NBER WORKING PAPER SERIES

INVESTMENT UNDER UNCERTAINTY: THEORY AND TESTS WITH INDUSTRY DATA

Robert E. Hall

Working Paper No. 2264

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1987

I am grateful to Michael Knetter and David Bizer for outstanding assistance and to numerous colleagues for helpful comments. This research was supported by the National Science Foundation. The data are available from the author upon receipt of an IBM-formatted diskette. 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.

NBER Working Paper #2264
May 1987

Investment under Uncertainty:

Theory and Tests with Industry Data

ABSTRACT

Under the assumption of constant returns to scale, there is a very simple, easily testable condition for optimal investment under uncertainty. Application of the test requires no parametric assumptions about technology and no assumptions about the competitiveness of the output market. The condition is that the expected marginal revenue product of labor equal the expected rental price of capital. The condition implies a certain invariance property for a modified version of Solow's productivity residual. Tests of the invariance property for U.S. industry data give very strong rejection in quite a few industries. The interpretation of rejection is either that the technology has increasing returns (possibly because of fixed costs) or that firms systematically over-invest.

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

Introduction

Since Dale Jorgenson's (1963) pioneering work, investment theory has been embedded in the theory of the optimizing firm. The firm acquires capital as part of a plan to produce output and purchase inputs so as to maximize the present discounted value of the firm. However, both Jorgenson and most of his succesors derived structural investment functions. Jorgenson's derivation was not explicitly stochastic and more recent work has been able to obtain closed-form investment equations only by making very strong assumptions. The purpose of this paper is to isolate the essential condition for optimal investment in an explicit framework of stochastic optimization. The condition is readily testable without parametric assumptions about the technology; the second half of the paper carries out simple tests with data on U.S. industries. I make no attempt to derive or estimate a structural investment equation. The test applied here should precede the development of an investment equation--only after a test of the type derived here can be passed should the research go on to the further step of estimating an investment equation.

The condition I examine is that the marginal revenue product of capital--the increase in revenue associated with an incremental unit of capital--be equated in expectation to the service or rental price of capital. The relevant expectation is one formed at or before the time that capital can be varied.

The reader may note the parallel in the objectives of this restatement of investment theory with my earlier work on consumption, Hall (1978).

For consumption, I derived a simple first-order condition that deals with the most central issue in consumption, the reaction of a consumer to new information about income. The first-order condition or Euler equation is readily testable. However, it is not a full statement of all the conditions for optimal consumption behavior and it does not yield a consumption function in closed form. Similarly, the condition derived here for optimal investment--the equality of the expected marginal revenue product of capital to the rental price of capital--does not result in a closed-form investment function under general conditions.

By setting a theory of investment within a formal stochastic model of the firm, it is possible to deal rigorously with the potentially important role of Jensen's inequality. For technologies in which the amount of capital determines the physical capacity to produce output, one would anticipate that the level of capacity would bear some relation to the expected level of output. However, as De Vany and Frey (1982) have noted, a firm may choose to invest in capacity which is frequently unutilized. But a firm in that situation should still equate the expected marginal revenue product of capital to its expected service cost.

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 maximizing expected profit given the probability distribution for future product demand. The optimal capital stock has a very simple property: The average value of the marginal revenue product of capital over a span of years should equal the average value of the rental price of capital.

The role of the assumption of constant returns to scale

The basic idea of this paper is that, under constant returns to scale, the actual value of the marginal revenue product of capital, *ex post*, can be computed as the residual profit per unit of capital. No further assumptions about technology are needed. The only other critical element of the computation is the removal from profit of the component generated by the market power of the firm.

The hypothesis of constant returns to scale is no novelty in this area--past research on investment has explicitly or implicitly adopted the hypothesis in most cases. Any model that concludes in a accelerator relation--proportionality between output and capital--rests on constant returns. Moreover, constant returns provides the easiest argument to support aggregation from the firm to the industry. The theory of investment under increasing returns, especially the type of increasing returns associated with fixed costs, involves rather different elements from those generally considered in investment theory.

The hypothesis tested in this paper is the joint hypothesis of constant returns and optimal choice of capital stock. The method employed does not permit the separation of the two parts of the joint hypothesis, because the use of residual profit to measure the realized value of the marginal revenue product of capital requires constant returns. The strong rejection of the joint hypothesis means either that the technology has increasing returns or that firms over-accumulate capital, or both.

My own view is that both explanations are important. The joint hypothesis is rejected because firms have too little profit to square with the evidence about their market power. The question is, what economic process dissipates the latent profit from market power? I think part of the answer is that there are fixed costs of labor, capital, advertising, and other factors needed to enter many markets. The other part of the answer is that market power attracts entry to the point where the capital stock is excessive, when judged by the comparison of the expected marginal revenue product to the rental price of capital. However, the method of this paper cannot separate the two.

The strategy

The paper proceeds in the following way: It characterizes the optimal capital of an optimizing firm (possibly with market power) under the assumption of constant returns to scale. Optimality involves the expected equality of the marginal revenue product of capital to the rental price of capital. The realized marginal revenue product and the rental price of capital will differ by an unpredictable error with mean zero, under optimal investment.

