Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 483

September 1994

CONSTANT RETURNS AND SMALL MARKUPS IN U.S. MANUFACTURING

Susanto Basu and John G. Fernald

ABSTRACT

We estimate that returns to scale are close to constant in two-digit gross output data. Value-added data appear instead to give significant increasing returns. We show why, with imperfect competition, value-added estimates are in general meaningless. We use data on intermediate inputs to correct the value-added estimates, and find that returns to scale again appear close to constant. Given that profits are small, our results imply that markups of price over marginal cost are also small.

Constant Returns and Small Markups in U.S. Manufacturing

Susanto Basu and John G. Fernald[1]

Why is productivity procyclical? That is, why do measures of labor productivity and total factor productivity rise in booms? The answer to this question sheds light on the relative merits of different models of business cycles. One recent class of explanations emphasizes the potential role of imperfect competition and increasing returns to scale. Robert Hall (1988, 1990), especially, has argued that departures from the traditional assumptions of perfect competition and constant returns are quantitatively important in explaining procyclical productivity.

These departures are potentially important in modeling business cycles. For example, recent papers that incorporate increasing returns and imperfect competition show that these features can magnify the propagation of shocks in an otherwise standard neoclassical model; they can explain why real wages and productivity rise in response to demand shocks such as an increase in government spending; and most strikingly, if increasing returns are large enough, they can lead to multiple equilibria, in which sunspots or monetary non-neutralities drive business cycles.[2] In assessing the merits of these models, one thus needs to know the empirical importance of increasing returns and imperfect competition.

---

[1] The authors are respectively: Assistant Professor of Economics, University of Michigan; and Staff Economist, Board of Governors of the Federal Reserve System. We thank Dale Jorgenson, Greg Mankiw, and an anonymous referee for helpful comments, and Barbara Fraumeni for help with the data. We thank Mike Woodford, especially, for extensive written comments on an earlier version of this paper. Basu is grateful to the National Science Foundation for financial support. This paper represents the views of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

[2] See, respectively, Rotemberg and Woodford (1991) and Hornstein (1993); Rotemberg and Woodford (1993) and Devereux, Head, and Lapham (1993); and Farmer and Guo (1994) and Beaudry and Devereux (1994).

We find that for 2-digit manufacturing industries returns to scale are close to constant and markups of price over marginal cost are small. This is true in both gross-output data and value-added data. Previous studies (notably Hall, 1990) report large apparent increasing returns in the same value-added data we use. The difference partly reflects specification error: the construction of value added implicitly assumes constant returns and perfect competition. With imperfect competition, the use of value added can lead to a spurious finding of increasing returns.

The intuition for this misspecification is straightforward. Real value added is constructed from gross output by subtracting the productive contribution of intermediate inputs. Gross output, for example, is books; value added is books lacking paper or ink. With perfect competition, this productive contribution is observable from factor payments to intermediate goods. With imperfect competition, however, the productive contribution of intermediate inputs is no longer observable, since marginal products exceed factor payments.

This paper has four sections. First, we review Hall's simple non-parametric method of estimating returns to scale, which essentially involves regressing output growth on input growth. We also relate estimates from gross output and value added, and show that it is unclear what is being estimated with value added. Second, we briefly discuss the data we use. Third, we present and discuss our results. Fourth, we conclude with thoughts for future research.

I. Relating Value Added and Gross Output Estimates of Returns to Scale[3]

In this section, we show why estimates of returns to scale from value-added data are likely to be unreliable. The construction of data on real value added assumes either perfect competition and constant returns, or else that intermediate inputs are used in fixed proportions to output. If neither

[3] The Appendix gives a complete and detailed derivation of the equations in this section.

assumption holds, then intermediate input use directly affects value added. Hence, in general one requires the same data to estimate returns to scale using value added as using gross output.[4]

We begin with an industry production function:

$$Y = F(K, L, M, T). \tag{1}$$

Y is gross output. K and L are primary inputs of capital and labor, while M is intermediate inputs of energy and materials. T indexes the state of technology.

Define s_J as the share of costs for input J in total revenue, and c_J as the share in total cost. We assume that all factors are freely variable and that firms are price takers in factor markets. Cost minimization then implies that the growth rate of output dy equals returns to scale γ multiplied by the cost-share-weighted growth in inputs dx, plus gross-output-augmenting productivity growth dt. That is, if dl, dk, and dm are the growth rates of L, K, and M, then

$$\begin{aligned} dy &= \gamma \cdot \left[c_L\, dl + c_K\, dk + (1 - c_L - c_K)\, dm \right] + dt \\ &\equiv \gamma \cdot dx + dt . \end{aligned} \tag{2}$$

This differs in two ways from the equation defining the Solow residual. First, the Solow method assumes constant returns, so that γ equals one. Second, the Solow method assumes perfect competition; together with constant returns this implies that there are no profits, so that capital's share is observable as a residual. If there are economic profits, however, then $s_K = 1-s_L-s_M-s_\pi$, where s_π is the share of profits. Required payments to capital are no longer observable as a residual, and hence we must instead estimate a required rental cost of capital, and allow shares in cost and revenue to differ.

