing output of an industry to increases in the quality and quantity of its labor and capital inputs. Since input-output relations are engineering concepts that apply to production processes of individual plants, the production function is most easily understood at the plant, rather than at the industry, level. However, production functions have been estimated generally from industry aggregates of plant data, despite the possibility that the estimates may not correspond to true plant production relations. This potential aggregation bias is recognized in the literature, but estimation from industry data has continued, largely due to the unavailability of plant statistics.

The causes of productivity growth in plants could be uncovered if establishment data were available to estimate production functions. These micro data would be useful because the sources of productivity

Benjamin Klotz is at Temple University. Rey Madoo is at Howard University. Reed Hansen is at The Urban Institute.

This study was supported by grants from both the National Commission on Productivity and The Urban Institute. The opinions and conclusions expressed are those of the authors and do not necessarily represent those of the Commission or the Institute.

advance might vary among the plants of an industry, and knowledge of any difference would help government officials assess growth policies more completely. For example, if the objective is to increase productivity growth in an industry where industry productivity is measured as a weighted average of individual establishments' productivities, the average may be elevated by shifting employment and output toward the most efficient plants, or by attempting to improve the efficiency of all plants themselves. The former effect would tend to occur naturally if efficient plants increase their share of industry sales. The latter effect, however, would require some knowledge of the technology structure in the industry. And it may be found that the productivity of some subgroup of establishments is easier to raise than that of other groups. In this instance, policymakers might find it desirable to target their efforts to the most responsive establishments.

The literature on economic growth has given much attention to the process by which new technology is carried into operation by the stream of new investment. This process supposes the existence of a spectrum of plant productivities that ranges from the best-practice establishments, using the newest techniques, to the worst-practice plants, presumably using the oldest methods. A key policy question arises in this situation: Given scarce resources for investment, should the government favor investments in "best-practice" plants or a policy that facilitates the improvement of lagging establishments? The former course has received most attention in the past, but to justify such a policy, knowledge about the sources and extent of interplant productivity variation is necessary.[1]

This study uses two complementary approaches to investigate labor productivity differences among manufacturing establishments.[2] Section 3.3 discusses the nature and significance of a plant-level data set employed in the analysis. The extent of plant productivity differences within industries is displayed and analyzed by both simple correlation and multiple regression methods. The results are examined to determine if high and low productivity can be attributed to the same factors, and if differences can be explained by postulating an underlying production function for the plants. Section 3.4 then introduces and estimates a very general production function, and we test whether the parameters of this function differ significantly between high and low productivity establishments. A corresponding cost function is also estimated. An attempt is made to incorporate such factors as monopolistic competition and plant disequilibrium into the analysis. Finally, some conclusions are

1. In the past, the USSR has also emphasized productivity advance through new construction, but it is now shifting attention toward the improvement of the plants it already has. See the *Wall Street Journal*, 12 June 1975, p. 4.

2. Throughout this paper value added per man-hour is the measurement concept of productivity that we refer to.

offered about the causes of productivity differences among plants and how establishment data could be organized to improve the analysis of such differences.

3.2 Census Data on Establishments

To date, little research about the nature of interplant differences in productivity exists. Census data on a wide variety of U.S. plants have been gathered for several years but this information has not been available for analysis because the Census Bureau has not had the resources to do an investigation itself, and because the Census Bureau is legally prohibited from disclosing establishment information to researchers who are not sworn census agents. Despite these limitations, a few studies of plant productivity differences have been done. Krishna (1967) studied plant production relations in four four-digit manufacturing industries, and Klotz (1970) analyzed seventeen industries using Cobb-Douglas and constant-elasticity-of-substitution production functions. Both investigations found that various proxies for measures of the capital-labor ratio were significant in explaining labor productivity. Furthermore, constant returns to scale seemed to be a central tendency in these industries, so that plant size was generally not a significant cause of productivity differences.[3]

Neither of these studies, however, surveyed a wide variety of industries, and neither tried to determine if variations in the capital-labor ratio were as successful in explaining high, as opposed to low, productivity levels in plants within industries. The high-low distinction may reveal that low productivity plants employ different production technologies from their high productivity competitors. Recent work at the Urban Institute investigated this question by using a special tabulation from the 1967 Census of Manufactures to inspect data for groups of high and low productivity plants within industries.[4] Madoo and Klotz (1975) applied simple correlation methods to 102 industries and found evidence suggestive of structural differences between the two groups of establishments. Using 40 industries, Jones (1975) found that the elasticity of demand for production workers depended upon plant size when low

3. Griliches and Ringstad (1971) found roughly the same results in Norwegian plant data, although some industries had increasing returns to scale. Implications about size have been invariably drawn from estimating homogeneous production functions and then checking the adding-up properties of proportionate changes in all inputs. But see Hanoch (1975) for remarks on the usefulness of this approach, and Madoo (1975) for treatment of the alternative formulation and estimation of the elasticity of scale along the cost-minimization expansion path.

4. These data were compiled by the Census Bureau for the National Commission on Productivity.

productivity establishments were examined, but that this dependence disappeared among the high productivity groups of plants.[5]

The rest of this paper contains a report of the findings of the first study (Madoo and Klotz 1973) and subsequent work with a larger sample of industries. But first we give an account of the special characteristics of the census tabulation and an explanation of how samples of industries were selected.

3.2.1 The Special Tabulation of 1967 Data

The set of data is based on information received by the U.S. Census Bureau from each establishment in 412 four-digit manufacturing industries. The Census Bureau used this information to rank establishments in each industry by their value added per production worker man-hour in 1967. The ranking was then divided into groups with an equal number of plants in each quartile.[6] Our goal is to explain quartile differences in value added per production worker man-hour. *We here call this concept productivity.*[7]

In each quartile, data are reported for production workers, nonproduction workers, gross book value of capital, payroll, man-hours, capital expenditures, value added, value of shipment, cost of materials, and inventories. These variables are constructed by summing the corresponding statistics of all plants in the quartile. The quartile data used in this study are divided by the number of plants in the sum and are thus arithmetic averages of the plant statistics in each quartile.

Sample Size Selection

For the various stages of empirical investigation attempted we were not able to use the quartile data of all 412 four-digit industries in our analysis. Missing data, especially those pertaining to gross book value of capital, combined with obviously incorrect values for some items forced us to work with a maximum set of only 195 industries. This sample of industries is not randomly chosen, but we consider it a representative sample because the distributions of both average hourly earnings and value added per man-hour are roughly the same in the 195

5. Size distinctions were made by grouping plants into two classes: those with 100 or more employees and those with less.

6. Decile grouping would allow more detailed analysis, but of fewer industries. Census Bureau disclosure rules are often violated in small industries when decile data pertain to only a few plants.

7. It is neither a pure efficiency nor a total factor productivity measure. Data limitations preclude using output per unit of input, where all inputs are quality-weighted and summed.

industries and in the 217 excluded from the sample.[8] The quartile data of the 195 industries are used in the multiple regression and production function analysis of this study. For simple correlation our results are based on a subset of the 195. We wanted industries whose quartile totals contained information on current capital expenditures as well as on the gross book value of assets. An additional criterion was to choose only industries in which each quartile was at least 89% specialized in the production of the industry's major commodity. Together, this screening by product specialization and the availability of gross assets and capital expenditures left only 102 usable industries for the simple correlation analysis;[9] they appear similar to the set of 195.

3.3 Productivity Differences among Establishments

3.3.1 Evidence from 102 Industries

Ranking and grouping establishments reveals some dramatic productivity differences among plants within industries. Figure 3.1 shows that value added per man-hour in the top quartile of plants is over twice as great as the corresponding industry average in 16 of the 102 industries; the top quartile in no industry is less than 125% of the average. Productivity in the typical top-quartile group of plants is about 65% greater than the industry average and 200% greater than the average of low-quartile establishments. On the other hand, productivity in the low quartiles is less than half the industry average in almost two-thirds of the 102 cases. This bottom quartile is always less than 70% of the comparable industry average. Value added per man-hour in the typical low-quartile establishment is only about 40% of the industry average. Value-added productivity for individual industries and quartiles is listed in the table in Appendix A for the interested reader. The table also contains measures of productivity spread within each industry.

3.3.2 Conjectures

Great productivity differences among plants in the same industry could be caused by a number of forces. First, establishments may not

8. In addition, average productivity in quartile 1 establishments is similar in both data sets. However, in quartiles 1 and 4 wages are 10–15% higher in our sample industries. In quartile 4 the excluded industries have higher average productivity but this is due to several extremely high observations caused by bad data. Little difference remains when median productivities are compared.

9. The data are discussed extensively in Madoo and Klotz (1973). The data do not, however, reveal the extent to which different five-digit products are produced in seemingly homogeneous four-digit industries.

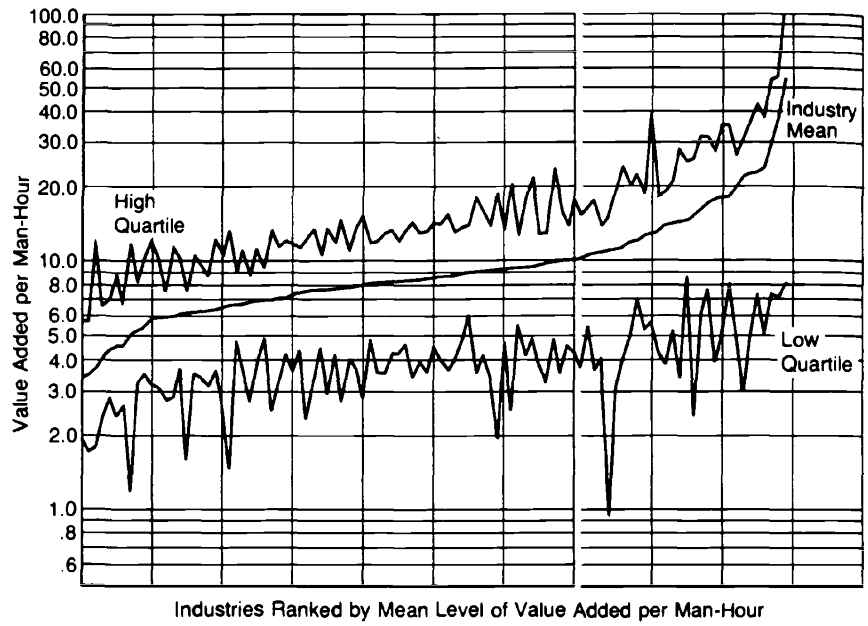

Fig. 3.1 Value added per man-hour of high and low quartiles to the industry mean

be using the same techniques to produce industrial commodities. Economists have been giving increasing attention to the implications of the idea that a wide spectrum of technologies, corresponding to capital equipment of different ages and efficiencies, can coexist in an industry at the same time. Since new vintages of capital make use of the latest and best industrial techniques, it follows that plant productivity will depend in part on the newness of its capital equipment. The "vintage" theory assumes that new technology is carried into practice by the current stream of capital investment expenditures.[10] This reasoning leads to the first conjecture about productivity: (1) Establishments with higher (lower) productivity invest more (less) in capital assets per worker than average plants.

A second source of establishment differences in value added per production worker man-hour is the likelihood that plants are not employing the same relative quantity of other factor inputs in combination with their production workers. Most economists consider capital assets and nonproduction labor as main factors of production (in addition to

10. The idea that a spectrum of technologies can coexist in practice was first extensively explored by Salter (1962).

production workers), so it would seem likely that top-quartile plants employ relatively more of both factors than average establishments, and that low-quartile plants employ relatively less, because output per man-hour is increased if laborers are able to work with more capital assets and more skilled nonproduction technicians.[11] This argument leads to the following conjectures: (2) High (low) productivity establishments employ more (less) capital assets per production worker man-hour than average plants, (3) High (low) productivity establishments employ more (less) nonproduction workers per production worker than average plants.

A third source of productivity differences among plants could be caused by any tendency they might have to hire production laborers of various qualities. Some establishments may prefer to pay high wages and thereby accumulate a highly skilled and efficient work force; other plants may be content with lower quality production workers whose productivity and wages would therefore be lower. Higher-wage plants could therefore be expected to have higher value added per production worker man-hour. Thus the conjecture: (4) High (low) productivity plants have high (low) quality production workers and therefore pay them higher (lower) wages than average establishments.

Fourth, plants may have productivity differences simply because they are not the same size. The current body of evidence on size and productivity is mixed; there are still different methodological approaches to this issue. For example, the engineering-information questionnaire approach may be contrasted with the estimation of production functions with census data.[12]

From the data available to us, plant size is measurable by assets, shipments, or employment.[13] But use of the shipments definition of size makes it difficult to distinguish any scale economies from disequilibrium effects. Firms recently awarded large contracts will probably be expanding shipments faster than their labor force. Since value added is defined as value of shipments minus cost of materials (adjusted for inventory change), expanding plants would tend to have both supranormal shipments volume and ratios of value added per man-hour, leading to the mistaken finding that value-added productivity is associated with scale

11. For surveys of production theory and empirical findings see Brown (1967), Jorgenson (1972, 1974), Kennedy and Thirlwall (1972), and Nadiri (1970).

12. See Scherer (1970, chap. 4; 1973), and Pratten (1971) for the questionnaire approach, and Griliches and Ringstad (1971), Klotz (1970), and Krishna (1967) for production function estimation. Also, see Madoo (1975) for an adaptation of the data set of this study to measuring the optimum plant size under a cost specification for eight (SIC two-digit) industries.

13. Capacity output is ideal, but unavailable.

of plant (shipments): i.e., economies of scale would appear to exist, but this appearance could be due purely to the transitory disequilibrium experienced by the fortunate plant.[14]

Errors in the computation of value of shipments will have the same effect as short-run production oscillations in introducing a transitory component to shipments and in creating a correlation between them and value added per man-hour.[15] Since both errors and oscillations will tend to increase the correlation when plant size is defined as value of shipments, it is better practice to define size in terms of the capital assets possessed by the plant because assets are not so likely to be strongly affected by transitory movements.