The condition just stated makes no assumption one way or the other about competition or market power. Under competition, the realized marginal product of capital could be measured directly as the realized rate of profit. In the presence of market power, the measurement of the marginal revenue product of capital is trickier. Measured profit contains an element which is chronically positive, the earnings from the typical firm's monopoly position. In addition, profit will contain the stochastic element with mean zero predicted by investment theory. In order to isolate the latter element, it is necessary to remove the monopoly element by valuing output at marginal cost rather than price.

If marginal cost were observed directly, then the calculations of this paper would be elementary. I would calculate the realized marginal revenue product of capital as the difference between output valued at marginal cost and actual input costs other than capital. Then I would compare the realized marginal revenue product 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 reject the joint hypothesis of constant returns and optimal investment.

Only a noisy measure of marginal cost is available directly from the data. My earlier work on the relation between price and marginal cost derived a measure of marginal cost based on changes in cost that occur from year to year as output changes. The approach I use in this paper implicitly makes use of the noisy direct measure of marginal cost. I show that the result of combining the hypothesis of optimal capital accumulation with my earlier method for measuring marginal cost can be expressed as an invariance condition for a certain measure of productivity growth. Measured productivity will be unchanged in the face of a shift which alters the levels of employment and output, provided that the shift has no effect on true productivity growth. In the presence of nonoptimal investment or increasing returns, on the other hand, measured productivity will not be invariant. Thus the relation between measured productivity growth and exogenous instrumental variables can form the basis for a test of the joint hypothesis.

My earlier work on testing the equality of price and marginal cost also derived an invariance result for productivity. I made use of the principle that Solow's productivity residual is invariant under exogenous shifts, provided that markets are competitive and the technology has constant returns to scale. The productivity measure used here is different; it does not consider the output price and its invariance does not depend on competition.

Relation to previous work on stochastic investment theory

The best-known exposition of investment theory under uncertainty is Lucas and Prescott (1971). A related development appears more recently in Abel and Blanchard (1986). Lucas and Prescott state the first order condition for optimal investment as the requirement that the discounted expected contribution of capital to next period's value of the firm be equal to the current cost of acquiring capital. As they note, this simplifies the problem from the point of view of the firm, but requires a full analysis of the determination of the value of the firm in general equilibrium. Abel and Blanchard express the first-order condition in the form that the present discounted value of the marginal contribution of current investment to all future profits be equal to the acquisition cost of capital. In other words, Abel and Blanchard substitute a simple theory of the value of the firm into Lucas and Prescott's condition.

Both Lucas-Prescott and Abel-Blanchard, and all writers in the less formal "q-theory" tradition, assume that there are adjustment costs in the investment process. By contrast, my own development adopts the "time to build" framework as in Kydland and Prescott (1982). An earlier paper of mine (Hall (1977)) explores the relation between the two approaches and notes their equivalence, in the following sense: Under the time-to-build constraint, there will be an exact relation between the value of the firm and the rate of investment of the type derivable from adjustment costs.

My earlier paper also shows in detail that the first-order condition relating the acquisition cost of capital to the present discounted value of the marginal revenue product of capital, as in Abel-Blanchard, is exactly equivalent to the first-order condition relating the rental price of capital to the current marginal revenue product of capital, as in this paper. In essence, the latter can be obtained by taking appropriate differences over time in the former. There is no conflict between the pre-Jorgenson view of investment which looks at present discounted values of the profit flows from an investment and Jorgenson's idea of relating the current profit contribution to the current rental price of capital.

In the usual specification of adjustment costs, it is not true that the expected marginal revenue product of capital is equated to the expected rental price of capital, when the expectation is formed some finite period before the two variables are realized. There is another term, the shadow price of adjustment. However, that term is positive half the time and negative the other half of the time. Hence the proposition investigated here, which asserts that the average gap between the marginal revenue product and the rental price is zero, holds under the adjustment cost specification as well as under the time-to-build specification.

I conclude that all formal models of investment are based on equivalent notions about the basic first order condition for optimal investment. The hypothesis tested here--expected equality of marginal revenue product and rental price of capital--is essentially a universal statement of optimality, provided that the expectation is evaluated over a long time perspective.

1. Theory

A firm uses capital K and labor N to produce output Q. Its production function is:

$$(1.1) \qquad Q = F_t(N,K)$$

The technology has constant returns to scale, so F_t is homogeneous of degree one in N and K.

The firm is uncertain about future demand and future factor prices, which are influenced by a random variable η. Its revenue function is $R(Q_t,\eta)$. The firm picks an employment strategy $N_t(\eta_1,\ldots,\eta_t)$ and a investment strategy $K_t(\eta_1,\ldots,\eta_{t-\tau})$ contingent on the observed realizations of η. Note that employment 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.

