

NBER WORKING PAPER SERIES

INVARIANCE PROPERTIES OF SOLOW'S PRODUCTIVITY RESIDUAL

Robert E. Hall

Working Paper No. 3034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1989

Presented at Robert Solow's 65th Birthday Celebration, April 28, 1989. I am grateful to James Conklin for outstanding assistance and to numerous colleagues for helpful comments. This research was supported by the National Science Foundation. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author not those of the National Bureau of Economic Research.

NBER Working Paper #3034
July 1989

INVARIANCE PROPERTIES OF SOLOW'S PRODUCTIVITY RESIDUAL

ABSTRACT

In 1957, Robert Solow published a paper that provided the theoretical foundation for almost all subsequent work on productivity measurement. Although most applications of Solow's method have measured trends over fairly long time periods, the method also has important uses at higher frequencies. Under constant returns to scale and competition, the Solow residual measures the pure shift of the production function. Shifts in product demand and factor supplies should have no effect on the residual. Tests of this invariance property show that it fails in a great many industries. Though other explanations may deserve some weight, it appears that the leading cause of the failure of invariance is increasing returns and market power. The empirical findings give some support to the theory of monopolistic competition.

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

Introduction

Solow [1957] developed what is now the dominant approach to the measurement of productivity growth. To measure the shift of the production function, Solow pointed out, one can simply subtract a Divisia index of input growths from output growth. Key assumptions of the derivation are competition and constant returns to scale. Solow thought of his method as a way to measure the trend in productivity. He took the average rate of growth of the Solow residual as the best measure of the average rate of growth of the Hicks-neutral multiplicative component of the production function. From the start, users of the Solow residual were aware that it tended to follow the business cycle; in years of expansion, the residual is unusually large; in years of recession, it is low or even negative. My purpose in this paper is to consider the lessons that can be learned from the year-to-year fluctuations in the Solow residual.

Under Solow's assumptions, the following theorem holds: *The productivity residual is uncorrelated with any variable that is uncorrelated with the rate of growth of true productivity.* The theorem is just a restatement of Solow's basic result that the residual measures the shift of the production function. It says, in particular, that productivity growth should be uncorrelated with any variable that is a driving force for output, provided that the variable is not one that shifts the production function. For example, in the face of an exogenous upward shift in the demand for a particular industry's output, the productivity of that industry should remain unchanged. Or, if the price of one of the factors used by the industry rises sharply, productivity should also remain unchanged.

Among U.S. industries, the Solow residual is correlated with exogenous product demand and factor price movements. The invariance property fails conspicuously. Most of this paper explores the alternative reasons for the

failure. Each explanation relates to a failure of the assumptions underlying Solow's original derivation. Two critical assumptions are competition and constant returns. It turns out that there is a variant of the Solow residual that handles the case of a firm or industry with market power but constant returns. Invariance of the modified Solow residual fails just as strongly as does invariance of the original Solow residual. Hence the evidence points in the direction of increasing returns, presumably coupled with market power.

Another important explanation for the observed failure of invariance of the Solow residual (original or modified) is that the supposedly exogenous movements in product demand or factor supplies are actually causally related to the stochastic shifts in the technology. This explanation comes in two variants. First, product demand shifts may be a response to changes in technology. If government purchases of the output of an industry are the exogenous variable used to check the invariance property, it is possible that exogeneity fails because the government purchases more output in times when the technology is unusually favorable. Second, movements of the exogenous variable may trigger shifts in the technology. Externalities are the obvious example. If one industry's production function depends on the output of other industries, through thick market effects, then an increase in government product demand will operate through the externality.

Mismeasurement of inputs and outputs is another explanation of the failure of the invariance of the Solow productivity residual. Errors in measuring output that are positively correlated with the exogenous variable would explain the failure of invariance. For example, if firms use their workers to build tangible and intangible capital when they are not busy making output, then invariance would fail; the rise in measured output following a stimulus to demand would overstate the rise in actual output because unmeasured investment would fall. Errors in measuring labor input could also explain the failure of invariance. Actual hours worked by salaried employees may fluctuate even though their reported hours are steady; because they are not paid by the hours, there is no reason to keep close track

of actual hours. Work effort per hour may also vary in a way that makes measured labor input less responsive to exogenous shifts in demand than true labor input. Similarly, errors in measuring the use of capital services may understate the variability of that input and explain the failure of invariance.

1. Solow's method

Consider a firm that produces output Q with a production function $\Theta F(K, N)$ using capital K and labor N as inputs. Θ is an index of Hicks-neutral technical progress. The firm faces a stochastic demand for its output, possibly perfectly elastic. It faces a labor market where the firm can engage any amount of labor at the same wage, w . Some time in advance of the realization of demand, the firm chooses a capital stock. I do not assume anything about the market for capital goods nor, for that matter, do I assume that the firm's investment policy is optimal. However, I do assume that the pure user cost of capital is zero—capital depreciates over time, not in relation to use. I also assume that the firm chooses its labor input so as to maximize profit and that the choice is made after the realization of demand. Finally, I assume that there is at least one observable variable that shifts the demand schedule, the labor supply schedule, or the level of capital used by the firm and is uncorrelated with the rate of technical progress.

Solow's [1957] famous paper derived a relationship involving output growth, product price, capital and labor input, and the wage rate, under the assumptions of competition and constant returns to scale. The relationship is

$$\Delta q_t - \alpha_t \Delta n_t - (1 - \alpha_t) k_t = \theta_t \quad (1.1)$$

where Δq is the rate of growth of output ($\Delta \log Q$), α is the factor share earned by labor (ratio of compensation wN to total revenue pQ), Δn is the rate of growth of labor ($\Delta \log N$), and θ is the rate of Hicks-neutral technical

progress($\Delta \log \Theta$). Solow recommended evaluating the left side in order to measure the rate of growth of productivity. This measure has come to be known as total factor productivity because, unlike measures that consider only output and labor input, it accounts for capital input and, in a more general form, for all other types of inputs.

The statistic on the left side of equation 1.1 is the "Solow residual." It is the difference between the rate of growth of output and the weighted rates of growth of the inputs. Under competition and constant returns, the observed share of labor is an exact measure of the elasticity of the production function with respect to labor. Without any further restriction on the production function, the elasticity can be read directly from the data on compensation and revenue. From constant returns, the elasticity of the production function with respect to capital input is just 1 minus the labor elasticity. Once the labor elasticity is known, the rate of growth of total input can be formed as the rate of growth of labor, weighted by the labor elasticity, plus the rate of growth of capital, weighted by the capital elasticity. Then the rate of productivity growth can be obtained by subtracting the rate of growth of total input from the rate of growth of output.

Solow had in mind the calculation of the rate of growth of productivity, θ_t , separately for each year. Because productivity growth seems to have a substantial random element, it is natural to view θ_t as the sum of a constant underlying growth rate, θ , and a random term, u_t . Then equation 1.1 becomes

$$\Delta q_t - \alpha_t \Delta n_t - (1 - \alpha_t) k_t = \theta + u_t \quad (1.2)$$

Now suppose there is a variable, say Δz_t , which is an important outside determinant of output and employment. It could be government purchases of the output of this industry, or a measure of the shift of labor supply to the industry, or something else that affects Δq and Δn . Suppose further that the variable Δz is exogenous to this equation; that is, it is uncorrelated with the

stochastic element of productivity growth, u_t . In other words, the variable Δz is of a type which is known from prior reasoning not to cause shifts in productivity nor to be influenced by productivity shifts that come from other sources. Later in the paper I will suggest that the change in military spending is one such variable. If the variable Δz_t has zero correlation with the right side of equation 1.2, it must have zero correlation with the left side as well. This establishes:

Proposition 1. Invariance of the Solow residual: Under competition and constant returns to scale, the Solow residual is uncorrelated with all variables known to be neither causes of productivity shifts nor to be caused by productivity shifts.

When a convincingly exogenous variable is found to be correlated with the Solow residual, it refutes the joint hypothesis of competition and constant returns. Later I will investigate the power of the test and the interpretation of rejection. I will demonstrate the following: For the case of an instrumental variable Δz which is positively correlated with output and employment, a positive correlation of Δz and the Solow residual could be a sign of market power. The evidence on the failure of invariance of the cost-based Solow residual will suggest that increasing returns is an important part of the story; there is not enough profit in most industries to reconcile market power with constant returns.

2. The modified Solow residual

Invariance of the original Solow residual may fail because of market power, increasing returns, or other reasons. In the case of market power, achieved through government restrictions on entry or by other means, together with constant returns, the earnings of the firms should include an

element of monopoly profit as well as the normal return to capital. A second explanation for the failure of invariance is that entry is free, but the technology has increasing returns. Then the equilibrium will involve just enough market power to pay for the inputs. Earnings will not exceed the normal return to capital.

A very simple strategy can tell the two explanations apart. Under constant returns to scale (no fixed costs), the telltale cyclical behavior of the Solow residual should disappear once a simple modification is made in the computation of the residual. The modification is to measure labor's share in relation to cost rather than revenue. Because cost will be lower than revenue, the cost-based share will exceed the revenue-based share; the cyclical-ity of the Solow residual will vanish once a higher share is applied to labor growth. On the other hand, with fixed costs and free entry, revenue and cost will be the same, so the cost-based Solow residual will have the same cyclical behavior as the original revenue-based one. When the cost-based Solow residual has almost as large a failure of invariance as the original residual, it means that the technology has increasing returns.

Later I will show that the second approach yields an estimate of the index of returns to scale. These estimates exceed three in quite a number of industries. Fixed costs or other types of increasing returns appear to be an important feature of some industries.

