Center for Research on Economic and Social Theory

and

Department of Economics

Working Paper Series

Procyclical Productivity: Overhead Inputs or Cyclical Utilization?

Susanto Basu

October, 1993
Number 93-25

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109-1220

Preliminary

Comments Invited

PROCYCLICAL PRODUCTIVITY:

OVERHEAD INPUTS OR CYCLICAL UTILIZATION?

Susanto Basu

Department of Economics
University of Michigan
611 Tappan Street
Ann Arbor, MI 48109-1220

internet: sbasu@econ.lsa.umich.edu

tel: 313-764-5359

February 1993

This Revision: October 1993

I would like to thank, without implicating, Robert Barsky, John Fernald, Miles Kimball, and Matthew Shapiro for very helpful discussions, and participants in the 1993 EF Summer Workshops at the NBER and the group of SWINE at the University of Michigan for comments.

It has long been argued that cyclical fluctuations in labor and capital utilization and overhead labor and capital are important for explaining procyclical productivity. Here I present two simple and direct tests of these hypotheses, and a way of measuring the relative importance of these two explanations. The intuition behind the paper is that materials input is likely to be measured with less cyclical error than labor and capital input, and materials are likely to be used in strict proportion to value added. In that case, materials growth provides a good measure of the unobserved changes in capital and labor input. I find that labor hoarding and cyclical capital utilization are quantitatively significant: the true growth of variable labor and capital inputs is, on average, almost twice the measured change in the capital stock or labor hours. More than half of that is caused by the presence of overhead inputs in production; the rest is due to cyclical factor utilization.

The idea that measured procyclical productivity results from unmeasured changes in labor effort and capital utilization is an old one. It has been present at least since the early discussions of Okun's Law and the short-run increasing returns to labor (SRIRL) puzzle. Recently, however, various authors have put forward alternative explanations of procyclical productivity. Researchers on "real business cycles" assume that fluctuations in measured productivity represent real changes in the technological production possibilities of the economy. However, the Solow residual still appears significantly procyclical in response to shocks that can be identified as not being technological in nature (Hall, 1988, 1990; Evans, 1992). This finding leaves two major competing hypotheses. One is that the correlation of the Solow residual with exogenous demand-side instruments is evidence of imperfect competition, increasing returns, or both: a position taken by Hall (1988, 1990). The other is the standard, older interpretation: that procyclical productivity results from cyclical factor utilization, and Hall's estimates of the markup and increasing returns are biased because they do not take this into account.

This criticism of Hall's work was made by Abbott, Griliches, and Hausman (1988) and has recently been restated by Gordon (1992). Shapiro (1993) has made a similar argument with respect to cyclical capital utilization. He argues that correcting for changes in the number of shifts worked over the business cycle eliminates the evidence for increasing returns that Hall cites.[1] Rotemberg and Summers (1990) show that a combination of rigid prices, rationed quantities, and labor hoarding is sufficient to explain procyclical productivity, even if firms do not have market power in the sense of facing a downward-sloping demand curve.

One of the problems with this literature is that the same evidence is often interpreted as favoring one or the other of these hypotheses, since they have very similar predictions for a wide class of phenomena. For example, Burnside, Eichenbaum, and Rebelo (1993) interpret results much like Hall's — the correlation between the Solow residual and government spending — as evidence of labor hoarding. Bernanke and Parkinson (1991) have interpreted the correlation of sectoral productivity with aggregate activity during the Great Depression — which they argue could not have been caused by technological regress — as evidence favoring labor hoarding. On the other hand, Caballero and Lyons (1990, 1992) and Bartlesman, Caballero, and Lyons (1991) have used a similar set of results as evidence for large productive externalities, a reinterpretation of Hall's finding of increasing returns. They argue that labor hoarding should be a function only of a sector's own activity, and hence should not be correlated with the activities of other sectors. However, Sbordone (1992) shows that the Caballero and Lyons results may well be a consequence of labor hoarding when the level of aggregate activity contains information about future sectoral demand.

My goal in this paper is to develop a new test of the labor hoarding hypothesis that nests imperfect competition and increasing returns as alternative possibilities.[2] Under a particular parameterization, the test also suggests a measure of the extent of unmeasured productive inputs of labor and capital.[3] The central insight of the paper is that firms may extract unmeasured services from their own capital stocks or from workers with whom they have a long-term relationship, but that in order to produce greater output, they need more materials input.[4] Thus, materials input is an index of unmeasured capital and labor input.

If the view that cyclical productivity is largely due to unobserved factor utilization is correct, then one would expect materials input to be more procyclical than measured inputs of capital and labor. This is in fact true. But this is not enough to confirm that there is cyclical factor utilization: that is only one of four possible explanations for this finding.

First, it might be the case that the business cycle is driven by technological shocks that affect the productivity of capital or labor (or both), but not materials. This possibility must be taken seriously: I

[1] However, Shapiro is careful to note that his results do not necessarily contradict Hall's hypothesis of imperfect competition, depending on what the compensation is for the extra shifts worked. If the marginal cost of higher factor utilization is zero, but output is sold at a price substantially higher than zero, Hall's (1988) hypothesis can explain the failure of invariance of the revenue-based Solow residual (though not of the cost-based Solow residual).

[2] I do not test Caballero and Lyons's hypothesis of productive externalities because I have shown in other work with John G. Fernald (Basu and Fernald, 1993b) that their finding of significant externalities does not survive the use of the correct gross output data in place of their value added data.

[3] Abbott et. al. (1988) do suggest a way to discriminate between labor hoarding and increasing returns, by including the measure of hours per worker as a right-hand-side variable that is supposed to be a proxy for labor hoarding. I suggest a different test.

[4] A similar idea is behind the use of electricity consumption data as a proxy for output in some industries in the monthly index of industrial production, and the use of electricity as a proxy for capital utilization by Jorgenson and Griliches (1967). I propose a procedure that is much broader, and which also allows me to investigate cyclical fluctuations in labor utilization.

control for it by using demand-side instruments similar to those advocated by Hall (1988, 1990) and Burnside et. al. (1991).

Second, materials growth is a complete measure of unobserved inputs of the other factors only if firms literally cannot substitute between capital and labor (or value added) and materials. If the production technology allows materials to be substituted for value added, one must use a more sophisticated procedure based on measuring changes in the price of materials relative to capital and labor. However, this too can be done without much difficulty.

Third, the procyclical materials-value added ratio might in fact be driven by unobserved growth in capital and labor inputs. Materials use is a convenient indicator of cyclical factor utilization because its input does not have an extra effort or time dimension. This is the idea exploited in the first test of the paper.

Fourth, it may be the case that the growth of materials input exceeds the growth of *total* capital and labor input, but only equals the growth of *variable* capital and labor. There may be a wedge between the two if production requires overhead capital and labor. If there are substantial overhead inputs of labor and capital, then one would expect to see materials input grow faster than total inputs of the other factors, but this would not necessarily imply procyclical utilization of those other factors. Another way of saying this is that if there is overhead capital and labor (but not materials), productivity might be procyclical because Hall's hypothesis of increasing returns is correct, but the production function is not homothetic. The first test proposed in the paper cannot distinguish between increasing returns and cyclical factor utilization if the increasing returns are concentrated mostly in the value-added production function and do not affect equally the gross-output production function. A second test is required, and one is proposed in the paper.

The major finding of the paper is that both labor hoarding and overhead capital and labor exist and are extremely important for explaining cyclical changes in productivity in U.S. manufacturing. The true increase in variable capital and labor input is at least 70 percent greater than the measured increase in labor hours and the capital stock. About 50 percent of this increase in variable inputs is attributable to the presence of overhead capital and labor, and the rest is due to procyclical factor utilization. Correcting for

the bias from cyclical capital factor utilization, the degree of returns to scale appears almost exactly constant. However, the combination of overall constant returns and large fixed costs of production implies that there must be diminishing returns to the variable inputs. The estimates in the paper suggest that, for fixed real input prices, marginal costs rise at least 0.2 percent for every one percent increase in output.

The paper is organized in six sections. Section I presents the method used to measure labor hoarding, under the assumption that there is no overhead capital and labor. Section II modifies the method of the previous section to make it robust to a frequently-cited type of non-separability in the gross-output production function. Section III introduces the test used to discriminate between overhead inputs and cyclical factor utilization. The data used for the estimation are discussed in Section IV. Section V presents the results. Section VI concludes with some comments, and indicates directions for future research.

I. METHOD

I start from an industry-level production function, treating all inputs, including intermediate goods, symmetrically:

$$Y_i = F(K_i, L_i, M_i, T_i) \tag{1}$$

Here Y is sectoral gross output (not value-added). K and L are primary inputs of capital and labor, while M is the quantity of materials input. I assume initially that the production function is homogeneous of degree γ (and hence homothetic) in capital, labor, and intermediate goods. The production function is not constrained to have constant returns to scale, so γ is allowed to be a free parameter.

In what follows, it will sometimes be useful to put some additional structure on the production function. I shall often assume that gross output is produced using a combination of value added and materials, where value added is produced by capital, labor, and technology:

$$Y_i = G(V(K_i, T_i L_i), H(M_i)). \tag{2}$$

Since F was assumed to be homothetic we can, without loss of generality, assume that V and H have constant returns to scale. I have assumed that technology is labor-augmenting to assure the existence of a steady state.