These arguments lead to the final conjectures: (5) High (low) productivity plants have more (less) capital assets and more (less) output than average establishments, (6) Due to transitory output movements, value-added productivity will be more strongly linked with value of shipments than with capital assets.

A crude test of the six conjectures appears in table 3.1. The first line compares average capital expenditure per employee across the 102 industries in 1967. Their top-quartile plants spent 50% more per employee than the industry average, so this comparison is consistent with conjecture (1). But bottom-quartile plants spent only 8% less than the average, seemingly weak evidence in favor of the conjecture. This weakness may appear because capital spending for only one year, 1967, is recorded, and such investment flows may be affected by a variety of short-term events that are unrelated to the basic long-run expenditure pattern of low-quartile plants.

The second and third conjectures are clearly supported by the data. Capital assets per production worker man-hour in top-quartile establishments are almost twice the industry average, while bottom quartiles are considerably below average.[16] These differences are even more pronounced when the variable in question is the ratio of nonproduction to production workers.

Conjecture (4) also fits the data: top (bottom) quartile establishments pay higher (lower) wages than their industry average, but these wage differences are not nearly as pronounced as the value-added productivity differences between top and bottom quartiles. Variations in

14. Conversely, unluckily, plants may lose contracts and experience a temporary depression in both their shipments and value added per man-hour. This also gives the appearance of scale economies because it creates a tendency for smaller plants to register less value-added productivity.

15. In the long run, transitory effects are averaged out so plants with temporarily high (and low) productivity will tend to rebound toward the industry average in subsequent years. Evidence of this rebound effect appears in Klotz (1966).

16. Recall that the capital measure is gross book value of capital; that is, historical cost rather than constant dollar values.

Table 3.1 **Average Establishment Production Statistics for 102 Manufacturing Industies, 1967**

Statistics	Bottom Quartiles	Industry Average	Top Quartiles
Capital expenditures per employee[a]	1.03	1.11	1.62
Assets per production worker man-hour[b]	6.85	9.29	17.31
Nonproduction worker per production worker	.24	.42	1.06
Hourly wage rate of production workers[b]	2.44	2.96	3.21
Capital assets per plant[c]	1.40	2.97	3.96
Shipments per plant[c]	2.46	6.67	9.82

[a]Dollars × 1000.
[b]Dollars.
[c]Dollars × 1,000,000.

production labor quality are therefore apparently not sufficient to explain the dispersion of value-added productivity within industries.

Row 5 of the table supports the fifth conjecture: top quartile establishments tend to be almost three times as large as their bottom-quartile counterparts when plant size is defined in terms of capital assets. And, consistent with conjecture (6), the difference is even greater (four to one) when size is defined by value of shipments. Apparently there are significant transitory components in shipments, but nevertheless (due to the good performance of the asset definition of size), the evidence appears to point toward some economies of large-scale production.

The findings of table 3.1, while interesting and suggestive, do not provide clear insights into the simple correlations between value-added productivity and other variables. These correlations are reported in the next section.

3.4 Simple Correlation Results for 102 Industries

Column (1) of table 3.2 shows that the value-added productivity of bottom-quartile establishments (relative to their industry average) has only a 0.15 correlation with their capital expenditures per employee (also relative to the industry average). The top-quartile correlation is only 0.23. Therefore, for both high and low-quartile observations, *variations* in relative capital expenditures per employee are not highly correlated with variations in relative productivity—in spite of the fact that conjecture (1) is true (the average *level* of productivity and expenditures per employee is higher among top-quartile establishments). The

Table 3.2 Correlation between Value Added per Man-Hour and Selected Variables (relative to the industry mean), 1967

Value Added per Man-Hour Relative to Industry Average	Relative Variable: Capital Expenditures per Employee	Assets per Production Man-Hour	Nonproduction Workers per Production Worker	Average Hourly Earnings of Production Workers	Assets per Plant	Shipments Per Plant
Top quartile[a]	.23	.58	.64	.18	−.43	−.30
Bottom quartile	.15	.07	.22	−.01	.24	.49

NOTE: Correlations greater than 0.2 in absolute value are statistically significant at the 95% confidence level because 102 observations are use.

[a]Excludes book publishing (SIC 2731) because of its extremely low reported man-hours.

low correlations, if interpreted as causal relations,[17] imply that increased spending will not strongly succeed in raising either top or bottom-quartile plant productivity in relation to the industry average. More likely, however, the low correlations may be due to the fact that capital expenditures contain many transitory effects. Cumulative expenditures might have much higher correlations. This possibility appears likely because assets could be a reasonable proxy for cumulated expenditures, and variations in the degree to which top quartile productivity exceeds its industry average is strongly (0.58) correlated with the degree to which its assets per man-hour exceed the average. If high correlation is again interpreted as causation, then the value-added productivity of top-quartile establishments can be raised by increasing their assets per man-hour. Top-quartile relative productivity is also highly correlated (0.64) with a relative abundance of nonproduction workers per production worker. Therefore, the proportions with which both capital and nonproduction labor are combined with production labor seem to have a definite connection with the productivity performance of top-quartile establishments.

The other variables tend to have less association with top-quartile productivity. The quality of labor (represented by the wage rate of production workers) has an unassuming 0.18 correlation. Plant size variables (assets and shipments) are negatively correlated with variations in value-added productivity: industries where top-quartile plants are farthest above average in productivity tend to be industries where these plants are the *least* above average in size. This finding does not support a belief in economies of scale in production.

The pattern of association exhibited by bottom-quartile establishments is considerably different from the top-quartile configuration. With one exception, variations in the degree to which low-quartile productivity falls short of its industry average is *not* associated with variations in the degree to which other variables fall short of the average. Relative plant size is the exception. Both size measures exhibit positive correlation with relative value-added productivity: assets per establishment and shipments per establishment show correlations of 0.24 and 0.49, respectively. The latter correlation may be greater due to transitory components (short-run disequilibria and data errors) in value of shipments which are transmitted to value added per man-hour through their effect on value added.[18]

17. With 102 industry observations, correlations above 0.2 are statistically significant at the 95% confidence level. The danger of attributing causation to correlations is discussed later.

18. Remember that value added is the residual obtained by subtracting materials cost from shipments.

Surprisingly, both the relative proportions of assets and the relative proportions of nonproduction labor (per production worker), which played such a powerful role in explaining top-quartile productivity variations, have no effect on the relative performance of bottom-quartile establishments.

This diversity of results between top and bottom-quartile correlations may be due to extremely poor data submitted by bottom-quartile plants, or it may be due to something more fundamental. Errors obviously exist in the data, but Salter (1962, p. 13) indicates that moderate errors do not distort correlations unduly. Data errors therefore would have to be truly monumental to cause the vastly different correlations experienced by top and bottom-quartile plants in table 3.2.

A more promising explanation of the differences might focus on unobservable variables left out of the correlation analysis. For example, low-quartile establishments could be using completely different technologies than their top-quartile counterparts within the same industry. This technological explanation may be sufficient reason for a different correlation pattern between the two types of plants, or it may reflect some unobservable factor even more fundamental, such as management quality or five-digit product mix not captured in the four-digit statistics. For example, if there are large differences in managerial quality among low-quartile establishments, these quality differences could easily lead to large variations in plant performance. The variations would likely influence the level of technology adopted in the plant and the speed of adjustment of the plant to changing economic conditions. Data on management quality are unfortunately not directly observed.[19]

For deeper insights into plant productivity differences we must turn to a multivariate analysis. For this analysis the data set is expanded to

19. One good proxy for quality in the census files is the ratio of nonproduction to production workers but, as we saw, this variable was not correlated with the value-added productivity of low-quartile establishments. A better proxy might be establishment profits but this quantity is not recorded in the census data file. What is recorded is the gross margin per dollar of value added: (value added—total payroll)/value added. This is a measure of how much is left after all wages are paid. Thus the gross margin includes profits, rental payments, depreciation, and many other minor items. It is an imperfect measure of management quality, but we expect it to be correlated with value-added productivity. We find, however, that a correlation of coefficient of .01 exists between value-added productivity and gross margin (both variables expressed relative to their industry average) across industries for the top-quartile establishments, but a dramatic 0.74 results for the bottom-quartile plants. If this measure is to be taken seriously, the bottom-quartile groups could have large differences in management quality (relative to their industry average) while the top-quartile groups do not. But gross margin cannot be accepted as an unambiguous measure of productivity performance without much more analysis.

195 industries by relaxing the criterion that industries have a high product specialization ratio. This ratio is measured explicitly.

3.5 Multiple Regression Analysis Results for 195 Industries

Both economic theory and the results of our simple correlations suggest at least five variables for analysis. Two of the five (the gross book value of capital per production worker man-hour and the ratio of nonproduction to production workers) are chosen to represent the volume of inputs that cooperate with production labor in the production process.[20]

A third variable, the wage of production workers, is designed to capture differences in the quality of production labor between high and low-productivity plants in the same four-digit industry. The fourth variable, a measure of scale, is plant size. It is measured by the level of production worker man-hours; other measures of size (such as assets or the value of shipments) were not used because they gave unstable results when they were simply correlated with productivity. The fifth analytical variable is the plant specialization ratio, the percent of plant shipments accounted for by the primary product produced. This ratio is important because different production mixes occur among plants even though they are in the same four-digit industry, and these differences affect average plant productivity to the extent that labor productivity differs by type of product.

Since the analysis seeks to explain both high and low-productivity performance, we focus on third-quartile establishments (i.e., those just below the high-productivity quartile) as a standard of comparison. These plants tend to cluster near the productivity average for the industry, so they are a convenient proxy for the average itself. The actual industry average is not used as a standard because it is a weighted average of the productivity levels of all four quartiles, and it is therefore influenced by the productivity levels in quartiles 4 (high-productivity) and 1 (low-productivity). This influence could cause spurious correlation in the regressions when either quartile 4 or 1 is compared with the industry average.

In the multiple regression analysis the dependent variable is the percentage by which the productivity of quartile 4 establishments exceeds

20. Capital expenditures per man-hour of labor is not used as a variable because there are gaps in the data but, in any case, we observed from the sample of 102 industries that its simple correlation with productivity was small. Also, expenditures in one year do not make a sufficient contribution to the capital stock in most cases to significantly increase productivity. The effect of expenditures on productivity, however, is implicitly included in the analysis because the capital-labor ratio is used as a variable.

the productivity of plants in quartile 3. (Also, the percentage difference in productivity between establishments in quartile 3 and establishments in quartile 1 is analyzed.) This percentage difference is hypothesized to be positively related to percentage differences in (1) gross book value per production worker man-hour (K/H); (2) nonproduction workers per production worker (N/L); (3) hourly wage of production workers (W); (4) production worker man-hours (H). A fifth variable, the product specialization ratio (S), is included. It may be positively or negatively related in quartile productivity differentials because labor productivity on major products may be either higher or lower than productivity on minor products.

The five productivity hypotheses are tested by multiple regression methods and the results are displayed in table 3.3. Strikingly little of the percentage productivity differences among quartiles (of plants in the same industry) can be explained by differences in the five variables in the regression equation. The relative variables for quartile 3 versus quartile 1 yield an equation with a coefficient of determination of 0.03: only 3% of the variation in the productivity ratio (quartile 3 to quartile 1) can be explained by the five variables. In addition, none of the five variables are significantly different from zero at the 95% confidence level because no t statistic exceeds 1.7 in value (with 195 observations, if the true coefficient of regression were zero the t statistic for the coefficient would exceed 2.0 in absolute value, by chance, 5% of the time). Defining size as the level of production worker man-hours, the 0.08 regression coefficient indicates that quartile 1 plants have 0.08 lower productivity than quartile 3 establishments for every one percent that they are smaller in size than these establishments. The positive elasticity seems to indicate that plants have low productivity partly because they are undersized.

Notably, percentage differences in neither capital per production worker man-hour nor nonproduction workers per production worker play much of a role in explaining the percentage difference in productivity between quartile 3 and quartile 1 plants. Neither of the two vari-

Table 3.3 **Multiple Regression Analysis of Relative Productivity Differences**

Percentage Difference in Productivity: VA/H	Percentage Difference in: Constant	K/H	N/L	W	H	S	R^2
Quartile 4 vs. Quartile 3	.56	.16	.07	−.03	.03	−.02	.08
(t statistic)	(3.6)	(3.6)	(1.5)	(−1.0)	(1.1)	(−o.1)	
Quartile 3 vs. Quartile 1	.39	.06	.07	.13	.08	.26	.03
(t statistic)	(1.9)	(1.1)	(0.9)	(1.1)	(1.7)	(1.5)	

ables has a t value above 1.1, indicating little confidence that their impact is significantly nonzero. The capital-labor coefficient is only 0.06. If this coefficient expresses a causal relation, it suggests that a 100% increase in the capital-labor ratio of quartile 1 plants would only reduce the productivity differential between quartile 1 and 3 plants by 6%. Differences due to product mix appear also to be unimportant.

The multiple regression results for the high-productivity quartile are not much more encouraging than those for quartile 1. The regression equation in table 3.3 is able to explain only 8% of the productivity variation between quartile 4 and quartile 3 plants. In striking contrast to the low-productivity equation, differences in capital per production worker man-hour are now significantly different from zero in explaining productivity differentials. However, the regression coefficient of the capital-labor ratio is small (0.16), suggesting that a 100% increase in the ratio for quartile 3 establishments would only lower the quartile 4–quartile 3 productivity differential by 16%. Other variables in the quartile 4–quartile 3 equation have even less impact on productivity differentials, and none of their effects are significantly different from zero at the 95% level of confidence.

The results of table 3.3 indicate that interquartile productivity differences, at a point in time, cannot be well explained by variations in the five variables examined in this study. On the other hand, our results are also consistent with the theory of production which suggests that differences in output per man-hour are positively related to differences in factor proportions among establishments. The strongest variables in the multiple regression equation for the quartile 4–quartile 3 differential were capital per man-hour and nonproduction workers per production workers. We conclude, therefore, that the interquartile productivity differentials may be better explained with a framework that uses a production function as the starting point. In the next section we will try to estimate the parameters of the recently developed translog production function (Christensen, Jorgenson, and Lau 1973).