Cost-based shares

As a prelude to the derivation of the modified, cost-based Solow residual, it is helpful to review some of the details of Solow's original derivation. Suppose output, Q , is produced by capital, K , and labor, N , in accord with the constant-returns production function, F :

$$Q = F(K, N). \quad (2.1)$$

Solow approximated the change in output as

$$\frac{\Delta Q}{Q} \doteq \frac{K}{Q} \frac{\partial F}{\partial K} \frac{\Delta K}{K} + \frac{N}{Q} \frac{\partial F}{\partial N} \frac{\Delta N}{N} . \quad (2.2)$$

As before, define lower-case letters as the logarithms of the corresponding upper-case variables. Then, still following Solow,

$$\Delta q \doteq \frac{K}{Q} \frac{\partial F}{\partial K} \Delta k + \frac{N}{Q} \frac{\partial F}{\partial N} \Delta n . \quad (2.3)$$

Let the firm be a price-taker in the capital services market at rental price r and in the labor market at wage w . Conditions for the minimization of cost are

$$\frac{\partial F}{\partial K} = \frac{r}{\lambda} \text{ and } \frac{\partial F}{\partial N} = \frac{w}{\lambda} . \quad (2.4)$$

Here λ is a Lagrangian interpreted as marginal cost. Solow assumed that λ was observable as the market price of output. Instead, I will treat λ as unobservable and use data on the service price of capital, which is not needed in Solow's approach. With r given, and with constant returns, it is possible to solve for the marginal products:

$$\frac{\partial F}{\partial N} = \frac{wQ}{rk + wN} = \alpha \frac{Q}{N} ; \quad (2.5)$$

$$\frac{\partial F}{\partial K} = \frac{rQ}{rk + wN} = (1 - \alpha) \frac{Q}{K} . \quad (2.6)$$

Here α is the share of labor cost, wN , in total factor cost, $wN + rK$. Solow measured the elasticity of output with respect to labor input by labor's share in revenue. That measure is appropriate under competition. In this approach, I measure the elasticity of output with respect to labor as labor's share in cost. No assumption of competition is required.

I can now state the basic relationship underlying the method:

$$\Delta q = \alpha \Delta n + (1 - \alpha) \Delta k + \theta \quad (2.7)$$

The percent change in output is the weighted percent changes in inputs. The weights for the inputs are the corresponding cost shares, α and $1-\alpha$. From equation 2.7 it is apparent that the cost-based Solow residual is invariant to exogenous changes in output. Only true shifts of the production function make the residual depart from zero. In comparison to the original Solow residual, the cost-based Solow residual has the important property that it measures the shift of the production function correctly in the presence of market power. Solow's original approach has the disadvantage of recording false movements of the production function for firms with market power, even with constant returns to scale. When revenue exceeds cost, because of pure monopoly profit, the revenue share of labor understates the elasticity of output with respect to labor input. When some exogenous event raises labor input relative to capital input, the revenue-based Solow residual fails to account for all of the increase in output, because it gives too little weight to labor.

Stochastic investment

The previous derivation assumed that there was an observed rental cost of capital, r , to which firms equated the marginal product of capital at all times. Because of lags between launching investment projects and putting new facilities into service, this assumption is unrealistic. The true cost share of labor, α , cannot be observed because it requires that capital cost be measured as the shadow cost of capital (capital's realized marginal product). But α has an observed counterpart,

$$\bar{\alpha} = \frac{wN}{rk + wN} \quad (2.8)$$

Here r is an observed cost of capital containing a random expectation error, ϵ . The expectation error arises from the lags in the investment process. The quantity of capital is set in advance, based on expectations of the demand schedule facing the firm and the interest rate and other determinants of the rental price of capital. The realized marginal product of capital differs by ϵ . Although ϵ has rational expectations properties, these properties cannot be exploited in this research, because ϵ and the random technology shift, θ , appear together.

Some algebra shows that the difference between the true labor share and the observed one is

$$\alpha - \bar{\alpha} = -(1 - \bar{\alpha})\alpha\epsilon \quad (2.9)$$

Then the relation among the observed variables is:

$$\Delta q = \bar{\alpha}\Delta n - (1 - \bar{\alpha})\Delta k + (1 - \bar{\alpha})\alpha\epsilon(\Delta n - \Delta k) + \theta \quad (2.10)$$

The factor $(1 - \bar{\alpha})\alpha$ is close to a constant. Suppose there is an instrumental variable that is a candidate for testing invariance—its movements do not cause changes in true productivity, θ , and changes in productivity do not cause its movements. Such an instrument is certainly correlated with the expectation error, ϵ , and with the change in the labor-capital ratio, $\Delta n - \Delta k$. Nonetheless, it is a reasonable identifying assumption that an instrument is uncorrelated with the product, $\epsilon(\Delta n - \Delta k)$, for the following reason. The product is positive in good times (when both its factors are positive) and is positive as well in bad times (when both its factors are negative). The instrument will be positive in good times and negative in bad times. Hence its correlation with the product will be close to zero. More generally, if the three random variables ϵ , $\Delta n - \Delta k$, and the instrument have a symmetric joint distribution, the correlation will be exactly zero, because the correlation is a third moment.

The Solow residual, when measured with the observed cost share, is

$$\Delta q - \bar{\alpha}\Delta n - (1-\bar{\alpha})\Delta k = -(1-\bar{\alpha})\alpha\epsilon(\Delta n - \Delta k) + \theta. \quad (2.11)$$

I have already argued that the other two terms should be uncorrelated with a properly chosen instrument. Consequently, the following result is established:

Proposition 2. Invariance of the cost-based Solow residual. Under constant returns to scale, the cost-based Solow residual is uncorrelated with an instrumental variable, irrespective of the amount of market power.

3. Value added

In addition to the labor and capital considered in the previous sections, firms use materials and other intermediate products as inputs to production. When time series data on other inputs are available, it is a simple matter to add additional terms to the Solow residual, each containing a factor share multiplying a rate of growth of an input. But it is also possible to make use of annual data on nominal and real value added in place of full input-output data. This section modifies the earlier analysis to deal with that case. In this section, variables with *s signify measures of the theoretical ideal: Q^* is true gross output, q^* is the log of Q^* , p^* is the actual price of output, β^* and α^* are the factor shares of materials and labor relative to the value of gross output, p^*Q^* , and θ^* is the rate of Hicks-neutral technical progress in the production function relating gross output to all inputs. Also, v is the price of materials, M is the quantity of materials employed and m is the log of the materials-capital ratio. With competition and constant returns, the Solow residual calculated with the full data is:

$$\Delta q^* - \alpha^* \Delta n - \beta^* \Delta m - (1 - \alpha^* - \beta^*) \Delta k = \theta^* \quad (3.1)$$

In the case at hand, the output measure that is available is not Q^* , gross output, but is Q , real value added. Nonetheless, it can be shown that the test based on the simple Solow residual, computed from real value added and employment growth, is a valid test. The rate of growth of the ratio of real value added is

$$\Delta q = \frac{\Delta Q}{Q} = \frac{p^* \Delta Q^* - v \Delta M}{p^* Q^* - v M} \quad (3.2)$$

$$= \frac{\frac{\Delta Q^*}{Q^*} - \frac{v M}{p^* Q^*} \frac{\Delta M}{M}}{1 - \frac{v M}{p^* Q^*}}$$

Recall that β^* is the share of materials in the value of gross output, vM/p^*Q^* . Thus

$$\Delta q = \frac{\Delta q^* - \beta^* \Delta m}{1 - \beta^*} \quad (3.3)$$

This relation can be used to eliminate the unobserved $\Delta q^* - \beta^* \Delta m$ from equation 3.1:

$$(1 - \beta^*) \Delta q - \alpha^* \Delta n - (1 - \alpha^* - \beta^*) \Delta k = \theta^* \quad (3.4)$$

or

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k = \theta \quad (3.5)$$

Here α is the labor's share in value added ($\alpha = \alpha^*/(1 - \beta^*)$) and θ is the rate of technical progress stated in labor-capital augmenting form ($\theta = \theta^*/(1 - \beta^*)$). Equation 3.5 says the following: Under competition and constant returns, the Solow residual calculated from value added will be equal to the rate of technical progress, appropriately defined. The covariance of the Solow residual with an exogenous instrument will be zero under competition and positive under market power. Thus we have

Proposition 3. Invariance of the Solow residual computed from value added: Under competition and constant returns, the Solow residual calculated from data on real value added as the measure of output and shares in nominal value added as measures of elasticities is uncorrelated with any instrument.

On the other hand, the cost-based Solow residual does not have an exact invariance property when calculated from value added, but invariance holds as a very close approximation. Consider equation 3.1 with the shares computed in relation to total cost, $wN + rK + vM$. Equation 3.3 becomes

$$\Delta q = \frac{\Delta q^* - \tilde{\beta} \Delta m}{1 - \tilde{\beta}} \quad (3.6)$$

where $\tilde{\beta}$ is the share of materials in revenue, vM/p^*Q^* . Then the cost-based Solow residual, using value-added data, is

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k = \beta^* \pi [\Delta m - \alpha \Delta n - (1 - \alpha) \Delta k] + \theta \quad (3.7)$$

Here π is the ratio of pure profit to value added,

$$\pi = \frac{p^*Q^* - wN - rK - vM}{p^*Q^* - vM} \quad (3.8)$$

Two comments are in order. First, the failure of invariance of the cost-based

Solow residual depends on the pure profit rate, π . If a firm or industry has little pure profit, then the failure of invariance of the cost-based residual cannot occur because of problems with the use of value added. The problems arise only in industries that earn profits in excess of all factor costs, including the market cost of capital. Second, even if profit is high, invariance fails only when materials do not move in proportion to output. If Δq and Δm are equal, then the term in brackets is itself the productivity residual. In that case, the residual is just a multiple of the true productivity shift, and so it will obey invariance. I summarize in

Proposition 4. Approximate invariance of the cost-based Solow residual computed from value added: If profit is low or materials are close complements to output, then the cost-based Solow residual computed from value added is close to uncorrelated with any instrument.

The discussion in this section made the implicit assumption that the change in real value added was computed each year using the previous year's prices as the base prices (see equation 3.2). In effect, it assumed the use of a Divisia index of real value added. In the U.S. national income accounts, base prices are changed about once a decade. I know of no reason to think that the low frequency of base changes has any important influence on the results obtained by the technique in this research.