The objective of an investment strategy is to maximize the expected discounted value of profit:

$$(1.2) \qquad \text{Max } E\,\{\sum D_t\,[R(F_t(N_t,K_t),\eta) - w_tN_t - r_t\,K_t\,]\,\}$$

Here D_t is the discount function, w_t is the wage, and r_t is the service or rental price of capital. 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 maximize the remaining future expected discounted profit as of any time period. Thus, it is not necessary to consider the conditional expectations midway through the process.

Under perfect competition and constant returns, it is well known that the investment and employment strategies of the firm have a knife-edge character. If the expected future price exceeds long-run marginal cost, investment and employment will be infinite, whereas if price is expected to fall short of that level, zero capital and zero labor will be used. At the point of equality, the scale of the firm is indeterminate. This paper considers primarily the case where each firm has market power, which is sufficient to eliminate the knife-edge problem. The basic condition--equality of expected marginal revenue product and rental price of capital--applies to the competitive case as well, but I will not burden the reader with an analytical apparatus that includes that special case.

Optimal capital accumulation under uncertainty

Let

$$(1.3) \qquad z_t = \frac{\partial R}{\partial Q_t}\frac{\partial F_t}{\partial K_t},$$

the marginal revenue product of capital. Then the first-order condition for optimal investment is

$$(1.4) \qquad E(z_t - r_t) = 0$$

The expectation is conditional on the same information available to the firm when it chooses its strategy. The basic message of this condition is simple: An investigator who calculates the excess of the marginal revenue product of capital over the service price of capital, after the fact, will find that its average value is zero. If its average value is consistently negative, the firm is holding too much capital to be consistent with profit maximization.

One could find more elaborate characterizations of optimal investment strategies. For example, the expectation of $z_t - r_t$ conditional on information available in year $t - \tau$ should also be zero. 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

Proposition 1 (Irrelevance of time to build).

For any value of τ, $E(z_t - r_t) = 0$ 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 revenue product of capital and not its correlation with other variables.

The basic condition examined here requires that the expected marginal revenue product of capital less the rental price of capital, r, equal zero. Equivalently, the actual value of the same variable should differ from zero by an error, ϵ, with mean zero:

$$z_t - r_t = \epsilon_t \tag{1.5}$$

Invariance of the cost-based productivity residual

My earlier work--Hall (1987)--derived a method for testing the equality of price and marginal cost from changes in cost and corresponding changes in output. The essential idea is expressed in the following proposition: Under competition, with price equal to marginal cost, the Solow productivity residual is invariant under changes in output and employment induced by some outside force that affects product demand or factor supply. On the other hand, with market power and a gap between price and marginal cost, the Solow residual will rise whenever an outside force raises output. The logic is simple: The Solow residual uses the ratio of compensation to the value of output to infer the elasticity of output with respect to employment. When the price is distorted upward by market power, the estimate of the elasticity is too low. Hence, the adjustment in the Solow residual for the change in employment is too small and the residual rises whenever employment and output rise in response to an outside stimulus.

This paper derives and uses a similar, but quite distinct invariance result. The reason that the Solow productivity measure fails in the presence of market power is its use of total revenue in the denominator of the estimate of the elasticity of output with respect to input. An alternative estimate of the elasticity is the ratio of labor compensation to total cost, where the latter is the sum of compensation and the service cost of capital, rK. I will demonstrate that the productivity residual calculated using this estimate of the elasticity retains the invariance property in the presence of market power. That is, measured productivity should not rise when an outside force raises output, provided that the productivity measure uses the cost-based elasticity.

The empirical results will show that most industries do not satisfy the invariance condition for the cost-based productivity measure. I show that the failure of the invariance condition could arise from excess capacity or from increasing returns to scale. Basically the cost-based productivity residual rises when output rises in response to an exogenous force because the cost-based elasticity is hardly larger than the revenue-based elasticity. Although most of the industries seem to be made up of firms with considerable market power, they are not exceptionally profitable. Their revenue is only modestly in excess of their costs, including the full service cost of capital. Hence the behavior of the cost-based residual is hardly different from the Solow residual; it rises whenever an exogenous force increases output.

The conclusion is that there is not nearly enough profit to make the amount of market power inferred from the behavior of the Solow residual consistent with constant returns to scale and optimal capacity. The question is what is absorbing the latent profit generated by the market power. One of the answers could be excess capacity. Entry stimulated by the profits associated with market power proceeds to the point that firms are all operating well below capacity. The other answer is fixed costs. Fixed costs are a failure of constant returns to scale and could also empirical findings.

The easiest way to see the major difference between the invariance condition tested in this paper and the one tested in my earlier paper is that the output price has a central role in the Solow residual, whose invariance tests the equality of price and marginal cost. On the other hand, the output price does not even appear in the calculation of the cost-based residual. Only the behavior of inputs, output, and costs determine the movements of the cost-based residual. Hence the cost-based residual is appropriate for testing the joint hypothesis of optimal investment and constant returns, both of which involve the technology and the choice of inputs, but not conditions in the output market.