3.6 Analytical Findings

3.6.1 Models of Production

In this section we investigate the extent to which productivity differences between high and low-quartile groups are due to their being on different production functions or at different points on the same production function. In particular, we focus on differences in the partial elasticity of substitution (S) among pairs of inputs. Our assumed functional form imposes no a priori restrictions of homotheticity or additivity. Thus, partial elasticity of substitution estimates can vary with output levels, and the cross effects of various combinations of inputs on output are not assumed to be zero.

At any level of output, ease of substitution (S) will depend on the type of production technology that translates inputs into outputs. A good deal of evidence in two-input models indicates that, although S can differ considerably among some manufacturing industries at the two-digit level, it tends to be on average near 1.0 in the aggregate (Jorgenson 1974).Unfortunately, most past estimates of S have made a number of restrictive assumptions about the form of the production function. Two of them will be relaxed in this study. First, only two inputs (aggregate labor and aggregate capital) are usually introduced in the analysis for each industry; few attempts are made to disaggregate labor into two or more classes.[21] But there is accumulating evidence that the elasticity of demand for labor varies by skill class and by other divisions, such as a split between production and nonproduction workers.[22] Second, most estimates of S assume that it is a constant regardless of the relative importance of capital and labor in the production process.[23] A more flexible procedure, as in Hildebrand and Liu (1965), allows S to vary with output and input levels so that the possibility of varying rates of substitution is considered. In this study we treat S as a variable, and we also distinguish between production and nonproduction labor.

Model Specification

Because output depends on physical inputs of raw materials and other produced products, as well as on labor and capital, the theory of production suggests that the physical volume of goods produced is the appropriate concept of output. But census data on raw material inputs and other produced products used by plants are incomplete or nonexistent. Thus, in practice, we separate materials from output and work with a concept of real value added as the output of an establishment. Econometrically, this definition of output removes measurement error from the right side of the production equation (we exclude materials inputs), where it necessarily causes a bias, and puts it on the left side (in value added), where it does not necessarily distort the regression coefficients of the production equation. Furthermore, this definition is theoretically permissible if the elasticities of substitution between mate-

21. Exceptions are Berndt and Christensen (1973), Crandall, MacRae, and Yap (1975, Gramlich (1972), and Hildebrand and Liu (1965).

22. See Berndt and Christensen (1974), Diewert (1969), and Gramlich (1972).

23. This assumption is behind the popular constant-elasticity of substitution production function which is examined in Jorgenson (1974) and Brown (1967).

rials and each included (labor and capital) input are the same. Arrow (1972) discusses this separability question further.[24]

When separability holds we can write the production function as a function of primary inputs alone, so that

$$\mathrm{VA} = f(L_1, L_2, K),$$

where VA is real value added, corresponding to the money value-added measures of our data set, and L_1, L_2, and K are the primary factors of production. Our estimation assumes that this real value-added production function is the *same* for all the plants of a given quartile across the 195 industries of our sample, but the function is allowed to vary among quartiles so that we can test for interquartile differences.

A Translog Production Function

The translog production function is given in equation (1) below. It contains three inputs and one output, and is a quadratic approximation in the logarithms of the variables for any arbitrary production function:

$$\begin{aligned} \log Q = {} & a_0 + a_1 \log L_1 + a_2 \log L_2 + a_3 \log K \\ & + a_4 (\log L_1)^2 \\ & + a_5 (\log L_2)^2 + a_6 (\log K)^2 + a_7 \log L_1 \log L_2 \\ & + a_8 \log L_1 \log K + a_9 \log L_2 \log K. \end{aligned} \tag{1}$$

The properties of this function are discussed elsewhere in great detail (Berndt and Christensen 1973, 1974; Christensen and Lau 1973). We will describe its estimation features here only to the extent necessary to furnish an understanding of how we use it to arrive at estimates of elasticity of substitution measures of L_1 with respect to the other inputs L_2 and K.

The coefficients of the translog function can be estimated directly from equation (1), but since there are nine variables (six of which are second-order terms involving squares of cross products), direct estimation risks multicollinearity. To avoid collinearity, an indirect estimation procedure can be used. If we can assume profit-maximization behavior for producers and that all markets are competitive, we can provide estimates of most of the parameters of the translog function indirectly by setting the set of three marginal productivity functions equal to their respective factor prices. Differentiating (1) with respect to log L_1, log

24. For more on these issues see Hall (1973) and Diewert (1973). Apart from these theoretical considerations there is some evidence that omitting materials as an input under a value-added weight of output may not be serious because they are used in nearly fixed proportions with output (Klotz 1970).

L_2, and log K, and applying the assumption that inputs are paid the value of their marginal product, gives the marginal productivity relations:

$$d(\log Q)/d(\log L_1) = a_{10} + a_{11} \log L_1 + a_{12} \log L_2 + a_{13} \log K, \tag{2}$$

$$d(\log Q)/d(\log L_2) = a_{20} + a_{21} \log L_1 + a_{22} \log L_2 + a_{23} \log K, \tag{3}$$

$$d(\log Q)/d(\log K) = a_{30} + a_{31} \log L_1 + a_{32} \log L_2 + a_{33} \log K, \tag{4}$$

where the a_{ij} coefficients are related to those of equation (1) in a simple manner.

Competitive profit maximization implies that $d/Q/dL_1 = P_L/P_Q$, where P_L and P_Q are the wage of labor L_1 and the price of output Q, respectively. Using the identity $d(\log Q)/d(\log L_1) = (dQ/dL_1)\ (L_1/Q)$, we have

$$d(\log Q)/d(\log L_1) = (P_L/P_Q)(L_1/Q) = M_1, \tag{5}$$

which is L_1's share in value added. Similar expressions hold for L_2 and K. Combining input (2)–(5) gives the share equations to be estimated:

$$\begin{aligned} M_1 &= a_{10} + a_{11} \log L_1 + a_{12} \log L_2 + a_{13} \log K, \\ M_2 &= a_{20} + a_{21} \log L_1 + a_{22} \log L_2 + a_{23} \log K, \\ M_3 &= a_{30} + a_{31} \log L_1 + a_{32} \log L_2 + a_{33} \log K. \end{aligned} \tag{6}$$

This set of three equations has some interesting properties that affect the way they can be estimated most efficiently. For one thing, the three cost shares in (6), M_1, M_2, and M_3, sum to unity by definition. Also, a change in log L_1 in each equation of (6) should not change the property that the cost shares sum to unity, so the sum of the three coefficients of log L_1 should be zero. The same zero-sum restriction holds if we change either log L_2 or log K. Thus there are three sets of restrictions on the a_{ij} in (6):

$$\begin{aligned} a_{11} + a_{21} + a_{31} &= 0, \\ a_{12} + a_{22} + a_{32} &= 0, \\ a_{13} + a_{23} + a_{33} &= 0. \end{aligned} \tag{7}$$

But because the three shares must add to unity, we also have the restriction that the intercepts of the equation must add to one. Thus:

$$a_{10} + a_{20} + a_{30} = 1. \tag{8}$$

These four restrictions are a characteristic of the translog share system. They are called the "homogeneity" restrictions because competitive cost shares sum to unity without a residual.

Estimates of the parameters of (6) help us compute the various partial elasticities of substitution among inputs. Using the notation

$$X_1 = L_1, X_2 = L_2 \text{ and } X_3 = K, \tag{9}$$

the technical definition of S is

$$S_{ij} = \sum_{h=1}^{3} f_h X_h |F_{ij}| / X_i X_j |F|, \tag{10}$$

where f_h is the partial derivative of the production function with respect to input h, F is defined as the bordered matrix of derivatives of (1),

$$F = \begin{bmatrix} 0 & f_1 & f_2 & f_3 \\ f_1 & f_{11} & f_{12} & f_{13} \\ f_2 & f_{21} & f_{22} & f_{23} \\ f_3 & f_{31} & f_{32} & f_{33} \end{bmatrix}, \tag{11}$$

f_{ij} ($i = 1, 2, 3$ and $j = 1, 2, 3$) is the partial derivative of the production function, first with respect to input i and then with respect to input j, and F_{ij} is the cofactor of F obtained by deleting its f_{ij} row and column (see Allen 1938, pp. 503–10). $|F|$ and $|F_{ij}|$ are the determinants of F and F_{ij}. In this formulation S_{ij} is the Allen elasticity of substitution (AES). Equations (5), (6), and (10) imply that S_{ij} depends on all production function coefficients and the relative levels of all three inputs.

A Translog Cost Function

Share system (6) gives the complicated expression (10) for the S_{ij}, so confidence bounds cannot be determined to see if, for example, they differ between the plants of various quartiles. To overcome this problem we can estimate a slightly different share system whose coefficients give the S_{ij} in a very simple and direct fashion. Instead of postulating a translog production function we suppose that there is a translog average cost function derived from an underlying production function, homogeneous of degree one. Then we can write

$$\log C/Q = b_o + \sum_{i=1}^{3} b_i \log p_i + \sum_{j=1}^{3} \sum_{i=1}^{3} b_{ij} \log p_i \log p_j, \tag{12}$$

where $C/Q =$ average cost, $p_i =$ price of the ith input, and prices and output are exogenous. The b's are coefficients to be estimated.

Differentiating (12) gives

$$d\log(C/Q)/d \log p_i = b_{io} + b_{i1} \log p_i + b_{i2} \log p_2 + b_{i3} \log p_3 \tag{13}$$

Using the results (14)–(16),

$$d\,(C/Q)/dp_i = L_i/Q, \tag{14}$$

(15) $$d\log(C/Q)/d\log p_i = [d(C/Q)/dp_i]p_i/(C/Q),$$

(16) $$C = VA$$

(by homogeneity of Q), we have

(17) $$d\log(C/Q)/d\log p_i = \frac{p_i L_i}{C} = \frac{p_i L_i}{VA} = M_i,$$

(18) $$M_i = b_{io} + \sum_{j=1}^{3} b_{ij} \log p_j \qquad (i = 1, 2, 3).$$

Since the Allen-Uzawa partial elasticities of substitution can be represented very simply by the derivatives of the cost function (Uzawa 1962), the share system (18) allows us to compute S_{ij} directly. Let the cost function be written as $C(Q_1, P_1, P_2, P_3) = Q\, g(P_1, P_2, P_3)$, then the average cost function can be written as

(19) $$C/Q = g\,(p_1, p_2, p_3)$$

and

(20) $$S_{ij} = [g\, \partial(\partial g/\partial p_i)/\partial p_j]/\,(\partial g/\partial p_i)\,(\partial g/\partial p_j).$$

For $i \neq j$ we have $S_{ij} > 0$; for $i = j$ we have $S_{ij} < 0$. Applying (20) to (13) gives

(21) $$S_{ij} = b_{ij}/M_i M_j + 1 \qquad (i \neq j) \text{ and}$$
$$S_{ii} = b_{ii}/M^2{}_i - M^{-1}{}_i + 1.$$

Of course, since the shares add to unity, estimation of (18) is also subject to the restrictions

(22) $$\sum_i b_{io} = 1 \text{ and } \sum_i b_{ij} = 0 \qquad (j = 1, 2, 3).$$

3.6.2 Estimating the Models

We estimate the parameters of the production function for the three different specifications just discussed. First we estimate the translog production function directly. Secondly, we estimate the share system of equation (6), and finally the share system (18) is used to provide a third set of estimates. For direct estimation of (1) we compare ordinary least-squares (OLS) and two-stage (2SLS) methods. For estimation of the translog parameters indirectly through (6) and (18), we use three-stage least-squares (3SLS). Each of the three estimation methods has its own advantages and disadvantages, depending upon whether we are estimating (1) or (6).[25]

A well known problem with OLS estimates of production function parameters is that they give inconsistent estimates of the coefficients if

25. A good analysis of the three methods can be found in Theil (1971).

the left-hand (dependent) and right-hand ("independent") variables are simultaneously determined. Bias arises in the structure we have imposed because the right-hand variables are not really independent of the disturbances in the regression equation, and this is the case for both the production function and the factor share system of (6). In (1), one clear chain of causation runs from the inputs on the right-side of the equation to output on the left side. In system (6), another chain runs from output back to the level of inputs through the labor demand equations represented by (6). This double chain of causation means that (1) cannot be estimated satisfactorily by OLS without bias. And in the case of (6), more efficient estimates may be obtained by estimating the system of equations jointly. The same argument for joint estimation applies to the cost system (18) as well.

Variables Used

The three inputs used in the estimation of the translog production function (1) are the man-hours of production workers (L_1), nonproduction workers (L_2), and capital (K). Data on the first two inputs are measured in physical units and the last, capital, are in dollars. For model (18), input prices are the wages of production workers (W_1), the wages of nonproduction workers (W_2), and the gross margin per unit of capital (value-added-less-payroll per dollar of gross book value of capital). In the directly estimated form (1), output is value added. Data on this variable are also given in dollar values. The problem of measurement-error bias in the direct estimates of the translog function (due to inclusion of variables measured in money units) is treated by dividing all money variables by similar money variables from neighboring quartiles. The assumption is that the output price index is the same for establishments in neighboring quartiles. All variables in the function are thus ratios of measured quantities, we hope with price effects purged.

Inputs (production workers, nonproduction workers, and gross assets, a proxy for capital) of the cost-share equation system (6) that appear on the right-hand side of the translog are left in their original measured form. The left-hand share variables represent the percentage share of total cost for each of the three inputs, with the sum of shares adding to one.

For purposes of 2SLS estimation we use nine instrumental variables: (1) the rate of growth in the price index for shipments from 1958 to 1967, (2) the rate of growth in the value of shipments from manufacturing industries from 1958 to 1967, (3) the rate of growth in the value of shipments per man-hour from 1958 to 1967, (4) the number of companies for each four-digit SIC industry, (5) the concentration ratio for each four-digit SIC industry, (6) establishments-company ratio for each four-digit SIC industry, (7) cost of materials for establishments,

(8) total beginning inventories for establishments, and (9) the specialization ratio.