4. Instrumental variables

The empirical results of my research and related work are of two types. The first show the failure of the invariance property by constructing Solow residuals and showing that they are positively correlated with instruments that themselves are positively correlated with output. The second measures the extent of market power and increasing returns by

estimating parameters that describe them.

The null hypothesis of invariance 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. The instrumental variables for the test should cause important movements in employment and output but be uncorrelated with the random fluctuations in productivity growth. Such exogenous variables could operate through product demand or through factor supplies. Lack of correlation with the random element of productivity growth involves two considerations: First, the instrument must not cause movements in productivity. Second, it must not respond to random variations in productivity growth.

It is a challenge to find instruments that are plainly exogenous under all views of macroeconomic fluctuations and which also have large enough influences on employment and output so that the test is powerful. Recent research has cast doubt on the exogeneity of all measures of monetary policy that are much correlated with output. On the fiscal side, only military spending is arguably unresponsive to the current state of employment and output. No single assumption is likely to appeal to all schools of thought about the relation between productivity growth and output fluctuations. A set of instruments suggested by Valerie Ramey has proven useful in this research.

Military spending

Military spending undergoes occasional large fluctuations that do not appear to be driven by the business cycle or by fluctuations in productivity. In addition, there is no reason to think that increases in government purchases of certain products should shift the production functions for the industries making those products, at least from one year to the next. Were

military spending sufficiently correlated with employment and output, it probably would be the most persuasive instrument for the purposes of this paper. In addition to government purchases of goods, which operate through product markets, changes in military employment help identify the equation through fluctuations transmitted via the labor market.

The world oil price

It is reasonable to assume that the historical pattern of shifts in the world price of oil has not been caused in any important way by fluctuations in U.S. productivity growth. The other part of the argument supporting the rate of change of the oil price as an instrument holds that shifts in oil prices do not cause changes in productivity. That hypothesis is more controversial. Its justification is that changes in factor prices do not shift production functions in the short run. Under this hypothesis, the observed tendency for measured productivity to fall when oil prices rise is the result of the negative response of output to that rise.

The political party of the President

Systematic differences in economic policies of the two political parties have caused differences in rates of expansion of the industries considered here, both over time and across industries. Output of services, durables, and regulated industries have risen noticeably faster under Democrats than under Republicans. Under the reasonable hypothesis that neither party has adopted policies that affect productivity growth in the short run, this systematic difference can be used to test the joint hypothesis of competition and constant returns.

5. Econometric method

Under the basic identifying hypothesis that true shifts in productivity are unrelated to movements of the instrumental variable, testing of the joint hypothesis of competition and constant returns is a simple matter of testing the hypothesis that the covariance of the Solow residual $\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k$ and the instrument Δz is zero. Although the test could be conducted with the raw covariance itself, an equivalent and more easily interpretable test is based on the regression coefficient of the Solow residual on the instrument. Thus, the tests are t -tests for the exclusion of the instruments from the regressions.

6. Data

I have obtained annual data for 7 one-digit industry groups and 26 industries at roughly the two-digit level for the years 1953-84. The industry detail is controlled by the labor input measure, which is an unpublished compilation of hours of work for all workers, including supervisory workers. The series are:

- Q : Real value added, 1982 dollars U.S. National Income and Product Accounts
- K : Net real capital stock, Bureau of Economic Analysis.
- p : Implicit deflator with indirect business taxes removed (ratio of nominal value added less IBT to real value added)
- N : Hours of work of all employees, U.S. NIPA
- w : Total compensation divided by N

Note that the data are chosen to eliminate tax wedges as a source of departures of marginal cost from price. The price level is measured net of sales and other taxes, and the wage is measured gross of social security, fringes, and other costs incurred by the employer. The industries chosen were the most detailed for which the NIPA report hours of all employees.

The only series used in the construction of the cost-based Solow residual that is not required for the original Solow residual 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:

$$r = (\rho + \delta) \frac{1 - k - \tau d}{1 - \tau} p_K . \quad (6.1)$$

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;
- 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.

The instrumental variables are:

Rate of increase of the world price of crude petroleum in dollars.

Rate of growth of military purchases of goods and services in real terms.

A dummy variable with the value of 1 when the President is a Democrat and zero when a Republican.

7. Summary of results

Table 1 summarizes the empirical evidence on the invariance of the original Solow residual and the cost-based modification. Much more detailed discussion of the same results appears in Hall [1988 and 1989].

TABLE 1

*Tests of invariance of original and cost-based Solow residuals:
marginal significance levels for one-digit industries*

Industry	Original			Cost-based		
	Military Spending	Oil Price	Political Party	Military Spending	Oil Price	Political Party
Construction	0.327	0.003 ^R	0.090	0.546	0.070	0.090
Durable Goods	0.500	0.029 ^R	0.357	0.272	0.018 ^R	0.175
Nondurable Goods	0.076	0.001 ^R	0.256	0.058	0.000 ^R	0.291
Transportation & Public Utilities	0.079	0.009 ^R	0.363	0.119	0.007 ^R	0.274
Trade	0.270	0.002 ^R	0.499	0.265	0.001 ^R	0.580
Finance, insurance & real estate	0.121	0.271	0.198	0.218	0.842	0.028 ^R
Services	0.193	0.082	0.043 ^R	0.040 ^R	0.769	0.092

Table 1 (continued)

Two-digit industries

Industry	Original			Cost-based		
	Military Spending	Oil Price	Political Party	Military Spending	Oil Price	Political Party
20: Food and kindred products	0.398	0.023 ^R	0.265	0.341	0.013 ^R	0.300
21: Tobacco manufacturing	0.087	0.231	0.366	0.334	0.268	0.242
22: Textile mill products	0.253	0.082	0.170	0.254	0.105	0.158
23: Apparel and other textile mill products	0.208	0.614	0.591	0.269	0.620	0.591
24: Lumber and wood	0.632	0.250	0.096	0.480	0.527	0.129
25: Furniture	0.043 ^R	0.063	0.447	0.036 ^R	0.126	0.486
26: Paper and allied products	0.191	0.004 ^R	0.271	0.073	0.008 ^R	0.367
27: Printing and publishing	0.068	0.081	0.500	0.136	0.013 ^R	0.440
28: Chemicals and allied products	0.184	0.001 ^R	0.291	0.104	0.000 ^R	0.392
29: Petroleum and coal	0.053	0.001 ^R	0.425	0.114	0.001 ^R	0.533
30: Rubber and misc. plastic products	0.285	0.237	0.246	0.607	0.052	0.135
31: Leather and leather products	0.141	0.494	0.161	0.160	0.631	0.140

Table 1 (continued)
Two-digit industries

Industry	Original			Cost-based		
	Military Spending	Oil Price	Political Party	Military Spending	Oil Price	Political Party
32: Stone, clay and glass products	0.357	0.002 ^R	0.379	0.177	0.005 ^R	0.565
33: Primary metals	0.155	0.341	0.221	0.272	0.369	0.233
34: Fabricated metals	0.265	0.092	0.304	0.075	0.001 ^R	0.143
35: Machinery, except electrical	0.748	0.065	0.624	0.311	0.041 ^R	0.242
36: Electric and electronic equipment	0.252	0.027 ^R	0.038	0.309	0.022 ^R	0.033
38: Instruments and related products	0.478	0.723	0.362	0.494	0.124	0.312
39: Miscellaneous manufacturing	0.452	0.144	0.075	0.195	0.001 ^R	0.076
48: Communication	0.356	0.216	0.674	0.341	0.144	0.665
49: Electric, gas and sanitary services	0.440	0.208	0.202	0.457	0.060	0.586
371: Motor vehicles and equipment	0.369	0.124	0.376	0.285	0.112	0.383
372-79: Other transportation equipment	0.455	0.557	0.669	0.688	0.024 ^R	0.441

The three columns on the left report the results of the joint tests of competition and constant returns, based on the original Solow residual. Each column corresponds to one of the three instruments. The three columns on the right report the results for the test of constant returns using the cost-based residual. An R means that the one-tailed test rejected at the 5 percent level in the direction of market power or increasing returns; that is, the covariance of the productivity residual and the instrument was significantly positive when the sign of the instrument was normalized so as to make its covariance with output growth positive.

In Table 1, it is evident that the results for the original and cost-based Solow residuals are very similar. The reason is simple. The difference between the two residuals depends on the level of pure profit—profit beyond the normal return to capital. Few industries have much pure profit, so the difference between the revenue share of labor and the cost share of labor is small. The absence of pure profit in most cases means that market power with constant returns is probably not a good explanation of the findings. If there is market power, it must be offset by fixed costs or other types of increasing returns.

The military spending instrument provides some evidence against invariance in most industries, though rarely at high levels of significance. In most industries, a change in demand induced by a change in military spending raises productivity if it raises output, and *vice versa*. This finding is inconsistent with the assumptions underlying the invariance theorems. The political party instrument, which also considers changes in product demand and labor supply, provides weak evidence against invariance in a number of industries. The oil price instrument yields the strongest evidence against invariance. In most of the one-digit industries and a large number of the two-digit industries, changes in the world oil price coincide with changes in productivity. Because factor prices do not shift production functions, this finding is a paradox within the assumptions of Solow's approach to productivity measurement.

8. Interpretation of rejection of the invariance hypothesis

Contrary to the predictions of theory, invariance of the Solow residual (either in original or cost-based form) fails. In most industries, expansions in response to outside forces involve a much larger increase in output than would be expected from the observed increase in labor input, based on the use of labor's share as an estimate of the elasticity of output with respect to labor input. This section considers a number of explanations of the failure of invariance.