Many researchers use data on value added instead of gross output. The preferred way to

---

[4] Our argument is unrelated to traditional arguments against the use of value added, which emphasize that a value-added function exists only if the production function satisfies certain separability properties (see, for example, Bruno (1978)). Our argument holds even if the production function is separable.

measure output net of intermediate inputs is to create a Divisia index.[5] The growth rate of the Divisia index of industry real value added dv is defined implicitly by assuming that dy is a weighted average of dv and dm, using revenue shares as weights: $dy \equiv (1-s_M)dv + s_M dm$. Hence,

$$dv \equiv \frac{(dy - s_M dm)}{(1-s_M)} = dy - \left(\frac{s_M}{1-s_M}\right)(dm - dy)\,. \tag{3}$$

Value added is like a partial Solow residual: the numerator subtracts the productive contribution of dm from dy. But as with Solow residuals, revenue-share weights measure this productive contribution correctly only under constant returns and perfect competition. The second identity in equation (3) indicates that if intermediate inputs are used in fixed proportions to output then the choice of weights is irrelevant: any weights that sum to one imply that value added grows at the same rate as gross output and intermediate inputs. In general, however, with imperfect competition real value added incorrectly weights the productive contribution of intermediate inputs, since this contribution exceeds the revenue share s_M.

We can express analytically the misspecification that arises from failures of constant returns and perfect competition. Define dx^v as the cost-weighted growth of primary inputs of K and L. Then dx equals $(1-c_M)dx^v + c_M dm$, and substituting equation (2) into equation (3), we find:

$$dv = \gamma \cdot \left[\frac{1-c_M}{1-s_M}\right] dx^v + \left(\frac{\gamma \cdot c_M - s_M}{1-s_M}\right) dm + \frac{dt}{1-s_M}\,, \tag{4}$$

We estimate this equation in Section III. It is easier to interpret, however, if we rewrite it in terms of the markup μ, the ratio of price to marginal cost. This markup and the degree of returns to scale are

[5] See Bruno (1978). NIPA data instead are primarily double-deflated, which as a measure of output suffers the same biases as the Divisia measure, plus an additional bias [See Bruno or Basu and Fernald (1994, App. I)]. For analytic simplicity, we focus here on the Divisia method.

not independent parameters, since they are related by the profit rate: $\gamma = \mu(1-s_\pi)$.[6] This in turn implies that γc_M equals μs_M, so that

$$dv = \gamma \cdot \left[\frac{1-c_M}{1-s_M}\right] dx^v + (\mu - 1)\left(\frac{s_M}{1-s_M}\right) dm + \frac{dt}{1-s_M}. \tag{5}$$

As expected, the growth of real value added depends on primary input growth and technological progress. Value-added growth also depends, however, on returns to scale, markups, and intermediate input growth. In particular, if price exceeds marginal cost ($\mu > 1$), then holding primary inputs fixed, the growth of intermediate inputs directly affects the growth of value added, since the full contribution of these inputs is not accounted for.

With imperfect competition, estimated returns to scale from value-added data are not estimates of the true gross-output returns to scale, i.e., the homogeneity of F. The bias can go either way. First, primary-input growth dx^v is multiplied by $\gamma(1-c_M)/(1-s_M)$. If there are economic profits, $(1-c_M)$ is less than $(1-s_M)$, providing a downward bias. Second, in the presence of markups the growth of intermediate inputs is an omitted variable correlated with primary input use, providing an upward bias.

It is not always the case, however, that the parameter of interest is the degree of returns to scale in gross output. Consider the following specialization of (1):

$$Y = G(V(K,L,T), H(M)). \tag{6}$$

Macroeconomists usually model sectoral production functions with a separable form as in (6) (Cobb-Douglas, for example) and assume that all returns to scale are in V, arising perhaps from overhead capital or labor. This requires that G be homogeneous of degree one in V and H, and that H be homogeneous of degree one in M. With increasing returns, value-added returns to scale γ^v then exceed gross-output returns to scale γ, since γ is a weighted average of γ^v and one. The relationship between

---

[6] The first-order conditions imply $F_J J/Y = \mu s_J$, J=L,L,M. Thus, $\gamma \equiv \Sigma(F_J J/Y) = \mu \Sigma s_J = \mu(1-s_\pi)$.

γ^v and γ is:

$$\gamma^v = \gamma\left(\frac{1-c_M}{1-\gamma\, c_M}\right). \tag{7}$$

As written, this is an accounting relationship. One source of economic intuition for γ^v is that under some circumstances it correctly captures "economy-wide" returns to scale, as small increasing returns at the plant level translate into larger increasing returns for the economy overall. Suppose, for example, that final output is produced at the end of a large number of stages. At each stage a representative firm with returns to scale γ uses all the output of the previous stage as intermediate input, and also uses primary inputs of capital and labor. Then in the limit as the number of stages goes to infinity, the percent change in national product (the output of the last stage) equals γ^v times the cost-share-weighted growth rate of aggregate primary inputs. Thus, γ^v is arguably the appropriate concept for calibrating returns to scale in a one-sector macroeconomic model.

Even if we want this value-added returns to scale, we still in general require data on intermediate inputs. The reason is that we do not observe V directly, but must infer it from observable gross output and intermediate inputs. The only cases in which we do not need intermediate input data to calculate γ^v are if there are constant returns and perfect competition, or if intermediate inputs are used in fixed proportions to output. One can show this with some algebra by rewriting value-added growth as

$$dv = \gamma^v \cdot dx^v + (\mu-1)\left[\frac{s_M}{(1-s_M)(1-\mu s_M)}\right](dm - dy) + \frac{dt}{1-\mu s_M}. \tag{8}$$

The second term on the right-hand side is zero if and only μ equals one or dm grows at the same rate as dy. With data on the price and quantity of intermediate input use, however, we can correct for omitted variable bias and estimate γ^v. In particular, we estimate the following equation:

$$dv + \frac{s_M}{1-s_M}dm = \left[\frac{\gamma^v}{1-c_M+\gamma^v c_M}\right]\left[\frac{(1-c_M)dx^V + c_M dm}{1-s_M}\right] + \frac{dt}{1-s_M} \tag{9}$$

This non-linear equation has two advantages over equation (8). First, it doesn't have dy--by definition an endogenous variable--on the right-hand side. Second, it is expressed in terms of a single parameter, γ^v. As before, it cannot be the case that both γ^v and μ are constant structural parameters, so it is inconsistent to estimate both.