Derivation of the invariance condition

The basic relation derived in my earlier paper can be written as

$$\begin{aligned} (1.6) \qquad \Delta q &= \frac{wN}{mQ}\Delta n + \theta \\ &= \sigma \Delta n + \theta \end{aligned}$$

Here Δq is the rate of growth of the output/capital ratio, Δn is the rate of growth of the labor/capital ratio, θ is the rate of Hicks-neutral technical progress, and σ is the share of labor cost, wN, in total cost, mQ. The relevant concept of cost is marginal cost, m, times total output, Q. Under constant returns, total cost mQ is just the sum of labor cost, wN, and capital cost, $(r+\epsilon)K$. Because the latter is unobservable, the share σ is also unobservable. Equation 1.6 is exactly the conclusion of

Solow's (1957) derivation of total factor productivity growth, except that he assumed the equality of price and marginal cost and so put p in place of m in the denominator of the labor share.

It is a simple matter to calculate a closely related observed share, $\sigma^\star$, which is labor's share in cost when the cost of capital is evaluated *ex ante* rather than *ex post*. That is,

$$(1.7) \qquad \sigma^\star = \frac{wN}{rK + wN}$$

The two shares differ because of the surprise difference between the anticipated cost of capital, rK, and the realized shadow cost of capital, $(r+\epsilon)K$.

I will be concerned with what I call the ***cost-based productivity residual***,

$$(1.8) \qquad \Delta q - \sigma^\star \Delta n$$

If this residual could be calculated with σ in place of $\sigma^\star$, it would obey an invariance property of the type exploited in my earlier paper. However, a little algebra shows that the cost-based residual has a second term involving the surprise, ϵ:

$$(1.9) \qquad \Delta q - \sigma^\star \Delta n = \theta - (1-\sigma^\star)\frac{\sigma}{r}\Delta n\, \epsilon$$

In a strong year, employment growth Δn will be positive and ϵ will be positive as well--capital's shadow value will exceed its rental price. Hence the second term will be negative. In a weak year, both Δn and ϵ

will be negative, and the second term will be negative again. Hence, the cost-based residual understates productivity growth.

The same property that makes the cost-based residual a poor measure of productivity growth provides the basis for a good test of the joint hypothesis of optimal investment and constant returns. Consider an instrumental variable, say x, that has a strong causal effect on output, but is uncorrelated with productivity growth, θ. The instrument is sure to be positively correlated with ϵ as well as with Δn, but its correlation with the composite disturbance $(1-\sigma^{\star})\,\sigma\,\Delta n\,\epsilon/r$ will be close to zero. The composite disturbance is generally positive in both good and bad times, whereas the instrument is positive in good times and negative in bad times. The correlation will be essentially zero. This establishes

Proposition 2 (invariance of the cost-based residual): Under the joint hypothesis of optimal investment and constant returns to scale, the cost-based residual is uncorrelated with an instrumental variable that is uncorrelated with the rate of productivity growth.

The cost-based productivity residual is similar to the one proposed by Solow (1957). 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, $\sigma^{\star}$. For a firm with significant pure profit derived from market power, α is considerably smaller than $\sigma^{\star}$. Solow's original form of the residual was the basis for the measurement of market power in my earlier work. Equation 1.9 says that the residual based instead on the cost share can test the joint hypothesis of constant returns and optimal investment. A firm with procyclical Solow productivity has market power. Under constant returns and optimal investment, the Solow residual but not the cost-based residual will be procyclical. When the productivity measure based on the cost share is also procyclical, the joint hypothesis is refuted. To put it a different way, the switch from the revenue share, α, to the cost share, $\sigma^\star$, would eliminate the cycle in productivity for a firm possessing market power with constant returns and optimal investment. A firm with market power and increasing returns or excess capacity would have procyclical productivity by both measures. In this discussion, procyclical means that when an exogenous force raises the firm's output, measured productivity rises as well.

Excess capacity

Now consider a firm that systematically over-invests. Its realized marginal revenue product of capital will generally be lower than the rental price of capital. The deviation, ϵ, will no longer be a pure surprise; it will have a negative mean. As a result of the negative mean, the composite disturbance in equation 1.9, $-(1-\sigma^\star)\,\sigma\Delta n\,\epsilon/r$, will be less negative in good times and more negative in bad times; it will be positively correlated with an instrument. This argument establishes

Proposition 3 (*effect of excess capacity*): For a firm with systematic overinvestment rather than optimal investment, the cost-based residual will be positively correlated with an instrument, when that instrument is itself positively correlated with the firm's output.

According to the proposition, one of the possible interpretations of the finding that the cost-based residual is positively correlated with an instrument is that the firm or industry generally holds excess capital.

Note that the concept of excess capacity in the proposition is strictly rooted in the optimizing condition. Excess capacity is defined as a negative mean of the difference, ϵ, between the actual marginal revenue product of capital and its target value, the rental price of capital. It is not based on any comparison of installed capacity to actual output.

Increasing returns to scale

Now consider a firm with increasing returns. It is easy to show that equation 1.6, the starting point for the analysis of this paper, becomes

$$(1.10) \qquad \Delta q = (1 + \gamma)\, \sigma\, \Delta n + \gamma \Delta k + \theta$$

under increasing returns. Here γ measures the degree of increasing returns; $1+\gamma$ is the elasticity of output with respect to equal proportional increases in all inputs. γ is a variable, not necessarily a constant. Δk is the rate of growth of the capital stock.