Explanation 1: Market power

The basic idea that Solow exploited in his development of the total factor productivity measure was that the observed share of labor cost in revenue is an exact measure of the elasticity of the production function under competition. For firms with market power, the corresponding measure is the ratio of labor cost to output valued at marginal cost. That is, the share should be measured as wN/xQ , where x is marginal cost. Under competition, where the firm equates marginal cost to the market price, the two measures are the same. With market power, the share of labor cost in revenue understates the elasticity of the production function with respect to labor, because revenue includes monopoly profit.

To see what happens to the Solow residual with market power, let μ be the markup ratio or ratio of price to marginal cost, $\mu = p/x$, and let α be the share of labor cost in revenue, as before ($\alpha = wN/pQ$). Then the elasticity of the production function with respect to labor input is $\mu\alpha$ and the rate of growth of output can be decomposed as:

$$\Delta q_t = \mu_t \alpha_t \Delta n_t + (1 - \mu_t \alpha_t) \Delta k_t + \theta_t \quad (8.1)$$

Here I have written each of the variables with a time subscript to emphasize that they can change over time. No assumptions of constancy of either μ or α is made. In what follows, α_t will always be considered time-series data. Under the null hypothesis of competition, μ has the constant value of one, but there is no assumption of constancy under the alternative hypothesis.

The Solow residual under market power is:

$$\Delta q_t - \alpha_t \Delta n_t - (1 - \alpha_t) k_t = (\mu_t - 1) \alpha_t (\Delta n_t - \Delta k_t) + \theta_t \quad (8.2)$$

The residual is no longer invariant in the presence of market power. Any outside event that raises the labor/capital ratio will make the residual positive. To simplify the discussion of the empirical test based on an exogenous instrument, Δz , let me assume, without loss of generality, that the instrument is positively correlated with the weighted growth of the labor/capital ratio, $\alpha_t (\Delta n_t - \Delta k_t)$. Then, from equation 8.2,

Proposition 5. Market power and invariance: In the presence of market power, the covariance of an instrument and the Solow residual is positive.

In Hall [1988], I review the detailed conditions required to make Proposition 5 universal. First, if the markup ratio, μ_t , is a constant, it is immediately apparent that the covariance will be positive if and only if μ exceeds one. Second, the validity of the test based on the covariance extends to cases of variable markup ratios. In particular, if the markup varies along with the instrument in a linear fashion, and weighted employment growth also varies linearly with the instrument, then market power will make the covariance positive except in very unusual circumstances.

Estimates of the degree of market power

Under the additional hypothesis that the markup ratio, μ , is roughly constant, its value can be estimated by applying instrumental variables to equation 8.1. However, these estimates often have very large standard errors. High apparent dispersion will occur whenever an instrument is strongly correlated with output but weakly correlated with weighted factor input. The high dispersion does not convey any uncertainty about the failure of invariance—that hypothesis would require that the covariances of output and weighted input with the instrument be the same. Rather, the uncertainty is over how much greater than unity is μ . A more informative procedure is to estimate the reciprocal, μ^{-1} . By mapping all values of μ greater than one into the interval between zero and one, the procedure of estimating the reciprocal gives a much more interpretable estimate of the sampling variation of the estimate of μ .

Table 2 gives estimates of $1/\mu$ with their standard errors. These estimates make use of all three instruments together by using the two-stage least squares estimator. The interpretation of these estimates must heed the warnings of section 3 with respect to the use of data on value added. The implicit estimate of μ measures the ratio of price less materials cost (the valued-added deflator) to marginal cost excluding marginal materials cost. Such an estimate always overstates μ^* , the ratio of price to full marginal cost. The estimates of μ in Table 2 range from a little under 2 to a little under 4. That is, of the total value added per unit of sales, only 25 to 55 percent is marginal cost; the rest is earnings from market power. The deviations from invariance of the Solow productivity residual documented in Table 2 correspond to economically significant amounts of market power.

TABLE 2
*Estimates of markup ratio
at one-digit level*

Industry	Estimate of reciprocal	Standard error	Estimate of markup ratio, $\hat{\mu}$
Construction	.455	.103	2.196
Durable Goods	.486	.111	2.058
Nondurable Goods	.323	.102	3.096
Transportation & Public Utilities	.313	.119	3.199
Trade	.264	.109	3.791
Finance, insurance & real estate	.303	.167	3.300
Services	.536	.187	1.864

Estimates of markup ratio at two-digit level

20: Food and kindred products	.189	.144	5.291
21: Tobacco manufacturing	.362	.193	2.766
22: Textile mill products	.388	.160	2.578
23: Apparel and other textile mill products	1.213	.592	.824
24: Lumber and wood	.555	.223	1.801
25: Furniture	.506	.118	1.977

TABLE 2 (continued)
Estimates of markup ratio
at two-digit level

Industry	Estimate of reciprocal	Standard error	Markup ratio, $\hat{\mu}$
26: Paper and allied products	.269	.060	3.716
27: Printing and publishing	.070	.294	14.263
28: Chemicals and allied products	.050	.067	20.112
29: Petroleum and coal	-.007	.122	-139.478
30: Rubber and miscellaneous plastic products	.663	.249	1.508
31: Leather and leather products	.476	.337	2.100
32: Stone, clay and glass products	.394	.090	2.536
33: Primary metal industries	.460	.100	2.172
34: Fabricated metal products	.607	.232	1.649
35: Machinery, except electrical	.700	.265	1.429
36: Electric and electronic equipment	.324	.175	3.086
38: Instruments and related products	.716	.540	1.397

TABLE 2 (continued)
Estimates of markup ratio
at two-digit level

Industry	Estimate of reciprocal	Standard error	Markup ratio, $\hat{\mu}$
39: Miscellaneous manufacturing	.223	.130	4.491
48: Communication	.028	.998	36.313
49: Electric, gas and sanitary services	.079	.290	12.591
371: Motor vehicles and equipment	.567	.191	1.763
372-79: Other transportation equipment	1.053	.413	.950
Transportation	.251	.196	3.976
Wholesale Trade	-.271	.366	-3.688
Retail Trade	.425	.109	2.355

Instruments: defense expenditures, price of oil, and political party

The lower part of Table 2 presents estimates of $1/\mu$ for the more detailed industries. Not every industry shows evidence of market power. For example, in apparel (SIC 23), $1/\hat{\mu}$ is slightly, but not significantly, greater than one.

Explanation 2: Increasing returns

A firm operating at a point of increasing returns must have market power to be viable; absent some monopoly profit, it cannot generate enough revenue to pay for its inputs. Hence it is appropriate to consider how the original Solow residual moves when there is a combination of market power and increasing returns. Let γ_t be the returns to scale index, that is, the elasticity of output with respect to both inputs:

$$\gamma_t = \frac{K}{Q} \frac{\partial F}{\partial K} + \frac{N}{Q} \frac{\partial F}{\partial N} . \quad (8.3)$$

Under constant returns, $\gamma=1$. It is not hard to show that the Solow residual has an extra term for increasing returns ($\gamma>1$):

$$\Delta q_t - \alpha_t \Delta n_t - (1 - \alpha_t) k_t = (\mu_t - 1) \alpha_t (\Delta n_t - \Delta k_t) + (\gamma_t - 1) \Delta k_t + \theta_t \quad (8.4)$$

If an exogenous instrument is positively correlated with investment, invariance of the Solow residual will fail both because of market power (the $\mu-1$ term) and because of increasing returns (the $\gamma-1$ term). The asymmetry in the Solow residual in equation 8.4 results from the original asymmetry in the construction of the residual, which uses labor's share in revenue to infer the elasticity with respect to labor and assumes that the elasticity with respect to capital is one minus labor's share.

Behavior of the cost-based residual under increasing returns

The cost-based residual of an optimizing firm with power in its output market and a constant-returns technology should obey invariance. The residual uses the cost share to measure the elasticity of output with respect to labor input. In the presence of fixed costs or other failures of constant returns, the cost share understates the true elasticity. Then, as a result of the understatement, the cost-based residual would incorporate too small an adjustment for variations in labor input and the residual itself would rise every time output rose.

To understand the effect of increasing returns on the cost-based Solow residual, it is useful to restate the later steps of the derivation of that residual without the assumption of constant returns. As before, let γ be the elasticity of the production function with respect to all of its inputs; $\gamma > 1$ in the case of increasing returns. Then the marginal products of labor and capital are:

$$\frac{\partial F}{\partial N} = \frac{w\gamma Q}{rk + wN} = \alpha\gamma \frac{Q}{N} ; \quad (8.5)$$

$$\frac{\partial F}{\partial K} = \frac{r\gamma Q}{rk + wN} = (1-\alpha)\gamma \frac{Q}{K} . \quad (8.6)$$

Solve as before to get:

$$\Delta q = \gamma[\alpha\Delta n + (1-\alpha)\Delta k] + \theta . \quad (8.7)$$

The percent change in output is the weighted percent changes in inputs, multiplied by the returns-to-scale index, γ . The weights for the inputs are the corresponding cost shares, α and $1-\alpha$. If γ is roughly a constant, then equation 8.7 is an estimating equation: γ can be estimated as the ratio of the actual change in output, Δq , to the amount by which output should change

under constant returns, $\alpha\Delta n + (1-\alpha)\Delta k$, when some exogenous event changes product demand or factor supplies. Equation 8.7 adds a productivity growth term, θ , as in my earlier discussion of the original Solow residual.

Under increasing returns, The cost-based Solow productivity residual is

$$\Delta q - \alpha\Delta n - (1-\alpha)\Delta k = (\gamma-1)[\alpha\Delta n + (1-\alpha)\Delta k] + \theta \quad (8.8)$$

Under market power and constant returns to scale ($\gamma=1$), the cost-based Solow residual is invariant to exogenous changes in output. Only true shifts of the production function make the residual depart from zero. On the other hand, with increasing returns ($\gamma>1$), the residual is positive when output rises, even when there has been no shift of the production function. The residual confuses increases in scale with shifts of the production function. Thus,

Proposition 6. Increasing returns and invariance: In the presence of increasing returns, the covariance of an instrumental variable and either version of the Solow residual is positive.