In applying 3SLS to the share system, the adding-up constraints (7) and (8) imply that the disturbance terms across equations sum to zero. Any one of the three equations is a linear combination of the other two, so the covariance matrix of the disturbances has a rank of only two. Since the full 3×3 matrix is singular, one equation of (6) must be eliminated and the 3SLS procedure applied to the remaining two equations.

In principle any pair can be picked because Berndt and Christensen (1974) indicate that iterative 3SLS estimates of (6) will converge regardless of which couplet is chosen.

The 3SLS procedure can test and impose cross equation restrictions on the parameters of (6). The most important additional restrictions worth testing are the symmetry conditions. Profit-maximization in competitive markets requires that $a_{12} = a_{21}$, $a_{13} = a_{31}$, and $a_{23} = a_{32}$. Unrestricted estimates of the parameters will not in general satisfy these equalities. Symmetry simultaneously tests three important hypotheses (existence of the translog, constant returns to scale, and profit-maximizing behavior by entrepreneurs). Thus the test of $a_{ij} = a_{ji}$ is critical. 3SLS estimation of the equation system (18) derived from the cost function is also approached in a similar way.

Estimates of the Translog Production Function

Table 3.4 summarizes the first phase of our sequence of estimates. Estimates of the translog function (1) using the OLS method are reported on line 1 for quartile 1 and on line 2 for quartile 4. The first set of estimates refer to the low-productivity group of establishments (Q_1/Q_2), and the second set refer to the high-productivity group (Q_4/Q_3). The function estimated relates differences in input levels to differences in output and is written as follows:

$$Y = a_o + \sum_{i=1}^{9} a_i X_i, \tag{1'}$$

where

$Y = \log VA_1 - \log VA_2$ (subscripts refer to quartiles 1 and 2),

$X_1 = (\log L_1)_1 - (\log L_1)_2,$

$X_2 = (\log L_2)_1 - (\log L_2)_2,$

$X_3 = \log K_1 - \log K_2,$

$X_4 = (\log L_1)^2{}_1 - (\log L_1)^2{}_2,$

Table 3.4 **OLS and 2SLS Regression Estimates of Translog Production Function**

	Constant	ln$L1$	ln$L2$	ln$K$	ln$L1^2$	ln$L1 \times L2$	ln$L1 \times K$	ln$L2^2$	ln$L2 \times K$	lnK^2	Dep. Var.	R^2
$Q1/Q2$ (OLS)	−370.77	44.69	60.68	2.77	27.57	−84.65	96.14	37.67	−51.38	−23.74	lnVA	0.806
(t-stat.)	(−0.896)	(0.718)	(0.995)	(0.043)	(−0.685)	(−1.508)	(1.659)	(0.875)	(−0.921)	(−0.543)		
$Q4/Q3$ (OLS)	−66.57	−21.91	24.05	18.23	30.75	−22.54	−16.23	−29.39	57.33	−29.59	lnVA	0.907
(t-stat.)	(−0.236)	(−0.472)	(0.558)	(0.405)	(0.929)	(−0.508)	(−0.412)	(−1.038)	(1.427)	(−0.991)		
$Q1/Q2$ (2SLS)	−10932.0	827.62	800.11	1538.0	207.28	481.62	−892.33	−567.39	−148.19	−248.76	lnVA	−0.627
(t-stat.)	(−0.441)	(0.337)	(0.283)	(0.258)	(−0.144)	(0.058)	(−0.105)	(−0.117)	(−0.081)	(−0.108)		
$Q4/Q3$ (2SLS)	−2957.4	1601.0	287.83	−1032.9	−2783.2	2911.9	1058.7	−745.86	−1705.8	837.70	lnVA	−11.095
(t-stat.)	(−0.017)	(0.077)	(0.056)	(−0.031)	(−0.184)	(0.114)	(0.149)	(−0.069)	(−0.198)	(0.053)		

$$X_5 = (\log L_2)^2{}_1 - (\log L_2)^2{}_2,$$
$$X_6 = (\log K)^2{}_1 - (\log K)^2{}_2,$$
$$X_7 = (\log L_1 \log L_2)_1 - (\log L_1 \log L_2)_2,$$
$$X_8 = (\log L_1 \log K)_1 - (\log L_1 \log K)_2,$$
$$X_9 = (\log L_2 \log K)_1 - (\log L_2 \log K)_2.$$

Since we assume that the production functions for quartile 1 and quartile 2 are the same, there is no problem of identifying the various parameters of the structure. Input and value-added data for quartile 4 and 3 establishments are combined in the same manner to form ratio variables and to provide estimates of parameters for the high-productivity group. The OLS estimates for quartile 1 (line 1 of table 3.4) indicate a good fit to the data in terms of R^2 but the t statistics indicate that none of the nine input variables is significant at the 95% level ($|t| > 1.96$). This result indicates that the input variables might be highly collinear. Quartile 4 estimates have slightly higher R^2 than was the case for quartile 1 but, again, no $|t|$ value exceeds 1.96.

The 2SLS estimates (lines 3 and 4) are also insignificantly different from zero despite a great increase in their average value over the OLS case. The 2SLS fit is even worse than the OLS case, but this is a characteristic of the 2SLS approach.[26]

The Translog Production Function Share Equations

Since direct estimation of the coefficients of the production function gave poor results, we did not use them to compute estimates of the AES. Instead we turned to estimation of the share equations (6). System (6) is derived from (1) by assuming profit-maximizing behavior by establishments. Three symmetric constraints on the share equation parameters are required:

$$a_{12} = a_{21},\ a_{13} = a_{31}, \text{ and } a_{23} = a_{32}. \tag{23}$$

Within the framework of 3SLS estimation, we can test (23). Because the share equations sum to unity, only two of the three share equations are independent. The capital share equation has the worst OLS fit in both quartile 1 and 4, so we chose to drop this equation.

Working with two share equations, only the symmetry restriction $a_{12} = a_{21}$ can be tested. Using 3SLS restricted estimation, we find that the hypothesis $a_{21} = a_{12}$ cannot be rejected at the 95% confidence level for either quartile 1 or 4. In addition, restrictions (7), (8), and (23)

26. The negative R^2s for the 2SLS cases are not reason for alarm because the formula for computed R^2 corrected for degrees of freedom can be highly negative when the true R^2 is close to zero.

imply that the sum of the three input coefficients in each share equation should sum to zero (row homogeneity): $a_{11} + a_{12} + a_{13} = 0$ and $a_{21} + a_{22} + a_{23} = 0$. Although this zero sum does not hold exactly for the unrestricted estimates, the actual coefficient sum does not depart significantly from zero at the 95% level of confidence.

Finally, imposing symmetry, we test to see if the production coefficients of quartile 1 are significantly different from those of quartile 4 at the 95% level of confidence. Estimates appear in table 3.5. The result is that we cannot reject the hypothesis that the production function coefficients are the same in both quartiles. In particular, although the fourth quartile intercepts appear different from the intercepts in the first quartile, the standard errors are so large that the hypothesis of structural equality cannot be rejected at the 95% level.

Computation of AES by Industry

We used equation (10) to compute the Allen partial elasticity of substitution S, for each of the 195 industries in quartile 1 and quartile 4. The calculation of individual elasticity estimates permits the examination of production behavior at each individual industry observation. Because industries may not have the same production structure, these calculations are a check on the appropriateness of the model.

Tables 3.6 and 3.7 present the own AES estimates for quartiles 1 and 4 based on coefficient estimates obtained using 3SLS with row homogeneity and cross equation symmetry restrictions imposed. We

Table 3.5 3SLS (with restrictions) Regression Estimates of Translog Two-Input Equation System

Equation	Constant	ln$L2$	ln$K$	ln$L1$	Dep. Var.
Quartile 1 (3 SLS with restrictions)					
Share 1	13.586	1.835	−14.675	12.841	Share L1
(t-stat.)	(0.902)	(0.417)	(−5.145)	(2.396)	
Share 2	43.207	−14.675	14.161	(0.515)	Share L2
(t-stat.)	(4.791)	(−5.145)	(4.555)	(0.209)	
Quartile 4 (3SLS with restrictions)					
Share 1	39.146	10.856	−3.700	−7.156	Share L1
(t-stat.)	(11.173)	(9.882)	(−4.531)	(−6.686)	
Share 2	38.198	−3.700	9.244	−5.544	Share L2
(t-stat.)	(15.482)	(—4.531)	(12.450)	(−7.726)	

NOTE: Each two-equation system is restricted for row homogeneity (the sum of independent variable coefficients is equal to zero) and cross equation symmetry.

Table 3.6 **Frequency Distribution of Allen Elasticity of Substitution (AES) across 195 Industries (based on 3SLS estimation with restrictions)**

AES Intervals		S_{11}	Quartile 1	S_{22}	Quartile 1	S_{33}	Quartile 1
	10.0+	6	(6)	52	(52)	8	(8)
5.01	10.00	4	(4)	5	(5)	1	(1)
0.01	− 5.00	42	(42)	1	(1)	3	(3)
− 0.99	− 0.0	13	(13)	0		2	(2)
− 1.99	− 1.0	47		0		28	(28)
− 2.99	− 2.0	45		3	(1)	52	(15)
− 3.99	− 3.0	14		4		39	(4)
− 4.99	− 4.0	5		7	(2)	20	(4)
− 5.99	− 5.0	4		5		9	
− 6.99	− 6.0	1		9	(1)	9	
− 7.99	− 7.0	1		6		1	
− 8.99	− 8.0	3		9		7	
− 9.99	− 9.0	1		8		3	
−19.99	−10.0	5		39	(2)	6	
	−20.0 or less	4		46		7	
Approx. Median AES		−1.68		−8.50		−3.09	

Note: Bracketed values represent the number of industries where calculated values for the bordered Hessians are positive and which are unacceptable for the specification of the model we have imposed. Unfortunately, we have no way of judging whether the interval around these estimates may contain negative values as well. S_{11}=production workers; S_{22}=nonproduction workers; S_{33}=capital.

note that the resulting own AES based on these coefficients vary over a wide range. The conditions of the model require that all own AES be negative but, for example, in quartile 1, 58 of the 195 own AES representing nonproduction workers, S_{22}, were positive. The same applies for quartile 4 in which there are 64 positive estimates associated with S_{22}. The remaining AES estimates also have numerous positive values. Taking these estimates on their face value, the evidence is overwhelming that the conditions for existence of the model we have imposed are not met in each industry.[27]

But the *median* values for all own AES estimates in both quartiles are negative and hence acceptable to the model specification. The bottom row of tables 3.6 and 3.7 indicates that median values for S_{11} and S_{22} are larger in quartile 4, while the S_{33} median is larger in quartile 1, suggesting greater substitution possibilities for labor in high-productivity establishments.

27. In this context we are speaking in the general sense of the suitability of our model of equilibrium behavior and the translog specification, or both.

Table 3.7 Frequency Distribution of Allen Elasticity of Substitution (AES) across 195 Industries (based on 3SLS estimation with restrictions)

AES Intervals		S_{11}	Quartile 4	S_{22}	Quartile 4	S_{33}	Quartile 4
	10.0+	20	(12)	58	(45)	1	(1)
5.01	10.00	1	(1)	5	(5)	0	
0.01	5.00	1	(1)	1	(1)	9	
− 0.99	0.0	1	(1)	0		49	(49)
− 1.99	− 1.0	2	(2)	1	(1)	115	(9)
− 2.99	− 2.0	5	(3)	0		14	(1)
− 3.99	− 3.0	7	(3)	1	(1)	5	(1)
− 4.99	− 4.0	11	(4)	3	(2)	0	
− 5.99	5.0	9	(4)	2	(1)	1	
− 6.99	6.0	20	(10)	2		0	
− 7.99	7.0	18	(2)	9		0	
− 8.99	8.0	18	(4)	7	(1)	0	
− 9.99	9.0	10	(1)	3		0	
−19.99	−10.0	46	(3)	48	(2)	1	
	−20.0 or less	27	(9)	60	(2)	0	
Approx. Median AES		−8.14		−12.19		−1.35	

NOTE: Bracketed values represent the number of industries where calculated values for the bordered Hessians are positive and which are unacceptable for the specification of the model we have imposed. Unfortunately, we have no way of judging whether the interval around these estimates may contain negative values as well. S_{11}=production workers; S_{22}=nonproduction workers; S_{33}=capital.

Cross-elasticity of substitution estimates based on the Allen formulation were also calculated, and are given in tables 3.8 and 3.9. In quartile 1 positive values occur for the median estimates of S_{12} and S_{23}, and one cross-elasticity term (S_{13}) is negative. In quartile 4, based on median cross-elasticity estimates, all pairs of inputs exhibit positive cross-elasticity effects. Positive effects imply that inputs are substitutes and negative effects imply that inputs are complements. Again, the evidence suggested by central measures is that the two quartile groups are different.

This completes the AES analysis, and we now turn to a study of the own and cross price elasticity of demand for production inputs.