Differentiating the production function (1), we find:

$$dy_i = \frac{F_K K}{Y} dk_i + \frac{F_L L}{Y} dl_i + \frac{F_M M}{Y} dm_i + dt_i \qquad (3)$$

Lower-case letters represent logs of their upper-case counterparts, so all of the quantity variables in (3) are log differences, or growth rates. The sum of the output elasticities equals the degree of returns to scale, γ. The elasticity of output with respect to technology is normalized to 1. Following Hall, I assume that output is potentially sold with an unknown markup, μ. Then we can show that:

$$\frac{P}{\mu} F_j = P_j \qquad \text{for } j = K, L, M \qquad (4)$$

and

$$\gamma = \mu \frac{P_K K + P_L L + P_M M}{PY} \qquad (5)$$

where the P_j's are the prices of the corresponding inputs and P is the price of output. Crucially, P_K must be defined as the required rate of return to capital or the rental price of capital in a competitive market, thus excluding observed monopoly profits as part of the *required* return to capital. As Hall (1988) has shown, we cannot use the share of each factor in revenue to measure the elasticities in (3). However, we can use (4) to eliminate the markup from the total differential of output. We find:

$$dy_i = \gamma[\alpha_i^l dl_i + \alpha_i^k dk_i + (1 - \alpha_i^l - \alpha_i^k)\, dm_i] + dt_i$$

$$= \gamma\, dx_i + dt_i \,. \qquad (6)$$

where α_i^l and α_i^k are the shares of labor and capital in total <u>costs</u> (not revenue). dx_i is a cost-weighted sum of the growth rates of the various inputs. Intuitively, equation (6) says that the growth rate of output equals the growth of the three inputs, each multiplied by the elasticity of production with respect to that input and the degree of returns to scale, plus the contribution of productivity shocks. If there are constant returns to scale and perfect competition — so that the cost shares are also revenue shares and $\gamma = 1$ — then

equation (6) is just the defining equation for the Solow residual. Estimates of (6) give me the baseline for my discussion of cyclical productivity.

The major difficulty with interpreting the estimate of γ from equation (6) as a consistent measure of returns to scale is that there are good reasons to believe that true inputs of capital and labor are not well measured by the capital stock and labor hours. This traditional explanation of cyclical productivity argues that cyclical capital and labor utilization (labor hoarding) make true inputs of capital and labor much more procyclical than is measured in the data. Clearly, if dx_i is measured as being less cyclical than is truly the case, then the resulting estimate of returns to scale, γ, will be biased upward.

The intuition behind this paper is that the available data on materials input make it possible to test for, and perhaps measure, the existence and extent of unmeasured inputs of materials and labor. The idea is a simple one: a worker putting in longer hours or more labor effort, or a machine that is being worked extra shifts, needs more materials in order to create more output. Materials use is a convenient indicator of cyclical factor utilization because its input does not have an extra effort or time dimension. An hour worked may represent very different amounts of labor input and a machine may be operated at different intensities, but a nail, a sheet of steel, or a piece of lumber always makes the same contribution to output: no amount of coaxing can make one nut fit on two bolts.

If capital and labor input are measured with cyclical errors, but materials input is free of such errors, we need to derive the relationship between the true (unmeasured) inputs of capital and labor and the measured input of materials. For this purpose, it will be useful to consider a production function of the type specified in equation (2).

For simplicity, I specialize (2) to the CES production function for gross output:[5]

$$Y_i = [\, V_i^{(\sigma-1)/\sigma} + M_i^{(\sigma-1)/\sigma}]^{\gamma\sigma/(\sigma-1)}$$

where (7)

$$V_i = V(K_i, T_i L_i),$$

[5] Since all the estimation will be done in growth rates, one can alternatively view the CES assumption as being simply a first-order log-linear approximation to any production function like (2).

and $\sigma \geq 0$ is the elasticity of substitution between value-added and materials.[6] Two reference values for σ are the Leontief case where materials are used in strict proportion to output ($\sigma = 0$), and the Cobb-Douglas case with unit elasticity of substitution ($\sigma = 1$).

The first-order conditions for cost minimization imply that

$$\left[\frac{V_i}{M_i}\right]^{-1/\sigma} = \frac{P_i^v}{P_i^m}, \qquad (8)$$

where P_i^v is the industry-specific cost of producing value added for a given technology (a Divisia index of the underlying prices of capital and labor input), and P_i^m is the price of materials, now allowed to vary by industry. Equation (8) implies that the growth rate of value-added equals the growth rate of materials, plus a substitution term that depends on the growth rate of the relative price of value added to materials and on the elasticity of substitution between the two inputs:

$$dv_i = dm_i - \sigma d\ln\left(\frac{P_i^v}{P_i^m}\right) \qquad (9)$$

$$\equiv dm_i - \sigma(dp_i^{v*} - dp_i^{m*}),$$

where dp_i^{v*} and dp_i^{m*} are the true changes in the prices of value added and materials.[7]

Assume that value-added is correctly defined as a Divisia index.[8,9] Then we have that

$$dv_i = \frac{\alpha_i^l(dl_i + dt_i) + \alpha_i^k dk_i}{\alpha_i^l + \alpha_i^k}, \qquad (10)$$

which implies

$$\alpha_i^l dl_i + \alpha_i^k dk_i = \left(\alpha_i^l + \alpha_i^k\right) dm_i - \sigma\left(\alpha_i^l + \alpha_i^k\right)\left(dp_i^{v*} - dp_i^{m*}\right) - \alpha_i^l dt_i. \qquad (11)$$

Substituting equation (11) into equation (6) gives

[6] Since we know that H(M) in equation (2) has constant returns, H must be linear in M, and by an appropriate choice of units we can set the coefficient to one.

[7] Here I leave open the possibility that the true prices of capital, labor, and materials are unobserved. This possibility is pursued below.

[8] Note that I have used cost shares, rather than the commonly used revenue shares, to define the Divisia index of value added. In cases when there is a difference between the two, the cost shares are the appropriate measure of the elasticity of output with respect to each input.

[9] In practice, the index of value added is usually calculated using the discrete-time Törnquist approximation to the continuous-time Divisia index. This procedure is exact if the production function is translog. See Diewert (1976).

$$dy_i = \gamma\left[dm_i - \sigma\left(\alpha_i^l + \alpha_i^k\right)\left(dp_i^{v*} - dp_i^{m*}\right)\right] - \gamma\alpha_i^l dt_i \qquad (12)$$

Equation (12) expresses the basic idea of the paper. In the absence of labor hoarding, estimates of (6) and (12) should yield the same estimate of γ. If, however, there is labor hoarding, then the estimate of returns to scale from (12) should be significantly smaller than those from (6), which are biased upward by the presence of unmeasured capital and labor inputs. The test is easiest to perform in the Leontief case: if $\sigma = 0$ the growth of gross output is just proportional to the growth of materials input, with the constant of proportionality given by the degree of returns to scale.

However, equation (12) shows that if the production function allows materials to be substituted for value added, estimating returns to scale properly corrected for labor hoarding depends on measuring accurately changes in the relative price of value added to materials. There are a number of reasons why the true costs of inputs may differ from their observed costs. The presence of long-term labor contracts — which are consistent with labor hoarding and may be either implicit or explicit — imply that the shadow wage might well differ from the observed wage. In a boom, where workers work longer hours, we would expect the shadow wage to exceed the observed wage. Also, in a boom the marginal cost of an "efficiency unit" of labor might well be significantly higher than the average cost. Bils (1987) argues that the mandated 50 percent overtime premium implies that the marginal wage increases much faster than the average wage.[10] Kydland and Prescott (1988) and Solon, Barsky, and Parker (1992) document the long-standing conjecture that the real wage is more procyclical than it appears because the marginal workers hired in an upturn have lower human capital than the average employed worker.[11] The true cost of capital may also be more cyclical than measured, especially if there is a significant pure user cost component — that is, if depreciation is related to use. Both of these considerations would argue that cyclical changes in dp_i^{v*} should be greater than measured. On the other hand, Carlton (1983, 1987) has argued that delivery lags and other reductions of service in upturns greatly increase the effective price of manufactured

[10] However, Trejo (1991) argues that these legally mandated overtime payments are not allocative, since most of them are offset by compensating changes in the base wage.

[11] The data set that I use controls for composition bias by weighting each type of labor input by its marginal product (as inferred from its relative wage). A similar procedure is used for capital and materials inputs. See Section II and the Appendix for a fuller description.

intermediate goods. These factors would argue that dp_i^{m*} should also be more positive than measured. I assume that the percent change in the true price of each input is a multiple of the observed change:

$$dp_i^{j*} = \beta^j dp_i^j, \qquad \text{for } j = v, m$$

where β^m, β^v are parameters. If the effects identified by Carlton are sufficiently important, β^m may exceed β^v by so much that the true relative price of materials to value added may be procyclical even if the measured price is countercyclical.[12]

Substituting the hypothesized relationship between the true and observed prices in equation (12) gives the first basic estimating equation of the paper:

$$dy_i = \gamma\left[dm_i - \sigma\left(\alpha_i^l + \alpha_i^k\right)\left(\beta^v dp_i^v - \beta^m dp_i^m\right)\right] + v_i . \qquad (13)$$

where v_i is just the composite error term defined in (12). Note that if $\beta^v = \beta^m$, changing β is equivalent to changing the elasticity of substitution between materials and value added.

