NBER WORKING PAPER SERIES

CHRONIC EXCESS CAPACITY IN U.S. INDUSTRY

Robert E. Hall

Working Paper No. 1973

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1986

I am grateful to Michael Knetter for outstanding assistance and to numerous colleagues for helpful comments. The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Working Paper #1973
July 1986

Chronic Excess Capacity in U.S. Industry

ABSTRACT

Previous research has suggested that firms in a number of industries have considerable market power, in the sense that their prices exceed their marginal costs. However, the observed profits of those industries are not nearly as high as would occur under full exploitation of the market power with a constant returns technology. Rather, because of fixed costs associated with a minimum scale of operation or for other reasons, industry equilibrium occurs at a point where no abnormal returns are earned, even though market power exists. This inference is supported by an empirical study that shows that most industries hold capital far beyond the point that would minimize cost given their actual output. In this sense, the industries have chronic excess capacity.

Robert E. Hall
Hoover Institution and
Department of Economics
Stanford University
Stanford, CA 94305

Introduction

Economists have long suspected that a number of industries hold excess capacity or capital stock. Over the 38 years from 1948 through 1985, the Federal Reserve Board's index of capacity utilization in manufacturing has averaged only 82 percent--in the typical year, output has stood 18 percent below the feasible level chosen to represent 100 percent. Only in one year, 1966, did capacity utilization reach as high as 90 percent. The last year in which utilization exceeded 85 percent was 1973.

Theorists have contributed two main explanations for chronic excess capacity: First, Chamberlin's (1933) model of monopolistic competition and its refinements such as Spence (1976) and Dixit and Stiglitz (1977) suggest that firms may each retain monopoly power even though entry proceeds to the point of zero expected profit. In the resulting equilibrium, firms operate with declining average cost. Under certain conditions, the equilibrium may involve excess capacity according to a reasonable definition. Second, Spence (1977) and Dixit (1980), with many successors, have shown that the incumbent firm enjoying profit from a monopoly or oligopoly position may choose to hold excess capacity in order to deter entry. The reserve capacity lowers the expected profit of the potential entrant, provided the incumbent's threat to use the capacity after entry is credible.

Existing evidence on excess capacity is far from conclusive. As De Vany and Frey (1982) have argued, the firm that chooses its capacity so as to minimize expected cost may well hold excess capacity most of the time. The large cost saving during the occasional period when it is profitable to produce large volumes of output may justify holding capacity that is idle most of the time. Moreover, the benchmark for the measurement of capacity in the first place is inherently ill-defined. Though the Federal Reserve and other compilers of capacity utilization data attempt to measure capacity in the sense of a fully practical level of production, there is always the possibility of an upward bias in their measures of capacity. Finally, if there is scope for smooth substitution between labor and capital, then the concept of capacity is not even defined as a matter of theory. The definition of capacity as the point of minimum short-run average cost, as proposed by Berndt and Morrison (1981), is essentially arbitrary; it does not emerge naturally from a theory of optimal choice of capacity.

My work on this question starts with an unambiguous definition of excess capacity: A firm has excess capacity when the expected marginal benefit of capital falls short of the service price of capital. Because equality of expected marginal benefit and service price is the first-order condition for cost-minimization, a firm with excess capacity, according to this definition, is one that is not minimizing expected cost.

By setting a definition of excess capacity within a formal stochastic model of the firm, it is possible to deal rigorously with the potentially important role of Jensen's inequality. As De Vany and Frey have suggested, a firm may rationally choose to hold a level of capacity in excess of its expected level of output, in which case its average level of capacity utilization may be well under 100 percent. However, a firm in that situation should still equate the expected marginal benefit of capital to its service price, unless one of the factors mentioned above causes it to hold true excess capacity.

The decision I examine here is the choice of capital stock as part of a general strategy that determines the level of output as well. The firm is viewed as minimizing expected cost given the probability distribution for future output generated by the strategy. The optimal capital stock has a very simple property: The average value of the marginal benefit of capital over a span of years should equal the rental price of capital.

Although I examine the investment decision conditional on the distribution of future output, I do not consider output an exogenous variable. Rather, I proceed in this way in order to isolate the investment decision from the other decisions made by the firm. The strategy is similar to Jorgenson's (1963) investment theory, except that he considers the investment decision conditional on the nominal value of the firm's sales, rather than on its real output.

The paper proceeds in the following way: It characterizes the optimal capacity of a cost-minimizing firm with market power, under the assumptions of constant returns to scale and the absence of any motive for holding capacity other than the minimization of cost. The hypothesis that U.S. industries hold the optimal amount of capacity, judged by this standard, is overwhelmingly rejected. My interpretation of the rejection is that either constant returns fails because fixed costs are important, or that firms hold capacity to deter entry or for some other reason other than to produce at least cost.

My characterization of the cost-minimizing, constant returns firm makes no assumptions about the functional form of the cost or production functions of the firm. The intuition of the method is the following: Think of the firm as divided into a production department and a marketing department. The production department sells its output to the marketing department at a price equal to marginal cost. The marketing department sells to the public at a higher price incorporating a markup. The markup ratio is the ratio of the price to marginal cost. The actual marginal benefit of capital is just the profit rate of the production department when its output is valued at marginal cost.

If marginal cost were observed directly, then the calculations of this paper would be elementary. I would calculate the realized profit of the production department each year, and compare the profit rate to the service price of capital as perceived earlier when the investment decision relevant for this year was made. If the realized profit rate was generally lower, I would conclude that the firm held excess capacity.

Only a noisy measure of marginal cost is available directly from the data. My earlier work on the relation between price and marginal cost, Hall (1986), derived a measure of marginal cost based on changes in cost that occur from year to year as output changes. It also showed how to estimate the markup ratio as a constant parameter. One of the methods used here infers marginal cost by dividing the observed price by the estimated markup ratio. The result is nothing more than a smoothed version of the marginal cost measure than can be derived directly from the data.

When the profit rate of the production departments of firms in manufacturing are calculated in this way, it turns out to be negative in every year since 1949. Under the maintained assumptions of this paper, the conclusion is unambiguous that these industries have chronic excess capacity. They are not choosing their capital stocks to minimize expected cost with constant returns. Rather, their productive units are larger than would be chosen under constant returns, because of a minimum scale requirement, or capital has some other benefit in addition to lowering the cost of production. It is held to deter entry, to attract customers, or for some other reason unrelated to production.