The cost-based Solow residual has the important property that it measures the shift of the production function correctly in the presence of market power. Solow's original approach has the disadvantage of recording false movements of the production function for firms with market power, even with constant returns to scale. When revenue exceeds cost, because of pure monopoly profit, the revenue share of labor understates the elasticity of output with respect to labor input. When some exogenous event raises labor input relative to capital input, the revenue-based Solow residual fails to account for all of the increase in output, because it gives too little weight to labor.

Comparison of the original and cost-based Solow residuals

With increasing returns and market power, the original Solow residual is

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k = (\mu - 1) \alpha (\Delta n - \Delta k) + (\gamma - 1) \Delta k + \theta \quad (8.9)$$

With α measured as a cost share rather than a revenue share, the cost-based residual is

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k = (\gamma - 1) [\alpha \Delta n + (1 - \alpha) \Delta k] + \theta \quad (8.10)$$

The original residual fails invariance under either market power or increasing returns; the cost-based residual fails invariance only under increasing returns. To put it differently, the original residual can provide information about market power, but the cost-based residual can provide information only about increasing returns.

Evidence of the magnitude of increasing returns

Measuring the precision of the estimate of γ has the same problem as for μ —when the evidence is strong that γ has a high value, the standard error is very large. Hence it is appropriate to estimate γ^{-1} rather than γ . These estimates appear in Table 3. They show that the failure of constant returns in many industries is quite profound. The estimated elasticity of output with respect to total input, γ , is above 1.5 in all one-digit industries in Table 3 save services. In three industries—nondurables, transportation-utilities, and trade—output rises by more than three percent when an outside force makes input rise by one percent. In Table 3, 7 of the 23 industries have returns-to-scale indexes of greater than three.

TABLE 3
Estimates of the index of returns to scale at one-digit level

Industry	Estimate of reciprocal	Standard error	Estimate of index, $\hat{\gamma}$
Construction	.597	.277	1.675
Durable Goods	.543	.128	1.841
Nondurable Goods	.322	.110	3.107
Transportation & Public Utilities	.100	.169	10.030
Trade	.224	.178	4.468
Finance, insurance & real estate	.353	.250	2.830
Services	.926	.220	1.080

Estimates of the index of returns to scale at two-digit level

20: Food and kindred products	.030	.132	33.557
21: Tobacco manufacturing	.256	.358	3.909
22: Textile mill products	.500	.152	1.999
23: Apparel and other textile mill products	.933	.271	1.072
24: Lumber and wood	.725	.276	1.379

TABLE 3 (continued)
Estimates of the index of returns to scale at two-digit level

Industry	Estimate of reciprocal	Standard error	Estimate of index, $\hat{\gamma}$
25: Furniture	.736	.141	1.359
26: Paper and allied products	.208	.079	4.810
27: Printing and publishing	.384	.165	2.605
28: Chemicals and allied products	.007	.091	138.889
29: Petroleum and coal	-.309	.187	-3.236
30: Rubber and misc. plastic prod.	.606	.144	1.650
31: Leather and leather products	.212	.238	4.710
32: Stone, clay and glass products	.461	.109	2.170
33: Primary metal industries	.351	.117	2.852
34: Fabricated metal products	.230	.249	4.352
35: Machinery, except electrical	.681	.168	1.469
36: Electric and elec- tronic equipment	.447	.166	2.237

TABLE 3 (continued)
Estimates of the index of returns to scale at two-digit level

Industry	Estimate of reciprocal	Standard error	Estimate of index, $\hat{\gamma}$
38: Instruments and related products	.474	.358	2.111
39: Misc. manufacturing	.182	.156	5.491
48: Communication	.834	.736	1.199
49: Electric, gas and sanitary services	.496	.206	2.016
371: Motor vehicles and equipment	.382	.189	2.621
372-79: Other trans- portation equipment	.886	.123	1.129

Instruments: defense expenditures, price of oil, and political party

Explanation 3: External technical complementarities

Although my discussion has considered increasing returns in the firm's own technology, the invariance properties will also fail when the firm's technology has constant returns but there is an externality that makes one firm's output complementary with other firms' output. The thick-market externality discussed by Diamond [1982] is a leading example. The extremely uneven geographical distribution of economic activity suggests that thick-market effects are strong. Efficiency is greater in places and at times when suppliers, workers, and customers are dense. The overall technology of an industry with a thick-market externality will have increasing returns even though each firm has constant returns. Caballero and Lyons [1989] use methods based on the Solow residual to measure the importance of externalities across U.S. manufacturing industries. They conclude that externalities rather than increasing returns within industries are the most important source of the failure of the invariance of the Solow residual.

Thick-market effects internal to the firm may be an important source of increasing returns for the firm, as well. Consider a package delivery service. When its customers become more numerous, its operations become more efficient because each truck can make more stops in each area, and deliver more packages per mile of driving.

Explanation 4: Chronic excess capacity

If firms consistently hold more than the optimal amount of capital, invariance of the original Solow residual will hold but invariance of the cost-based residual will fail. 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. The revenue share of a

competitive, constant-returns firm or industry will measure the elasticity correctly. Some theories of the strategic interaction of firms have suggested the desirability of capacity above the cost-minimizing level for the realized distribution of output. Excess capacity makes credible a threat to revert to competition.

Explanation 5: Unmeasured fluctuations in work effort and hours

A significant problem of measurement of labor input arises if labor has two dimensions, hours and effort. Suppose, for concreteness, that hours, h , and effort, f , multiply to form labor input, n . If fluctuations in effort are ignored in computing Δn , invariance of the Solow residual will occur even with competition and constant returns. The residual with a correct measure of the change in labor input measures technical progress correctly:

$$\Delta q - \alpha \Delta n - (1-\alpha) \Delta k = \theta \quad (8.11)$$

Suppose that only a fraction ψ of variations in labor input take the form of changes in hours; the rest take the form of variations in work effort. Then $\Delta h = \psi \Delta n$ and the Solow residual based on hours is not a proper measure of productivity growth:

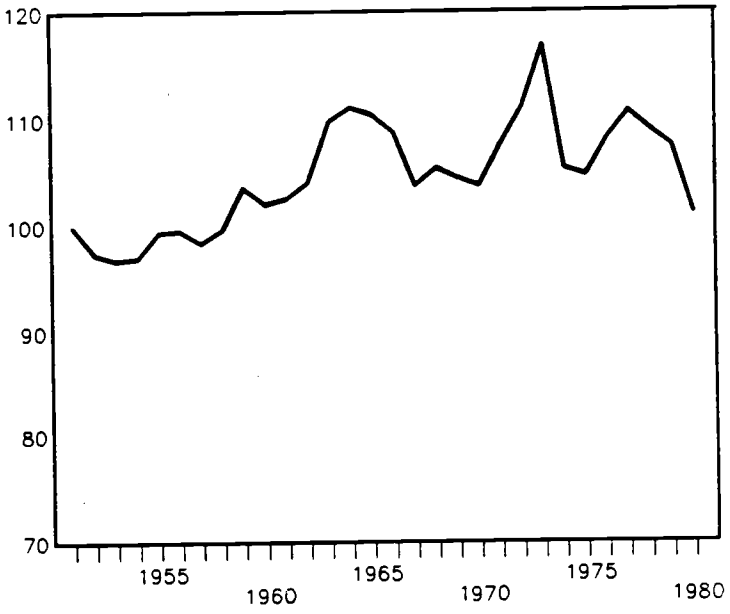
$$\Delta q - \alpha \Delta h - (1-\alpha) \Delta k = \frac{1-\psi}{\psi} \alpha \Delta h + \theta \quad (8.12)$$

The residual based on hours rather than total labor input will have a positive covariance with an exogenous instrument because of the extra term on the right-hand side. The error in measuring labor input leads to a false rejection of invariance.

Figure 1 shows the magnitude of the fluctuations in work effort needed to explain the measured fluctuations in work effort per hour needed to

Figure 1 Work effort implied by invariance

Index of effort per hour



explain the measured fluctuations in productivity in a competitive, constant-returns setting. I calculated the effort index by solving equation 8.11 for the change in total labor input under the assumption that θ is a constant. Then I subtracted the measured change in hours to get the implied change in work effort. According to Figure 1, work effort would have to have been more than 10 percent above normal for three successive years in the mid-1960s, as one example of the size of the unmeasured fluctuations in work effort needed to explain the observed movements of the Solow residual.

One piece of direct evidence suggests that cyclical fluctuations in work effort is not an important explanation of fluctuations in measured productivity. Fay and Medoff [1985] surveyed almost two hundred managers of manufacturing plants. They asked the managers whether the work effort of blue collar workers increased or decreased when output fell from a peak to a trough. Somewhat more of the managers reported an *increase* in work effort during a slump than reported a decrease.

The work effort variable, f , can be interpreted as accomplishments per hour. Under this interpretation, the assumption that output depends on the product of accomplishments per hour and hours of work says that the unit of labor input is the accomplishment. In a competitive labor market under this assumption, workers would be paid a piece rate per accomplishment equal to the marginal value of an accomplishment. One of the ways of appraising the unmeasured-work-effort explanation of the finding of positive covariances of the Solow residual with instruments is to ask about its implications for labor supply. With the piecework technology, firms are indifferent between various combinations of hours and effort that yield the same volume of accomplishments. The split between hours and effort is based purely on the preferences of workers. In this setting, the parameter ψ is interpreted as the ratio of the elasticity of the supply of hours with respect to the piece rate to the elasticity of the supply of accomplishments with respect to that wage. A value of ψ of 0.5, for example, means that a decline in the piece rate brings equal percentage declines in effort and hours. Workers with those preferences

respond to lower piecerates by working less intensively. They could reduce their hours twice as much by continuing the same level of effort, but instead they choose more leisure on the job.