If we calculate the cost-based productivity residual for the firm, we get:

$$(1.11) \qquad \Delta q - \sigma^{\star}\Delta n = \theta + \gamma(\sigma\Delta n + \Delta k) - (1-\sigma^{\star})\frac{\sigma}{r}\Delta n \epsilon$$

In addition to the composite term related to the surprise in the shadow value of capital, there is another term involving γ. Two amendments must be made to ensure that the cost-based residual properly measures productivity growth under increasing returns. First, the cost share, σ, understates the actual elasticity of output with respect to labor, which is actually $(1+\gamma)\sigma$. Hence the term $\gamma\sigma\Delta n$ appears on the right. Second, the rate of growth of capital appears separately with coefficient γ.

By assumption, the instrument is positively correlated with Δn. In addition, it seems likely that it would be positively correlated with Δk; times of rising output are also likely to be times of high investment. On both accounts, the correlation between the cost-based residual and the instrument will be positive. Hence,

Proposition 4 (*effect of increasing returns*): With increasing returns to scale, the correlation of the cost-based residual and the instrument will be positive.

A finding of positive correlation could as well be the result of increasing returns as the result of excess capacity.

2. *Econometric method and choice of instruments*

The invariance proposition tested in this paper is very similar in form to the one tested in my earlier paper, Hall (1987). The null hypothesis is refuted by finding a positive correlation between the productivity residual and an exogenous instrument. Econometrically, the simplest way to test for the absence of correlation is to calculate the regression coefficient of the productivity residual on the instrument and use the t-test for inference.

To be useful as an instrument, a variable must be the cause of important movements in the output and employment of an industry, but not a cause or an effect of shifts in its productivity. Here I use the same three instruments as in my earlier work: the rate of growth of military spending, the rate of change of the price of crude oil, and the political party of the President. All can be shown to be correlated with the output and employment of at least some of the seven industries studied here. For a more extensive defense of their exogeneity with respect to random productivity shifts, see my earlier paper.

3. *Data and results*

Most of the data used in this study are the same as described in my earlier paper (Hall (1987)). These include 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.

4. *Results for 7 U.S. industries*

Table 1 shows the basic data for nondurables manufacturing, one of the seven industries considered. The first column is the rate of growth of output per unit of capital; the second is hours growth per unit of capital, and the third column is labor's share in total cost. The cost-based residual in the fourth column is obtained by multiplying hours growth by the labor share and subtracting the product from output growth. The last two columns show the values of two instruments--the rate of growth of the price of crude oil and the rate of growth of military spending. There is a noticeable negative correlation between each of the instruments and the growth of output, on the one hand, and the cost-based residual, on the other hand. Oil price increases in 1957, 1973-75, and 1978 were associated with low or negative rates of growth of output and measured productivity residuals. Oil price declines in 1959, 1963-65, and 1972 were coupled with high output growth and large measured productivity residuals. For military spending, increases in 1966-67 came at the same time as low or negative growth rates of output and measured productivity residuals. Declines in military spending in 1955 and 1971-73 coincided with high measured productivity growth. The evidence based on military spending is more mixed; for example, in 1954, a large decline in military spending was associated with a decline in output but measured productivity growth was only slightly below normal.

The regressions to carry out formal tests of the invariance of the cost-

Table 1. Data for nondurables

(percent change or percent)

	Output growth	Hours growth	Labor share	Cost-based residual	Instruments: Oil	Instruments: Military
1953	-1.9	-0.6	78.8	1.6	7.0	-1.7
1954	-4.4	-6.5	80.0	2.0	2.9	-12.3
1955	5.5	2.5	81.2	4.0	0.1	-11.8
1956	0.9	-2.5	80.6	2.3	0.5	-1.8
1957	-1.8	-5.2	79.6	1.4	9.7	.0
1958	-4.8	-5.2	80.0	3.0	0.2	-1.9
1959	3.9	4.5	81.5	5.0	-3.4	-1.6
1960	0.8	-2.0	81.3	0.7	0.4	-0.2
1961	-0.3	-2.8	81.9	2.5	-0.7	2.6
1962	2.2	-0.1	82.9	3.3	0.2	0.5
1963	2.9	-2.5	83.2	7.1	-0.4	-0.9
1964	2.1	-2.4	83.7	3.2	-0.4	-4.0
1965	3.7	-3.2	83.4	1.9	-0.1	-0.2
1966	3.2	-4.2	82.1	1.2	0.7	13.2
1967	-1.5	-5.8	81.8	-1.8	1.1	12.3
1968	2.9	-3.1	81.9	3.4	0.8	1.7
1969	-0.1	-3.4	80.2	2.0	4.3	-6.0
1970	-1.9	-7.8	78.1	1.2	0.9	-9.1
1971	-0.4	-5.0	79.3	5.2	7.7	-8.4
1972	2.5	-0.9	80.1	5.5	-0.7	-7.3
1973	0.7	-0.9	79.6	7.3	10.2	-7.8
1974	-2.9	-7.4	76.8	-3.3	51.9	-4.6
1975	-0.9	-10.7	75.5	0.7	14.8	-3.0
1976	2.3	0.8	76.9	2.9	3.2	-2.7
1977	2.4	-2.0	76.0	2.0	7.8	1.3
1978	2.9	-1.4	74.4	0.4	9.0	1.0

based productivity measure in nondurables are:

$$\Delta q - \sigma^{\star}\Delta n = \underset{(.0045)}{.0306} - \underset{(.040)}{.115}\, x_{OIL}$$