The second approach I use in this paper makes use of the noisy direct measure of marginal cost. Estimation and hypothesis testing is set up in a formal stochastic framework that takes account of both the noise in the measure of marginal cost and the error in the marginal benefit of capital. I estimate the average value of the marginal benefit of capital over the sample period; in almost all industries the average falls short of its theoretical level and, indeed, is negative in the great majority. The hypothesis of cost-minimization with constant returns is rejected decisively in 12 of the 20 industries. Then I go on to characterize the magnitude of the failure of cost minimization by defining a parameter that measures the shortfall of the marginal benefit of capital from its theoretical value, expressed as a fraction of the gap that would exist under zero expected profit. In the majority of industries, the marginal benefit of capital is found to be between 70 percent and 100 percent of the way from its value under cost minimization to its value under zero profit.

The controversial element in these calculation is the imputation of marginal cost. There do not seem to be any other aspects that are particularly vulnerable to specification errors, data errors, or other sources of bias. Hence, the persuasive power of this paper hangs on the argument in my earlier paper that marginal cost has been correctly measured.

Figure 1 illustrates the relationship among the issues and tests considered in this and my earlier paper. The two axes describe the two quantitative dimensions of the studies. On the horizontal axis is market power, measured by the markup of price over marginal cost. On the vertical axis is the degree of excess capacity, measured by the shortfall of the marginal benefit of capital from its cost-minimizing level. The diagonal line divides the plane into a region of negative profit, above and to the lefht, and a region of positive profit. Firms are unlikely to be observed for any length of time in the region of negative profit. A firm that is able to protect its market power by effective methods might be observed to be chronically in the region of postive profit. If a firm is unable to deter entry, or can only do do by surrendering all of its monopoly profit, then the firm will wind up along the diagonal, with zero profit. One of the most interesting questions answered in the current paper is whether the process of competing away profit also eliminates market power. If so, firms should be observed mainly near the origin, where there is no excess capacity or market power. I show that in most inudstries, pure profit is close to zero, but market power is considerable. That is, the typical firm examined here and in my previous paper occupies a position near the x in Figure 1.

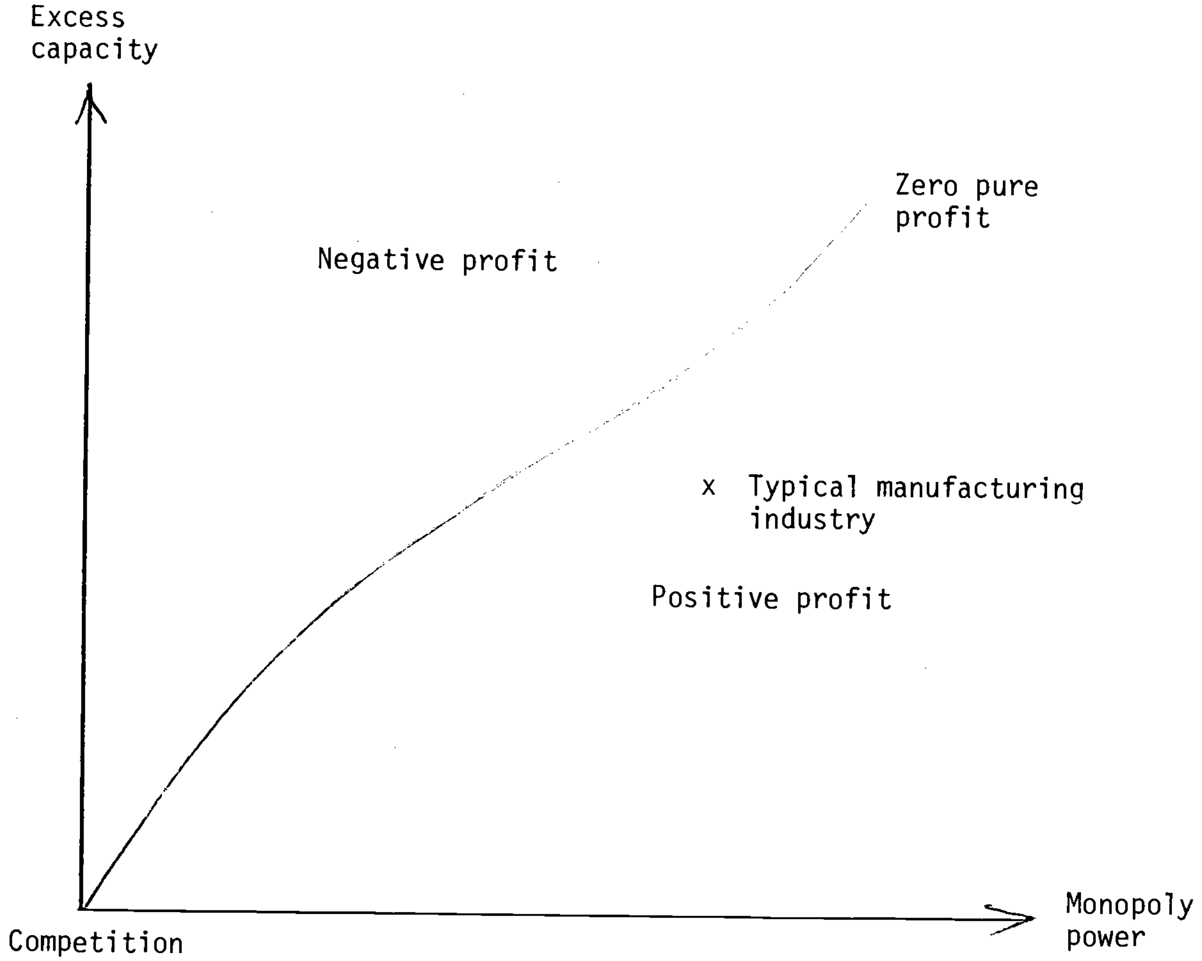


Figure 1. Two dimensions of industry structure.

Earlier research measured the horizontal position of a number of industries, measuring monopoly power by the markup of price over marginal cost. The present paper measures the vertical position of the same industries, measuring excess capacity by the shortfall of the marginal benefit of capital from its cost-minimizing level.

Firms hold substantial amounts of excess capacity. I offer the conjecture that most excess capacity arises from the fixed costs of a minimum scale of production, but the results presented here do not try to distinguish among the alternative sources of excess capacity.

1. *Theory*

A firm uses capital K and labor N to produce output Q. Its short-run cost function is:

$$(1.1) \qquad C(Q,K,w) = w\,L(Q,K)$$

$L(Q,K)$ is the labor requirement to produce Q with capital K; w is the wage. The firm has constant returns to scale, so that $L(Q,K)$ is homogeneous of degree one in Q and K.