The simplest and most direct test of the labor hoarding hypothesis comes from the Leontief case: if $\sigma = 0$ then materials are used in strict proportion to value added, and the growth of gross output depends only on the growth of materials and on the degree of returns to scale. If the estimate of γ from equation (6) is significantly smaller than that from (13), we can infer that there is labor hoarding.

If, however, materials are substitutable for capital and labor, then equation (13) shows that the estimate of returns to scale must also include a correction for cyclical changes in the relative price of materials; we must incorporate a substitution term into the estimation. Here, however, there is a problem with unobservables. We do not directly observe the elasticity of substitution. Nor do we necessarily observe the true change in the relative price of value added (β times the observed change). The first strategy I follow is to take reasonable baseline values for each of these parameters and then experiment with changing their values over a plausible range. Second, I simply estimate the unknown parameters (actually, the product of σ and the β's, since they cannot be identified separately from (13)).

12 Murphy, Shleifer, and Vishny (1989, Table 4) argue that materials prices are strongly procyclical relative to labor costs, which are the main cost of producing value added. I return to this issue in Section V.

Suppose, however, that we do not wish simply to test the hypothesis that labor hoarding exists, but also to measure its extent. It is intuitively plausible that the extent of labor hoarding can be inferred by comparing the sizes of the estimated γs from equations (6) and (13). Under a particular simple parameterization of cyclical factor utilization (essentially cyclical utilization *qua* measurement error) this is in fact true. In the first appendix I discuss the conditions under which this parameterization may be reasonable.

Suppose that the true growth rate of capital and labor is a multiple of the observed growth rate:

$$dj^* = \chi\, dj \qquad \text{for } j = k, l$$

where χ is greater than 1. The intuition for this relationship is that inputs of capital and labor have both an observed time component (the number of labor hours, for example) and an unobserved intensity component (an effort index corresponding to work per labor hour). For the labor hoarding story to be sensible, variations along both the extensive and the intensive margin must be costly to the firm: variations in labor hours because of hiring and firing costs and shift premia, and variations in labor effort because workers must be compensated for the (presumably) increasing marginal disutility of work intensity. It seems reasonable, therefore, that in response to an unexpected shock to demand, firms will increase both the number of labor hours used and work per hour, leading to a relationship between the true input of labor and the observed input of labor hours similar to the one hypothesized above.[13]

Under this relationship between true and observed inputs, we can use equations (6) and (9) to show that the probability limit of γ from estimating (6) (denoted γ^c) exceeds the true γ even when the endogeneity of inputs in response to technology shocks (the "transmission problem") is solved by instrumenting:

$$\text{plim}(\gamma^c) = \gamma\left(1 + (\chi - 1)\,\frac{\alpha^v\left(1 + \chi(1 - \alpha^v)\right)}{\left(\alpha^v + \chi(1 - \alpha^v)\right)^2}\right) \qquad (14)$$

13 I have assumed that the mismeasurement parameter is the same for both capital and labor. This may be reasonable depending on the degree of complementarity between capital and labor in production. For example, if intensity of labor effort is related to line speed in manufacturing, it is clear that capital and labor effort must be increased simultaneously and by the same degree. On the other hand, if capital and labor can be worked independently, there is no reason for the same mismeasurement parameter to apply to both capital and labor. In that case, the χ estimated from the data will be a weighted average of the parameters for capital and labor, where the weights are the cost shares of each input. This issue is discussed further in Appendix 1; Sbordone (1993) justifies a similar approach in a dynamic model.

where $\alpha^v \equiv \alpha^k + \alpha^l$ is the share of value added in total cost and is taken to be a constant. Intuitively, equation (14) shows that the extent of the bias depends on the average degree to which the true input of value added exceeds the measured input $(\chi - 1)$, and the use of value added in production, α^v. If $\chi = 1$ or if $\alpha^v = 0$, then there is no bias. As argued above, we can obtain a consistent estimate of the true γ by estimating equation (13), and of course the estimate of equation (6) provides the baseline value for the biased γ. Inserting the estimates of γ from (6) and (13) into (14) will then allow us to infer the average extent of labor hoarding, as measured by χ. The results are reported in Section V.

II. PROBLEMS WITH A VALUE-ADDED PRODUCTION FUNCTION

Much of the empirical literature estimating production functions from gross-output data tests and rejects the conditions needed for the existence of a value-added function: see, e.g. Jorgenson, Gollop, and Fraumeni (1987). Of course, once one admits that the data upon which this estimation is based contains systematic cyclical errors, it is difficult to know how one should interpret the rejection of the necessary separability conditions. But the intuition for the rejection is not difficult to understand. A production function like (2) requires that the elasticity of substitution between materials and labor have the same sign as the elasticity of substitution between materials and capital. The reason is that materials cannot interact with capital and labor separately but only jointly through the value-added production function V. The sign of the elasticity of substitution between materials and either capital or labor must be the same as that of the elasticity of substitution between materials and value added.

The separability restriction is typically rejected because energy use, which is one component of overall materials consumption, is generally found to be a complement with respect to capital, but a substitute for labor. However, taking the nonseparability of energy into account, it is reasonable to assume that non-energy inputs can be separated from inputs of all the other factors: capital, labor, and energy. While cheap oil may lead to production using more machines and fewer workers, and expensive energy lead firms to economize on machines and use more labor, in either case one needs a certain amount of raw materials to create output. Under this new separability assumption, we can write (2) as:

$$Y_i = G(V(K_i, E_i, T_i L_i), H(N_i)). \qquad (2')$$

where E is the input of energy and N is all non-energy materials input. In this case V no longer has the interpretation of value added, but the essential idea of exploiting the relationship between V and H remains unchanged. Following a sequence of steps exactly analogous to those detailed in Section II, we can show that this new formulation leads to a slightly different estimating equation:

$$dy_i = \gamma \left[dn_i - \sigma\left(\alpha_i^l + \alpha_i^k + \alpha_i^e\right)\left(\beta^{ve} dp_i^{ve} - \beta^n dp_i^n\right) \right] + v_i . \qquad (13')$$

Now the price index for V includes energy while the materials price index does not, and the cost share of energy is added to the shares of capital and labor. The growth rate of materials is now defined over non-energy materials only. But this equation is also easy to estimate, and the estimate of γ has just the same interpretation as before. If we can estimate (13) and (13') without much difference in the results, we can be reasonably confident that the failure of separability is not a major problem for the approach suggested here.

III. INCREASING RETURNS IN THE VALUE-ADDED FUNCTION

So far, the analysis has proceeded under the assumptions laid out in Section I, notably that the gross-output production function (F in equation 1) is homothetic. (Actually, we assumed the stronger condition that F was homogeneous.) Thus, we could assume that both the V and H functions in equation (2) have constant returns; this was important later for measuring value added in (10).

Suppose, however, that homotheticity does not hold. Then it is possible that the V and H functions do not have constant returns. In particular, it may be the case that the value-added function has greater returns to scale than the materials function. In this case, we could well have materials growth exceeding the growth of value added without inferring the existence of cyclical factor utilization: it would simply be the case that marginal inputs of capital and labor create value added more efficiently than marginal inputs of intermediate goods create materials input.

It is important to recognize that there are economically relevant cases where the assumption of homotheticity will not hold. The most interesting is if production requires overhead capital and labor but not overhead materials.[14] In this section, I relax the assumption of homotheticity. I assume that the V function has constant returns in the *variable* inputs of capital and labor, but allow for the possibility of overhead capital and labor that lead to increasing returns overall. (Of course, V may have increasing returns for reasons other than the presence of overhead inputs, and my test will detect these as well. The existence of overhead inputs is, however, the most compelling reason for increasing returns in the value-added function.) Under these assumptions, the value-added production function is

$$V_i = V[(K_i - K^*), (T_i (L_i - L^*))] \qquad (15)$$

where K* and L* are overhead inputs of capital and labor. The overall production function continues to be:

$$Y = G(V_i, H(M_i)). \qquad (2)$$

Since H and V are not directly observed, one cannot estimate separately the returns to scale of the three functions G, H, and V. In the previous sections the assumption of a constant homogeneity of G allowed us to normalize the homogeneity of both V and H to one. By relaxing the assumption that G is homothetic, we can investigate the relative homogeneities of G and V with respect to H. But one normalization is still required. Thus, I continue to assume that H has constant returns to scale: that is, there are no "overhead materials." Neither of the main conclusions of the paper — the extent of cyclical factor utilization or the true degree of returns to scale — depends on this normalization. The only inference that is affected is the calculation of the homogeneity of G versus that of H.[15] This issue is discussed further below.

If production is characterized by equations (2) and (15), then, even in the absence of procyclical factor utilization, a firm would always require smaller percent increases in capital and labor than in materials to produce extra output — a situation that heretofore has been taken as *prima facie* evidence of

[14] If production requires the same average ratio of overhead to total inputs for all inputs (including materials) then the production function is again homothetic and the method for detecting variable utilization of capital and labor presented in Section I is exactly correct.

[15] Clearly one cannot distinguish between two models, one saying that firms are very efficient at creating "materials input" from materials but very inefficient at using "materials input" and value added to create output, and the other holding that firms are inefficient at creating "materials input" but efficient at using it and value added to create output.

labor hoarding.[16] In this case, the estimate of γ from equation (13) would not be the true degree of returns to scale, but rather would reflect the degree of homogeneity of G. True returns to scale would be higher than this estimated γ because of the presence of increasing returns in the value-added function.