Own Price Elasticities of Demand

The own AES estimates in tables 3.6 and 3.7 lead directly to the computation of own price elasticities (OPE) of demand: $\mathrm{OPE}_{ii} = M_i\mathrm{AES}_{ii}$. The distribution of OPE estimates for all inputs (not shown) is less dispersed than that of the AES estimates, the dispersion being compressed by the share weighting factor. The median OPE estimates

Table 3.8 **Frequency Distribution of Allen Elasticity of Substitution (AES) across 195 Industries (based on 3SLS estimation with restrictions)**

AES Intervals		S_{12}	Quartile 1	S_{13}	Quartile 1	S_{23}	Quartile 1
	10.0+	30		6	(6)	16	
5.01	10.00	40		3	(3)	19	
0.01	5.00	65	(5)	56	(56)	107	(12)
− 0.99	− 0.0	6	(6)	87		15	(15)
− 1.99	− 1.0	6	(6)	18		7	(7)
− 2.99	− 2.0	6	(6)	7		8	(8)
− 3.99	− 3.0	6	(6)	4		3	(3)
− 4.99	− 4.0	3	(3)	1		2	(2)
− 5.99	− 5.0	3	(3)	2		3	(3)
− 6.99	− 6.0	3	(3)	1		3	(3)
− 7.99	− 7.0	2	(2)	1		1	(1)
− 8.99	− 8.0	3	(3)	0		2	(2)
− 9.99	− 9.0	0		1		0	
−19.99	−10.0	12	(12)	4		4	(4)
	−20 or less	10	(10)	4		5	(5)
Approx. Median AES		2.89		−0.37		2.08	

NOTE: Bracketed values represent the number of industries where calculated values for the bordered Hessians are positive and which are unacceptable for the specification of the model we have imposed. Unfortunately, we have no way of judging whether the interval around these estimates may contain negative values as well. S_{11}=production workers; S_{22}=nonproduction workers; S_{33}=capital.

shown in table 3.10 are mostly larger in quartile 4 than in quartile 1. They differ substantially in the case of nonproduction labor and mildly in the case of capital. In the case of production labor (E_{11}) they are reasonably concentrated about the median of —0.94 for quartile 1, while they are somewhat more dispersed about the median of —1.15 for quartile 4. The median value for nonproduction labor (E_{22}) is —1.23 for quartile 1 and —2.00 for quartile 4. The median OPE for capital is —0.58 for quartile 1 and —0.63 for quartile 4.

Cross Price Elasticities of Demand

The cross price elasticity of demand estimates are computed for pairs of inputs by a generalization of the OPE formula, and are denoted by CPE:

$$\text{CPE}_{ij} = M_i \text{AES}_{ij}. \tag{24}$$

We mention only a few of the size comparisons for median values of quartiles among factors, since OPE will take on the sign of the respective AES. Median CPE estimates for quartile 4 are all positive. All

Table 3.9 **Frequency Distribution of Allen Elasticity of Substitution (AES) across 195 Industries (based on 3SLS estimation with restrictions)**

AES Intervals		S_{12}	Quartile 4	S_{13}	Quartile 4	S_{23}	Quartile 4
	10.0+	42		6	(4)	9	(1)
5.01	10.0	40		9	(6)	20	(6)
0.01	5.0	52		156	(42)	129	(27)
− 0.99	0.0	1	(1)	7	(2)	19	(17)
− 1.99	− 1.0	2	(2)	6	(3)	3	(1)
− 2.99	− 2.0	2	(2)	1	(1)	1	(1)
− 3.99	− 3.0	5	(5)	4	(3)	3	(3)
− 4.99	− 4.0	1	(1)	1		1	(1)
− 5.99	− 5.0	6	(6)			0	
− 6.99	− 6.0	4	(4)			3	(1)
− 7.99	− 7.0	3	(3)			1	
− 8.99	− 8.0	5	(5)			1	
− 9.99	− 9.0	3	(3)			3	(1)
−19.99	−10.0	8	(8)	2		2	(2)
	−20.0 or less	21	(21)	3			
Approx. Median AES		+3.51		+2.35		+2.35	

NOTE: Bracketed values represent the number of industries where calculated values for the bordered Hessians are positive and which are unacceptable for the specification of the model we have imposed. Unfortunately, we have no way of judging whether the interval around these estimates may contain negative values as well. S_{11}=production workers; S_{22}=nonproduction workers; S_{33}=capital.

Table 3.10 **Median Elasticity of Substitution of 195 Industries (based on 3SLS estimation with restrictions)**

Elasticity Measure	Quartile 1	Quartile 4
Own price elasticity of substitution		
Production workers (E_{11})	−0.94	−1.15
Nonproduction workers (E_{22})	−1.23	−2.00
Capital (E_{33})	−0.58	−0.63
Cross price elasticity of substitution		
Production workers–nonproduction workers (E_{12})	1.87	0.98
Production workers–capital (E_{13})	−0.34	0.49
Nonproduction workers–capital (E_{23})	0.40	0.39

inputs are substitutes. In quartile 1 the median CPE for production workers and capital (E_{13}) is negative at —0.34. Thus, unlike quartile 4, capital and production workers are complements. The median CPE for production workers and nonproduction workers is positive at 1.87, indicating that they are substitutes. The median estimate of the CPE for nonproduction workers and capital is 0.40 in quartile 1.

In the next section we report the estimates from the analysis of a translog average cost function which uses price variables as inputs.

The Share Equations of a Translog Cost Function

In this part of the study we estimate the parameters of a translog cost function due to Christensen, Jorgensen, and Lau (1973), under the assumption of constant returns to scale. As demonstrated in (12)–(18), the estimating forms of the share equations also appear in the logs of the variables, but, unlike the profit-maximizing model, with the production function as the starting point, they are functions of input prices and not functions of input levels.

We separately tested for cross equation symmetry and linear homogeneity and found that the test restrictions were not rejected at the 95% level. Symmetry and homogeneity restrictions were then imposed and the system reestimated to provide estimators of Allen partial elasticities of substitution (AES). Table 3.11 summarizes the results of the regression estimation. The coefficient estimates are in general poor.

We computed the Allen partial elasticities anyway and evaluated them at the mean level of a quartile class according to formula (22):

$$S_{ij} = \frac{B_{ij}}{M_i M_j} + 1,$$

when

$$i = j, S_{ii} = \frac{B_{ii}}{M_i^2} + 1 - \frac{1}{M_i}.$$

Our results on the own and cross AES estimates are reported in table 3.12. Since all cross elasticity estimates are positive, no significant complementarity between inputs is indicated. In addition, all own-elasticity measures have the appropriate negative sign.

In table 3.12, the own AES estimates differ considerably between the two quartiles. Estimates corresponding to S_{11} and S_{22} in quartile 4 are 6 times and 2 times larger than their first-quartile counterparts. The value for S_{33} in quartile 1, however, is 6 times greater than in quartile 4. Table 3.13 gives the corresponding estimates of the OPE for the three inputs, and for comparison we repeat the results of OPE estimates obtained via the production function route in table 3.14. The estimates from both specifications are remarkably close. The clear pattern that emerges, except for E_{33}, is that quartile 4 effects are more elastic than quartile 1. This completes the analysis of translog specifications of technology differences.

Table 3.11 **3SLS (with restrictions) Regression Estimates of Translog Two–Input Price Shares Equation System**

	Constant	ln W1	ln W2	ln VK	Dep. Var.
Quartile 1 (3 SLS with restrictions)					
Share 1	48.555	−11.006	11.722	−0.716	Share L1
(t–stat.)	(8.863)	(−1.307)	(1.386)	(−4.117)	
Share 2	25.261	11.722	−11.608	−0.113	Share L2
(t–stat.)	(5.287)	(1.386)	(−1.366)	(−1.108)	
Quartile 4 (3SLS with restrictions)					
Share 1	36.823	−12.140	2.826	9.314	Share L1
(t–stat.)	(4.123)	(−2.525)	(0.575)	(2.115)	
Share 2	28.501	2.826	−9.255	6.429	Share L2
(t–stat.)	(3.540)	(0.575)	(−1.441)	(1.880)	

NOTE: Each two-equation system is restricted for row homogeneity (the sum of independent variables coefficients is equal to zero) and cross equation symmetry.

Table 3.12 **Allen Partial Elasticities of Input Substitution across 195 Industries (based on 3SLS estimation with restrictions)**

Quartile	Own	Cross
1	$S_{11} = -\ 0.79$	$S_{12} =\ 1.82$
	$S_{22} = -\ 5.99$	$S_{13} =\ 0.91$
	$S_{33} = -\ 6.31$	$S_{23} =\ 0.96$
4	$S_{11} = -\ 6.214$	$S_{12} =\ 1.99$
	$S_{22} = -11.70$	$S_{13} =\ 1.66$
	$S_{33} = -\ 0.03$	$S_{23} =\ 1.74$

NOTE: OPE estimates obtained from two-input price shares equation system.

Within the cost function framework, an attempt was made to see if economies of scale could be a possible explanation for differences in productivity between the two quartiles. In carrying out this test we added a proxy variable for scale, the log of man-hours of production workers, to the share equations.[28] Using 3SLS with all restrictions im-

28. This new share equation system would result if we postulated a translog average cost function with nonconstant returns to scale $[C/Q = f(p_1, p_2, p_3, Q)]$, and derived its share equations as in (12)–(18).

Table 3.13 **Own Price Elasticity of Demand for Inputs across 195 Industries (based on 3SLS estimation with restrictions): Cost Function Estimates**

Quartile 1	$E_{11}=-0.51$	$E_{22}=-1.31$	$E_{33}=-0.81$
Quartile 4	$E_{11}=-1.35$	$E_{22}=-1.56$	$E_{33}=-0.02$

NOTE: OPE estimates obtained from two-input price shares equation system.

Table 3.14 **Own Price Elasticity of Demand for Inputs across 195 Industries (based on 3SLS estimation with restrictions): Production Function Estimates**

Quartile 1	$E_{22}=-1.23$	$E_{11}=$ 0.94	$E_{33}=-0.58$
Quartile 4	$E_{22}=-2.00$	$E_{11}=$ 1.15	$E_{33}=-0.63$

NOTE: OPE estimates obtained from two-input price shares equation system.

posed, none of the coefficients of the scale variable were significant, suggesting that scale does not explain differences in labor shares and differences in labor productivity among quartiles or industries.

3.6.3 Monopoly and Growth Considerations: A Further Single-Equation Experiment

The 3SLS estimates of the three translog share equations (6) indicate that none of the corresponding coefficients differ significantly (at the 95% level) between quartiles 1 and 4. However, in general, our estimates are not very precise, and we had to appeal to average measures in many cases for the experiment to make economic sense. The theoretical specification appears to be too rich for the data we have on our hands. We speculate, therefore, that quartile 1 data, especially, may contain much "noise" and are not explainable by a static production model. The divergence of coefficients may indicate that corresponding parameters of the translog production function really differ between the quartiles, or perhaps it may indicate that the input shares of establishments in the two quartiles did not arise from long-run equilibrium conditions in 1967. This latter possibility is worth investigating because, due to the ranking of plants by their productivity, establishments may appear in the top quartile not only because they have normally high productivity but also, as mentioned previously, because they may be the beneficiary of favorable economic events which have added a positive transitory component to their value added per production worker man-hour. Similarly, bottom-quartile plants may, on the average, have some negative transitory components in their 1967 productivity. A positive component in the value-added productivity of quartile 4 plants will lift their capital share above, and reduce their labor shares below, the long-run levels. Conversely, a negative transitory element in quartile 1 establishments will depress their capital share below, and push their labor shares above, true equilibrium amounts.

The disequilibrium hypothesis cannot be checked directly because a longitudinal sample of plant data is unavailable to us. Instead, as proxies for quartile disequilibrium, we need unpublished Bureau of Labor Statistics data on the past (1958–67) trend of growth rates in industry shipments, shipments per man-hour (productivity), and shipment prices. In this instance output was defined as value of shipments deflated for price changes, and productivity was defined as deflated shipments per man-hour of production workers. In order to discover which of the three growth-rate variables was the most important indicator of disequilibrium we added all three to our previous multiple regression equation intended to explain interquartile productivity differentials. The equation is reproduced in table 3.15 along with the effects of adding the growth-rate variables. Table 3.15 indicates that the past rate of productivity increase in an industry is significantly related to its productivity differentials, not only between quartiles 4 and 3 but also between quartiles 3 and 1. This suggests that a group of leading plants experience productivity surges that tend to outstrip the industry average and to drag up the average as well. And, in addition, the accelerating average leaves the low-productivity establishments even further behind. The strong effect of past productivity advance upon productivity differentials suggests that, at some point, the plants that fall into either quartiles 1 or 4 are out of equilibrium, the former group being below their long-run level of productivity and the latter set being above it.[29]

The past rate of price increases in an industry also has a significantly positive influence on interquartile productivity differences. Perhaps this is due to unequal shifts in the demand for specific plants' products that allow these establishments to increase prices by more than their competitors. Or price increases may be due to monopoly power, in which case they would be associated with productivity differentials flowing from the same source. We will discuss this possibility shortly.

The third growth-rate variable, that of shipments, has a significantly negative effect on productivity differentials. More rapid expansion of demand for an industry's products may allow productivity laggards to catch up somewhat with their higher productivity competitors, perhaps because of a relatively faster expansion of sales which lifts their capacity utilization and their labor productivity.

All three growth-rate variables have an influence on productivity differentials between quartiles 1 and 4, and these differentials in turn

29. This conclusion is consistent with the erosion of plant productivity differentials through time noticed in seven of eight industries studied by Klotz (1966).

Table 3.15 **Multiple Regression Analysis of Interquartile Productivity Differentials**

Productivity Differential	Constant	$\%\Delta(K/H)$	$\%\Delta(N/L)$	$\%\Delta(W)$	$\%\Delta(H)$	ΔS	$\%\Delta P$	$\%\Delta$VS	$\%\Delta$(VS/H)	log(P/C)	log(Con.)	R^2
Quartile 4 to	.559	.158	.066	−.034	.029	−.015						.080
quartile 3	(3.6)*	(3.6)	(1.5)	(−1.0)	(1.1)	(−0.1)						
Percentage	.105	.154	.079	−.032	−.007	−.013	.045	−.009	.048	−.085	.046	.170
difference	(0.5)	(3.7)	(1.8)	(−1.0)	(−0.2)	(−0.1)	(3.3)	(−1.9)	(4.2)	(−2.0)	(1.7)	
Quartile 3 to	.388	.061	.067	.130	.078	.261						.028
quartile 1	(1.9)	(1.1)	(0.9)	(1.1)	(1.7)	(1.5)						
Percentage	.179	.042	.134	.108	.059	.157	.047	−.021	.070	.049	.067	.107
difference	(0.6)	(0.8)	(1.8)	(0.9)	(1.3)	(0.9)	(1.9)	(−2.6)	(3.5)	(0.7)	(1.4)	

*t values based on 195 observations: $t = 1.65$ (90% confidence); $t = 1.97$ (95% confidence)

Variables:
- K/H = capital per production worker man-hour
- N/L = nonproduction workers per production worker
- W = hourly wages of production workers
- H = production workers man-hours
- S = product specialization ratio
- P = price of shipments
- VS = value of shipments
- VS/H = shipments per production worker man-hour
- E/C = plants per company
- Con. = Concentration ratio (percent of shipments in four largest companies)

are associated with disparities in input shares between the quartiles. Thus, the growth-rate variables should be incorporated into the translog share equations in some manner. But the differential impact of the variables suggests that they may have a multiplicative, rather than an additive, effect on input shares. In this instance, each input variable in the share equations should be multiplied by some correction factor, which would be a weighted average of the past growth rates of industry output, productivity, and prices. The problem with this approach is that the weights are unknown.