The firm knows its future factor costs but is uncertain about future demand, which is influenced by a random variable η. The firm picks an output strategy $Q_t(\eta_1,\ldots,\eta_t)$ and a investment stragegy $K_t(\eta_1,\ldots,\eta_{t-\tau})$ contingent on the observed realizations of η. Note that output can respond to the most recent information but there is a lag, τ, in the response of capacity to new information; τ is the time to build. One of the criteria for an optimal strategy is that the investment strategy minimize the expected discounted value of total cost given the output strategy:

$$(1.2) \qquad \text{Min } E \{ \sum R_t [w_t L_t (Q_t , K_t) + r_t K_t] \}$$

The expectation is conditional upon all information known to the firm at the time it picks the strategy. A fully optimal strategy will be time-consistent--it will minimize the remaining future expected discounted cost as of any time period. Thus, it is not necessary to consider the conditional expectations midway through the process. The work presented in this paper derives from the following obvious result:

Theorem. Let

$$(1.3) \qquad z_t = - \frac{w_t}{r_t} \frac{\partial L_t}{\partial K_t}$$

which I will call the marginal benefit of capital. Then

$$(1.4) \qquad E(z_t) = 1$$

The proof follows immediately by considering perturbations around the optimal investment strategy, taking derivatives with respect to the perturbations, and then setting the perturbations to zero.

The expectation is conditional on the same information available to the firm when it chooses its strategy. The basic message of the theorem is simple: An investigator who calculates the marginal benefit of capital after the fact will find that its average value is one. If its average value is consistently below one, the firm is holding too much capital to be consistent with cost minimization and constant returns.

One could find more elaborate characterizations of optimal investment strategies. For example, the expectation of z_t conditional on information available in year $t - \tau$ should also be one. However, the results obtained here rejecting even the simplest characterization are so strong that there is no good reason to examine other characterizations. The advantage of my procedure is expressed in the following

Corollary (***Irrelevance of Time to Build***). For any value of τ, $E(z_t) = 1$ for all periods in which output is produced.

Thus, the troublesome issue of lags in the investment process can be sidestepped by looking only at the average of the marginal benefit of capital and not its correlation with other variables.

The basic condition examined here requires that the expected marginal benefit of capital (in labor units) is equal to the rental price of capital, r, divided by the wage. Equivalently, the ratio of the marginal benefit of capital to the rental cost differs from one by an error, ϵ, with mean zero:

$$(1.5) \qquad -\frac{w}{r}\frac{\partial L}{\partial K} = 1 + \epsilon$$

The marginal benefit of capital can be measured directly from the data without making assumptions about the functional form of the cost function. Under constant returns,

$$(1.6) \qquad \frac{Q}{N}\frac{\partial L}{\partial Q} + \frac{K}{N}\frac{\partial L}{\partial K} = 1$$

Let x denote marginal cost:

$$(1.7) \qquad x = w \frac{\partial L}{\partial Q}$$

Inserting this into equation 1.6 and solving for the marginal benefit of capital, z, gives:

$$(1.8) \qquad z = -\frac{w}{r}\frac{\partial L}{\partial K} = \frac{xQ - wN}{rK}$$

The realized marginal benefit of capital is gross profit per unit of capital rental cost, where output is valued at marginal cost instead of price.

Call the markup ratio of price to marginal cost μ. Then x can be replaced in equation 1.8 by p/μ, yielding:

$$(1.9) \qquad z = \frac{\frac{1}{\mu} p Q - w N}{r K}$$

Minimization of expected cost implies that $E(z) = 1$--the departures of marginal benefit from rental cost have mean zero. If the technology has a minimum practical scale, or some factor other than cost minimization motivates investment, so the firm has chronic excess capacity, then z will be consistently below one.

My earlier work--Hall (1986)--estimated values of the markup ratio μ for a number of industries. One of the measures of chronic excess capacity computed later in this paper uses those estimates in equation 1.9 to estimate the marginal benefit of capital.

The basic finding of this paper is that a number of U.S. industries operate with chronic excess capacity. In those industries, marginal cost is low; the typical firm can expand output without encountering a capacity constraint or a steeply rising part of its marginal cost curve. A significant amount of labor input is overhead labor. Capacity consistently has a negative marginal benefit because it is staffed with labor but is not fully utilized.

The technique to be used to test hypotheses about excess capacity does not make the assumption that the markup ratio, μ, is a constant. Rather, it uses the measure of marginal cost that underlies my earlier estimates of μ. Equation 1.4 from my earlier paper states a measure based on first differences of inputs and output. In the notation of this paper, it is:

$$(1.10) \qquad x = \frac{w \Delta N + r z \Delta K}{\Delta Q - \theta Q}$$

The first term in the numerator is the actual change in wage cost less the part of the change attributable to change in the wage. The second term is the imputed cost of the change in the capital stock, abstracting from changes in the rental price. The denominator is the actual change in output less the amount attributable to technical progress at rate θ.

Equation 1.8 gives the marginal benefit or shadow value of capital in terms of marginal cost. Equation 1.10 gives marginal cost in terms of the shadow price of capital. The two equations can be solved for the two variables. For z, the result is:

$$(1.11) \qquad z = \frac{\alpha^\star}{1 - \alpha^\star} \frac{\Delta n + \theta - \Delta q}{\Delta q - \theta}$$

Here $\alpha^\star$ is labor's cost share:

$$(1.12) \qquad \alpha^\star = \frac{wN}{wN + rK}$$

Also, Δn is the proportional rate of change in the labor-capital ratio ($\Delta n = \Delta \log (N/K)$ and Δq is the proportional change in the output-capital ratio. The hypothesis of cost minimization requires that the z in equation 1.11 differ from unity by an error with mean zero.

Let $\bar{z}$ be the mean of z. Estimation of $\bar{z}$ and its standard error will provide the test of the hypothesis of no excess capacity. The statistical model is

$$(1.13) \qquad \frac{\alpha^\star}{1-\alpha^\star}\,\frac{\Delta n + \theta - \Delta q}{\Delta q - \theta} = \bar{z} + \epsilon$$

Because θ is a random variable, this form is not satisfactory for estimation. Multiplying by $\Delta q - \theta$ and doing a little additional algebra yields

$$(1.14) \qquad \frac{1}{1-\alpha^\star}(\Delta q - \alpha^\star \Delta n) = (1-\bar{z})\,\Delta q + \frac{1}{1-\alpha^\star}\theta$$
$$- (1-\bar{z})\,\theta - (\Delta q - \theta)\,\epsilon$$

The left-hand side of this equation is a productivity residual similar to the one proposed by Solow (1956). However, Solow used labor's share in in total revenue, $\alpha = wN/pQ$, as an estimate of the elasticity of output with respect to labor input, whereas this measure uses labor's share in cost, $\alpha^\star$. For a firm with significant pure profit derived from market power, α is considerably smaller than $\alpha^\star$. Solow's original form of the residual was the basis for the measurement of market power in my earlier work. Equation 1.14 says that the residual based instead on the cost share can answer the question of excess capacity. A firm with procyclical Solow productivity has market power. One where the productivity measure based on the cost share is also procyclical has excess capacity. To put it a different way, the switch from the revenue share, α, to the cost share, $\alpha^\star$, would eliminate the cycle in productivity for a firm possessing market power without chronic excess capacity. A firm with market power and excess capacity would have procyclical productivity by both measures. In all this discussion, procyclical means that when an exogenous force raises the firm's output, measured productivity rises as well.