II. The Data

We combine two sources of data on inputs and output for 21 manufacturing industries, at roughly the two-digit SIC level. For both data sets, we omit petroleum and coal products (SIC 29), giving us 20 industries. First, we use Robert Hall's NIPA data on real value added and hours worked and BEA data on industry capital stocks. NIPA constructs a double-deflated measure of real value added, and the two input measures make no adjustments for changes in quality or composition over time.[7]

Second, we use unpublished data provided by Dale Jorgenson and Barbara Fraumeni on industry-level gross output and inputs of capital, labor, energy, and materials. These sectoral accounts seek to provide accounts that are, to the extent possible, consistent with the economic theory of production. These data are available both with and without an adjustment for input quality. The quality adjustment essentially involves taking account of changes over time in input composition. Computers, for example, give a higher service flow per dollar than factories, since they depreciate faster. Similarly, engineers and janitors make a different marginal contribution to output, and one can use information on relative factor payments to adjust for the differences.

[7] Hall (1990, pp82-83) describes the data sources. NIPA value added for petroleum and coal products is extrapolated from gross output, not double-deflated. The resulting data are unsuitable for our purpose, so we omit this sector. This has no substantive effect on our gross-output results. The other difference from two-digit classification is that we divide transport equipment (SIC 37) into "motor vehicles" (SIC 371) and "other transport equipment."

Without the quality adjustment, the Jorgenson measures of labor and capital input are roughly the same as Hall's measures--hours worked and the capital stock.[8] Quality-adjusted inputs are preferable, however. Solon, Barsky, and Parker (1994) show that failing to adjust input prices for changes in quality (i.e., composition) makes real wages appear acyclical when they are in fact strongly procyclical. We make the dual point: failing to adjust inputs for quality change affects estimated returns to scale.

We use the Jorgenson data to estimate returns to scale directly, as in equation (2). In addition, we use equations (4) and (9) and combine the intermediate-input data with NIPA data to assess the importance of the value-added biases we discussed. In order to compare the two types of data directly, we construct two measures of real value added for Jorgenson's data. We create the theoretically preferred Divisia index using the Tornquist approximation to the continuous time formula in equation (3), using log-changes in place of derivatives and using average shares in adjacent time periods. We also calculate double-deflated value added as $V^{DD} = Y - M$, where all quantities are calculated in 1982 dollars.

We estimate payments to capital following Hall and Jorgenson (1967) and Hall (1990), computing a series for the user cost of capital. See Basu and Fernald (1994) for details.

III. Results

We find that returns to scale are very close to constant with the theoretically preferred data, i.e., gross output and quality-adjusted inputs. NIPA value-added data show much stronger evidence of

---

[8] The capital stock data differ in two ways. First, Jorgenson and Fraumeni include land and inventories. Second, in creating stock estimates from the underlying investment data, Jorgenson and Fraumeni assume that the underlying assets depreciate geometrically while the BEA net capital series Hall uses assume straight-line depreciation over the estimated service life of the asset. See Basu and Fernald (1994, Appendix II) for an overview of the Jorgenson-Fraumeni data. For a complete description, see Jorgenson, Gollop, and Fraumeni (1987).

increasing returns. We estimate that about two-thirds of this apparent increasing returns results from the value-added biases we identified in Section I and the theoretical shortcomings of the NIPA data. The remaining extent of increasing returns seems quantitively small.

In all tables, we summarize returns to scale in manufacturing by estimating the 20 equations jointly, constraining returns to scale to be the same in all industries. We present results both uninstrumented and instrumented.[9] As an alternative summary of the data, Table 1 shows the median estimate of returns to scale from estimating the 20 equations separately.

Table 1 presents our bottomline estimates of returns to scale for gross output from equation (2). Returns to scale are consistently close to constant. In the uninstrumented constrained estimates, with quality-adjusted inputs, returns to scale are 1.13. One expects an upward bias, however, because of the positive correlation between technology shocks and input use. Indeed, when we instrument, gross output returns to scale fall to 1.03, no longer statistically greater than one. The median estimates are very similar to the constrained estimates, and also indicate that returns to scale are close to constant.[10]

Table 1 also shows value-added results, using the value-added analogue of equation (2) with no correction for the biases identified in Section I. These results are less consistent. The constrained uninstrumented results for both data sets indicate sizeable and statistically significant *decreasing* returns to scale, even in the NIPA data. Instrumenting with non-quality adjusted inputs, however, gives strong

---

[9] Because input use is likely to be correlated with technology shocks, we seek demand-side instruments for input use. We use as instruments the growth rate of the price of oil deflated by the GDP deflator; the growth rate of real government defence spending, along with one lag; and the political party of the President, also along with one lag. Hall (1990) discusses these instruments.

[10] The 20 individual point estimates for gross-output returns to scale have a standard deviation of 0.22. Durable-goods industries show some evidence of increasing returns: instrumented, the constrained estimate is 1.06. For non-durables, the instrumented estimate is 0.84.

evidence of increasing returns: the NIPA estimate is 1.17, while our Divisia estimate is 1.12.[11] (Adjusting inputs for quality change reduces this point estimate to 0.95. Hence, the quality adjustment--which is clearly preferred on theoretical grounds--makes a large difference to value-added results.)