We can distinguish failures of homotheticity from cyclical factor utilization by looking at dynamic evidence, however. If measured materials growth is greater than measured capital and labor growth because of labor hoarding, then this should be a temporary phenomenon. Over a long enough period of time, a shock to demand should not change the ratio of measured capital and labor to materials. (Another way of saying this is that along a steady-state growth path the degrees of capital and labor utilization should be stationary variables; this is shown in the first appendix) On the other hand, if there are increasing returns to scale in the production of value added, a permanent increase in output should permanently increase the ratio of materials input to capital and labor inputs.[17] This prediction is the basis of the test developed here.

However, this test is not quite dispositive as it stands. Suppose that there are overhead requirements of capital and labor and not of materials. Then it is true that if each firm were operating at a permanently larger scale, the ratio of materials to labor or capital should also change permanently. However, if the estimation is done with industry data, we may fail to detect this outcome. Suppose the following story describes industry evolution. Initially, there is a permanent demand shock. Because of overhead capital and labor, materials input grows faster than capital and labor input. But then, as the industry makes supernormal profits, new firms enter. Ultimately, the industry returns to long-run equilibrium with a larger number of firms but the same amount of output per firm. Clearly, the ratio of capital and labor to materials returns to its long-run value, which might be interpreted as evidence that the short-run excess growth of materials is driven by labor hoarding.

[16] This would also be the case if the value-added function had increasing returns without overhead inputs — that is, if the sum of the exponents on capital and labor exceeded 1. The method proposed here is equally valid to test that possibility.

[17] This can be seen by substituting a log-linear approximation of equation (15) into equation (9).

To avoid this bias in the test proposed above, I examine all of the variables on a per-establishment basis. With this transformation of the data, we can test the prediction that the average ratio of materials to capital or labor per firm is unaffected by demand-induced changes in the quantity of output per firm.

To test the prediction, I estimate equations of the form:

$$dm_i - dj_i = \delta(L)\, dy_i^* + \varepsilon_i \qquad \text{for } j=k,l, \qquad (16)$$

where y^* is the amount of output per establishment, and the dependent variables are (log changes in) the per-establishment materials-capital and materials-labor ratios. All of the regressions also include an industry-specific constant, a time trend, a post-1973 dummy, and a trend interacted with the post-1973 dummy. These are meant to control for secular changes in productivity that might change the ratio of materials to the other inputs, especially labor. The hypothesis to be tested is that $\delta(1) = 0$: this is the hypothesis that there are no overhead inputs of capital and labor.

If we reject this hypothesis, the size of $\delta(1)$ provides information about the average ratio of overhead capital and labor to total inputs of capital and labor.[18] Using this information, we can then calculate the percent change in *variable* capital and labor input for a one percent change in total inputs. This will tell us how large a χ we should expect even in the absence of labor hoarding. Comparing this with the χ computed from estimates of (6) and (13) and equation (14), we can then calculate the fraction of the increase in the variable inputs that is due to overhead inputs and the fraction that is due to cyclical factor utilization. Finally, knowing the ratios of overhead inputs to total inputs (or, more generally, the degree of homogeneity of V) and also the degree of homogeneity of G (from our estimates of equation (13) or (13')), we can then calculate the true degree of returns to scale, where "returns to scale" is defined in the usual fashion: the percent increase in output resulting from a one percent increase in *total* inputs (as opposed to a one percent increase in only the variable inputs). If there are overhead inputs this returns to scale will not be our estimate of γ from equation (13). But the estimate of γ from (13) is still a very interesting quantity. Given that it is the degree of homogeneity of G, it answers the question: "What is the percent increase in *variable* inputs required to produce a one percent increase in output?" Thus, for given real factor prices, $1/\gamma$ gives the elasticity of marginal cost of production with respect to output.

[18] The sum of the coefficients $\delta(1)$ gives the ratio of overhead capital (labor) to total capital (labor), times the percent change in materials required to produce one percent more output. The latter quantity is just $1/\gamma$.

IV. THE DATA

Performing the tests proposed here requires that we examine the usage of materials relative to value added, so I obviously need a data set that has data on gross output and materials input, as well as the usual capital and labor input. I use unpublished data provided by Dale Jorgenson and Barbara Fraumeni on industry-level inputs and outputs. The data consist of a panel of U.S. manufacturing industries, at approximately the 2-digit S.I.C. level, for the years 1953-84.[19] In the following paragraphs I highlight a few key features of the data; in the appendix I discuss the underlying sources and methods at greater length.[20] For a complete description, see Jorgenson, Gollop, and Fraumeni (1987) or Jorgenson (1990).

These sectoral accounts seek to provide accounts that are, to the extent possible, consistent with the economic theory of production. Output is measured as gross output, and inputs are separated into capital, labor, energy, and materials. For our purposes, an essential aspect of the data is their inclusion of intermediate inputs.

Jorgenson's input data are available both with and without an adjustment for input quality. Without the quality adjustment, the Jorgenson measures of labor and capital input are essentially the standard ones — hours worked and the capital stock. From the perspective of a firm, however, the relevant measure of its input of, say, labor is not merely labor hours. The firm also cares about the relative productivity of different workers. In creating a series for labor input, Jorgenson, Gollop, and Fraumeni assume that wages are proportional to marginal products. This allows them, in essence, to calculate quality-adjusted labor input by weighting the hours worked by different types of workers (distinguished by various demographic and occupational characteristics) by relative wage rates. Note that

[19] The only major difference between the industry groupings in the Jorgenson data and the standard 2-digit S.I.C. classification is that Motor Vehicles (S.I.C. 371) are separated from other transportation equipment (S.I.C. 372-79). (This is the standard practice of the BLS, though not of the BEA.) Thus, there are 21 manufacturing industries in the panel rather than the usual 20.

[20] The data description and Appendix 2 are slight modifications of their counterparts in Basu and Fernald (1993a).

the validity of this procedure is unaffected by markup pricing, because markups affect the level of wages, not relative wages. Hence, labor input can increase either because the number of hours worked increases, or because the "quality" of those hours increases. Similarly, Jorgenson, Gollop and Fraumeni adjust inputs of capital and intermediate goods for changes in quality. This quality-adjustment is particularly important to adjust the wage series for the labor force composition effect stressed by Solon, Barsky, and Parker (1992).

The Jorgenson data report the quantity of materials purchased per year, not actual materials usage. Since, as Ramey (1989) documents, materials and work-in-process inventories are procyclical, substituting materials purchases for usage can bias downward the γ estimated from equation (13). Thus I have modified the Jorgenson data by adjusting the materials data for changes in inventories.[21]

To estimate the payments to capital, I follow Hall and Jorgenson (1967), Hall (1986, 1990), and Caballero and Lyons (1992), and compute a series for the user cost of capital r. The required payment for any type of capital is then rP_KK, where P_KK is the current-dollar value of the stock of this type of capital. In each sector, I use data on the current value of the 51 types of capital, plus land and inventories, distinguished by the BEA in constructing the national product accounts. Hence, for each of these 53 assets, I compute the user cost of capital as

$$r_s = (\rho + \delta_s) \frac{(1 - ITC_s - \tau d_s)}{(1 - \tau)}, \quad s = 1 \text{ to } 53.$$

ρ is the required rate of return on capital, and δ_s is the depreciation rate for this asset. ITC is the investment tax credit, τ is the corporate tax rate, and d is the present value of depreciation allowances. I follow Hall (1986, 1990) in assuming that the required return ρ equals the dividend yield on the S&P 500. From Jorgenson and Yun (1990), I obtained data on ITC and d that is specific to each type of capital good. Given required payments to capital, it is then straightforward to calculate the cost shares.

The price series for composite capital and labor input is constructed as a Divisia index from the underlying quality-adjusted wage series and the series on required payments to capital, where the weights are the cost shares of capital and labor in value added. The aggregate materials price is similarly constructed from the underlying energy and non-energy materials price series.

21 I am extremely grateful to John Fernald for providing me with the adjusted data.

Since input use is likely to be correlated with technology shocks, I seek demand-side instruments for input use. To solve the "transmission problem" of endogeneity between productivity shocks and input growth, I use the instruments advocated by Ramey (1989) and Hall (1988) as modified by Caballero and Lyons (1992): the growth rate of the price of oil deflated by the price of manufacturing durables; the growth rate of the price of oil deflated by the price of manufacturing nondurables; the growth rate of real government defence spending; and the political party of the President. I use the current value of each instrument, as well as one lag of defence spending and the party of the President.

V. RESULTS.

First I estimate equation (6) as it stands, using the measured values of all three inputs.[22] I use Seemingly Unrelated Regressions to estimate the equations for all the industries together, imposing the constraint that the degree of returns to scale, γ, is equal across industries. The resulting estimate of returns to scale is shown in Table 1. The estimate is 1.09, indicating statistically significant evidence of increasing returns to scale. The estimate is also quite large economically, though much smaller than those reported by Hall (1990). Basu and Fernald (1993a) show that Hall's (1990) very large estimates of γ result from his inappropriate use of value-added data. Estimates of returns to scale using value-added data are subject to a number of significant biases that tend to exaggerate the degree of increasing returns.

Next I perform the simplest test of the labor hoarding hypothesis. This corresponds to the case of a homogeneous production function with Leontief technology in the production of gross output, so that output growth is just proportional to materials growth. The results are reported in the first column of Table 2. Here the results are very different. The estimate of γ is 0.83. Clearly this is significantly different, both statistically and economically, from the estimate in Table 1. We can decisively reject the null hypothesis that variable capital and labor inputs are accurately measured by changes in the capital

22 The Jorgenson data set separates intermediate inputs into energy and non-energy materials. Since this distinction is often not germane for my purpose, I typically combine both types of inputs into a single Divisia index of materials input. Naturally, I do not follow this procedure in estimating (13').

stock and labor hours. The calculated χ, reported in the second line of Table 2, shows that the true growth of variable capital and labor is on average 68 percent greater than the measured growth.