The units of the measure $\bar{z}$ are the same as those of z--dollars of marginal benefit per dollar of rental cost of capital. A value of $\bar{z}$ of zero means that an additional unit of capital does not decrease expected cost at all and falls 100 percent short of covering its own cost. A value of $\bar{z}$ of -1 (which I actually find for a number of industries including total manufacturing) means that capital adds to cost by as much as its own rental cost, so it falls 200 percent short of covering its own cost.

Zero expected pure profit

The introduction indicated that the hypothesis of zero expected pure profit figured in an important way in the issues considered in this research.. Models where entry takes place up to the point of zero profit may have the character that overinvestment in capacity dissipates the potential profit available when price exceeds marginal cost. In competition, where marginal cost and price are equal, cost-miminization and zero expected profit are equivalent. Where marginal cost falls short of price, profit is positive when the capital stock minimizes expected cost. Zero profit involves a capital stock in excess of the cost-minimizing level.

The condition for zero expected profit is

$$(1.15) \qquad \pi = \frac{pQ - wN}{rK} = 1 + \epsilon$$

The difference between zero profit and cost minimization can be seen by comparing equation 1.15 to 1.8. Cost minimization uses marginal cost, x, to value output whereas zero profit uses actual price, p. Since π can be observed directly, testing the hypothesis of zero expected profit is a simple matter of calculating $\bar{\pi}$, the sample average of π, and testing the hypothesis that it equals one.

Estimating the degree of excess capital relative to zero profit

Recall that $\bar{z}$ measures the marginal benefit of capital in its own units; in Figure 1, $1 - \bar{z}$ is the vertical coordinate. Another way to think about the extent of excess capital is to measure the shortfall in the marginal benefit of capital as a fraction of the shortfall that would drive profit to zero. I will call this fraction γ. In Figure 1, the fraction is the ratio of the height of the point describing a firm or industry to the height of the diagonal line directly above the point.

Substituting the condition $E(\pi) = 0$ into the definition of z yields the value of $\bar{z}$ for which expected profit is zero:

$$(1.16) \qquad 1 - \bar{z} = E\,[\,\frac{1}{\beta}\,(1 - \frac{\alpha\,\Delta\,n}{\Delta\,q - \theta}\,)]$$

Here β is the ratio of capital cost to revenue, rK/pQ. I have written this condition in terms of the shortfall in the marginal benefit, $1 - \bar{z}$, because that is the quantity on the vertical axis in Figure 1. Equation 1.16 describes the diagonal in Figure 1.

If the shortfall of the marginal benefit of capital is a fraction γ of the amount that would drive profit to zero, then

$$(1.17) \qquad 1 - z = \gamma\,\frac{1}{\beta}\,(\,1 - \frac{\alpha\,\Delta\,n}{\Delta\,q - \theta}) - \epsilon$$

Putting this on the right-hand side and formula 1.11 on the left-hand side and multiplying by $\Delta\,q - \theta$ yields the estimating equation for γ:

$$(1.18) \qquad \frac{1}{1 - \alpha^{\star}}\,(\Delta\,q - \alpha^{\star}\Delta\,n) = \gamma\,\frac{1}{\beta}\,(\Delta\,q - \alpha\,\Delta\,n)$$

$$+ \;\theta\,\frac{1}{1 - \alpha^{\star}} - \gamma\theta\,\frac{1}{\beta} - (\Delta\,q - \theta)\,\epsilon$$

The expression in parentheses on the right-hand side, $\Delta\,q - \alpha\,\Delta\,n$, is the Solow productivity residual or index of the increase in total factor productivity. The similar expression on the left-hand side, $\Delta\,q - \alpha^{\star}\Delta\,n$, is the productivity residual computed with the cost share $\alpha^{\star}$, in place of the revenue share, α. The equation provides information about the parameter γ from the different behavior of the two productivity measures, as follows:

Case I. Competition. Neither productivity measure can be shifted by any outside influence, and no information is available about γ. The distinction captured by γ is meaningless in competition--both cost minimization and expected zero profit must occur simultaneously. In Figure 1, this corresponds to the origin.

Case II. Monopoly power with cost minimization. The productivity measure with the revenue share, α, changes with output. Because monopoly power means that labor is paid less than the value of its marginal product, α understates the true elasticity of output with respect to labor input. Even though true productivity does not change when output and employment rise, productivity measured in this way does rise. On the other hand, the productivity measure on the left, with the cost share $\alpha^\star$, measures productivity correctly, assuming cost minimization. Hence, an exogenous change that makes output rise increases the right-hand side of equation 1.18 but leaves the left-hand side unchanged. The parameter γ is revealed to be zero. This corresponds to the points along the horizontal axis in Figure 1.

Case III. Monopoly power with zero expected profit. Because cost and revenue are equal, on the average, the cost and revenue shares, $\alpha^\star$ and α, are equal, on the average, and the two measures of productivity are essentially the same. Monopoly power makes both measures rise by the same amount when output rises. Hence γ is shown to be equal to one. This corresponds to points along the diagonal in Figure 1.

2. *Identification and estimation*

As in my earlier work, it seems reasonable to portray the rate of productivity growth, θ, as the sum of a constant and a random element, u. To keep the notation uncluttered, I will redefine θ to be the constant. Now let

$$(2.1) \qquad f = \frac{1}{1 - \alpha^\star} (\Delta q - \alpha^\star \Delta n)$$

$$(2.2) \qquad g = \frac{1}{\beta} (\Delta q - \alpha \Delta n)$$

$$(2.3) \qquad h = \frac{1}{1 - \alpha^\star}$$

$$(2.4) \qquad j = \frac{1}{\beta}$$

The equation for estimating the average value of the marginal benefit of capital, $\bar{z}$, is

$$(2.5) \qquad f = (1 - \bar{z}) \Delta q + \theta h - (1 - \bar{z}) \theta + h u - (\Delta q - \theta) \epsilon + u\epsilon$$

The shortfall, $1 - \bar{z}$, is simply the coefficient of the rate of change of output when the left-hand variable is the productivity residual computed with the cost share, $\alpha^\star$.