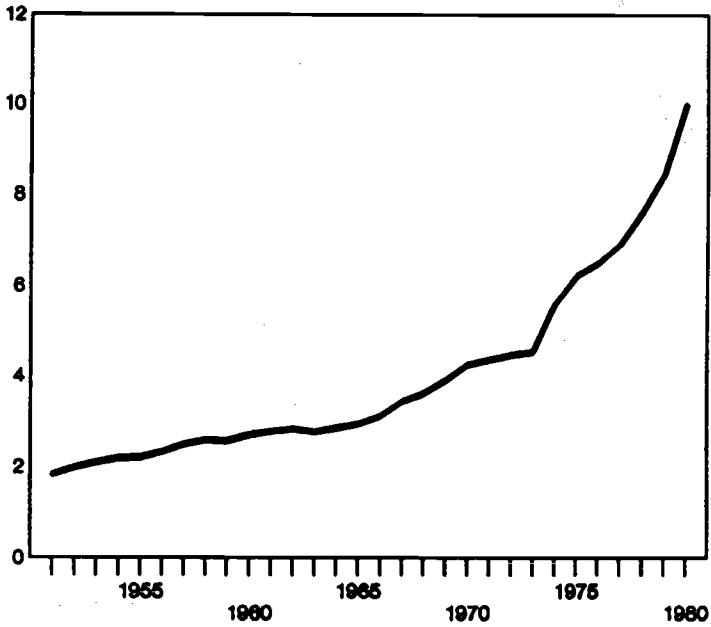
Figure 2 shows that the unmeasured-work-effort explanation of the failure of the invariance of the Solow residual must rely heavily on the theory of wage smoothing. The figure shows the actual hourly wage and the hourly wage computed as the ratio of compensation to the adjusted measure of labor input underlying Figure 8.1. The inferred wage is to be interpreted as compensation per accomplishment. Movements of the inferred wage are implausible. In the expansion of the mid-1960s, the inferred piecerate actually declined; in the highly inflationary expansion of the early 1970s, it remained level. It is highly unlikely that the market-clearing piecerate wage moved along the inferred path. Rather, the unmeasured-work-effort explanation of the failure of invariance must assert that compensation is decoupled from hours of work or work effort. The bulges of intense work effort in Figure 8.1 were not paid for on a current basis by employers. Instead, workers provided the extra labor input in accord with long-term agreements, if the unmeasured-work-effort story is to be believed.

Measurement error in hours

Purely random measurement errors in the change in labor input, Δn , uncorrelated with the instrumental variable, do not lead to rejection of invariance of the Solow residual. However, the hypotheses that spring to mind about errors in Δn suggest they would be negatively correlated with Δz . Suppose, for example, that some workers always report 40 hours of work per week even though they work more hours when demand is strong and fewer when it is weak. The correlation is plainly negative. Such a negative correlation could explain the finding of positive covariances of the Solow residual with the instrument, because labor growth is subtracted in calculating the residual. In formal terms, if $\hat{\Delta n}$ is an erroneous measure of

Figure 2. Wage adjusted for work effort

Hourly wage, current dollars



Δn , such that a fraction $1-\psi$ of movements in Δn are omitted from $\Delta \bar{n}$, then the situation is formally the same as in the previous case of unmeasured fluctuations in work effort. An industry with constant returns and competition would be diagnosed as having increasing returns and market power, in that situation, when in fact the problem was the understatement of fluctuations in labor input. All of my earlier remarks on unmeasured fluctuations in work effort also apply to measurement errors in hours of work: The measurement errors would have to be quite large in order to explain the movements of the Solow residual, and it is untenable that workers are being paid on a current basis for the extra hours they work that are not recorded in the data.

The likely source of errors in measuring total employee-hours is presumably in hours per worker, rather than in the count of workers. However, changes in the number of workers account for most of the variation in employee-hours. In the data used in this paper for non-durables, the standard deviation of the annual first difference of weekly hours per worker is just 1.2 percent. On the other hand, the standard deviation of the measurement error in hours needed to eliminate fluctuations in the Solow residual is 5.0 percent. In order to interpret the movements of the residual as caused by measurement errors in hours of work, rather than a failure of competition and constant returns, the unmeasured fluctuations in hours would have to dwarf the measured fluctuations, which I find implausible.

A second reason to be skeptical of major measurement errors in hours of work is that the data make maximal use of all available information on hours. For workers paid by the hour, the data rely on data reported by employers. For workers paid on a salary basis (without reporting of hours), data on hours are taken from the Current Population Survey data from households.

Explanation 6: Errors in measuring capital

An important implicit assumption of this line of work is that capital input is correctly measured. The measure of capital used 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 and let $\Delta \bar{k}$ be the change in measured capital ($\Delta \bar{k} = \Delta k + \Delta v$). Then the Solow residual, calculated with measured rather than actual capital, under constant returns, will be:

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta \bar{k} = -(1 - \alpha) \Delta v \quad . \quad (8.13)$$

Because capital measurement errors are likely to be negatively correlated with output changes (an increase in output raises unmeasured capital utilization and lowers v), the errors are likely to contribute to a failure of the invariance condition in the direction found in this paper. For example, suppose that the change in capital measurement error is proportional to the change in labor input per unit of measured capital,

$$\Delta v = -\phi(\Delta n - \Delta \bar{k}) \quad . \quad (8.14)$$

Strict complementarity of work hours and capital hours would mean that ϕ had the value of one. Then the Solow residual is

$$\Delta q - \alpha \Delta n - (1 - \alpha) \Delta \bar{k} = \phi(1 - \alpha)(\Delta n - \Delta \bar{k}) \quad (8.15)$$

Estimation of ϕ by instrumental variables answers the following question: What magnitude of measurement error would be required to explain the observed failure of the invariance condition under constant returns? The answer turns out to be a very large magnitude, well above the intuitive maximum of $\phi=1$. For nondurables, the results of estimation with the three instruments are:

$$\Delta q - \bar{\alpha} \Delta n - (1 - \bar{\alpha}) \Delta k = .0549 - 5.034 (1 - \alpha)(\Delta n - \Delta \bar{k}) \quad (8.16)$$

(.012) (1.606)

In order to explain the magnitude of the correlation of the Solow residual with the instruments, the elasticity of the measurement error with respect to the change in labor input must be implausibly large—around 5. The simple model in which capital and labor fluctuate in proportion, with $\phi=1$, is not nearly enough to explain the findings of the research.

Explanation 7: Errors in measuring output

The hoarding of labor during cyclical contractions is probably an important element of the explanation of cyclical fluctuations in productivity, though, as I will show later, such an explanation almost certainly involves a failure of competition and constant returns. However, labor-hoarding could induce a measurement error in output that would cause the method of this paper to overstate the extent to which invariance of the Solow residual fails. Specifically, hoarded workers may be put to work on projects other than the

production of output. They may repair equipment, build new facilities, train themselves or others, and engage in many other investment activities. Though the NIPA data attempt in principle to include some of these items in output, many are no doubt unmeasured.

Fay and Medoff [1985] document the importance of investment activities during slumps in output. In the typical contraction suffered by their respondents, output fell by 31 percent, while the increase in investment activities by workers was about 3 percent of their normal hours. The resulting upward bias in $\hat{\gamma}$ is the ratio of these two numbers, or about 0.1. An industry with constant returns with the amount of unrecorded output found by Fay and Medoff would have an estimated returns to scale index of 1.1, far below the estimates reported earlier. Only a small fraction of the observed failure of invariance of the Solow residual can be attributed to unobserved output.

Explanation 8: Monopsony power in the labor market.

A basic maintained hypothesis of this research is that the firm chooses an optimal level of employment. All of the derivations make the assumption that the marginal revenue product of labor is equated to the wage. An alternative is that the firm employs too few workers, on the average. Then the measured cost share of labor would understate the true elasticity because of the understatement of effective labor cost, and the covariance of the cost-based residual and an instrument would 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 lump-sum 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. 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.

Non-explanations

Although quite a variety of conditions might explain the widespread failure of invariance of the Solow productivity residual, some of the conditions that come to mind are actually incapable of explaining the finding. Among these are labor hoarding and overhead labor, wage smoothing and other cyclical errors in measuring wages, adjustment costs for labor input.

Non-explanation 1: Overhead labor and labor hoarding

A number of authors have suggested that labor hoarding is an alternative explanation of the failure of invariance of the Solow residual. That suggestion is incorrect. The overhead labor technology, which is the simplest model that generates labor hoarding, fits all of the assumptions of the invariance theorems including constant returns. Hence the Solow residual is invariant for a cost-minimizing firm with an overhead labor technology. Of course, residuals for industries with obvious signs of labor hoarding frequently fail invariance. The correct conclusion is that one of the assumptions underlying the invariance theorems does not hold. The leading possibility, in

my opinion, is increasing returns. A second possibility is that firms do not minimize expected costs and instead hold excess capacity. Data problems are a third.

The following example will show how the original Solow residual is invariant for an overhead labor technology. A price-taking firm has capacity K . In order to produce any output at all, it must hire λK overhead workers. In addition, for each unit of output, it must hire ϕ workers. Thus, to produce a level of output Q , it must have a K at least as large as Q , and employment of $\lambda K + \phi Q$. The firm's marginal cost is $w\phi$ whenever $Q < K$ and can be taken to be any number above $w\phi$ when the capacity constraint is binding. In competitive equilibrium, $p = w\phi$ whenever $Q < K$ and $p \geq w\phi$ whenever $Q = K$. Note that the technology has constant returns to scale (the fixed component of labor is proportional to capacity, not absolutely fixed) so a competitive equilibrium is possible. Now consider the Solow residual for a period when output is below capacity. Labor's share will be

$$\alpha = \frac{wN}{pQ} = \frac{w(\phi Q + \lambda K)}{\phi w Q} \quad (8.17)$$

Because the competitive firm operates at a loss whenever its output is below capacity, the share exceeds one. The Solow residual is:

$$\begin{aligned} \Delta q - \alpha \Delta n &= \frac{\Delta Q}{Q} - \alpha \frac{\Delta N}{N} \\ &= \frac{\Delta Q}{Q} - \frac{w(\phi Q + \lambda K)}{\phi w Q} \frac{\phi \Delta Q}{\phi Q + \lambda K} \\ &= 0 \end{aligned} \quad (8.18)$$

Thus the Solow residual remains unchanged when an outside force alters the levels of output and employment. The covariance of the Solow residual and an exogenous instrument is zero and the proposed test will reveal, correctly, that the firm is competitive. Even though the variation in labor input itself

may be very small, because most workers are overhead workers, the competitive value of α exceeds one by enough to make $\alpha\Delta n$ equal Δq . The mere existence of overhead labor does not lead to the rejection of competition.