Does allowing for the possibility that materials can be substituted for value added change the conclusion that the unobserved inputs of labor and capital are very large? To see how the estimate of χ changes when we allow for the possibility that firms substitute towards using more materials in a boom, I now estimate equation (13) allowing for substitution between materials and value added.

In part the outcome of this test depends on whether the price of materials relative to value added is procyclical. The answer to that question depends on the cyclicality of the relative price of materials. If the price of materials is less procyclical than real labor (and capital) costs, then firms will use proportionately more materials in a boom and our measure of unobserved inputs will be an overestimate. Of course, if materials prices are more procyclical than the cost of producing value added, the reverse will be true. In the Jorgenson data, I find that the cost of value added relative to the price of materials is significantly procyclical. The results do not depend on whether "materials" are defined as including energy; the coefficients are very similar in both cases.

It may be surprising to find that relative materials prices are countercyclical. Conventional wisdom holds that raw commodities prices are highly procyclical, while the real wage, much the larger input to the price of value added, is acyclical or only mildly procyclical (Murphy, Shleifer, and Vishny, 1989). Both parts of this statement, however, need careful interpretation. While it is true that commodities prices are volatile and procyclical, raw commodities are only a small fraction of the total input of intermediate goods. Since "materials" are correctly classified by use and not by type of good, an industry's materials input includes all of its purchases from other industries (and from itself). For example, Boeing's purchase of a computer to fit into its newest jet is a purchase of an intermediate good, although a computer is typically classified as a high-technology finished good. The Jorgenson data set uses the input-output tables for the U.S. to incorporate such purchases into the definition of materials, along with more conventional materials inputs like raw commodities. Second, when one takes account of the composition bias in labor input pointed out by Solon, Barsky, and Parker (1992) and others, real wages become much more procyclical. The labor cost data that are used for the estimation are adjusted for

the fact that labor quality is highly countercyclical, making the cost of an efficiency unit of labor strongly procyclical. These two considerations together lead to the finding that materials prices are countercyclical relative to capital and labor costs. However, it should be emphasized that while the data set corrects for one significant source of error, the cyclical composition bias of labor (and other inputs), it does not correct for the factors emphasized by Carlton that would work in the other direction, increasing the effective price of materials in a boom. It seems best to allow the data to speak on the issue of whether the effective price is more cyclical than the observed price; this is one of the strategies that I follow below.

First I follow the strategy of assuming plausible parameter values. Since the effect of allowing for substitution between materials and value added depends on the size of σ, the major question, of course, is the right σ to assume. Rotemberg and Woodford (1992, Appendix) use a version of equation (9) to estimate the elasticity of substitution. They find $\sigma = 0.7$. Bruno (1984) surveys a number of papers and reports a consensus range for σ of 0.3-0.4.[23] Clearly, cyclical changes in the relative price of materials can overturn the initial findings only if σ is relatively large. Therefore, since the bias from an overestimate works against the results of the paper, I use 0.7 as the baseline value of σ.

I estimate (13) for two non-zero values of σ: $\sigma = 0.7$, and the Cobb-Douglas case with unit elasticity of substitution, $\sigma = 1$. I use the average share of value added over the sample period, and the observed change in the ratio of the cost of value added to the price of materials.[24] Note that for non-zero values of σ, one must also choose values of the β's. One way of regarding the estimates for $\sigma = 1$ is that they reflect an elasticity of substitution of 0.7, but allow the true relative price of value-added to be 1.4 times as cyclical as the data indicate.

The estimates for the non-zero value of σ are in the second and third columns of Table 2. There is almost no change in the estimated γ's when we allow for substitution: the estimates are 0.83 and 0.84,

---

23 Of course, the data these papers use for their estimation may be flawed because of cyclical measurement error. The presence of labor hoarding would tend to bias upward the estimate of the elasticity of substitution, since in the data one sees output growth being accompanied by low capital and labor growth but high materials growth, implying a high degree of substitution of materials for capital and labor. As I have argued here, the better interpretation of that observation is that the increase in materials growth is accompanied by unobserved increases in the growth of variable capital and labor inputs. So the cited elasticities of substitution should be regarded as upper bounds.

24 Since (13) is a log-linear approximation that holds for small deviations from the steady-state growth path, one should in fact use the average shares to weight price changes. The results are basically unaffected by the choice of contemporaneous versus average shares.

depending on the value of σ. Based on these estimates, the true growth of variable capital and labor inputs is between 63 percent and 68 percent greater than the measured growth of these inputs. For plausible parameter values, the results are barely sensitive to variations in the parameters governing substitution between value added and materials. So even when we allow for the possibility that materials can be substituted for value added, we obtain the preliminary result that the economy is characterized by a considerable degree of cyclical factor utilization.

Next, in Table 3, I present estimates of equation (13'), which controls for the most common problem with the existence of a value-added function. By separating energy from non-energy materials inputs, we allow capital and energy to be complements while labor and energy may be substitutes. The estimates of γ from this specification are even smaller than those in Table 2, implying even higher estimates of unmeasured factor inputs. It is clear that the results are not being driven by this type of non-separability.

Finally, rather than trying to impose reasonable values of σ and β, I estimate them. This is particularly important because it allows the data to speak on the issue of whether the true changes in the prices of materials and value added are more or less cyclical than the observed changes. The results are found in the first line of Table 4. They confirm the hypothesis of unmeasured factor inputs and Carlton's conjecture that the effective price of materials is very procyclical: when we do not impose a priori values of σ and β, the regression chooses β^m to be large relative to β^v, and hence the degree of returns to scale to be smaller than even in the case when σ was constrained to be zero. As noted previously, the interpretation of a large β^m is that the effective price of materials is much more procyclical than observed: that is to say, the effects suggested by Carlton are relatively more important. This is reasonable, because the price of value added computed from the data already contains a correction for the major bias identified in the literature, the cyclical change in the composition of the labor force. Thus, the major omitted factor that the regression estimates is the change in the effective price of materials. So it appears that the conventional wisdom may be right after all — the price of materials may be procyclical relative to the price of capital and labor — but not for the reasons usually given. These results imply an even greater degree of labor hoarding than previously found, strengthening the findings of the paper. The estimate of γ

using materials and energy is 0.75. So allowing for substitution, the estimated degree of variable factor utilization actually rises. The second line of Table 4 shows that the same is true when we allow for the possibility that the value-added production function is not separable and include energy as one of the inputs to the production of V.

The next possibility to examine is whether much of this unmeasured increase in variable inputs is driven by the presence of large overhead inputs. An noted above, if there are large overhead inputs of capital and labor, then a one percent increase in total input represents a much greater than one percent increase in the variable input. Assuming that materials input is proportional to the variable input of capital and labor, we would see large increases in materials input without equal increases in capital and labor input. But it would not be correct to infer that the increase in materials relative to labor and capital signals unmeasured factor utilization; this observation might be consistent with having the variable inputs of all three factors increasing at the same rate.

As argued in Section III, the way to test for this possibility is to use dynamic evidence from estimating an equation like (16), and see whether the materials to labor ratio and materials to capital ratio change permanently in response to a permanent, demand-driven shock to per-firm output. I use the current and six lags of output growth as the right-hand-side variables in (16). To estimate the δ's parsimoniously and with fewer collinearity problems, I use a polynomial distributed lag of the type proposed by Almon (1962). I do not impose endpoint restrictions because, as Dhrymes (1971) points out, these are restrictions not just on the first or last δ's, but through them on the coefficients of the lag polynomial, and hence on all the other δ's as well.

Estimates of equation (16) for changes in the materials/labor and the materials/capital ratios are found in Table 5. I have recovered the δ's from the estimated coefficients of the lag polynomial and listed them, along with the associated standard errors. The test statistics for the hypothesis that $\delta(1) = 0$ are reported in the last line. For both ratios, the hypothesis is rejected at any reasonable significance level.

This result shows that our conjecture was justified: some of the unmeasured increase in variable inputs does indeed come from the presence of overhead capital and labor. We can ask if the amount of overhead inputs implied by the estimates in Table 5 are sufficient to explain the results obtained earlier in

Tables 2-4. The sum of the coefficients in the first column gives the average ratio of overhead capital to total capital times the percent change in materials input required to increase output by one percent. Taking our estimates from Table 2 as the benchmark, we know that a one percent increase in output requires an increase of approximately 1.2 percent in materials input. This then implies that overhead capital averages about 44 percent of the total capital stock in manufacturing, while overhead labor is about 31 percent of the total labor force.[25] These figures imply that on average, about 50 percent of the unmeasured growth in variable inputs can be attributed to overhead factors. So from overhead factors alone, we would expect to see a χ of about 1.50. But the results of Tables 2-4 imply values of χ that are distinctly higher, with a minimum of 1.63 and a median value of 1.83. Thus, while overhead labor is a major cause for the finding that materials growth exceeds the growth of capital and labor inputs, cyclical factor utilization also seems to play a significant role in the explanation of procyclical productivity.