The equation for estimating the normalized measure of the degree of excess capacity, γ, is

$$(2.6) \qquad f = \gamma g + \theta h - \gamma \theta j + (h - \gamma j) u - (\Delta q - \theta) \epsilon + u \epsilon$$

If capacity is chosen to minimize expected cost, then γ will be zero. If there is chronic excess capacity, so the marginal benefit of capital is consistently below the rental price of capital, then γ will be positive.

Identification of $\bar{z}$ and γ hinges on the availability of an observed variable that causes important changes in employment and output in the industry but is not correlated with the disturbances in equations 2.5 and 2.6. With respect to the industry's own productivity shift, u, the issues here are the same as in my earlier paper estimating the markup ratio. If the main sources of overall economic fluctuations are shifts in product demands and factor supplies, not in productivity, then a macro aggregate, specifically the change in real GNP, is suitable as an instrument. On the other hand, if a a major cause of fluctuations is a pattern of correlated shifts in productivity among many industries, then real GNP is not a suitable instrument for any industry. My untested identifying hypothesis is that the productivity disturbance, u, is uncorrelated with the change in real GNP.

Another part of the disturbance in the two equations is the product of output growth and the surprise in the marginal benefit of capital, $(\Delta q - \theta)\epsilon$. Both of these variables are highly correlated with the change in real GNP. Unexpected increases in product demand are probably the most important source of favorable surprises about the marginal benefit of capital. However, the product of the two variables is only slightly correlated with the change in real GNP, Δy. If the mean of the growth rate of the output-capital ratio, Δq, were exactly θ, and if $\Delta q - \theta$, ϵ, and Δy all had symmetric distributions, then the correlation would be exactly zero. Basically, $(\Delta q - \theta)\,\epsilon$ is positive for both negative and positive surprises, whereas Δy changes sign, so the expectation of the product is zero. Hence the change in real GNP satisfies the conditions needed for eligibility as an instrument under very general conditions.

The estimator I use is Amemiya's (1977) nonlinear three-stage least squares estimator, with the contemporaneous change in real GNP as the instrument. The estimator is applied to the two-equation system consisting of the equation from my earlier paper and either equation 2.5 or equation 2.6. Bivariate estimation is required because the parameter θ appears in both equations but is hardly identified in equations 2.5 or 2.6.

3. *Data and results*

Most of the data used in this study are the same as described in my earlier paper (Hall (1986)). These include nominal and real value added, compensation and total hours of work, and the real capital stock. The only series used here that was not part of the earlier work is the rental price of capital.

Construction of the rental price follows Hall and Jorgenson (1967). The formula relating the rental price to its determinants is:

$$(3.1) \qquad r = (\rho + \delta) \frac{1 - k - \tau d}{1 - \tau} p_K$$

The determinants are:

ρ: The firm's real cost of funds, measured as the dividend yield of the S&P 500 portfolio;

δ: The economic rate of depreciation, 0.127, obtained from Jorgenson and Sullivan (1981), Table 1, p. 179;

k: The effective rate of the investment tax credit, from Jorgenson and Sullivan, Table 10, p. 194;

d: The present discounted value of tax deductions for depreciation, from Jorgenson and Sullivan, Table 6, pp. 188-189;

τ: The statutory corporate tax rate, from Auerbach (1983), Appendix A;

p_K: The deflator for business fixed investment from the U.S. National Income and Product Accounts.

Use of the dividend yield as the real cost of funds is justified by two considerations: First, the great bulk of investment is financed through equity in the form of retained earnings. Second, the use of a market-determined real rate avoids the very substantial problems of deriving an estimated real rate by subtracting expected inflation from a nominal rate. The dividend yield is a good estimate of the real cost of equity funds whenever the path of future dividends is expected to be proportional to the price of capital goods. For the typical firm, this is an eminently reasonable hypothesis. Of course, for firms with low current dividend payouts and high expected growth, the dividend yield understates the real cost of funds. But these firms are counterbalanced by mature firms whose payouts are high and whose growth rates are below the rate of inflation.

Table 1 shows the basic calculations of the marginal benefit of capital in the manufacturing sector given an estimate of the markup ratio, μ. The estimate, $\mu = 1.67$, is taken from my earlier paper. The first column gives nominal value added. The second column reduces value added to an estimate of production revenue valued at marginal cost, by dividing by 1.67. The third column shows the level of compensation. In every year, compensation exceeds estimated production revenue, which means that the implicit earnings of capital are negative. The fourth column shows the total rental value of the capital stock, rK. The fifth column shows z, the ratio of the implicit earnings to the rental value. Under cost minimization, z should fluctuate above and below one. Instead, it ranges from -0.6 to -1. There is not a single year when z is even positive. The evidence lends no support to the hypothesis of cost minimization.

Table 1 does not fully support the alternative hypothesis of zero expected profit, although that hypothesis fares better than does cost minimization. Under zero expected profit, nominal value added, in the first column, would equal the sum of compensation, in the third column, and the rental value of capital, in the fourth column. In fact, value added exceeds the sum in every year, though often not by much. In 1963, for example, nominal value added was \$154 billion, while compensation was \$112 billion and the rental value of capital was \$19 billion, for a total of \$131 billion. The mean of profit per unit of capital income, π, is 1.75 with a standard error of .06, so the hypothesis of zero pure profit (mean

Table 1. Marginal benefit of capital in manufacturing

Year	Nonminal value added	Value at marginal cost	Compensation	Rental value of capital	Marginal benefit of capital
1949	65.8	39.4	47.1	11.4	-0.67
1950	76.8	46.0	53.6	12.0	-0.64
1951	91.4	54.7	63.7	13.4	-0.66
1952	94.7	56.7	68.8	14.1	-0.86
1953	103.2	61.8	76.4	15.3	-0.95
1954	97.9	58.6	72.9	14.5	-0.98
1955	111.5	66.7	80.0	14.6	-0.91
1956	116.5	69.8	86.1	16.6	-0.99
1957	120.9	72.4	90.2	18.7	-0.95
1958	113.5	67.9	86.4	18.4	-1.00
1959	130.2	78.0	95.7	18.3	-0.97
1960	131.8	78.9	99.4	19.3	-1.06
1961	132.1	79.1	99.6	18.9	-1.08
1962	144.9	86.8	107.8	18.7	-1.12
1963	153.7	92.0	112.3	19.0	-1.07
1964	164.9	98.7	119.8	19.4	-1.08
1965	182.6	109.3	129.3	21.1	-0.95
1966	201.9	120.9	144.3	25.2	-0.93
1967	207.2	124.1	151.2	27.5	-0.99
1968	225.8	135.2	165.2	29.9	-1.00
1969	238.0	142.5	179.5	36.3	-1.02
1970	232.4	139.1	180.9	42.8	-0.98
1971	244.7	146.5	185.0	41.1	-0.94
1972	271.0	162.3	203.6	42.2	-0.98
1973	303.2	181.6	230.4	47.4	-1.03
1974	317.4	190.0	250.1	60.4	-0.99
1975	334.1	200.1	252.4	66.7	-0.78
1976	384.9	230.5	286.1	69.2	-0.80
1977	437.6	262.0	322.9	80.4	-0.76
1978	490.4	293.6	365.1	97.3	-0.73