It is a striking anomaly that value-added returns to scale appear much larger instrumented, given that one expects the reverse. This anomaly is easily explained, however. Section I identified two biases in value-added estimates: any profits tend to bias the direct estimate of γ down, whereas the omitted materials term tends to bias it up. Our results suggest that the net bias is positive for the instrumented estimates, but negative for the uninstrumented ones. Reassuringly, we note the reverse and expected pattern for the gross-output results: the estimate of γ always falls in the instrumented regressions.

We conclude that using the correct data gives a result very different from the ones cited in the literature: returns to scale are constant, not strongly increasing.

The value-added estimates in Table 1 have no clear interpretation, since in general they are not unbiased estimates of either γ or γ^{v}. Table 2 augments the value-added data as required by equation (4) to recover the gross-output returns to scale. The first line uses the theoretically preferred Divisia value added and quality-adjusted inputs. As expected, this essentially replicates the first line of Table 1, giving almost exactly constant returns to scale. The second line uses our closest approximation to the NIPA data, i.e., double-deflated value-added and non-quality adjusted inputs. Returns to scale instrumented rise to 1.05.[12] The third line shows NIPA double-deflated value-added and non-quality

---

[11] Hall (1990) uses this same data and sample period; but his median estimate of returns to scale is 2.2. Hall obtains these estimates by running the regression "in reverse"--regressing dx on dy and estimating $1/\gamma$. Bartelsman (1993) shows that with more than one instrument, there is a sizeable small-sample bias to Hall's method, which in this case leads to spuriously large estimates of returns to scale. We confirmed Bartelsman's claim: with both value-added and gross-output data, running the reverse regression produces stronger evidence of increasing returns.

[12] The movement from Divisia to double-deflated value-added adds approximately 0.02 to the estimate of returns to scale, while the movement from quality-adjusted to non-quality adjusted inputs adds about 0.03.

adjusted inputs. Returns to scale instrumented are 1.09, and significantly exceed one.

Overall, we can thus account for perhaps two-thirds of the apparent increasing returns in the NIPA data. Correcting for the value-added biases cuts apparent increasing returns in half, from 1.17 to 1.09. Table 2 then suggest that the theoretical shortcomings of the NIPA data--double deflation and the lack of quality adjustment--can account for about 0.05 of the deviation from constant returns. Hence, our best estimate of gross-output returns to scale using NIPA value added is 1.04 (1.09 minus 0.05). This is close to our gross-output estimate of 1.03; it is also not sizeable.

Our method is simple and transparent. This is a virtue, but one might worry that perhaps the simplifying assumptions (e.g., no quasi-fixed inputs) matter. It is thus interesting to compare our results with those of Morrison (1990), who also estimates returns to scale in U.S. manufacturing using gross-output data. She estimates a complex eight-equation structural system with a very large number of parameters. The complexity makes it hard to tell whether her results are driven by her use of gross output, or by the novel elements of her technique. Morrison uses non-quality adjusted inputs for 17 two-digit manufacturing industries, and finds returns to scale of about 1.12. Using the same industries and non-quality adjusted inputs, our estimate is 1.09—statistically indistinguishable from hers. Hence, it appears that our simplifying assumptions make little difference; it is the choice of data that matters.

As discussed in Section I, for some calibration exercises one may want value-added returns to scale, not gross-output returns to scale. The value-added estimates in Table 1 are not, however, estimates of γ^v. Intermediate inputs account for about 2/3 of costs in our data. Hence, using equation (7), our gross-output γ of 1.03 translates into a value-added γ^v of 1.09. (Note, again, that this estimate is not significantly different from one). At the high end, the unadjusted NIPA estimate of 1.09 translates into a γ^v of 1.30. Our double-deflated non-quality-adjusted estimate of 1.05 translates into a γ^v of 1.17. These figures are calculated, but if γ^v is the true structural parameter we should estimate it directly. Table 3 reports the results of estimating γ^v from equation (9). The results agree almost perfectly with our calculations; in all cases the estimated γ^v is not statistically distinguishable from its predicted value. Note that the uncorrected value-added estimates in Table 1 are smaller than the estimates in Table 3: one cannot estimate γ^v with value-added data alone. Note also that most of the large estimate of γ^v with NIPA data is explained by the non-quality adjustment of the inputs.

Why do our results for returns to scale (be it γ or γ^v) matter? As noted in the introduction, a wide variety of recent papers use the twin assumptions of increasing returns and imperfect competition to generate striking results. But striking results are of considerably less interest if they rely on parameter values that are empirically irrelevant. Indeed, measurement is ahead of business-cycle theory: given our results, the papers we cited earlier uniformly presuppose returns to scale that are too large to be plausible.

One might hope that some of the appealing theoretical results can be saved by dropping the large increasing returns but keeping a high degree of imperfect competition, as measured by a high markup of price over marginal cost. Unfortunately, such an approach counterfactually requires very large pure profits. Returns to scale and markups are linked by the profit-rate: $\gamma = \mu(1\text{-}s_\pi)$. We estimate that s_π averages about 0.04 over our sample period. (This is of course sensitive to the calculation of the user cost of capital; if anything, it is likely to be too high, not too low.) Our preferred point estimate of γ of 1.03 implies an average markup of about 1.08. Hence, if U.S. manufacturing industries have essentially constant returns to scale then markups must be small.

Domowitz et al. (1988) estimate that the markup of price over marginal cost in U.S. manufacturing averages about 1.6 in gross-output data. We, however, find this markup prima facie unreasonable. Their method assumes constant returns, but if γ is 1 and μ is 1.6, then the implied profit rate is 38 percent. Alternatively, note that with constant returns, the output elasticities of all factors sum to 1. The sum of output elasticities with respect to labor and materials is $\mu(s_L+s_M)$. Since (s_L+s_M) is about 0.8, if μ is 1.6 then the output elasticity with respect to capital is *negative*: a one percent increase in capital lowers output by about 0.3 percent! Thus, a markup of 1.6 is not defensible. With constant returns, the highest markup consistent with even a zero output elasticity of capital is about 1.2.