Finally, given the average ratios of overhead capital and labor to total inputs of these factors, we can calculate the degree of returns to scale. We use the average cost share of each input as the measure of its relative output elasticity, and multiply that by a γ of 0.83 to convert that to an absolute output elasticity. This calculation gives a returns to scale parameter of 0.98 — very close to constant returns. But the homogeneity of G is 0.83 — significantly less than 1 — indicating that there are increasing marginal costs. (Indeed, marginal costs must be increasing in order for locally constant returns to scale to coexist with large overhead inputs.) As noted above, for constant input prices, the elasticity of marginal cost with respect to output is about 1.2. This conclusion, however, as well as the calculation of the ratios of overhead to total capital and labor, depends on the assumption that there are no overhead materials. If there were overhead materials, the homogeneity of G would be smaller (implying a steeper marginal cost curve), and the ratios of overhead capital and labor to total capital and labor would be those calculated

25 One should note that the problems with measuring long-run changes in capital input are severe, since estimates of the real capital stock rely on estimated rates of depreciation (or, equivalently, of embodied technological progress) that are very uncertain. Gordon (1990) argues that, over roughly the sample period of this paper, the growth of real capital input is understated by using official figures that do not correct properly for quality improvements in equipment. Gordon's findings may well imply that the ratio of overhead to total capital is smaller than calculated.

plus the ratio of overhead to total materials. But none of these adjustments would affect the central calculations of the paper: the true returns to scale, or the extent of labor hoarding.

VI. CONCLUSION

It has long been argued that overhead inputs and cyclical factor utilization are important for explaining procyclical productivity. Here I present two simple and direct tests of these hypotheses, and a way of measuring their relative importance as explanations of cyclical productivity. I find that both are very important. My estimates suggest that the true input of variable capital and labor inputs grows between 70 and 120 percent faster than one would calculate from changes in labor hours and the capital stock. About 50 percent of this increase is due to the presence of overhead capital and labor, while the rest can be attributed to procyclical factor utilization. Based on these results, I calculate that the true degree of returns to scale is almost exactly one, a far cry from the very large returns to scale reported by Hall (1990). It is true that I find larger returns to scale in the value-added function, which may explain some of the gap between my results and those of Hall, who uses value-added data. However, since firms produce and sell actual output not real value added, which is an economic index number without physical interpretation, it is the smaller estimate of returns to scale that seems relevant for most economic purposes.

The combination of constant returns and overhead inputs implies that there must be diminishing returns to the variable factors of production. This is indeed true. Based on the assumption that all of materials input is variable, I find that the production function is homogeneous of about degree 0.83 in variable inputs, i.e. the elasticity of marginal cost with respect to output is about 1.2. If some of the total materials input is a fixed cost then total overhead inputs are larger, and the marginal cost curve is even steeper.

These findings resolve some puzzles and raise others. They provide an explanation of cyclical productivity that draws upon some elements of Hall's hypothesis — there are increasing returns, but only in the production of value added — and combines them with the older story of procyclical factor utilization. As a result, overall returns to scale are approximately constant. Since there are approximately constant returns, equation (5) implies that the markup of price over marginal cost must be approximately equal to the profit rate. The data used to construct the cost shares imply that total costs are about 90 percent of total revenue, which means that the markup must be about 1.10. This is very close to the markup of 1.15 estimated by Basu and Fernald (1993a), using completely different methods. These findings — constant returns, small markups, and cyclical factor utilization — are quite different than those of Hall (1988, 1990) and his followers, and much more consistent with a traditional view of production technology and competitive structure.

Some puzzles remain, however. Chief among them is the question of why there are diminishing returns in the variable inputs: why, absent increases in factor prices, should marginal costs of production increase with output? Second, if there are indeed increasing marginal costs, why is there not stronger evidence of production smoothing? Miron and Zeldes (1988) and Ramey (1991) fail to find evidence of production smoothing in the data at either seasonal or cyclical frequencies. However, Fair (1989) and Krane and Braun (1991) notes that there is evidence of production smoothing once one examines data on physical quantities.

There is reason to believe that cyclical factor utilization and production smoothing are closely linked issues. In the empirical literature on capacity utilization (e.g. Shapiro, 1993), one often finds the puzzling result that estimated shift premia, presumed to be the major cost of utilizing capacity more fully, are not very large. A limited number of studies have also failed to find much evidence that economic depreciation is related to use (e.g. Hulten and Wykoff, 1981). But absent either of these costs, one cannot explain the low degree of capacity utilization in a perfectly competitive setting. A competitive firm takes price as given, and absent increasing costs of utilization, should be willing to use its capacity fully at all times. Imperfect competition can help explain this paradox: an imperfectly competitive firm, even one with low marginal cost of increasing capacity utilization, will not necessarily produce to the full extent of its (given) capacity lest it drive down the price of its product. But even the imperfectly competitive firm should minimize costs over the long horizon, when capacity is itself a choice variable: if the firm finds itself carrying too much capacity on average, it should reduce the size of its capital stock and its labor

force, and use inventories to smooth production. So even the imperfect competition explanation for low capacity utilization must rest on an assumption or finding that it is very costly to hold inventories. If holding inventories is costly, however, then under uncertainty it may pay the imperfectly competitive firm to hold excess capacity on average. If there are large profits to be had at times of high demand, and production smoothing is not cost effective, it may pay a firm to utilize capacity fully only in rare cases, even if the marginal cost of capacity utilization is not high. This story suggests, however, that if there is imperfect competition in product markets, we can explain large cyclical variations in utilization based not on increasing marginal costs of utilization, but on the uncertainty of demand, the cost of carrying inventories, and the fear of spoiling one's market. At the least, it suggests that capacity utilization, costs of investment in capital and labor, costs of changing the work intensity of capital and labor, the cost of carrying inventories, motives for smoothing or bunching production, and imperfect competition in product markets all interact in a complex way over the business cycle, and explaining cyclical factor utilization is likely to require an investigation of all these issues.

APPENDIX 1: THE DYNAMICS OF LABOR HOARDING

In this appendix I discuss briefly the implications of a simple dynamic model of capacity utilization for the type of estimation in the paper and for the general issue of labor hoarding.

Assume that imperfectly competitive firms maximize the present discounted value of future cash flows:

$$\text{Max} \int_0^{\infty} e^{-rt} \left[P(Y)Y - w(E)L - P_m M - X\left(1+\varphi\left(\frac{X}{L}\right)\right) - I\left(1+\phi\left(\frac{I}{K}\right)\right)\right] dt \qquad \text{(A1)}$$

subject to

$$Y = F(KU, EL, M) \qquad \text{(A2)}$$

$$\frac{dK}{dt} = I - \delta(U)K \qquad \text{(A3)}$$

$$\frac{dL}{dt} = X - sL \qquad \text{(A4)}$$

$$K_0, L_0 \text{ given.} \qquad \text{(A5)}$$

K, L, M, U, and E are, respectively, capital, labor (hours), materials, the degree of capital utilization, and the work intensity of labor. w(E) is the wage paid per unit of labor, where the wage depends on the degree of labor effort required. δ(U) is the depreciation rate of capital, which depends on the intensity of use; the penalty for greater capacity utilization is faster depreciation. The functions φ and ϕ represent the adjustment costs of labor and capital. We assume $w' > 0$, $w'' > 0$, $\delta' > 0$, $\delta'' > 0$, $2\varphi' + \frac{X}{L}\varphi'' > 0$, and $2\phi' + \frac{I}{K}\phi'' > 0$. P_m is the price of materials. s is the exogenous separation rate of workers. The production function F is not constrained to have constant returns to scale, although it is taken to have diminishing marginal product in all three inputs. I assume the conditions on the production function and the elasticity of demand necessary to ensure that the profit-maximization problem is well-defined.

A few points about problem (A1-A5) are worth noticing. First, it assumes that capital and labor utilization can be varied independently. In certain situations this may not be an appropriate assumption; one can then add an additional constraint that E=U. Second, for the purposes of comparing the results of the derivations here to the setup of the paper, I have identified K and L with the observable inputs — the

capital stock and labor hours — and U and E with the unobservable inputs. Given that L therefore represents worker hours and not the number of workers, the use of a single adjustment cost function for L implicitly assumes that the cost of adjusting hours by changing of the number of workers — hiring and firing costs — equals the cost of changing the number of hours per worker — perhaps due to shift premia. In general, one would want to distinguish the factors influencing a firm's decision to employ more workers from those prompting it to work each employee longer hours. Finally, I have allowed for imperfectly competitive behavior, but assumed that at every instant the firm acts like a static monopolist; this abstracts from considerations of dynamic monopoly (see Tirole, 1988, ch. 1) or supergames in markets with a number of imperfectly competitive firms (Green and Porter, 1984; Rotemberg and Saloner, 1986).

Solving the maximization problem yields five static first order conditions, two Euler equations for the state variables, and a pair of associated transversality conditions. I use the first order conditions to describe the solution to the problem, focusing on the relation between unobserved capacity utilization and observable changes in the quantities of the associated inputs. They are:

$$U = a\left(\frac{PF_K}{\mu q}\right) \qquad \text{(A6)}$$

$$E = b\left(\frac{PF_L}{\mu}\right) \qquad \text{(A7)}$$

$$\frac{I}{K} = c(q) \qquad \text{(A8)}$$

$$\frac{X}{L} = d(\lambda) \qquad \text{(A9)}$$

$$q_t = \int_t^{\infty} e^{-(r+\delta)(x-t)} \left[\frac{P}{\mu}F_K U + \left(\frac{I}{K}\right)^2 \phi'\left(\frac{I}{K}\right)\right] dx \qquad \text{(A10)}$$

$$\lambda_t = \int_t^{\infty} e^{-(r+s)(x-t)} \left[\frac{P}{\mu}F_L E - w + \left(\frac{X}{L}\right)^2 \varphi'\left(\frac{X}{L}\right)\right] dx \qquad \text{(A11)}$$

where μ is the markup of price over marginal cost. The assumptions made above ensure that the functions a, b, c, and d are increasing in their arguments.