of π equal to one) is clearly rejected.

The conclusion expressed in Table 1 is highly sensitive to one of the table's ingredients, the markup ratio, and hardly sensitive at all to the others. Cost minimization fails badly because the estimated markup ratio, $\mu = 1.67$, is so high that the value of output based on imputed marginal cost is extremely low, below even the cost of labor. Nothing in the calculation of the rental value of capital, for example, much affects the conclusion. No matter what series was used for the rental value, as long as it was positive in each year, the estimated marginal benefit of capital, z, would be negative in every year.

For 1963, values of z for alternative markup ratios μ are:

μ	z
1.00	2.18
1.17	1.00
1.37	0.00
1.67	-1.07

The standard deviation of the estimate of μ reported in my earlier paper is 0.10, so it is unlikely that sampling error alone could account for the finding of negative z. A rigorous treatment of this question follows shortly.

Because of the central importance of the finding that the markup ratio in total manufacturing (and in most two-digit manufacturing industries) considerably exceeds one, I think it is useful at this point to review the empirical basis for that finding. Marginal cost is inferred from the actual change in cost from one year to the next, in comparison to the change in output. In many industries, and in manufacturing as an aggregate, the change in cost is quite small in comparison to the change in output. That is, marginal cost is low relative to price.

The cost that enters these calculations is labor cost, so another way to express the finding is that the variation in labor input is small relative to the variation in output. Labor productivity is procyclical. The standard explanations of procyclical productivity are harnomious with my conclusion that marginal cost is low. First, if a significant fraction of the work force has an overhead function, then the marginal labor requirement is low in comparison to the average labor requirement and productivity is procyclical. Second, if workers are hoarded during temporary cyclical downturns, then the availability of idle workers makes the marginal cost of labor low during any episode when employment is not growing.

Of the various specification errors that may have biased the estimate of the markup ratio upward, the only one that seems to have the potential to reverse the conclusion of chronic excess capacity is the following, considered at length in the earlier paper: There are unmeasured variations in work effort that are positively correlated with output. A proper measure of marginal cost would count the cost of extra effort and might reverse the conclusion that marginal cost is well under price. A number of considerations convince me that unmeasured fluctuations in effort cannot explain a bias in the estimate of the markup ratio large enough to bring the calculated marginal benefit of capital up to its theoretical value of unity. First, the magnitude of the fluctuations would have to be large. Figure 2 of my earlier paper shows that the effort of the typical worker would have to have been almost 10 percent above normal for a sustained period in the 1960s, for example. Second, survey evidence collected from employers by Fay and Medoff (1985) suggests that effort is slightly negatively correlated with output, not strongly positively, as required to give an upward bias in the estimated markup ratio. Third, the fluctuations in effort needed to rationalize the observed fluctuations in productivity are inconsistent with the observed behavior of compensation. Work effort rises so much in a boom that the wage, corrected for changes in effort, actually falls. I find this implausible. The only way to rescue the hypothesis of large fluctuations in work effort is to invoke the theory of wage smoothing, in which workers are not paid on a current basis for their labor input, but rather receive compensation based on the average level of work over an extended period.

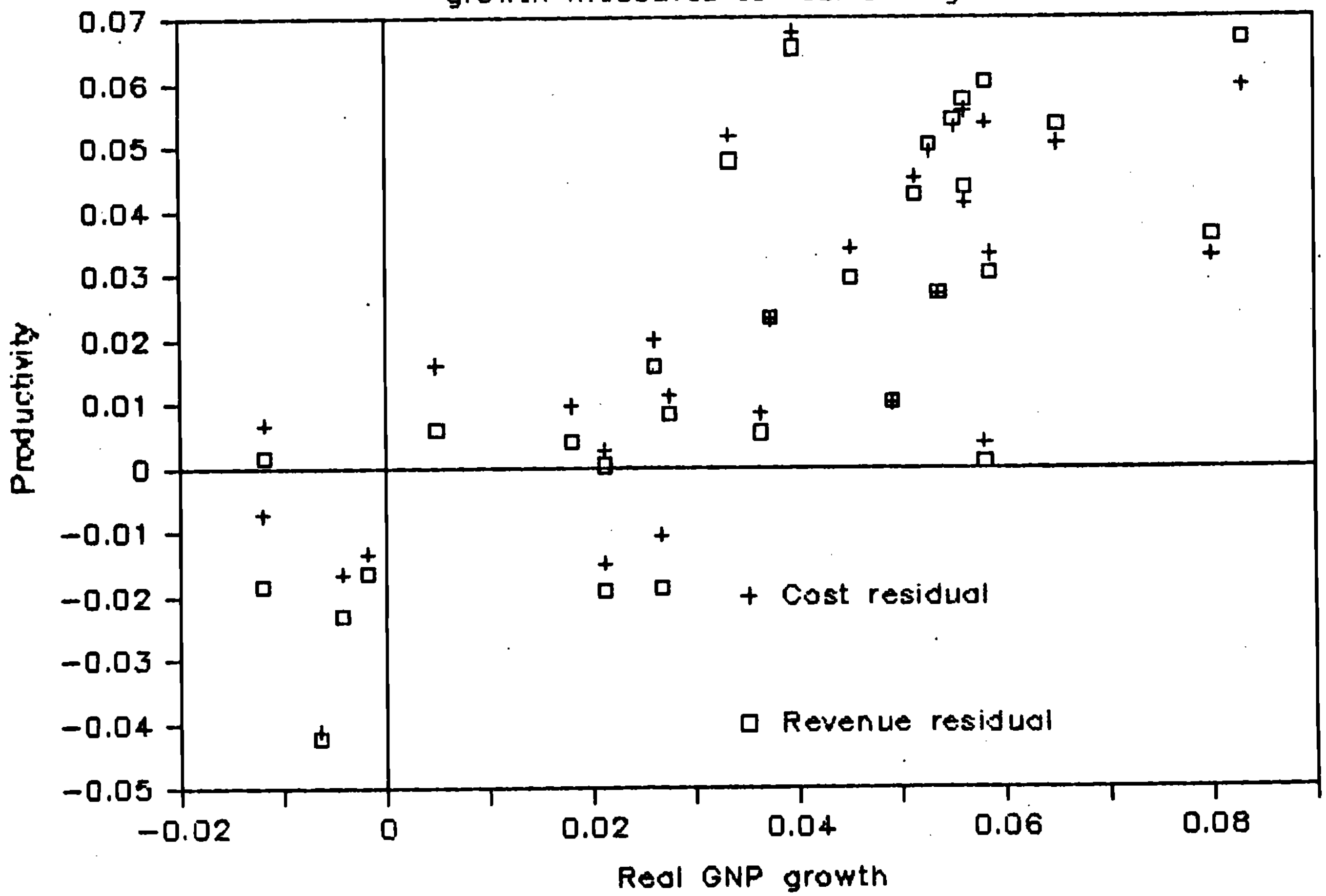
Testing the hypothesis of cost minimization without assuming a constant markup ratio