One last way of rescuing the assumption of large increasing returns is to argue that small increasing returns at the level of each productive unit translates to a large degree of economy-wide increasing returns. As noted earlier, the resulting "economy-wide increasing returns" is exactly what we call the "value-added returns to scale," γ^v. As we showed, for our preferred estimate of γ of 1.03, γ^v is still quite small: about 1.10 or less. Thus, although the theoretical point is valid, in practical terms there is probably little difference in calibrating a model with γ^v rather than γ.

IV. Conclusion.

We have shown that returns to scale in U.S. manufacturing are approximately constant. Given that pure profits are small, the markup of price over marginal cost must then be small as well. Previous results to the contrary are primarily due to specification error in the use of value-added data.

Caballero and Lyons (1992) have argued that returns to scale internal to firms might be constant or decreasing; but productive spillovers to output lead to overall increasing returns that are external to firms. A number of papers (e.g., Baxter and King, 1991) invoke external increasing returns to explain fluctuations. In a related paper (Basu and Fernald, 1994), we show that these apparent productive spillovers also reflect specification error from the use of value-added data.

Our results paint a picture of the macroeconomy that is quite different from the one proposed by Hall (1988, 1990) and others, and they are important for rethinking the now-standard calibration of many macroeconomic models. As we have argued, many issues depend on accurate measures of the degree of returns to scale and the size of the markup. These include the fundamental question with which we began: why is productivity procyclical? Our evidence indicates that one plausible answer is unlikely to be the main reason: productivity is not procyclical because of increasing returns to scale.

BIBLIOGRAPHY

Basu, Susanto and John G. Fernald, 1994. "Are Apparent Productive Spillovers a Figment of Specification Error?" International Finance Discussion Paper no. 463. Board of Governors of the Federal Reserve System.

Beaudry, Paul and Michael Devereux, 1994. "Monopolistic Competition, Price Setting and the Effects of Real and Monetary Shocks." mimeo.

Bruno, Michael, 1978. "Duality, Intermediate Inputs, and Value-Added." In M. Fuss and D. McFadden, eds., *Production Economics: A Dual Approach to Theory and Practice*, Volume 2 (North-Holland, Amsterdam).

Baxter, M. and R. King, 1991. "Productive Externalities and Business Cycles." Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics discussion paper #53.

Caballero, Ricardo and Richard Lyons, 1992. "External Effects in U.S. Procyclical Productivity." *Journal of Monetary Economics* 29, 209-226.

Devereux, Michael, Allen Head, and Beverly Lapham, 1993. "Monopolistic Competition, Increasing Returns, and the Effects of Government Spending." Institute for Economic Research Discussion Paper no. 894. Queen's University.

Domar, Evsey, 1961, "On the Measurement of Technological Change." *Economic Journal* 71: 709-729.

Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen, 1988. "Market Structure and Cyclical Fluctuations in U.S. Manufacturing." *Review of Economics and Statistics* 55-66.

Farmer, Roger and Jang-Ting Guo, 1994. "Real Business Cycles and the Animal Spirits Hypothesis." *Journal of Economic Theory* 63: 42-72.

Hall, Robert E., 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96, 921-947.

Hall, Robert E., 1990. "Invariance properties of Solow's Productivity Residual." in Peter Diamond, ed., Growth/Productivity/Employment (MIT Press, Cambridge).

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

Hornstein, Andreas, 1993. "Monopolistic Competition, Increasing Returns to Scale, and the Importance of Productivity Shocks." *Journal of Monetary Economics* 31:299-316.

Jorgenson, Dale, Frank Gollop, and Barbara Fraumeni, 1987. *Productivity and U.S. Economic Growth.* (Harvard University Press, Cambridge).

Morrison, Catherine, 1990. "Market Power, Economic Profitability and Productivity Growth Measurement: An Integrated Structural Approach." NBER Working Paper no. 3355.

Rotemberg, Julio and Michael Woodford, 1991. "Markups and the Business Cycle." in Blanchard, O. and S. Fischer, eds. NBER Macroeconomics Annual 1991 (MIT Press, Cambridge).

Rotemberg, Julio and Michael Woodford, 1993. "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets." NBER Working Paper no. 4502.

Solon, Gary, Robert Barsky, and Jonathan A. Parker, 1994. "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics* CIX:1-25 (February).

Table 1
Estimated Returns to Scale in Gross Output and Value Added

Data Set	Estimation Method	Uninstrumented		Instrumented	
		Non-qual adj. inputs	Qual. adj. inputs	Non-qual adj. inputs	Qual. adj. inputs
1. Gross Output	a. Constrained System	1.13 (0.01)	1.13 (0.01)	1.05 (0.02)	1.03 (0.02)
	b. Indiv. Eqn. Medians	1.04	1.03	1.04	1.01
2. Divisia VA	a. Constrained System	0.85 (0.03)	0.68 (0.03)	1.12 (0.06)	0.95 (0.06)
	b. Indiv. Eqn Medians	1.04	1.00	1.12	1.15
3. NIPA VA	a. Constrained System	0.77 (0.03)	--	1.17 (0.06)	--
	b. Indiv. Eqn. Medians	1.10	--	1.30	--

Sample is 20 2-digit manufacturing industries from 1953-1984. Constrained estimates are SUR/3SLS. Standard errors are in parentheses.