In practice, for those industries that appear to have large overhead labor requirements and small variable labor requirements, the behavior of the labor share α and the resulting covariance of the Solow residual and an instrument are not at all what is described by the competitive model just summarized. Rather, when such an industry operates below capacity, its price remains far above the cost of the variable component of labor. Profit often remains positive, so labor's share, α , is less than one. The ratio of Δq to $\alpha\Delta n$ is, say, three, not one. The appropriate conclusion is that price exceeds marginal cost and there may be increasing returns as well. The Solow residual rises sharply whenever an outside force causes employment and output to rise.

The most widely advocated explanation of the positive correlation of output and productivity is that firms carry workers through slumps, because discharging them would dissipate the value of their job-specific human capital. In a simple version of the labor-hoarding model, the firm would lay off only $\phi\Delta Q$ of its workers, if a slump caused by adverse external developments caused output to fall by ΔQ . Additional workers would be kept on even though they were idle. The economics are then identical to the first example. Marginal cost is $w\phi$ and the competitive price should fall to this level. Then α will be well above one, so that the Solow residual is invariant to a shift in employment and output, even though the change in output is much larger than the change in employment. However, if the firm is not competitive, so that price does not fall enough to make α large, then the Solow residual will rise when an outside force raises employment and output.

A related result is that invariance of the cost-based Solow residual holds for the noncompetitive firm with a constant-returns overhead labor technology. The following example illustrates the point. Suppose that the

technology is such that the level of employment required to produce output Q is $\lambda K + \phi Q + \rho \max(Q - K, 0)$. That is, with a capital stock of K , overhead labor of λK , and variable labor ϕQ , it is possible to produce $Q \leq K$ units of output. Additional output requires an increment of $\phi + \rho$ 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 $(\rho - \lambda)w$ when Q exceeds K —in that regime, more capital requires more overhead workers but reduces the requirement for the incremental labor described by ρ . Let β be the probability that output will exceed K . Then the expected shadow value of capital is $(\beta\rho - \lambda)w$. At the optimum capital stock, the expected shadow value of capital equals the service price of capital, r . Hence, $\beta = (r/w + \lambda)/\rho$. Suppose that the fluctuations in output are in a small region above and below K , and, for convenience, scale output and capital so that $K=1$. The level of employment is close to $\lambda + \phi$. The cost share, α , will be close to $(\lambda + \phi)/(\lambda + \phi + r/w)$. 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 $\phi/(\lambda + \phi)\Delta Q$ and the cost-based residual is equal to $[1 - \alpha\phi/(\lambda + \phi)]\Delta q$. Thus the relation between the residual and the instrument has slope $1 - \alpha\phi/(\lambda + \phi)$. When output is above K , the change in employment, ΔN , is $(\phi + \rho)\Delta Q$. Hence, the rate of growth of employment, Δn , is approximately $(\phi + \rho)/(\lambda + \phi)\Delta q$. The slope of the relation between the cost-based residual $\Delta q - \alpha\Delta n$ and Δq is $1 - \alpha(\phi + \rho)/(\lambda + \phi)$. The average slope is $(1 - \beta)[1 - \alpha\phi/(\lambda + \phi)] + \beta[1 - \alpha(\phi + \rho)/(\lambda + \phi)]$. Inserting the values for β and α 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 (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. As a general matter, the condition for optimal choice of capacity guarantees the invariance of the cost-based Solow residual.

Labor hoarding and overhead labor are probably important phenomena in many industries. 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 research. However, labor hoarding is not an alternative explanation of the failure of the invariance property. Fixed costs or other types of increasing returns are likely to underlie chronic operation in a labor-hoarding regime. 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.

Cyclical errors in measuring labor's share

Errors in measuring the value of α that are correlated with the instrument but which do not affect the mean value of α are benign in this framework. Examples of measurement errors with this character are (1) payment of workers under wage smoothing arrangements, where the wage equals the long-run opportunity cost of time, but does not track short-run fluctuations in labor-market conditions, (2) adjustment costs in employment, where the full marginal cost of incremental hours of work fluctuates above and below the observed wage, and (3) price rigidity, where prices are set at the long-run average of marginal cost.

Under wage smoothing, workers receive less than their marginal products in good times and more in bad times. Hence the share α is understated in good times and overstated in bad times. When the instrument is positive and consequently output growth and employment growth are also positive, the Solow residual measured with too small an α is also positive. A positive term enters the covariance of the instrument and the residual. On the other hand, when the instrument, output growth, and employment growth are all negative, the Solow residual is positive as well--the term $-\alpha\Delta n$ is overstated in a positive direction because Δn is negative. A negative term enters the covariance of the instrument and the residual. In data with an approximately equal mixture of good and bad times, the covariance will turn out to be zero. That is, a competitive industry with wage smoothing will not generate data that reject the invariance property.

Adjustment costs in employment have the same character as wage smoothing. Half the time, the shadow cost of labor to the firm exceeds the wage, and the measured value of α understates the true value. These are times when the current change is in a direction that adds to adjustment costs. The other half of the time, the shadow cost of labor falls short of the wage because the current change in employment conserves adjustment costs. There is no bias in labor's share, α , in the long run, but there are measurement errors correlated with the instrument. But the errors cancel out and there is no reason to expect to find a correlation of the Solow residual and the instrument in a competitive industry with labor adjustment costs.

Price rigidity could arise in a competitive industry if firms find it necessary to post prices before observing current demand. If firms stand ready to serve all demand, then the same type of symmetry prevails as described earlier. When demand is strong, α is overstated because the pQ in the denominator understates the true value of output based on marginal cost. When demand is weak, α understates labor's share in marginal cost. But there is no resulting correlation of the Solow residual and an instrument correlated with demand. Rotemberg and Summers [1987] examine this case in

more detail. They show that a positive covariance of the Solow residual and an instrument would arise if the firm's behavior is asymmetric, serving all demand in the low demand states but rationing output if marginal cost exceeds the predetermined price.

A general argument shows that none of these specification errors could explain the failure of the invariance of the Solow residual. The first two errors make the measured wage differ from the true effective wage over the cycle, but not in the long run. Such errors contribute to failure of invariance only in a certain second-order way, and, in any case, the bias is toward competition and constant returns. Similarly, price rigidity creates only a tiny bias against invariance and could not explain an important element of the observed failure of invariance.

Suppose that all of the assumptions hold that are needed to ensure the invariance of the Solow residual. Then the true Solow residual, $\Delta q - \alpha \Delta n - (1 - \alpha) \Delta k$, will be uncorrelated with the instrument, Δz . Now suppose that an erroneous labor share, $\hat{\alpha}$, is used in place of α in forming the Solow residual. It may differ from α because of wage smoothing, adjustment costs, price rigidity, or any other reason. I will assume that $\hat{\alpha}$ is constant over time, which is generally a close approximation. Then the covariance of the residual and the instrument is

$$\text{Cov}[\Delta q - \hat{\alpha} \Delta n - (1 - \hat{\alpha}) \Delta k, \Delta z] = -\text{Cov}[(\hat{\alpha} - \alpha)(\Delta n - \Delta k), \Delta z] \quad (8.19)$$

Consider the following representation of the factors in the covariance:

$$\hat{\alpha} - \alpha = -b_1 x + \eta_1 \quad (8.20)$$

$$\Delta n - \Delta k = -a_2 + b_2 x + \eta_2 \quad (8.21)$$

$$\Delta z = b_3 x + \eta_3 \quad (8.22)$$

Here x is the systematic part of the instrument. It has mean zero and variance one, without loss of generality. I also assume that x has zero third

moment. A sufficient condition for the last property is that the distribution of x is symmetric. The random variables η_i have zero means and are independent of each other and of x . The measurement error in the labor share, $\hat{\alpha} - \alpha$, has a negative correlation with x , controlled by the coefficient b_1 . The rate of growth of the labor/capital ratio, $\Delta n - \Delta k$, has a negative mean and a positive correlation with x , controlled by the coefficient, b_2 . Under these assumptions, the covariance of the Solow residual and the instrument is

$$-Cov\{(\hat{\alpha} - \alpha)(\Delta n - \Delta k), \Delta z\} = -a_2 b_1 b_3 . \quad (8.23)$$

All of the other terms vanish because they are third moments, which have been assumed to be zero. The one remaining term is a second moment times a first moment. The basic point can be seen most easily by assuming that the mean growth of the labor/capital ratio, $\Delta n - \Delta k$, a_2 , is zero. In that case, when demand rises and x is positive, the error in the labor share, $\hat{\alpha} - \alpha$, is negative. Growth in the labor/capital ratio is also negative, so the product, $(\hat{\alpha} - \alpha)(\Delta n - \Delta k)$, is positive. The instrument, Δz , is positive, so the contribution to the covariance is positive. On the other hand, when demand falls and x is negative, the error in the labor share, $\hat{\alpha} - \alpha$, is positive. Growth in the labor/capital ratio is also positive, so the product, $(\hat{\alpha} - \alpha)(\Delta n - \Delta k)$, is positive. The instrument, Δz , is negative, so the contribution to the covariance is negative. On net, the covariance is zero.