The findings just presented do not amount to a formal test of the hypothesis that firms choose their capital to minimize expected cost. They rely on an outside estimate of the markup ratio in order to infer the level of marginal cost. Both because of the lack of formal consideration of the sampling properties of the estimated markup ratio, and because of the lack of any strong economic foundation for the assumption that the markup ratio is a constant, it is desirable to carry out a self-contained test. The basis for the test was developed in equation 1.14 in section 1. In essence, the method diagnoses cost minimization by exploiting its implications for cyclical fluctuations in productivity. All imperfectly competitive industries show procyclical productivity when productivity is calculated as recommended by Solow, where the elsticity of output with respect to labor input is inferred from the share of compensation in total revenue. However, in the cost-minimizing industry, productivity can be measured accurately by using labor's share in total cost, where cost is the sum of actual compensation and the rental value of the capital stock. Under cost minimization, productivity measured with the cost share will not be procyclical.

Only those changes in productivity caused by an exogenous shift in product demand or labor supply are relevant for this calculation. Shifts in the true underlying rate of productivity growth need to be omitted from consideration. In other words, estimation of the slope must use an instrumental variable. My instrument is the change in real GNP. Again, my fundamental identifying hypothesis is that movements of real GNP are dominated by factors other than the common element of productivity shifts in individual industries.

Figure 2 shows the evidence on the two measures of productivity growth in manufacturing, graphed against the change in real GNP. Both measures are quite procyclical. The upward slope of the relation between the Solow measure (plotted with squares), which uses the revenue share, and real GNP reflects the conclusion of my earlier paper that markets are imperfect and price exceeds marginal cost. The upward slope for the measure based on the cost share (plotted with pluses) leads to the conclusion that the marginal benefit of capital consistently falls short of its theoretical cost-minimizing value.

Figure 2. Relation of two productivity growth measures to real GNP growth



The results of estimating equation 2.5 by nonlinear three-stage least squares are:

(3.2)
$$\frac{1}{1-\alpha^\star}(\Delta q - \alpha^\star \Delta n) = \underset{(.24)}{(1 + .93)}\,\Delta q + \underset{(.004)}{.035}\,\frac{1}{1-\alpha^\star} - (1.93)(.035)$$

Durbin Watson statistic: 1.76

The hypothesis of no excess capacity is overwhelmingly rejected--the estimated average value of the marginal benefit of capital, $\bar{z}$, is -.93 with a standard error of .24. This value is completely consistent with the calculations in Table 1.

Equation 2.6 provides a way to interpret the finding of a negative marginal benefit of capital in relation to the hypothesis of zero expected pure profit. The right-hand variable is normalized so that its coefficient is zero if there is no chronic excess capacity and one if excess capacity is sufficient to extinguish all latent profit arising from market power. The normalized right-hand variable is just Solow's productivity measure. If the two measures move together with close to unit slope, the degree of excess capacity is almost enough to eliminate all of the latent pure profit from market power. In that case, the estimate of γ will be close to one. On the other hand, if capacity is held to the cost-minimizing level, the lefthand variable will not move along with the right-hand variable and the estimate of γ will be close to zero.

In fact, as Figure 2 shows, the two measures of productivity growth are very similar. Although the cost shares are a little higher than the revenue shares, the differences in the rates of productivity growth are not large enough to make the slope of the one based on the cost share much less than the slope of the one based on the revenue share. The two measures are almost equally procyclical. The results are unfavorable to cost minimization and tend to support the alternative of zero expected profit.

The results of estimating equation 2.6 by nonlinear three-stage least squares are:

$$(3.3) \qquad \frac{1}{1-\alpha^\star}(\Delta q - \alpha^\star \Delta n) = \underset{(.026)}{0.791} \frac{1}{\beta}(\Delta q - \alpha \Delta n)$$

$$+ \underset{(.004)}{.035} \frac{1}{1-\alpha^\star} - (.791)(.035)\frac{1}{\beta}$$

Durbin Watson statistic: 1.83

The results suggest that the marginal benefit of capital is about three-quarters of the way from its theoretical value of one (under cost minimization) to its value if all latent monopoly profit is dissipated in excess capacity.

Results for two-digit industries

Table 2 presents similar results for 20 two-digit industries. The first column gives the estimate of the markup ratio of price over marginal cost from my earlier study, Hall (1986). The second column shows the marginal benefit of capital, estimated from equation 2.5. In 16 of the industries, the marginal benefit is negative. In tobacco, where the marginal benefit of capital is absurdly high, the problem is extreme sampling error in the estimate of μ. The third column shows the Durbin-Watson statistic for the estimation.

The fourth column of Table 2 shows the average level of profit per unit of capital, $\bar{\pi}$, as defined in equation 1.15. Comparison of $\bar{\pi}$ to its standard error, reported beneath it, yields a test of the hypothesis of zero pure profit. In most industries, that hypothesis is rejected decisively. Although profit is not has high as it would be under constant returns and cost minimization, it is still higher than it would be absent market power. The test here is biased toward the finding of profit, however, because the only element of capital cost considered is the rental cost of fixed capital. If a full accounting were made for inventories and financial capital, the calculated values of $\bar{\pi}$ would be lower.