Table 2
Corrected Gross Output Returns to Scale (γ) from Value-Added Data

Data Set	Uninstrumented (SUR)	Instrumented (3SLS)
Divisia V.A. with Qual-Adj Inputs	0.98 (0.02)	1.00 (0.04)
Double-Defl. V.A. with Non-Qual Adj Inputs	0.99 (0.01)	1.05 (0.02)
NIPA V.A.	1.00 (0.01)	1.09 (0.03)

Sample is 20 2-digit manufacturing industries from 1953-1984. Constrained estimates are SUR/3SLS. Standard errors are in parentheses.

Table 3
Corrected Value-Added Returns to Scale (γ^{V})

Data Set	Uninstrumented (SUR)		Instrumented (3SLS)	
	Non-qual. Adj. Inputs	**Qual.-Adj Inputs**	**Non-qual. Adj. Inputs**	**Qual.-Adj Inputs**
Divisia Value Added	1.01 (0.03)	0.99 (0.02)	1.17 (0.05)	1.05 (0.05)
NIPA	1.00 (0.03)	--	1.26 (0.06)	--

Sample is 20 2-digit manufacturing industries from 1953-1984. Constrained estimates are SUR/3SLS. Standard errors are in parentheses.

Appendix:
Detailed Derivations of Equations in Section I

This appendix derives the equations in the Section I of the paper in detail. We begin with the following production function for an industry:

$$Y = F(K, L, M, T). \tag{A1}$$

Y is gross output. K and L are primary inputs of capital and labor, while M is intermediate inputs of energy and materials. T is an index of the state of technology. We assume that the production function is homogeneous of degree γ in capital, labor, and intermediate goods.

Differentiating the production function (A1), we can express the growth rate of output as:

$$dy = \left(\frac{F_K K}{Y}\right)dk + \left(\frac{F_L L}{Y}\right)dl + \left(\frac{F_M M}{Y}\right)dm + dt. \tag{A2}$$

Lower-case letters represent logs of their upper-case counterparts, so all of the quantity variables in (A2) are log differences, or growth rates. F_J represents the derivative of F with respect to argument J. For convenience, we normalize to unity the elasticity of output with respect to technology T. The sum of the output elasticities equals the degree of returns to scale γ, so that

$$\gamma = \left(\frac{F_K K}{Y}\right) + \left(\frac{F_L L}{Y}\right) + \left(\frac{F_M M}{Y}\right). \tag{A3}$$

Suppose that firms have some degree of monopoly power in the goods markets but are price takers in factor markets. The first-order conditions for cost minimization then imply:

$$P_J = \lambda F_J \text{, } J = K,L,M,$$

where λ is a Lagrange multiplier with the interpretation of marginal cost, and P_J is the shadow value of the Jth input as perceived by the firm. (As discussed below, this shadow value may or may not be observable as the input price.) By definition, the markup μ of the output price over marginal cost is P/λ. Hence, we can rewrite the above equation as:

$$\left[\frac{F_J J}{Y}\right] = \mu\left[\frac{P_J J}{PY}\right] \equiv \mu s_J, \qquad J = K, L, M. \tag{A5}$$

Thus, the elasticity of output with respect to any factor J equals a markup μ multiplied by the share of that input in total revenue, s_J. Note that the price of capital, P_K, must be defined as the rental cost of capital. It does not include possible profits, which generally are also payments to capital. With perfect competition, where $\mu = 1$, equation (A5) just states that the elasticity of output with respect to any input equals the input's share in revenue. Under imperfect competition, the elasticity of output exceeds the revenue share.

Equations (A3) and (A5) together imply that

$$\gamma = \mu\left[\frac{P_K K + P_L L + P_M M}{PY}\right] = \mu(1-s_\pi), \tag{A6}$$

where s_π is the share of profits in total revenue. Equation (A6) then implies that the output elasticities, which equal μs_J, also equal γc_J, where c_J is the share of payments to input J in total costs. Hence, we can write the total differential of output as:

$$\begin{aligned} dy &= \gamma\cdot\left[c_L dl + c_K dk + (1 - c_L - c_K)\, dm\right] + dt \\ &\equiv \gamma\cdot dx + dt. \end{aligned} \tag{A7}$$

dx is a cost-weighted sum of the growth rates of the various inputs. We weight factors by their cost shares rather than revenue shares, since as Hall (1988) shows, with markup pricing the measured Solow residual is procyclical, even with constant returns and no change in technology. Note that nothing in the above derivation relies on profit maximization: it depends solely on the production function and cost minimization.

Suppose that all factors are freely variable, so that there are no quasi-fixed factors. Then we can observe P_L and P_M as the market prices of L and M, and construct P_K as the steady-state rental rate of capital following Hall and Jorgenson (1967) and Hall (1990), allowing us to create the cost shares. If, however, capital is quasi-fixed (sunk in the short run) then the marginal product of capital does not equal its steady-state rental rate but rather its current shadow rental. That is, the user cost of capital in the Hall-Jorgenson formula should be multiplied not by the price of investment goods, which is the usual procedure, but by the shadow value of capital, marginal q. (The shadow rental also includes expected capital gains.)

This problem is not significant for estimates of returns to scale, however, for two reasons. First, as argued, quasi-fixity affects only the period-by-period computation of the input shares, not the growth rate of capital (or any other quasi-fixed input). Since these shares are constant to a first-order Taylor approximation, any errors caused by failure to track the movements of the shares is likely to be small. Second, mismeasurement of the rental rate of capital has its strongest effect on capital's share. But since the growth rate of capital is almost uncorrelated with the business cycle, errors in measuring capital's share are unlikely to cause significant biases in a study of cyclical productivity. Caballero and Lyons (1989) present simulations indicating that maximum biases from quasi-fixity are likely to be on the order of 3% of the estimated coefficients.