In addition to disequilibrium elements, monopoly power may cause differences in the estimated coefficients of comparable share equations between quartiles 1 and 4. If quartile 4 establishments tend to have less elastic product demand than their quartile 1 competitors, then the capital share (measured as a residual in this study) will be larger in quartile 4, even if both groups of plants use the same technology and factor proportions. Conversely, because factor shares add to unity, the labor share of the quartile 4 group of plants will be less than in quartile 1.

Although monopoly affects the capital and labor share equations, we cannot incorporate it directly into the share estimates because the elasticity of product demand is unknown. However, two proxies for monopoly power were chosen for analysis: the industry concentration ratio (the fraction of industry shipments accounted for by the four largest companies) and the intensity of multiplant companies (the ratio of establishments to firms in the industry). Both should be positively related to the degree of monopoly in an industry, and the greater this degree the greater the chance that productivity differentials could occur. Table 3.15 indicates that the concentration ratio was significantly related (with 90% confidence) to the magnitude of the productivity differential between quartiles 4 and 3, but the ratio was less successful in explaining the quartile 3–quartile 1 discrepancy. This result suggests that high-productivity plants may belong to firms with market power while their low-productivity competitors have little market impact and may tend to act more like pure competitors: quartile 4 plants may belong to companies who are price makers, while quartile 1 establishments may tend to be owned by firms who are price takers. This explanation is consistent with the previous finding that, *ceteris paribus*, productivity differentials are wider in industries with larger past rates of price advance. The multiplant variable for monopoly power did not perform as well as the concentration ratio, being insignificant in the quartile 3–quartile 1 comparison and having a negative influence on the quartile 4–quartile 3 difference.

Summarizing the multiple regression results of table 3.15, we note that, although R^2s were low in all cases, the addition of both equilibrium and monopoly variables doubled the goodness of fit of the top-quartile equation and quadrupled that of the bottom quartile. These added variables did not appreciably alter the coefficients of the factor proportion variables (capital per man-hour and nonproduction workers per production worker) in the top-quartile equation, and they made these coefficients more significant statistically. On the other hand, the added variables decreased the coefficient of capital per man-hour, while increasing that of nonproduction workers per production worker, in the bottom-quartile equation; the t statistics moved accordingly. In addition, the returns to scale proxy variable (production worker man-hours) becomes insignificant in both quartile equations with the addition of monopoly and disequilibrium variables. This behavior seems to suggest that a production function explanation of the top-quartile productivity differential is more reliable than a similar explanation of the bottom-quartile difference. Our results also indicate that the incorporation of disequilibrium and monopoly elements into the translog share equations might move the estimated coefficients of comparable share equations, between high and low-productivity plants, closer together. Supposing the parameters of comparable share equations to be the same, the only technical difference among plants would then occur in the intercept terms a_0 of their translog production function (1). This term, which is not estimated by the share equations, would be an index of technical ability rather than allocative wisdom. In this case the three major sources of interquartile productivity dispersion would be differences in pure technical efficiency, transitory disturbances in establishment productivity, and monopoly power.

3.7 Summary and Conclusions

Estimates from a theoretical formulation based on the translog production function and multiple regression analyses both indicate that factor proportions, represented by capital per man-hour and nonproduction workers per production worker, contribute toward an explanation of high productivity in manufacturing establishments. But these factors are less successful in explaining the level of low-productivity plants. Monopoly power also seems to be more important in explaining high, as opposed to low, productivity.

In addition to factor proportions and monopoly power effects, both high and low-productivity establishments in 1967 appear to be out of equilibrium. Their outputs, and possibly their inputs, seem to contain significant transitory elements that depend on the past growth rates of industry output, productivity, and prices. These elements appear to be strong enough to cast doubt on any static formulation of productivity differences.

In all of our regression experiments with interquartile productivity differences, unexplained factors buried in the residual were most noticeable. The combined effect of factor proportions and monopoly power explained only 17% of the quartile 4–quartile 3 productivity variation, and only 11% of the quartile 3–quartile 1 differential, across 195 industries. Differences in managerial quality and in product (at the five and seven-digit level of disaggregation) may be responsible for much of the residual variance. On the other hand, the low R^2s of the productivity equations may have been due to our poor proxies for measuring disequilibrium effects.

Klotz (1966) found that industries differ in the extent to which their high and low-productivity plants move toward the industry mean through time. This regression can be due to a competitive tendency to equalize their factor proportions, plus the attrition of their initial transitory components in output and inputs. These two effects can only be isolated and measured by tracing specific groups of high and low-productivity establishments in an industry over a period of years, while relating their differential productivity growth to their initial productivity level and the changes in their factor proportions. Ideally, for this undertaking, the analyst needs a longitudinal data set on each industry in which annual production statistics on specific groups of plants are recorded for a number of years. Such a data set would permit investigation of the dynamics of plant productivity growth, and knowledge of the dynamics of the situation would allow separation of the long-run causes of establishment productivity differentials from the transitory disturbances. The long-run causes are of most interest because the transitory forces are probably random and uncontrollable. We conclude, therefore, that due to the power of short-run disturbances in plant productivity, cross-section data for one year, such as those analyzed in this study, are of limited use for analyzing establishment differentials.

Since the preparation of data at the plant level is an expensive operation and census studies are years apart, we conclude with a few remarks as to how the usefulness of cross-section data can be improved. First, to analyze interquartile productivity differentials, the statistical theory of ranking bias (Harman and Burstein 1974) requires that plants be ranked *not by a productivity measure but by a variable that is a prime cause of productivity*. Such a ranking would reduce the difficulty of measuring transitory forces affecting particular groups. The best candidate for a causal variable may be capital per production worker man-hour. We therefore suggest that *establishments be ranked by their capital per production worker man-hour in any future tabulations designed to analyze productivity differences*. In addition, according to the ranking theory in Harman and Burstein (1974), the best ordering of plants is by a variable most strongly related to long-run productivity but not correlated with the transitory component. This variable might be a measure of plant productivity predicted from an equation estimated by regressing actual labor productivity against a number of causal variables at the individual establishment level. When this is done, the regression can be provided the analyst, along with the quartile or decile groupings of the plant data, without violating Bureau of Census rules about disclosure of individual establishment information.

Second, most empirical production-function forms suggest a double-log relation between productivity and its causal variables. This implies that *geometric as well as arithmetic averages of the data of individual plants comprising the quartile should be reported.* The arithmetic averages now derivable from census tabulations do not allow rigorous testing of production function relations which require geometric averages.

Much of the unexplained variation in quartile productivity might be due to differences in product specialization at the five-digit level, and managerial and other quality differences in inputs. Therefore, we suggest, thirdly, that *information on five-digit product specialization be included in future compilations of plant data.* Qualitative factors might be represented by the size or other attributes of the parent company as well as the work force in establishments. It is easy to provide identification codes that describe specific economic attributes of companies along with the plant information, and these might be the key to uncovering how differences arise.

Appendix A

Table 3.A.1 **Value Added per Man-Hour** (in dollars)

SIC	Rank	Ind. Mean ($=X$)	Quartile Means 1	Quartile Means 2	Quartile Means 3	Quartile Means 4	Industry Dispersion: Range (Q4–Q1) ($=s_1$)	Industry Dispersion: Range (Q3–Q1)/2 ($=s_2$)	Coefficients of Variation: S1/X	Coefficients of Variation: S2/X
2731	1	56.76	8.44	34.57	132.70	2143.00	2134.57	62.13	37.61	1.09
2087	2	54.11	7.28	14.10	22.33	109.76	102.48	7.53	1.89	0.14
2095	3	38.38	7.47	16.76	26.55	57.30	49.83	9.54	1.30	0.25
2911	4	31.31	5.18	14.01	25.16	55.12	49.94	9.99	1.59	0.32
2822	5	24.10	7.39	15.03	24.41	38.81	31.42	8.51	1.30	0.35
2085	6	23.16	5.33	13.67	23.78	43.79	38.46	9.22	1.66	0.40
2084	7	23.00	3.11	12.19	22.85	37.14	34.03	9.87	1.48	0.43
3861	8	22.26	4.82	8.19	10.39	32.97	28.15	2.79	1.26	0.13
2082	9	20.12	8.23	13.00	17.70	27.20	18.97	4.73	0.94	0.24
2026	10	18.60	5.24	11.19	18.63	36.03	30.79	6.70	1.66	0.36
3573	11	18.44	4.13	8.10	17.34	35.97	31.84	6.61	1.73	0.36
2851	12	18.06	7.88	12.86	17.20	28.42	20.54	4.66	1.14	0.26
2086	13	17.66	6.37	11.36	16.44	31.99	25.62	5.03	1.45	0.29
2042	14	16.37	2.64	7.66	13.30	32.66	30.02	5.33	1.83	0.33
3241	15	15.47	8.92	12.50	16.83	26.55	17.63	3.95	1.14	0.26
2041	16	14.97	3.56	8.18	12.23	25.87	22.31	4.33	1.49	0.29
2024	17	14.89	5.37	9.92	14.89	29.42	24.06	4.76	1.62	0.32
2647	18	14.36	3.92	8.05	12.61	21.45	17.53	4.35	1.22	0.30
2711	19	14.20	4.34	7.32	11.63	19.86	15.52	3.65	1.09	0.26
3011	20	13.36	5.82	9.66	12.57	18.51	12.70	3.38	0.95	0.25
2951	21	13.13	5.23	11.18	18.31	40.61	35.08	6.39	2.67	0.49

Table 3.A.1 (continued)

SIC	Rank	Ind. Mean ($=X$)	Quartile Means 1	2	3	4	Industry Dispersion: Range (Q4–Q1) ($=s_1$)	Range (Q3–Q1)/2 ($=s_2$)	Coefficients of Variation: S1/X	S2/X
3275	22	12.92	7.21	11.04	14.08	19.29	12.07	3.43	0.93	0.27
2893	23	12.33	5.03	11.11	15.09	23.74	18.71	5.03	1.52	0.41
3662	24	12.12	4.21	7.91	11.57	20.45	16.25	3.68	1.34	0.30
2831	25	11.74	3.21	7.94	13.05	24.40	21.19	4.92	1.81	0.42
3356	26	11.39	0.97	8.33	12.26	19.57	18.60	5.64	1.63	0.50
3537	27	11.20	4.27	6.75	10.16	15.28	11.01	2.94	0.98	0.26
2052	28	11.01	3.73	5.83	7.84	14.03	10.31	2.06	0.94	0.19
3351	29	10.92	5.62	9.07	11.62	18.22	12.60	3.00	1.15	0.27
2051	30	10.90	3.72	6.95	10.10	16.93	13.21	3.19	1.21	0.29
3585	31	10.49	4.65	7.42	10.29	15.83	11.18	2.82	1.06	0.27
3295	32	10.42	4.35	7.44	11.02	18.65	14.30	3.34	1.37	0.32
3519	33	10.39	4.69	7.54	9.35	14.48	9.79	2.33	0.94	0.22
2083	34	10.33	3.71	8.24	11.57	16.96	13.25	3.93	1.28	0.38
3411	35	10.13	5.01	8.61	12.13	23.62	18.61	3.56	1.84	0.35
3522	36	10.01	3.35	5.74	8.13	13.25	9.90	2.39	0.99	0.24
3612	37	9.91	3.87	6.60	9.05	13.09	9.23	2.59	0.93	0.26
2952	38	9.65	5.07	7.82	10.61	22.30	17.23	2.77	1.79	0.29
3843	39	9.62	4.33	7.20	9.94	19.13	14.79	2.80	1.54	0.29
3566	40	9.61	5.31	7.83	9.91	13.46	8.16	2.30	0.85	0.24
2094	41	9.48	2.63	6.33	9.43	20.68	18.05	3.40	1.90	0.36
2621	42	9.45	4.83	7.53	9.79	13.83	8.99	2.48	0.95	0.26
2034	43	9.33	2.03	5.57	8.97	19.31	17.28	3.47	1.85	0.37
3352	44	9.29	3.49	6.44	8.90	14.19	10.70	2.71	1.15	0.29

Table 3.A.1 (continued)