SE: 2.1% DW: 1.58

$$\Delta q - \sigma^{\star}\Delta n = \underset{(.0047)}{.0220} - \underset{(.075)}{.144}\, x_{MIL}$$

SE: 2.3% DW: 2.01

In both cases, the correlation of the instrument with output growth is negative, so the evidence is unambiguous that an event such as an oil price decline or cut in military spending, which stimulates nondurables sales, raises the cost-based measure of productivity. The failure of the theoretical invariance property is attributable to some failure of its underlying assumptions. My interpretation is that either the assumption of optimal investment or the assumption of constant returns to scale fails. I will return later to a fuller discussion of the implications of the rejection of the invariance proposition.

Tables 2 and 3 present the evidence for all seven industries and for the oil, military, and political instruments. The entries in Table 2 are the marginal significance levels for the 21 tests for invariance. That is, each number is the probability that a covariance at least as positive as the one actually found might have arisen purely by chance. All industries except services reject the invariance hypothesis at the 10 percent level and all except services and finance-insurance-real estate reject at the 5 percent

Table 2. Test statistics for 7 industries and 3 instruments.

Industry	Oil	Military Spending	Party
Services	0.263	0.049	0.025
FIRE	0.071	0.255	0.357
Durables	0.045	0.709	0.100
Nondurables	0.004	0.034	0.389
Trade	0.007	0.240	0.782
Construction	0.007	0.342	0.191
Trans. and utilities	0.015	0.271	0.029

Notes:

The statistics are the marginal significance levels for a one-tailed test of the hypothesis that the covariance of the cost-based residual and the instrument is positive. The sign of the instrument is normalized so that its covariance with output growth is positive.

Table 3. Results for 7 industries and 3 instruments.

Coefficient, standard error, and Durbin-Watson statistic for each regression

Industry	Oil		Military Spending		Party	
Services	-0.012	2.40	0.053	2.63	0.0075	2.60
	(0.018)		(0.031)		(0.0036)	
FIRE	0.029	1.33	0.024	1.35	0.0016	1.34
	(0.020)		(0.036)		(0.0043)	
Durables	-0.108	2.14	-0.063	2.09	0.0176	2.28
	(0.061)		(0.114)		(0.0133)	
Nondurables	-0.115	1.58	-0.144	2.01	-0.0028	1.88
	(0.040)		(0.075)		(0.0098)	
Trade	-0.084	2.44	-0.045	2.34	0.0059	2.31
	(0.031)		(0.062)		(0.0075)	
Construction	-0.197	1.52	-0.060	1.44	-0.0156	1.51
	(0.073)		(0.147)		(0.0174)	
Trans. and utilities	-0.076	2.10	0.039	2.11	0.0142	2.33
	(0.033)		(0.063)		(0.0071)	

level. The oil price is a factor price and a source of demand shifts for each of the 7 industries. On the assumption that neither role should shift the production function, that is, that oil price fluctuations are uncorrelated with true productivity growth, I conclude that invariance fails in the direction predicted by excess capacity or increasing returns to scale.

In two industries, services and nondurables, fluctuations in military spending show evidence of failure of the invariance hypothesis in the direction predicted by excess capacity or increasing returns. All industries except durables show some evidence of failure in that direction.

The political party of the President is suitable as an instrument to the extent that the differences in policies of the two parties create differences in output growth rates but not in true productivity growth. In fact, real growth has generally been greater under Democrats that under Republicans. Under the assumption that the growth was achieved through differences in monetary and fiscal policy and not through differences in policies affecting the production function, the political dummy is a good instrument. I consider this assumption eminently reasonable. The third column of Table 2 shows that the political dummy gives strong rejection of invariance for services and transportation-utilities and somewhat weaker rejection for durables. Results for other industries are indeterminate mainly because output growth has hardly been different under the two parties.

5. *Interpretation of rejection of invariance of the cost-based productivity residual*

As a matter of theory, an optimizing firm with a constant returns technology should obey invariance--its cost-based productivity measure should be uncorrelated with any outside force that changes output but does not change its production function. I have shown earlier that excess capacity or increasing returns to scale could explain the failure of invariance in the direction found here. With chronic excess capacity, the firm's costs would be higher than appropriate, so the cost share of labor would understate the true elasticity of output with respect to labor input. Similarly, fixed costs or other failures of constant returns would make the cost share understate the true elasticity. Then, as a result of the understatement of the elasticity, the cost-based productivity residual would incorporate too small an adjustment for variations in labor input and the residual itself would rise every time output rose.