This simple model provides several insights about the relationship between labor hoarding and observed changes in inputs. We focus first on the conditions governing labor utilization and changes in labor hours, equations (A7), (A9), and A(11). The first equation says that the decision to utilize labor more intensively is a function only of the current marginal revenue product of labor. The investment decision, on the other hand, concerns a state variable, and so is based on the present discounted value of an unit of labor. So there need not be a relationship between labor utilization and labor hours: an instantaneous shock to the current marginal revenue product of labor that is of finite size will lead the firm to increase labor effort instantaneously but not increase its stock of workers, since the expected present value of a worker does not change in response to the momentary shock. And there can certainly be changes in the value of an "installed worker" — equation (A11) shows that one likely source is changes in the time path of expected wage payments — that do not change the rate of labor effort demanded. Nevertheless, it seems reasonable that shocks will, on average, change both the current marginal revenue product of labor and the present value of an hour of labor in the same direction, leading to the relationship between observed hours and unobserved effort of the type hypothesized in the paper.[26]

One cannot be so sanguine when it comes to capital utilization, however. Equations (A8) and (A10) show that investment in physical capital has much the same determinants as investment in labor. However, (A6) shows that the capital utilization decision differs significantly from its counterpart for labor. The difference is that the shadow value of installed capital, q, enters the utilization decision. In fact, *ceteris paribus*, an increase in q increases investment but decreases capital utilization. The intuition is straightforward: an increase in q that is not accompanied by an increase in the current marginal revenue product of labor means that capital will be more valuable in the future than it is today. Since the cost of using capital more intensively is that it depreciates faster, an increase in the anticipated value of capital should lead firms to decrease capital utilization and save their capital stock for the future. An analogous situation does not arise with labor because we take the separation rate of workers to be independent of the

[26] I should emphasize that this is true only for unanticipated shocks. An anticipated shock to the marginal revenue product of labor that is expected to last for some non-infinitesimal length of time will cause the firm to start hiring workers from the time it receives that information. But as the number of workers rises, the marginal product of labor will fall, and the firm will tend to actually *decrease* the degree of utilization of labor.

labor effort the employer demands.[27] The negative relationship between capital utilization and q is a problem for the method of the paper, which assumes that observable physical investment and unobservable capital utilization move together. There are certainly shocks for which this is true: if changes in the marginal revenue product of capital are mostly unanticipated and temporary (but not instantaneous), then it is likely that investment and utilization will move together. But this is by no means guaranteed.

So far, however, we have treated the utilization rates of capital and labor as independent decision variables for the firm. But if the two are linked so that it is not possible to change one without changing the other, then the single utilization rate will be more heavily influenced by current factors (since the labor utilization decision depends only on the current marginal revenue product), and it is more likely that the utilization rate, physical investment, and labor hours will all move together. This is what is assumed in the paper.

The appendix has shown that under certain conditions on the timing and sources of shocks, one can justify the assumption in the paper that labor hoarding and labor hours move together. This provides some grounds for the parameterization of labor hoarding assumed in Section I. However, even this simple model shows the limitations modeling labor hoarding and capital utilization as cyclical measurement error. It also points out a link between observables — investment and labor hours — and unobservables — capacity utilization and labor intensity. This link can potentially be estimated using a tightly parameterized structural model. There are formidable difficulties with such a course of action, however, for the link between the observables and unobservables depends, for a specific shock, on the entire change in the time path of the marginal revenue products of capital and labor produced by the shock or, on average, on the stochastic process governing the transmission of shocks. And in the case when the shock is purely to future marginal products, there need not be any relationship between current utilization and current investment.

[27] Of course, one might imagine that one of the costs of working one's employees too hard is that they will leave such a "Simon Legree." In that case, the capital and labor utilization decisions would be similar.

APPENDIX 2: THE JORGENSON DATA SET

The data provided by Jorgenson and Fraumeni are part of Jorgenson's long-term research effort, with a number of co-authors, aimed at creating a complete set of national accounts for both inputs and outputs at the level of individual industrial sectors as well as the economy as a whole. Their purpose is to allow Jorgenson to allocate U.S. economic growth to its sources at the level of individual industries. Because the goal is to account for growth in production, these data seek to provide measures of inputs and output that are, to the extent possible, consistent with the economic theory of production. A complete description of the data is contained in Jorgenson, Gollop, and Fraumeni (1987), henceforth referred to as JGF.

Output is measured as gross output, and inputs are separated into capital, labor, energy, and materials. Each of these inputs is, in turn, a composite of many more subcategories of these inputs. For example, capital input includes the contribution of hammers, computers, office buildings, and many others forms of physical capital. For both national accounting and productivity measurement, an important issue is how these many subcategories should be combined to form a small, manageable set of categories, in this case inputs of capital, labor, energy, and materials.

Gross output by industry, in current and constant prices, comes from the Office of Economic Growth of the Bureau of Labor Statistics. The main thing to note is that output is not measured as value added. From the perspective of a firm or industry, the proper measure of output is gross output, so that intermediate inputs are treated symmetrically with primary inputs. Value added by industry, as published in the National Income and Product Accounts, is useful for national accounting, since it sums to national expenditure. It does not, in general, have an interpretation as a measure of output.

Conceptually, JGF divide labor input into hours worked, and average labor quality. NIPA provides hours worked by industry. From the point of view of a producer, however, the proper measure of labor input is not merely labor hours: firms also care about the relative productivity of different workers. Fraumeni and Jorgenson use data from household surveys to disaggregate total hours into hours worked by different types of workers, categorized by demographic variables such as sex, age, and

education. JGF then assume that workers are paid proportionately to the value of their marginal products. This assumption allows them to calculate labor input as essentially a weighted sum of the hours worked by different types of workers, weighting by relative wage rates. For the economy as a whole, labor-input growth from 1947 to 1985 averaged 1.81 percent; the growth of labor hours averaged 1.18 percent, while the growth of labor quality averaged 0.63 percent.

JGF measure capital input analogously to labor input, attempting to weight the input of different types of capital by relative efficiencies. The BEA provides industry data for 27 categories of producer durables, including trucks and autos, and 23 categories of nonresidential structures. In addition to these, Jorgenson and Fraumeni include data on the stock of inventories and land by industry. A simple capital stock measure, with no adjustments for quality, can be calculated as an unweighted sum of the stocks of all types of capital. As an input measure, this is analogous to labor hours. Capital input from the point of view of a producer, however, should weight the different types of capital by relative productivities. Just as we need wage rates to calculate labor input, we require rental rates to calculate capital input. These rental rates are not directly observed. JGF assume that there are either constant returns and competition or monopolistic competition, however, so that total payments to capital are observed as property compensation, a residual after all other factors have been compensated. JGF use this to back out the implied rental rates for each type of capital, based on knowledge of the stock of each type of capital and its depreciation rate, as well as tax parameters such as the corporate income tax and investment tax credits.

Note that the largest determinant of relative rental rates is differences in depreciation rates. Computers, for example, depreciate at an estimated rate of 27 percent per year; office buildings depreciate at an estimated rate of 2.5 percent per year. Hence, a one-dollar investment in computers must provide a higher flow of services than a one-dollar investment in office buildings.

The data on intermediate inputs of energy and materials is the most problematic. Conceptually, the task is straightforward: for each year, we observe payments to intermediate factors as the difference between nominal gross output and nominal value-added; we want to divide this nominal intermediate input into indices of price and quantity. Note that we also observe the output price for each of the types of intermediate input, from the gross output price data. Thus, to construct an appropriate deflator for intermediate goods, we simply need to know how to weight the price of each component of the intermediate goods index.

The difficulty arises from the low quality of the underlying data: the BEA compiles comprehensive input-output tables only about every five years (1947, 1958, 1963, 1967, 1972, 1977, and 1982). Intermediate inputs into any sector include inputs from all sectors. As with capital and labor, these disaggregated inputs should be weighted by marginal productivities in order to calculate a composite intermediate input. This requires consistent annual input-output tables in current and constant prices. In brief, for the benchmark years the data are adjusted to make the definitions consistent over time; they are then aggregated to the 35-industry level. These benchmarks are converted to shares of industry output, and then these shares are interpolated from benchmark to benchmark. This gives an estimated input-output table for each year. This then allows us to create an appropriate price deflator for nominal payments to intermediate factors in each year.

Is it a problem that I use data constructed from these interpolated input-output tables for regressions at annual frequencies? Probably not. First, the weights on the different components of the intermediate-goods index do not change much, even over a five to ten year period. Second, even when the weights on some component do change, they tend to change only gradually. Hence, a linear interpolation probably provides a reasonable approximation to the true weights.

REFERENCES

Abbott, Thomas, Zvi Griliches, and Jerry Hausman (1988). "Short Run Movements in Productivity: Market Power versus Capacity Utilization." Mimeo.

Almon, Shirley (1962). "The Distributed Lag between Capital Appropriations and Expenditures." *Econometrica*, 30, 407-23.