It is convenient to define a cost-weighted sum of primary inputs of capital and labor. Since this is the "value-added" analogue to dx, we call it dx^v, and define it as:

$$dx^{v} = \left(\frac{c_K}{c_K+c_L}\right)dk + \left(\frac{c_L}{c_K+c_L}\right)dl. \tag{A8}$$

It is clear that we can rewrite (A7) as:

$$dy = \gamma \cdot \left[(1-c_M)dx^{v} + c_M dm\right] + dt. \tag{A9}$$

Many researchers use data on value added instead of gross output, where measures of real value added attempt to subtract from gross output the productive contribution of intermediate goods. Hence, gross output is shoes, while value added is "shoes lacking leather, made without power" (Domar 1961, p716). Despite its unintuitive nature, researchers use value added for at least two reasons. First, since sectoral value added sums to national output, it is a natural simplification for macroeconomists to focus, even at a sectoral level, on value added and primary inputs of capital and labor. Second, because national accounting data are widely distributed, data on sectoral value added tend to be better known than the gross-output data used to construct them.

The preferred way to measure output net of intermediate inputs is to create a Divisia index. The growth rate of the Divisia index of industry real value added dv is defined as

$$\frac{dV}{V} \equiv dv = \frac{1}{1-s_M}dy - \frac{s_M}{1-s_M}dm, \tag{A10}$$

where s_M is the share of materials in revenue. Substituting from equation (A9) for dy, we find:

$$dv = \gamma \cdot \left[\frac{1-c_M}{1-s_M}\right]dx^{v} + \left(\frac{\gamma \cdot c_M - s_M}{1-s_M}\right)dm + \frac{dt}{1-s_M}, \tag{A11}$$

Since γc_M equals μs_M, this equation can also be written as:

$$dv = \gamma \cdot \left[\frac{1-c_M}{1-s_M}\right] dx^v + (\mu-1)\left(\frac{s_M}{1-s_M}\right) dm + \frac{dt}{1-s_M}. \tag{A12}$$

If there are constant returns to scale and no markups, then the growth rate of value added equals the growth rate of primary inputs plus technological progress. That is,

$$dv = dx^v + \frac{dt}{(1-s_M)}. \tag{A13}$$

One implication of equation (A13) is that if s_M is constant, then as Hall (1990) notes, under competition and constant returns the productivity residual ($dv - dx^v$) calculated from a Divisia index of value added is uncorrelated with any variable that neither causes productivity shifts, nor is caused by productivity shifts. In addition, dx^v can be calculated with either cost or revenue shares, since there are no profits.

In the presence of markups, however, equation (A12) shows that the growth of materials and energy directly affect value-added growth. Intuitively, value added is calculated by subtracting from gross output the productive contribution of intermediate goods, assuming that the elasticity of output with respect to intermediate inputs equals its revenue share. With markups, this elasticity exceeds its revenue share. Hence, some of the contribution of materials and energy is attributed to value added. As a result, the value-added productivity residual is correlated with any variable that is correlated with intermediate-goods use, regardless of whether it is correlated with technology. Thus, Hall's argument that under constant returns the cost-based Solow residual should be invariant is true in gross-output data, but, apart from special cases, is not true with value-added data. In particular, if markups exceed 1, then a rise in energy prices that causes energy use to fall affects the growth of value added.

As we noted in the text, it is not always the case that the parameter of interest is the degree of returns to scale in gross output. Consider the following specialization of (A1):

$$Y = G(V(K,L,T), H(M)) . \quad \textbf{(A14)}$$

Macroeconomists usually model sectoral production functions with a separable form as in (A14)--Cobb-Douglas, for example. There are three possible sources of increasing returns in Y: increasing returns in V, increasing returns in H, or increasing returns in how V and H are combined to produce G. Examples of these different sources include overhead labor and capital, overhead intermediate inputs such as advertising, or declining marginal cost. Suppose one assumes that all returns to scale are in V. One might hope that estimates of returns to scale with value-added data are estimates of the homogeneity of V.

The sum of output elasticities γ with respect to K, L, and M is then

$$\gamma = \frac{G_V V}{Y}\left(\frac{V_K K}{V} + \frac{V_L L}{V}\right) + \frac{G_H H}{Y}\frac{H_M M}{H}. \quad \textbf{(A15)}$$

Let G be homogeneous of degree one in V and H, and H be homogeneous of degree one in M. Then $G_V V/Y$ equals $(1-G_M M/Y)$, and the first order conditions tell us that the output elasticity with respect to materials is γc_M. Hence, we can rewrite this equation as

$$\gamma = (1-\gamma c_M)\gamma^v + \gamma c_M \quad \textbf{(A16)}$$

The relationship between γ^v and γ is then:

$$\gamma^v = \gamma\left(\frac{1-c_M}{1-\gamma c_M}\right). \quad \textbf{(A17)}$$

With increasing returns, value-added returns to scale γ^v exceed gross-output returns to scale γ, since

γ is a weighted average of γ^v and one.