SIC	Rank	Ind. Mean ($=X$)	Quartile Means 1	2	3	4	Industry Dispersion: Range (Q4–Q1) ($=s_1$)	Range (Q3–Q1)/2 ($=s_2$)	Coefficients of Variation: S1/X	S2/X
3443	45	9.27	4.30	6.42	8.32	16.35	12.05	2.01	1.30	0.22
2061	46	9.23	3.66	5.75	9.03	18.16	14.50	2.69	1.57	0.29
3391	47	9.04	5.63	7.67	9.54	14.21	8.58	1.96	0.95	0.22
3452	48	9.03	4.94	7.29	9.82	14.00	9.06	2.44	1.00	0.27
3429	49	8.91	4.16	6.34	8.20	13.30	9.14	2.02	1.03	0.23
2013	50	8.82	3.75	6.67	9.14	15.74	11.99	2.70	1.36	0.31
3423	51	8.82	4.06	6.12	8.03	14.34	10.28	1.99	1.17	0.23
3292	52	8.79	4.58	6.77	9.16	14.51	9.93	2.29	1.13	0.26
2542	53	8.66	3.67	5.47	7.88	13.45	9.78	2.10	1.13	0.24
3582	54	8.58	4.00	7.31	9.29	13.26	9.26	2.65	1.08	0.31
3634	55	8.54	3.49	5.76	8.17	14.83	11.34	2.34	1.33	0.27
3691	56	8.50	4.68	6.96	8.65	13.42	8.73	1.98	1.03	0.23
2654	57	8.41	4.37	6.23	8.40	12.11	7.74	2.01	0.92	0.24
3642	58	8.39	4.34	6.28	8.22	13.55	9.21	1.94	1.10	0.23
3231	59	8.37	3.62	5.54	7.24	13.10	9.48	1.81	1.13	0.22
3621	60	8.34	3.64	5.98	8.03	12.27	8.63	2.19	1.04	0.26
3562	61	8.26	4.94	7.63	9.49	12.01	7.07	2.28	0.86	0.28
2033	62	8.14	2.78	5.19	7.83	15.21	12.42	2.52	1.53	0.31
2433	63	8.04	3.74	5.79	8.15	13.95	10.21	2.21	1.27	0.27
3461	64	7.98	4.18	6.04	7.99	11.21	7.03	1.90	0.88	0.24
2272	65	7.95	2.89	5.41	7.94	15.16	12.27	2.52	1.54	0.32
3255	66	7.91	4.31	6.46	8.23	12.27	7.96	1.96	1.01	0.25
3951	67	7.80	3.00	5.04	6.39	13.73	10.73	1.69	1.38	0.22

Table 3.A.1 (continued)

SIC	Rank	Ind. Mean ($=X$)	Quartile Means 1	2	3	4	Industry Dispersion: Range (Q4–Q1) ($=s_1$)	Range (Q3–Q1)/2 ($=s_2$)	Coefficients of Variation: S1/X	S2/X
3491	68	7.78	4.55	6.44	7.98	10.82	6.27	1.72	0.81	0.22
3272	69	7.73	3.49	5.59	7.96	13.89	10.40	2.24	1.35	0.29
3651	70	7.66	2.40	5.07	6.64	12.29	9.89	2.12	1.29	0.28
3544	71	7.63	4.44	6.44	8.05	11.63	7.20	1.81	0.94	0.24
3715	72	7.56	3.62	5.90	7.37	11.96	8.34	1.87	1.10	0.25
3451	73	7.24	4.33	6.21	7.92	12.16	7.83	1.79	1.08	0.25
3629	74	7.20	3.37	5.56	8.06	11.82	8.45	2.34	1.17	0.33
3742	75	7.06	2.59	6.74	9.68	13.74	11.16	3.55	1.58	0.50
3221	76	7.05	4.93	6.14	7.19	9.53	4.60	1.13	0.65	0.16
3674	77	6.94	4.00	5.73	8.53	11.47	7.47	2.27	1.08	0.33
2396	78	6.90	2.89	4.65	5.93	9.12	6.24	1.52	0.90	0.22
3481	79	6.76	3.61	5.36	7.20	11.62	8.01	1.80	1.18	0.27
2642	80	6.73	4.76	5.96	7.22	9.19	4.43	1.23	0.66	0.18
2022	81	6.67	1.53	4.15	6.37	13.66	12.14	2.42	1.82	0.36
3949	82	6.57	2.78	4.20	6.14	10.90	8.11	1.68	1.24	0.26
3479	83	6.42	3.71	5.11	6.76	12.32	8.61	1.52	1.34	0.24
3321	84	6.37	3.22	4.83	5.99	8.94	5.72	1.38	0.90	0.22
3259	85	6.27	3.42	5.11	6.73	9.77	6.35	1.66	1.01	0.26
3471	86	6.21	3.54	5.06	6.73	10.58	7.04	1.60	1.13	0.26
2121	87	6.20	1.73	4.00	5.43	7.73	5.99	1.85	0.97	0.30
3731	88	6.19	3.69	5.33	6.84	10.24	6.54	1.57	1.06	0.25
3111	89	6.01	2.88	4.87	6.55	11.76	8.88	1.83	1.48	0.31
3253	90	6.01	2.78	4.17	5.66	7.62	4.84	1.44	0.81	0.24

Table 3.A.1 (continued)

			Quartile Means				Industry Dispersion		Coefficients of Variation	
SIC	Rank	Ind. Mean ($=X$)	1	2	3	4	Range (Q4–Q1) ($=s_1$)	Range (Q3–Q1)/2 ($=s_2$)	S1/X	S2/X
3931	91	5.96	3.12	4.48	5.87	10.27	7.15	1.37	1.20	0.23
2431	92	5.95	3.20	5.01	6.81	12.09	8.89	1.81	1.49	0.30
2397	93	5.46	3.55	4.39	5.52	10.25	6.70	0.99	1.23	0.18
2512	94	5.32	3.23	4.36	5.50	8.24	5.01	1.13	0.94	0.21
2141	95	5.08	1.28	3.78	5.38	11.80	10.52	2.05	2.07	0.40
3141	96	4.57	2.67	3.75	4.64	6.81	4.13	0.98	0.90	0.21
3199	97	4.54	2.41	3.71	5.07	8.92	6.51	1.33	1.43	0.29
2852	98	4.49	2.85	3.68	4.59	7.03	4.18	0.87	0.93	0.19
2251	99	4.17	2.38	3.45	4.18	6.62	4.24	0.90	1.02	0.22
2321	100	3.72	1.81	2.62	4.28	12.03	10.22	1.23	2.75	0.33
2381	101	3.52	1.76	2.88	3.33	5.76	4.00	0.79	1.14	0.22
2426	102	3.42	2.02	2.96	3.89	5.82	3.80	0.93	1.11	0.27

Appendix B

Grouping Bias

Because the analyst is forced to work with grouped data, it is natural to wonder if the data accurately reflect relations occurring at the plant level. Under most general conditions, estimates of grouped micro data will cause biased estimates of the micro (i.e., plant) parameters to result. Theil (1971, chap. 11) has shown that the coefficients of linear regression equations using grouped data are weighted averages of the corresponding micro coefficients, but that this bias vanishes if all micro parameters are equal (i.e., all plants in the industry have the same production function parameters), or if the weights and the micro parameters are uncorrelated. Hannan and Burstein (1974), on the other hand, consider the case where the micro parameters are equal, but where the micro observations are ranked and grouped by some criterion, and a regression is performed using each group average as an observation point. In a simulation experiment, for random grouping of plants, the macro coefficient was found to be an unbiased estimate but a very inefficient estimator of the micro coefficient. An unbiased estimator of high efficiency resulted when micro observations were ranked and grouped by the values of the independent variable in the causal equation to be estimated. Conversely, grouping by values of the dependent variable lead to biased estimation.[30] The situation is worse if the micro relation to be estimated is log linear. In this case the grouped data reported should be a geometric mean of the micro data, but in practice arithmetic means are reported and this causes bias, unless the variance of the micro data is uncorrelated with the mean of the data.

Recall that for each of the 412 four-digit manufacturing industries, the census plant data used in this study have been ranked by the plant's productivity in 1967 and the ranking has been grouped into quartiles. Arithmetic sums of the quartile data are reported so that only arithmetic averages of the data pertaining to plants in the quartile could be constructed. If we were to attempt to explain productivity differences by comparing, say, capital-labor differences among the four quartiles of a given industry, then, according to Hannan and Burstein, we would

30. When there are several independent variables, the micro units might be ranked and grouped on the basis of a variable that is highly correlated with the combined effect of all the independent variables. The best such variable seems to be the value of the dependent variable estimated by regressing it against all independent variables, using the micro data. But this micro regression can be computed, its parameter values can be furnished to the analyst directly, obviating the need to use grouped data to estimate the micro parameters indirectly. Supposing the micro regression cannot be run (due to, say, undue cost); then the grouping might best be done on the basis of the most important explanatory variable.

obtain a biased estimate of the influence of this, presumably causal, variable. The ranking leads to an overestimate of the true differences due to the capital-labor ratio because this ratio is correlated with transitory productivity forces. The top (bottom) quartile of plants would appear to experience the greatest positive (negative) disturbance to their productivity since they tend to have the highest (lowest) capital-labor ratio.[31]

It makes sense, therefore, to carry out our investigation in relative comparisons to minimize ranking bias. If we compare productivity and capital-labor ratios in the top or bottom quartiles *across* four-digit industries, though some of this differential contains a positive transitory element, the transitory fraction of a difference can be either large or small (depending on how near the *industry* is to long-run equilibrium in input and product markets) but may be independent of the size of the differential. So, when comparing across industries, there is no special reason for industries with the highest differentials to have the largest transitory fractions.

This framework of comparing quartile productivity differentials across industries differs from a comparison of quartiles *within* industries. The latter matching is suspect because the observation with the largest differential with respect to the average is the top quartile of plants, and it also is likely to contain the greatest transitory disturbance. On the other hand, matching across industries does not force the observation with the largest differential (the industry whose top quartile of establishments is most above the average of its own industry) to have the largest transitory fraction in its differential.

When productivity differentials and transitory fractions of these differentials are uncorrelated across industries, then some theories of productivity behavior can be tested without distortion by ranking bias. For example, we hypothesized that the productivity differential between the top-quartile plants and the average establishments of an industry is positively related to their capital-labor differential. The differential measure we examined between quartile 1 and quartile 4 also contains a transitory element, but, since the fraction is not likely to be related to the differential, the element is probably not proportionately greater in industries which have the largest interquartile differential in their

31. Had the plants been ranked by their capital-labor ratio rather than their productivity, then the Wald-Bartlett method (Kendall and Stuart 1961, p. 404) could have been used to compute the effect of the capital-labor ratio. This method, designed to overcome the effect of measurement errors in the variables, involves ranking the data by the independent variable and joining the midpoints of the top and bottom 30% of the data points by a line whose slope is an estimate of the marginal impact of the independent variable. But this estimate is itself biased if, as is very likely, variables other than the capital-labor ratio influence productivity.

capital-labor ratios. A lack of correlation, therefore, between any transitory productivity element and the capital-labor differential means that a regression of productivity differentials on capital-labor differentials across industries will not lead to biased estimates of the latter's effect.[32]

References

Allen, R. G. D. 1938. *Mathematical analysis for economists.* London: Macmillan.

Arrow, Kenneth. 1972. *The measurement of real value added.* Technical report no. 60, Institute for Mathematical Studies in the Social Sciences, Stanford.

Berndt, Ernst, and Christensen, Laurits. 1973. The translog production function and the substitution of equipment, structures and labor in U.S. manufacturing 1929–68. *Journal of Econometrics* 1:81–114.

———. 1974. Testing for the existence of a consistent aggregate index of labor inputs. *American Economic Review* 64:391–404.

Brown, Murray, ed. 1967. *The theory and empirical analysis of production.* New York: Columbia University Press, for National Bureau of Economic Research.

Crandall, Robert; MacRae, Duncan; and Yap, Lorene. 1975. An econometric model of the low-skill labor market. *Journal of Human Resources* 10:3–24.

Christensen, Laurits; Jorgenson, Dale; and Lau, Laurence. 1973. Transcendental logarithmic production frontiers. *Review of Economics and Statistics* 55:28–45.

Diewert, W. E. 1969. *Canadian labor markets: A neoclassical econometric approach.* Project for the evaluation and optimization of economic growth. Berkeley: University of California Press.

———. 1973. Hicks' aggregation theorem and the existence of a real value added function. Technical report no. 84, Institute for Mathematical studies in the Social Sciences, Stanford.

Gramlich, Edward. 1972. *The demand for different skill classes of labor.* Washington: Office of Economic Opportunity.

Griliches, Zvi, and Ringstad, Vidar. 1971. *Economies of scale and the form of the production function.* Amsterdam: North-Holland.

Hall, Robert. 1973. The specification of technology with several kinds of output. *Journal of Political Economy* 81:750–60.

32. However, the intercept of the regression will be biased upward, because the average transitory element in top-quartiles productivity is positive, but this distortion is unimportant for explaining *variation* in the productivity differential.

Hanoch, G. 1975. The elasticity of scale and the shape of average costs. *American Economic Review* 65:492–97.

Harman, Michael, and Burstein, Leigh. 1974. Estimation from grouped observations. *American Sociological Review* 39:374–92.

Hildebrand, George, and Liu, T. C. 1965. *Manufacturing production functions in the United States*. Ithaca: Cornell University Press.

Johnston, John. 1971. *Econometric methods*. 2d ed. New York: McGraw-Hill.

Jones, Melvin. 1975. A statistical analysis of inter- and intra-industry elasticity of demand for production workers in plants with large and small employment size. WP 982–01. Washington: The Urban Institute.

Jorgenson, Dale. 1972. Investment behavior and the production function. *Bell Journal of Economics and Management Science* 3:220–51.

———. 1974. Investment and production: a review. In David Kendrick, ed., *Frontiers of quantitative economics*, vol. 2. Amsterdam: North-Holland.

Kendall, Maurice, and Stuart, Alan. 1961. *The advanced theory of statistics*, vol. 2. London: Griffin.

Kennedy, Charles, and Thirlwall, A. P. 1972. Technical progress: a survey. *Economic Journal* 82:11–72.

Klotz, Benjamin. 1966. Industry productivity projections: a methodological study. In *Technical papers*, vol. 2. Washington: National Commission on Technology, Automation and Economic Progress.

———. 1970. *Productivity analysis in manufacturing plants*. Bureau of Labor Statistics Staff Paper 3. Washington: Department of Labor.

Krishna, K. L. 1967. Production relations in manufacturing plants: an explanatory study. Ph.D. diss., University of Chicago.

Madoo, R. B. 1975. Production efficiency and scale in U.S. manufacturing: An inter-intra industry analysis. Ph.D. diss., University of California, Berkeley.

Madoo, Rey, and Klotz, Benjamin. 1973. Analysis of differences in value added per manhour among and within U.S. manufacturing industries, 1967. WP 2703–1. Washington: The Urban Institute.

Nadiri, M. I. 1970. Some approaches to the theory and measurement of total factor productivity: a survey. *Journal of Economic Literature* 8:1137–77.