Other conditions could cause the failure of invariance. These include unmeasured fluctuations in work effort, suboptimal levels of employment because of monopsony power in the labor market, unmeasured fluctuations in capital utilization, and mis-measurement of labor input.

Labor hoarding

Before considering the various specification errors in turn, I should discuss one major phenomenon that has an important role in the story told by the data but is *not* an alternative explanation of the findings. I refer to labor hoarding and overhead labor. The following example shows how the invariance property of the cost-based residual holds in the presence of overhead labor:

Suppose that the technology is such that the level of employment required to produce an output Q is $\lambda K + \phi[Q - K]^+$. That is, with a capital stock of K and overhead labor of λK, it is possible to produce up to K units of output. Additional output requires an increment of ϕ units of labor for each unit of output above K. The shadow value of capital is $-\lambda w$ when output is below K because the firm could produce just as much output with lower overhead labor if its capital were lower. The shadow value of capital is $(\phi - \lambda)w$ when Q exceeds K--in that regime, more capital requires more overhead workers but reduces the requirement for the incremental labor described by ϕ. Let β be the probability that output will exceed K. Then the expected shadow value of capital is $(\beta\phi - \lambda)w$. At the optimum capital stock, this equals the service price of capital, r. Hence, $\beta = (r+\lambda w)/\phi w$. Suppose that the fluctuations in output are in a small region above and below K. The cost share, $\sigma^\star$, will be close to $w\lambda/(w\lambda+r)$. Because there is no true productivity change, the actual change in output, Δq, is a valid instrument itself. Suppose that the capital stock does not change over time. When output is below K, the change in employment is zero and the cost-based residual is equal to Δq. Thus the relation between the residual and the instrument has a unit slope. When output is above K, the change in employment, ΔN, is $\phi\Delta Q$. The level of employment is close to λQ, because Q is only a little over K. Hence the rate of growth of employment, Δn, is approximately $(\phi/\lambda)\Delta q$. The slope of the relation between the cost-based residual $\Delta q - \sigma^\star\Delta n$ and Δq is

Unmeasured fluctuations in work effort

Of the various specification errors that may have biased the covariance of the cost-based residual and an instrument upward, the only one that seems to have the potential to reverse the negative conclusion about investment theory with constant returns is the following, considered at length in the earlier paper: There are unmeasured variations in work effort that are positively correlated with output. When an outside force drives up output and employment, measured productivity rises for a reason unrelated to increasing returns or excess capacity. There is no question that the method of this paper is vulnerable to such measurement errors; the only question is the numerical importance of the errors.

A number of considerations convince me that unmeasured fluctuations in effort cannot explain all of the correlation I find between the cost-based residual and various instruments. 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

$1 - \sigma^{\star}\phi/\lambda$. The average slope is $1-\beta + \beta(1 - \sigma^{\star}\phi/\lambda)$. Inserting the values for β and $\sigma^{\star}$ derived above shows that the average slope is zero.

In the example, it is true that when the firm is in the labor-hoarding regime, when Q is below K, the covariance of the cost-based residual and the instrument would be strongly positive. However, this is exactly counterbalanced by a negative covariance when output is above K. What if a firm spent most of its time in the labor-hoarding regime and had output above K only in times of extreme demand? Isn't this the normal case for most firms? The answer is that such a firm is not satisfying the condition for optimal investment; it has excess capacity. If it is the normal condition, it simply proves the point of this paper, that the evidence suggests that excess capacity or fixed costs are important.

Labor hoarding and overhead labor are probably important phenomena in a number, if not all, of the industries studied in this paper. When a firm is in a labor-hoarding regime, its cost-based residual will be positively correlated with an instrument. In that respect, labor hoarding is an essential part of the explanation of the findings of this paper. However, labor hoarding is not an alternative explanation to excess capacity or increasing returns for the failure of the invariance property of the cost-based residual. A firm with a constant returns technology and an optimal investment strategy, no matter how ridden with forecasting errors, will spend enough time in a labor-shortage regime to offset the time spent in the labor-hoarding regime. As the example shows, the condition for optimal investment amounts to stating that the two regimes combine in such a way as to eliminate any covariance of the cost-based residual with an instrument.

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.

Other labor issues

A basic maintained hypothesis of this paper is that the firm chooses an optimal level of employment. The derivation of equation 1.6 makes the assumption that the marginal revenue product of labor is equated to the wage. An alternative is that the firm employs too many workers, on the average. Then the measured elasticity, $\sigma^{\star}$, would exceed the true elasticity, σ, because the observed wage would overstate the shadow value of labor. Hence the covariance of the measured cost-based productivity residual with an instrument would be negative. Excess labor could not explain the findings of the paper.