The fifth column of Table 2 gives the nonlinear three-stage least squares estimates of γ. Recall that $\gamma = 0$ corresponds to the choice of capacity to minimize cost and $\gamma = 1$ corresponds to zero expected profit. In the majority of cases, the estimate of γ is between .5 and 1, and is significantly different from both polar values.

Table 2. Results for two-digit industries.

SIC code	Description	Markup ratio	Est. $\bar{z}$	Durbin-Watson	Est. π	Est. γ	Durbin-Watson
20	Food and	3.09	-1.56	1.82	1.542	0.841	2.26
	beverages	(1.64)	(0.70)		(0.057)	(0.074)	
21	Tobacco	1.28	4.35	2.24	6.724	-0.633	2.08
		(2.14)	(4.25)		(0.29)	(1.057)	
22	Textiles	1.05	0.65	1.89	1.386	0.084	1.87
		(0.27)	(1.89)		(0.081)	(2.536)	
23	Apparel	1.30	-1.90	2.06	4.07	0.642	2.13
		(0.24)	(3.76)		(0.225)	(0.226)	
24	Lumber	1.00	2.53	2.22	2.749	-2.127	1.74
		(.21)	(1.92)		(0.086)	(5.180)	
25	Furniture	1.38	-1.54	1.94	2.516	0.723	2.22
		(.17)	(1.15)		(0.142)	(0.091)	
26	Paper	2.68	-1.00	1.72	1.517	0.852	2.07
		(.33)	(0.19)		(0.054)	(0.024)	
27	Printing and	1.61	-1.83	1.57	2.537	0.670	1.73
	publishing	(.66)	(2.32)		(0.079)	(0.178)	
28	Chemicals	3.39	-1.03	2.38	1.881	0.771	1.94
		(0.78)	(0.29)		(0.087)	(0.035)	
30	Rubber	1.41	-0.72	2.28	1.563	0.797	2.43
		(.20)	(0.63)		(0.088)	(0.059)	
31	Leather	1.59	-5.60	2.76	2.807	0.792	2.81
		(.33)	(3.15)		(0.162)	(0.079)	
32	Stone, clay,	1.81	-0.79	2.10	1.586	0.812	1.67
	and glass	(0.22)	(0.29)		(0.08)	(0.048)	
33	Primary	2.06	-0.78	1.75	1.212	0.923	2.29
	metals	(.15)	(0.12)		(0.076)	(0.036)	
34	Fabricated	1.39	-0.73	2.33	1.756	0.800	1.64
	metals	(.13)	(0.38)		(0.076)	(0.048)	
35	Machinery	1.39	-0.11	2.28	2.212	0.478	2.30
	(non-elec.)	(.10)	(0.44)		(0.094)	(0.071)	
36	Machinery	1.43	-0.86	2.49	2.36	0.733	2.22
	(electrical)	(.15)	(0.86)		(0.122)	(0.077)	
38	Instruments	1.29	-0.70	2.52	2.756	0.556	2.51
		(.15)	(0.92)		(0.174)	(0.556)	
39	Miscellan-	1.52	-1.84	2.42	2.507	0.668	2.63
	eous	(.55)	(1.74)		(0.084)	(0.182)	
48	Communi-	1.43	0.09	1.95	0.711	0.829	2.15
	cation	(.64)	(0.42)		(0.029)	(0.086)	
49	Elec., gas,	10.18	-0.13	1.11	0.562	1.123	2.58
	and sanitary	(9.09)	(0.07)		(0.018)	(29.025)	

4. *Conclusions*

The data strongly refute the combination of two hypotheses: Constant returns to scale and a level of capacity that minimizes expected cost. The refutation is conditional on the identifying hypothesis that correlated shifts in productivity growth across industries are not a major moving force for total GNP.

Although the method of this paper does not make it possible to demonstrate which of the two hypotheses fails most conspicuously, I believe that it is probably constant returns. If many technologies have minimum practical scales, a robust type of theory can explain why the interaction of rationally managed firms and intelligent consumers will generate the findings of this paper. With a minimum practical scale, entrepreneurs will build new productive units in every market where costs can be covered, even though capacity is excessive by the definition used in this paper. Where markets are distinguished by geographical location or differentiated products, the equilibrium will have a multiplicity of productive units, each with a marginal benefit of capital below one.

It is a subtle question, one which I will not pursue here, whether the equilibrium consistent with these empirical findings is socially optimal. Competition is infeasible under the conditions just sketched. The social optimum involves a proliferation of differentiated products, with the costs of excess capacity covered by some system of taxes or charges.

References

Takeshi Amemiya, "The Maximum Likelihood and the Nonlinear Two-Stage Least Squares Estimator in the General Nonlinear Simultaneous Eequation Model,"*Econometrica 45*: 955-968, 1977

Alan J. Auerbach, "Corporate Taxation in the United States," *Brookings Papers on Economic Activity* 2:1983, pp. 451-505

Ernst Berndt and Catherine Morrison, "Capacity Utilization Measures: Underlying Economic Theory and an Alternative Approach," *American Economic Review Papers and Proceedings 71*: 48-52, May 1981

Edward Chamberlin, *The Theory of Monopolistic Competition* (Cambridge: Harvard University Press), 1933

Avinash Dixit, "The Role of Investment in Entry-Deterrence," *Economic Journal* 90:95-106, March 1980

Avinash Dixit and Joseph Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review 67*: 297-308, June 1977

Arthur De Vany and Gail Frey, "Backlogs and the Value of Excess Capacity in the Steel Industry," *American Economic Review* 72:441-451, June 1982

Jon Fay and James Medoff, "Labor and Output over the Business Cycle," *American Economic Review* 75:638-655, September 1985

Robert E. Hall, "The Relation between Price and Marginal Cost in U.S. Industry," Stanford, May 1986

Robert E. Hall and Dale W. Jorgenson, "Tax Policy and Investment Behavior," *American Economic Review* 57:391-414, June 1967

Dale W. Jorgenson, "Capital Theory and Investment Behavior," *American Economic Review Papers and Proceedings* 53:247-259, May 1963

Dale W. Jorgenson and Martin A. Sullivan, "Inflation and Corporate Capital Recovery" in ***Depreciation, Inflation, and the Taxation of Income from Capital***, Charles R. Hulten (ed.) (Washington DC: Urban Institute) 1981, pp. 171-237

Robert M. Solow, "Technical Change and the Aggregate Production Function, " *Review of Economics and Statistics* 39: 312-320, August 1957

A. Michael Spence, "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies* 43: 217-235, June 1976

____________, "Entry, Capacity, Investment, and Oligopolistic Pricing," *Bell Journal of Economics* Autumn 1977:534-544