Pratten, C. F. 1971. *Economies of scale in manufacturing industry*. Cambridge University Press.

Salter, W. E. G. 1962. *Productivity and technical change*. Cambridge: Cambridge University Press.

Scherer, F. M. 1970. *Industrial market structure and economic performance*. Chicago: Rand McNally.

———. 1973. The determinants of industry plant sizes in six nations. *Review of economics and statistics* 55: 135–45.
Theil, Henri. 1971. *Principles of econometrics.* New York: Wiley.
Uzawa, Hirofumi. 1962. Constant elasticity of substitution production functions. *Review of Economic Studies* 29:291–99.

Comment   Irving H. Siegel

This paper by Klotz, Madoo, and Hansen provides a wholesome reminder of the importance of the "establishment" as (1) the basic site of productive activity and (2) the source, therefore, of "atomic" information required for productivity (and other) measurement, analysis, and policy at both the micro and macro levels. Such a reminder is in order because so much of the community of quantitative economists is concerned nowadays with "the big picture," with models and aggregates pertaining to the whole economy or to components no smaller than a four-digit industry. In particular, the "postindustrial" evolution of our society has diminished the probability of early or prolonged professional exposure to the mysteries of the Census of Manufactures, which has for much more than a century been identified with the term and concept of "establishment." The census long ago also innovated the term and concept of "value added" for gauging the economic contribution of a manufacturing establishment. This census notion is the prototype of "income originating" in an industry, which is estimated by the Bureau of Economic Analysis from company, rather than establishment, data.

The authors examine closely the relationship of value added per production worker man-hour to other establishment variables for only one year, 1967, and they properly conclude from their efforts that longitudinal studies would yield more satisfying results. The tracking of value-added productivity through time would, for example, permit better assessments of transitory "noise," of the persistence of early productivity dominance, and of the relevance of market power and scale of production than the authors were able to hazard on the basis of only one year's data. At this juncture, we should recall that a promising program of direct productivity reporting by companies was inaugurated by the U.S. Bureau of Labor Statistics shortly after World War II, and that it did not long survive. Mention ought also to be made here of a current venture by the Department of Commerce to encourage companies to set up batteries of continuing productivity measurements for

Irving H. Siegel, economic adviser, Bureau of Domestic Business Development, U.S. Department of Commerce, until July 1979, is now a private consulting economist in Bethesda, Maryland.

key organizational units. This initiative, and the imitation it has inspired, should help improve the data base for longitudinal establishment studies.

Although the authors make additional recommendations concerning the design of future inquiries into value-added productivity, they omit two that merit consideration. One of these is the grouping of the establishments in each four-digit industry into a larger number of categories —into deciles, say, rather than quartiles. A finer-grain classification would permit a more sensitive analysis of interrelationships, especially at the lower end of the productivity spectrum, where heterogeneous "small businesses" tend to be concentrated.

A second needed refinement in subsequent studies is the discrimination of establishments in the same industry, insofar as possible, according to process of manufacture. Unexplained interquartile differences in productivity are surely attributable, in some degree, to differences in technology that are hardly reflected in, say, the dollar values of capital assets.[1] The authors acknowledge, in their remarks on simple correlation coefficients computed from plant data for 102 four-digit industries, "that low-productivity establishments could be using completely different technologies from their top-quartile counterparts within the same industry." Nevertheless, in their recommendations, they are silent on the need for coding of plants by process even though they would welcome information on five-digit product mix.

The patient statistical experiments and exercises of the authors, however admirable, do not encourage belief that more advanced econometric tools have much to add to the hints given by simpler ones, experience, and common sense. In particular, they offer little hope, if any, for the development from census data of reliable production functions for establishments at the various productivity levels. They show that a "causal" analysis of interplant productivity differences cannot be successfully pursued for any distance. Indeed, a summary of their attempts to wring more out of the data than is told in table 3.1—by means of simple correlation, multiple regression, the fitting of transcendental-logarithmic (translog) production functions (ordinary, two-stage, and three-stage least-squares), and the computation of Allen elasticities of substitution—would make an instructive, cautionary introductory chapter for an econometric primer. Any reader of the paper who stays the course not only feels sadder and wiser at the end but is also inclined to congratulate the data for withstanding the torments of advanced technique without confessing what they did not really know and there-

1. Such dollar values should, ideally, be expressed in the "same" prices for different establishments—an impossible feat. It should also be observed that, even if two establishments have the "same" technology, a difference in degree of technical integration could lead to a difference in price per "unit" of capital assets and in value added per man-hour.

fore could not tell. The following three paragraphs, which highlight the report's findings, elaborate these statements.

Pearsonian coefficients of correlation between the value-added productivity of establishments and other variables (all referred to corresponding industry means) indicate dissimilar patterns of association for top-quartile and bottom-quartile plants (table 3.2). For high-productivity plants, productivity is perceptibly correlated with both capital assets available per man-hour of production workers and with the ratio of nonproduction to production workers. For low-productivity establishments, however, the two coefficients are minuscule. The authors suggest that other variables that could not be taken into account would have substantial explanatory value—e.g., managerial quality, process technology, and product specialization. Their subsequent statistical odyssey, however, adds little new insight.

A multivariate investigation of interquartile differences in productivity in 195 industries employs five presumably "causal" variables as regressors: gross book value per production worker man-hour, nonproduction workers per production worker, hourly wages of production workers, production worker man-hours (a measure of plant size), and product specialization (the percentage of plant shipments comprised by primary products). The coefficient of determination (R^2) for the equation connecting these five variables with percentage differences in productivity between the top and third quartiles is only 0.08. The corresponding coefficient for the equation comparing the productivity rates of the third and bottom quartiles is still smaller, only 0.03 (table 3.3). The individual regression coefficients are also small.

Despite the weak apparent explanatory value of the variables, a brave try is made to learn something from translog production (and cost) functions and Allen elasticities of substitution. The nine-parameter equations (subject to subsidiary constraints) are fitted to logarithms of production worker man-hours, nonproduction workers, and gross assets. Negative R^2s are obtained for the two-stage least-square equations when degrees of freedom are taken into account; and many of the Allen elasticities have the wrong sign. The heroic undertaking seems to confirm that top-quartile and bottom-quartile plants have different dependency profiles; and it indicates that bottom-quartile data, in particular, may suffer from significant transitory distortion. Additional test computations suggest that "disequilibrium" vitiates low-quartile relationships and that "monopoly" affects top-quartile relationships. Factor inputs and monopoly, however, seem to explain only 17% of the productivity variation between the top and third quartiles, and they account for only 11% of the productivity differential between the third and bottom quartiles. Again the authors cite managerial quality and product specialization as pertinent, though omitted, explanatory variables; they do not this time mention the relevance of process of manufacture.

The gist of various marginal notes prompted by comments made by the authors may be of interest or of use to them and to other readers of their report. Accordingly, a few of these notes have been combined and restated for offering below as observations on concepts and measurement.

1. Apart from the omission of variables, it should be recorded that census information on the included variables leaves little leeway for experiment in the measurement of establishment performance. Neither production workers nor nonproduction workers are occupationally fungible; and establishment differences in compensation of production workers, which are reported, do not reflect qualitative differences with respect to such germane labor attributes as morale. Furthermore, census figures for gross assets are only crude measures of capital supply; they include a variable price element and make no allowance for age or depreciation of plant. The different plant ages, incidentally, also affect the mutual adaptation of labor and capital—a "learning-curve"' phenomenon that augments both factors in terms of "efficiency units," rather than a contribution of management.

2. The reasonableness of appealing to additional external, even nonquantitative, information for appraising the "disequilibrium" and "monopoly" distortions of census data for a particular year should not be overlooked.

3. A production function for an establishment is really an "average" of imaginable, though not necessarily computable, elemental functions relating to more detailed products. The latter functions would require the estimation of inputs that in fact are joint—such as the services of various nonproduction workers. In principle, however, the inputs of "direct" (production-worker) labor and materials can be matched readily with the quantities of detailed products.

4. The choice of value added or some other net-output concept for a production function does not require validation by a "separability" theorem. It is justified, rather, by a plausible historic interest in "economic" production functions, which are intended to "explain" output levels and income shares simultaneously by reference to inputs of remunerable factors. To imply that net output belongs to a "second-best" class of concepts is as whimsical as to say that Leontief tables of gross transactions are inherently preferable to a system of national income and product accounts.

5. An "engineering" production function is not more characteristic of measurement at a plant level than is an "economic" function. It may refer either to net or gross output, but its independent variables are not

confined to the remunerable factor inputs. Thus, it may include cost elements such as materials and energy, gifts of nature, and noneconomic variables reflecting product specifications. A hybrid engineering-economic function of special interest substitutes capital services for capital supply, and it uses energy (usually purchased) as a proxy for such services.

6. A study concerned with interquartile (or interdecile) differences in "real" value added per production worker man-hour ought ideally to (a) distinguish between quantity and price in the numerator and (b) suitably "fix" the price component. Thus, for a comparison of top and bottom quartiles (deciles), prices of one or the other or industry averages should be used in both, if feasible. Failure to make a price adjustment in the measurement process should be taken into account in interpretation.[2]

7. Since unadjusted dollar figures of value added are interpretable as both net-output and factor-input values, it matters just what "deflators" are used. For productivity measurement, of course, "real" value added should reflect output, so price should be stabilized for (a) sales adjusted for inventories and (b) subtracted energy, materials, etc.

8. Even if no price adjustment of dollar figures for value added is feasible, it would seem desirable, when interquartile (interdecile) comparisons of productivity are sought, to weight establishment ratios by production worker man-hours.[3]

9. The availability of census information for value-added productivity and other variables affords an opportunity for design, if not full construction, of systems of algebraically compatible index numbers. Establishments occupying the same ranks in different quartiles (deciles) would be treated as "identical" for the computation of comparative numbers. (As in temporal comparisons, Fisherian or Divisian principles of index-number design might be invoked, and the two approaches could even be harmonized to some degree.)

10. For the analysis of interquartile (interdecile) differences in value added and associated variables, it may be useful to start with definitional identities, then perturb all the variables, and keep the terms containing second-order (and higher) "deltas." Arc elasticities could be computed; and they could also be adjusted, if desired, to include portions of symmetrically distributed interaction terms. The perturbed equation, still an identity, is highly respectable, being an exact Taylor (difference-

2. The points made in this paragraph and the next are related to those made by another commentator (Lipsey), of which the writer has first become aware on prepublication review of edited copy.

3. This remark, referring to appropriate aggregation of establishment ratios, should not be confused with another ideal desideratum: the weighting of intra-establishment man-hours according to hourly pay.

differential) expansion without remainder. It could be modified by the introduction of behavioral statements connecting the variables (e.g., a production function) or of other simplifying relationships. (The same approach could be used, if desired, with an initial production function —say, of the Cobb-Douglas variety. The function could be perturbed without arbitrary sacrifice of discrete interaction terms, which need not be negligibly small.)

A concluding optimistic comment is warranted. Despite some introductory remarks by the authors on the utility of mathematical functions for policy, the limited success of their painstaking inquiry hardly means that programs for deliberate advancement of productivity will be frustrated. Engineering and management consultants and business and government officials can still pinpoint opportunities for the improvement of plant operations, even by the manipulation of variables that the present study may indicate to be unpromising. The Department of Commerce (or any other) program of encouraging company productivity measurement should have a salutary effect on industry averages, even if it fails to reduce the gap between low-productivity and high-productivity establishments. (Indeed, the differential ability or willingness of firms to install a measurement system could itself be an indicator of variation in management quality.) The prospect of toning up productivity at the micro level, however, does not detract from the importance of a breakthrough on the macro level. Government could still influence the acceleration of productivity most decisively if it knew how to curb inflation without inducing or prolonging economic sluggishness, and how to maintain "stable" growth of employment and production ever after—in the spirit of the Employment Act of 1946, as amended, and according to the most ambitious interpretations thereof.[4]

4. Achievement of substantial disinflation without recession would, for example, improve the outlook for (1) private bond and equity financing and (2) private spending on research and development, two significant sources of productivity gain.

The Full Employment and Balanced Growth (Humphrey-Hawkins) Act of 1978, which amounts to a "most ambitious interpretation" of the Employment Act of 1946, offers no encouragement of greater governmental success in achievement of price stability and productivity acceleration. See *Economic Report of the President: 1979*, pp. 106ff.

Comment Robert E. Lipsey

The results of this study are ambiguous because there is a basic flaw in the data: the measure of "productivity" used, value added per production worker man-hour, is not really an efficiency measure but is more like a proxy for factor proportions.[1] A high value reflects high inputs of physical or financial capital, or nonproduction workers, or skilled workers per unit of unskilled labor input. There is no reason to say that high values of any such ratios imply efficient production. Since this is a type of factor proportions ratio, it is not surprising that the authors then find it correlated with other measures of or proxies for the capital-labor ratio, such as assets per production worker man-hour or nonproduction workers per production worker, or the hourly wage rate of production workers. No inferences about the effects of factor proportions on efficiency or prescriptions about methods of improving efficiency can properly be drawn. All that can be said is that value added per production worker man-hour is correlated with other measures of capital intensity.

Aside from the unsuitability of this productivity measure, cross-sectional studies of efficiency at the micro level using census data are subject to other problems that make it difficult to draw conclusions about efficiency. Value added is affected by indirect taxes and by erratic variations in profitability which can produce an impression of large differences in productivity when none exist. Furthermore, many of the individual establishments that are the units of observation are parts of larger enterprises which may, for tax or other reasons, influence the value added by manipulating such variables as the price paid by an establishment for products purchased from another unit of the same enterprise that is in another industry. Particularly in an industry in which value added is small compared with sales, such practices are another possible source of spurious variability in value added per man-hour which does not reflect differences in efficiency.

Robert E. Lipsey is with Queens College and the National Bureau of Economic Research.

1. See, for example, the use of a similar construct to measure the ratio of capital (including human capital) to labor in production in Hal B. Lary, *Imports of Manufactures from Less Developed Countries* (NBER, 1968).