By the same token, if the shadow value of labor exceeds the observed wage most of the time, the findings could be explained. For example, if the typical firm has strong monopsony power in its labor market, a failure of the invariance property would occur in the observed direction. But the conditions under which this could be expected to persist for long periods are strenuous. First, if there is bilateral bargaining with a labor union, one would not expect to find a shadow value of labor in excess of the observed wage. Both parties could be made better off by attracting a worker from the open market and paying the worker the prevailing union wage. And if the union has much monopoly power, it is likely to succeed in pushing the observed wage above the shadow value, by extracting a lumpsum component of compensation as part of an efficient bargain.

Second, the firm has a strong incentive to overcome its monopsony position in the labor market by attracting workers from more distant markets. That is, when it can only get more work from its own local market by driving up every worker's wage, it will turn to other markets. What matters is the elasticity of labor supply from the entire labor market to the one firm in the long run. It is hard to believe that this elasticity is anything less than a very large number for most firms.

Mismeasurement of capital

An important implicit assumption of my work is that capital input is correctly measured. The measure of capital I use is the amount of capital available for use. As long as capital has no pure user cost, it is reasonable to assume that all capital available is in use. If there is a pure user cost--if capital depreciates in use rather than just over time--then the situation is different. There is a capital supply decision similar to the labor supply decision and presumably fluctuations in capital input occur in parallel to fluctuations in output. I should note at the outset that if capital is out of use because it is redundant--its shadow value is zero--then there is no bias in my procedure. The dangerous case is when capital has a positive shadow value and there are unmeasured fluctuations in utilization.

Though it is not possible to dispose of this hypothesis as a complete or partial explanation of the failure of invariance, it is possible to show that it calls for rather extreme movements of the true capital stock, corresponding to substantial pure user costs of capital. Let Δv be the change in measurement error of capital actually in use. Then the true relation between Δq and Δn, when these are calculated with measured rather than actual capital, will be:

$$(5.1) \qquad \Delta q = \sigma \Delta n - (1 - \sigma)\Delta v + \theta$$

Suppose that the measurement errors are related to the change in employment, as they would be if they arose from unmeasured fluctuations in capital utilization:

$$(5.2) \qquad \Delta v = -\phi \Delta n$$

Strict complementarity of work hours and capital hours would mean that ϕ has the value of one.

In this setup, the cost-based residual is

$$(5.3) \qquad \Delta q - \sigma \Delta n = \phi(1 - \sigma)\Delta n + \theta$$

The variable $(1 - \sigma)\Delta n$ can be formed (using σ^* in place of σ, which only adds an innocuous ϵ term) and then the parameter ϕ can be estimated by instrumental variables.

For nondurables, the results of estimation are:

$$\Delta q - \sigma^{\star}\Delta n = \underset{(.003)}{.025} + \underset{(.52)}{2.43}\,(1 - \sigma^{\star})\,\Delta n$$

DW: 1.98 SE: .013

Instrument: Change in oil price

$$\Delta q - \sigma^{\star}\Delta n = \underset{(.005)}{.025} + \underset{(3.19)}{4.97}\,(1 - \sigma^{\star})\,\Delta n$$

DW: 2.02 SE: .028

Instrument: Change in military spending

In effect, these estimates interpret the observed correlation of the instruments with the cost-based productivity residual as arising from measurement errors in capital utilization. In order to explain the magnitude of the correlation, the elasticities of the measurement errors with respect to the change in labor input must be implausibly large--around 2.5 or 5. The simple model in which capital and labor fluctuate in proportion, in which ϕ is one, is not nearly enough to explain the findings of the paper.

6. *Conclusion*

The data strongly refute the combination of two hypotheses: Constant returns to scale and a capital stock that maximizes expected profit. Taken at face value, the findings preclude the development of a structural investment equation along conventional lines. Since Jorgenson, the centerpiece of investment equations has been profit-maximizing choice of the capital stock by the firm. But that hypothesis requires the recognition of substantial departures from constant returns to scale, in the direction of increasing returns. Investment theory then must enter the thicket of industry equilibrium with fixed costs or other sources of increasing returns. No simple relation between output, the rental price of capital, and the capital stock can be expected to describe the behavior of investment. For example, output and the capital stock will expand in proportion if the expansion is accompanfied by a proportional expansion in the number of productive units, so that the scale of each unit remains unchanged. Capital will grow less than in proportion to output if the number of units remains unchanged and each achieves a greater scale.

References

Andrew B. Abel and Olivier J. Blanchard, "The Present Value of Profits and Cyclical Movement in Investment," *Econometrica 54*:249-274, March 1986

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

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, "Investment, Interest Rates, and the Effects of Stabilization Policies," *Brookings Papers on Economic Activity 1977:1*, pp. 61-103

Robert E. Hall, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy* 86:971-988, December 1978

Robert E. Hall, "The Relation between Price and Marginal Cost in U.S. Industry," Stanford, March 1987

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

Finn E. Kydland and Edward C. Prescott, "Time to Build and Aggregate Fluctuations," *Econometrica* 50:1345-1370, November 1982

Robert E. Lucas and Edward C. Prescott, "Investment under Uncertainty," *Econometrica* 39:659-682, 1971.

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