When the average growth of the labor/capital ratio is negative ($a_2 > 0$), the covariance of the Solow residual and the instrument is negative. Rejection of invariance in the form of a positive covariance with an instrument cannot be explained by cyclical measurement errors in the labor share that are negatively correlated with the instrument, that is, understatement of the labor share when demand is strong. Even measurement errors that overstate the labor share when demand is strong have only a slight effect on the covariance, because in all industries, the trend growth rate of the labor/capital ratio, a_2 , is quite small. I summarize in

Proposition 7. Irrelevance of cyclical measurement errors: Under the assumptions stated above, cyclical errors in measurement of the labor share cannot cause a positive covariance of the Solow residual with an instrument.

Non-explanation 2: Wage smoothing

I will consider two sources of measurement error in the wage. For the revenue-based share, the proportional error in the share equals the proportional error in the wage, assuming the other elements of the share to be measured correctly. For the cost-based share, the proportional error in the share is smaller than the proportional error in the wage, because the wage appears in both the numerator and the denominator.

One potential source of measurement error in the wage is wage smoothing. Martin Neil Baily's [1974] pioneering paper pointed out the advantage to workers of earning smoothed wages. When workers cannot use credit markets as easily as employers can, then it makes sense to decouple earnings from labor-market fluctuations. In the extreme case, workers receive a predetermined real annual income, unrelated to the amount of work they do and unrelated to the value of their time. Though such an arrangement could be examined with the aid of Proposition 7, I have taken a less extreme view. Suppose that workers receive a guaranteed hourly wage, but the wage is paid only for hours actually worked. Those hours are determined not by equating the value of the marginal product of labor to the wage, but rather by equating the value of the marginal product to the marginal value of time. The error in measuring the share arises because the effective wage is the value of time, but the measured wage is the predetermined contract wage. However, if the market for employment contracts is competitive, the smoothed wage will equal the marginal value of time averaged over the

duration of the typical contract.

When an increase in product demand raises output and employment, it raises the value of time and hence raises the true share of labor. A share computed from the smoothed wage understates the movement of the true share; this is true both for the revenue- and cost-based shares. Proposition 7 applies—the consequence is a slight downward bias in the covariance of the Solow residual and the instrument. The cyclical error in the labor share does not bias the covariance.

Non-explanation 3: Adjustment costs

A second potential source of cyclical departures of the effective marginal cost of labor from the quoted wage could arise from adjustment costs for labor. Mark Bilal [1987] has investigated the ways that adjustment costs enter the effective marginal cost. In a simple setup where costs of adjustment are quadratic in the proportional change in effective labor input, the marginal cost of adding an hour of labor input in year t is

$$w_t = \hat{w}_t(1 + g\Delta n_t - g\Delta n_{t+1}). \quad (8.24)$$

Here \hat{w} is the measured hourly wage and g is a parameter with the following interpretation: If g is 1, at a time when employment has risen by 10 percent, the marginal adjustment cost of labor is 1/10 of the direct labor cost.

With this specification of adjustment costs, the error in the labor share is

$$\hat{\alpha} - \alpha = -g\hat{\alpha}(\Delta n_t - \Delta n_{t+1}). \quad (8.25)$$

In almost all industries, the serial correlation of Δn_t is negative, so the covariance of $\Delta n_t - \Delta n_{t+1}$ and the instrument is almost certainly positive. Thus the error in the labor share is negatively correlated with the instrument

and Proposition 7 shows that there is a slight downward bias in the covariance of the instrument with the Solow residual. Adjustment costs cannot explain the finding of a positive covariance.

Non-explanation 4: Price rigidity

A good deal of research has reached the conclusion that product prices fluctuate less as demand changes than is predicted by the competitive model. Do rigid prices explain the failure of the invariance of the Solow residual? The answer is no. If price rigidity is itself unbiased—if marginal cost spends as much time below price as above—then the Solow residual will be uncorrelated with an exogenous demand instrument. Recall that the competitive labor share is

$$\frac{wN}{pQ} = \alpha \quad (8.26)$$

Substitute $Y=N/Q$, the labor/output ratio, and solve for the price:

$$p = \frac{1}{\alpha} wY \quad (8.27)$$

This is the “price equation” under competition—the price is proportional to cost, measured by the wage, w , and increases with the labor/output ratio. Under a simple characterization of price rigidity, the variable labor/output ratio, Y , is replaced by a long-run average, \bar{Y} :

$$\hat{p} = \frac{1}{\alpha} w\bar{Y} \quad (8.28)$$

Under the assumption that the firm supplies whatever quantity of output is demanded by its customers at the price \hat{p} , the proper measure of the labor share for the Solow residual is still α —the price p is marginal cost. If the residual is calculated mistakenly with the share $\hat{\alpha}$ based on \hat{p} , it is easy to show that the

resulting error in the share is

$$\hat{\alpha} - \alpha = \alpha \frac{Y - \bar{Y}}{\bar{Y}} . \quad (8.29)$$

Note that this error in the share is positively correlated with the demand instrument, so the Solow residual will be positively correlated with the instrument, even with competition and constant returns. However, the correlation is very small.

To keep the discussion simple, I will assume that the labor share can be treated as a constant. The natural metric for the covariance of the instrument and the residual is the covariance of the instrument and the weighted change in the labor/capital ratio, $\alpha(\Delta n - \Delta k)$, because the ratio of the two covariances is the estimate of $\mu - 1$. In terms of the representation presented earlier in this section, the ratio is

$$-\frac{a_2 b_1}{\alpha b_2} . \quad (8.30)$$

Some manipulations reveal that

$$b_1 = -Cov(\alpha \Delta y, \Delta x) \quad (8.31)$$

$$b_2 = Cov\left(\frac{\Delta y + \theta}{1 - \alpha}, \Delta x\right) \quad (8.32)$$

where $y = \log Y$. Let

$$C_i = Cov(y_t, \Delta x_{t-i}) \quad (8.33)$$

Then

$$\frac{b_1}{b_2} = -\alpha(1 - \alpha) \frac{C_0 - C_{-1}}{2C_0 - C_{-1} - C_1} \quad (8.34)$$

If the two cross-covariances, C_{-1} and C_1 , are close to each other and smaller than C_0 , then

$$\frac{b_1}{b_2} \doteq -\frac{1}{2}\alpha(1-\alpha) \quad (8.35)$$

and the bias in the estimate of μ is

$$\frac{1}{2}(1-\alpha)a_2 \quad (8.36)$$

This bias is tiny. For example, if α is 0.7 and a_2 is .02, the bias is 0.003. Price rigidity cannot explain any meaningful amount of the failure of invariance.

9. *Subsequent research*

The results reported here are strongly confirmed by subsequent research in the framework developed here carried out by Domowitz, Hubbard, and Petersen [1987]. D-H-P have a rich body of data on extremely detailed industries. The data report gross output and materials inputs, so that it is not necessary to work with value added. By pooling industries within two-digit categories, D-H-P are able to achieve much greater power than the tests of this paper. They find extremely strong rejection of competition in most manufacturing industries.

Shapiro [1987], using data similar to those of this paper, extends this framework to estimate the elasticity of market demand jointly with the ratio of price to marginal cost. He confirms the basic finding of market power in numerous industries.

10. *Conclusions*

The assumptions Solow made in developing the now-standard approach to productivity measurement are clearly false. Under the assumptions, productivity growth should be uncorrelated with exogenous variables that induce changes in output but do not shift the production function. In fact, productivity growth is highly correlated with oil prices, quite correlated with military spending, and somewhat correlated with the political party of the president. The correlation does not arise simply because of market power; a simple modification of the Solow residual eliminates the bias in factor shares caused by monopoly profit, but the modified Solow residual has the same failure of invariance as does the original residual.

Of the many explanations considered in this paper, one of the most salient is monopolistic competition. With free entry and fixed costs, firms will reach a zero-profit equilibrium in which both the original and cost-based Solow residuals will fail invariance. Increasing returns is the basic explanation in this case.

A second leading explanation is massive errors in the data. What matters is not measured hours of work, but work effort. People work harder when there is more work to do. Employment arrangements allocate valuable work effort in a way that is not reflected in current wages. The understatement of fluctuations in labor input in response to changes in output induced by the exogenous instruments is the explanation of the failure of invariance.

A third explanation, harmonious with a major new strand of thought in macroeconomics, is that firms enjoy thick-market external benefits when output is high. When an exogenous event stimulates economic activity, it shifts the firm's production function upward.

All of these leading explanations represent important departures from standard economic thought. Failure of invariance of the Solow residual is a fact that needs to be taken into account in micro and macroeconomic models

of firms, industries, and markets. Models with constant returns and no unobserved movements in work effort are simply inconsistent with the data.

References

- Martin Neil Baily, "Wages and Employment under Uncertain Demand" *Review of Economic Studies* 41:37-50, January 1974
- Mark Bills, "Cyclical Behavior of Marginal Cost and Price" *American Economic Review* December 1987
- Ricardo J. Caballero and Richard K. Lyons, "The Role of External Economies in U.S. Manufacturing" Discussion Paper 431, Department of Economics, Columbia University, May 1989
- Peter A. Diamond, "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, vol. 90. pp. 881-894, 1982.
- Ian Domowitz, R. Glenn Hubbard, and Bruce Petersen, "Market Structure and Cyclical Fluctuations in U.S. Manufacturing," Northwestern University, January 1987 (forthcoming, *Review of Economics and Statistics*)
- Jon Fay and James Medoff, "Labor and Output over the Business Cycle," *American Economic Review* 75:638-655, September 1985
- Robert E. Hall, "Increasing Returns: Theory and Measurement with Industry Data" Stanford, 1989.
- Robert E. Hall, "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, vol. 96, October 1988.
- Robert E. Hall and Dale W. Jorgenson, "Tax Policy and Investment Behavior," *American Economic Review*, vol. 57, pp. 391-414, June 1967.

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.

Julio Rotemberg and Lawrence Summers, "Procyclical Productivity as a Consequence of Inflexible Prices", MIT and Harvard University, 1987

Matthew Shapiro, "Measuring Market Power in U.S. Industry," Yale University and NBER, March 1987

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