Bartlesman, Eric Ricardo J. Caballero, and Richard K. Lyons (1991). "Short- and Long-Run Externalities." NBER Working Paper #3810.

Basu, Susanto and John G. Fernald (1993a). "Constant Returns and Small Markups in U.S. Manufacturing" Mimeo, University of Michigan.

______ and ______ (1993b). "Productive Externalities in U.S. Manufacturing: Do They Exist, and Are They Important?" Mimeo, University of Michigan.

Bernanke, Ben S. and Martin L. Parkinson (1991). "Procyclical Labor Productivity and Competing Theories of the Business Cycle: Some Evidence from Interwar U.S. Manufacturing Industries." *Journal of Political Economy* 99 (June) 439-59.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. "Labor Hoarding and the Business Cycle." *Journal of Political Economy* 101 (April) 245-73.

Bils, Mark (1987). The Cyclical Behavior of Marginal Cost and Price." *American Economic Review* 77 (Dec.) 838-855.

Caballero, Ricardo J., and Richard K. Lyons (1990). "Internal Versus External Economies in European Industry." *European Economic Review* 34, 805-30.

______, and ______ (1992). "External Effects in U.S. Procyclical Productivity." *Journal of Monetary Economics* 29 (April) 209-226.

Carlton, Dennis W. (1983). "Equilibrium Fluctuations when Price and Delivery Lag Clear Markets." *Bell Journal of Economics* 14 (no. 2) 562-72.

______ (1987). "The Theory and the Facts of How Markets Clear: Is Industrial Organization Useful for Understanding Macroeconomics?" In R. Schmalensee and R. Willig (eds.) *Handbook of Industrial Organization*. Amsterdam: North:Holland.

Diewert, Erwin (1976). "Exact and Superlative Index Numbers." *Journal of Econometrics* 4, 115-45.

Dhrymes, Phoebus J. (1971). *Distributed Lags: Problems of Estimation and Formulation.* San Francisco: Hoden-Day.

Evans, Charles L. (1992). "Productivity Shocks and Real Business Cycles." *Journal of Monetary Economics* 29 (April) 191-208.

Fair, Ray C. (1989). "The Production-Smoothing Model is Alive and Well." *Journal of Monetary Economics* 24 (November) 353-70.

Gordon, Robert J. (1990). *The Measurement of Durable Goods Prices.* Chicago: University of Chicago Press.

______ (1992). "Are Procyclical Productivity Fluctuations a Figment of Measurement Error?" Mimeo, Northwestern University.

Hall, Robert E. (1986). "Market Structure and Macroeconomic Fluctuations." *Brookings Papers in Economic Activity* 2, 285-322.

______ (1988). "The Relation Between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96 (Oct.) 921-947.

______ (1990). "Invariance Properties of Solow's Productivity Residual." In Peter Diamond (ed.) *Growth, Productivity, Employment* (Cambridge: MIT Press).

______ and Dale W. Jorgenson (1967). "Tax Policy and Investment Behavior." *American Economic Review* 57 (June) 391-414.

Hulten, Charles R. and Frank C. Wykoff (1981). "The Estimation of Economic Depreciation Using Vintage Asset Prices: An Application of the Box-Cox Power Transformation." *Journal of Econometrics* 15 (Aug.) 367-96.

Jorgenson, Dale W. (1990). "Productivity and Economic Growth." In Ernst Berndt and Jack Triplett (eds.) *Fifty Years of Economic Measurement* (Chicago: University of Chicago Press).

______, Frank Gollop, and Barbara Fraumeni (1987). *Productivity and US Economic Growth*. Cambridge: Harvard University Press.

______ and Zvi Griliches (1967). "The Explanation of Productivity Change." *Review of Economic Studies* 34, 249-283.

Krane, Spencer D. and Steven N. Braun (1991). "Production Smoothing Evidence from Physical-Product Data."*Journal of Political Economy* 99 (June) 558-581.

Kydland, Finn E. and Edward C. Prescott (1988). "Cyclical Movements of the Labor Input and Its Real Wage." Federal Reserve Bank of Minneapolis Working Paper no. 413.

Mankiw, N. Gregory (1989). "Real Business Cycles: A New Keynesian Perspective." *Journal of Economic Perspectives* 3 (Summer) 79-90.

Miron, Jeffrey A. and Stephen P. Zeldes (1988). "Seasonality, Cost Shocks, and the Production Smoothing Model of Inventories." *Econometrica* 56 (July) 877-908.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. "Building Blocks of Market Clearing Business Cycle Models" in Olivier J. Blanchard and Stanley Fischer (eds.) *NBER Macroeconomics Annual* (Cambridge: MIT Press).

Ramey, Valerie A. (1989). "Inventories as Factors of Production and Economic Fluctuations." *American Economic Review* 79 (June) 338-54.

______ (1991). "Nonconvex Costs and the Behavior of Inventories." *Journal of Political Economy* 99 (April) 306-334.

Rotemberg, Julio J., and Garth Saloner (1986). "A Super Game Theoretic Model of Price Wars during Booms." *American Economic Review* 77 (Dec.) 917-926.

______ and Lawrence H. Summers (1990). "Inflexible Prices and Procyclical Productivity." *Quarterly Journal of Economics* 105 (Nov.) 851-874.

______ and Michael Woodford (1991). "Markups and the Business Cycle" in Olivier J. Blanchard (ed.) *NBER Macroeconomics Annual* (Cambridge: MIT Press.)

______ and ______ (1992). "The Effects of Energy Price Increases on Economic Activity." Mimeo.

Sbordone, Argia. (1992). "Procyclical Productivity: External Economies or Labor Hoarding?" Mimeo, Federal Reserve Bank of Chicago.

______ (1993). "A Dynamic Model of Labor Hoarding." Mimeo, Federal Reserve Bank of Chicago.

Shapiro, Matthew D. (1993) "Cyclical Productivity and the Workweek of Capital." *American Economic Review* 83 (May) 229-33.

Sims, Christopher A. (1974). "Output and Labor Input in Manufacturing." *Brookings Papers in Economic Activity* 5 (no. 3) 695-728.

Solon, Gary, Robert Barsky, and Jonathan Parker (1992). "Measuring the Cyclicality of Real Wages: How Important is Composition Bias?" Mimeo, University of Michigan.

Trejo, Stephen J. (1991). "The Effect of Overtime Pay Regulation on Worker Compensation." *American Economic Review* 81 (Sept.) 710-40.

Table 1. Estimate of Returns to Scale from Equation (6):

$$dy_i = c_i + \gamma\, dx_i + dt_i$$

γ	1.090 (0.015)

Standard errors in parentheses.

Table 2. Estimates of Returns to Scale from Equation (13) [$\beta^v = \beta^m$]:

$$dy_i = c_i + \gamma\left[dm_i - \sigma\beta\left(\alpha_i^l + \alpha_i^k\right)\left(dp_i^v - dp_i^m\right)\right] + \nu_i .$$

	$\sigma = 0$	$\sigma\beta = 0.7$	$\sigma\beta = 1$
Estimate of γ	0.83 (0.017)	0.84 (0.021)	0.883 (0.022)
Implied χ	1.68	1.63	1.68

Standard errors in parentheses.

Table 3. Estimates of Returns to Scale from Equation (13') [$\beta^v = \beta^m$]:

$$dy_i = c_i + \gamma\left[dn_i - \sigma\beta\left(\alpha_i^l + \alpha_i^k + \alpha_i^e\right)\left(dp_i^{ve} - dp_i^n\right)\right] + \nu_i$$

	$\sigma = 0$	$\sigma\beta = 0.7$	$\sigma\beta = 1$
Estimate of γ	0.80 (0.017)	0.81 (0.020)	0.79 (0.019)
Implied χ	1.87	1.80	1.94

Standard errors in parentheses.

Table 4. Estimates of Returns to Scale and the Cyclicality of Prices:

$$dy_i = c_i + \gamma dm_i - \gamma\sigma\beta^v\left[\left(\alpha_i^l + \alpha_i^k\right) dp_i^v\right] + \gamma\sigma\beta^m\left[\left(\alpha_i^l + \alpha_i^k\right) dp_i^m\right] + \nu_i$$

	γ	$\gamma\sigma\beta^v$	$\gamma\sigma\beta^m$	Implied β^m/β^v	Implied χ
Materials and Energy	0.75 (0.015)	0.47 (0.064)	0.80 (0.065)	1.70	2.31
Materials only	0.72 (0.015)	0.40 (0.071)	0.97 (0.046)	2.42	2.75

Standard errors in parentheses.

Table 5. Test for Constancy of Materials/Labor-Capital Ratio

$$dm_i - dj_i = c_i + d73_i + t + d73 \cdot t + \delta(L)\, dy_i^* \qquad j = k, l$$

	Dependent Variable $dm_i - dl_i$	Dependent Variable $dm_i - dk_i$
δ_0	0.28 (0.017)	0.57 (0.033)
δ_1	0.036 (0.012)	0.14 (0.019)
δ_2	-0.060 (0.013)	-0.084 (0.019)
δ_3	-0.048 (0.010)	-0.14 (0.017)
δ_4	0.024 (0.013)	-0.11 (0.022)
δ_5	0.072 (0.014)	0.0048 (0.020)
δ_6	0.072 (0.020)	0.11 (0.037)
p-value for $\delta(1)=0$	0.00	0.00

Standard errors in parentheses.