One source of economic intuition for γ^v is that under some circumstances, it correctly captures "economy-wide" returns to scale, as small increasing returns at the plant level translate into larger increasing returns for the economy overall. Suppose, for example, that final output is produced at the end of an infinite number of stages. At each stage a representative firm with returns to scale γ uses all the output of the previous stage as intermediate input, and also uses primary inputs of capital and labor. Then the percent change in national income--the output of the last (nth) stage--is

$$\begin{aligned} dy_n &= \gamma(1-c_M)dx_n^v + \gamma c_M dy_{n-1} \\ &= \gamma(1-c_M)dx_n^v + \gamma c_M \gamma(1-c_M)dx_{n-1}^v + (\gamma c_M)^2 dy_{n-2}. \end{aligned} \tag{A18}$$

We can substitute into this equation for dy_{n-j}, and let j go to infinity. Since each firm is identical, dx_i^v is the same for all i as for the aggregate. This gives an infinite sum, which sums to γ^v times the cost-share-weighted growth rate of aggregate primary inputs. Thus, γ^v is plausibly the appropriate concept for calibrating the degree of returns to scale of the production function in a one-sector macroeconomic model.

Even if we want this value-added returns to scale, we still in general require data on intermediate inputs. The reason is that we do not observe V directly, but must infer it from observable gross output and intermediate inputs. Note that in equation (A14), with the homogeneity assumptions we've made, the growth of true value added, which we will denote dv^c, is implicitly defined as follows:

$$dv^c = dy - \left(\frac{G_M M}{G_V V}\right)(dm - dy). \tag{A19}$$

Comparing this to the equation defining standard value added dv, and substituting for the output elasticities with respect to materials and value-added, we can write dv as

$$dv = dv^{c} + \left(\frac{\mu s_M}{1-\mu s_M} - \frac{s_M}{1-s_M}\right)(dm-dy). \tag{A20}$$

The growth of this corrected value-added is $\gamma^{v}dx^{v} + (V_T T/V)dT/T$. Since gross-output augmenting technical change (which we defined as dt) is $(G_V V/Y)(V_T T/V)dT/T$, the true value-added-productivity-growth term $dt/(1-\mu s_M)$. Substituting for dv^c in the above equation and rearranging slightly gives the equation in the text:

$$dv = \gamma^{v}\cdot dx^{v} + (\mu-1)\left[\frac{s_M}{(1-s_M)(1-\mu s_M)}\right](dm-dy) + \frac{dt}{1-\mu s_M}. \tag{A21}$$

The second term on the right-hand side is zero if and only μ equals one or dm grows at the same rate as dy.

With data on the price and quantity of intermediate input use, however, one can correct for omitted variable bias and estimate γ^{v}. As before, we want to estimate an equation that has only one parameter, γ^{v} or μ, in it, since the parameters are related by profits. Also, we prefer to estimate an equation that doesn't have dy on the right-hand side. Rearranging equation (4) in the text gives the following:

$$dv + \frac{s_M}{1-s_M}dm = \gamma\left[\frac{1-c_M)dx^{v} + c_M dm}{1-s_M}\right] + \frac{dt}{1-s_M}. \tag{A22}$$

Substituting for γ in terms of γ^{v} gives the following equation, which we estimate:

$$dv + \frac{s_M}{1-s_M}dm = \left[\frac{\gamma^v}{1-c_M+\gamma^v c_M}\right]\left[\frac{(1-c_M)dx^V+c_M dm}{1-s_M}\right] + \frac{dt}{1-s_M} \tag{A23}$$

The disturbance term is approximately, but not exactly correct. If technical change is value-added augmenting, then the disturbance term in (A21), $dt/(1-\mu s_M)$ is invariant to shocks, say, to s_M or μ. If μ is close to one, however, then (A23) gives a good approximation. (This issue of whether technical change is gross-output or value-added augmenting is probably not crucial, since our results are virtually unaffected by whether the regression has dt or $dt/(1-s_M)$ as the disturbance term.)

To summarize, in this appendix (and in the theory section in the paper) we show the relationship between estimates of returns to scale and markups using data on gross output and value added. With imperfect competition, the data required to estimate returns to scale and markups with value-added data are the same as the required with gross output data. This is true even if one wants to estimate the degree of returns to scale to a value-added function. In particular, one cannot measure returns to scale and markups using data on primary inputs of capital and labor alone. One requires data on intermediate inputs as well.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	1994	
483	Constant Returns and Small Markups in U.S. Manufacturing	Susanto Basu John G. Fernald
482	The Real Exchange Rate and Fiscal Policy During the Gold Standard Period: Evidence from the United States and Great Britain	Graciela L. Kaminsky Michael Klein
481	The Debt Crisis: Lessons of the 1980s for the 1990s	Graciela L. Kaminsky Alfredo Pereira
480	Who Will Join EMU? Impact of the Maastricht Convergence Criteria on Economic Policy Choice and Performance	R. Sean Craig
479	Determinants of the 1991-93 Japanese Recession: Evidence from a Structural Model of the Japanese Economy	Allan D. Brunner Steven B. Kamin
478	On Risk, Rational Expectations, and Efficient Asset Markets	Guy V.G. Stevens Dara Akbarian
477	Finance and Growth: A Synthesis and Interpretation of the Evidence	Alexander Galetovic
476	Trade Barriers and Trade Flows Across Countries and Industries	Jong-Wha Lee Phillip Swagel
475	The Constancy of Illusions or the Illusion of Constancies: Income and Price Elasticities for U.S. Imports, 1890-1992	Jaime Marquez
474	The Dollar as an Official Reserve Currency under EMU	Michael P. Leahy
473	Inflation Targeting in the 1990s: The Experiences of New Zealand, Canada, and the United Kingdom	John Ammer Richard T. Freeman
472	International Capital Mobility in the 1990s	Maurice Obstfeld
471	The Effect of Changes in Reserve Requirements on Investment and GNP	Prakash Loungani Mark Rush

---

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	1994	
470	International Economic Implications of the End of the Soviet Union	William L. Helkie David H. Howard Jaime Marquez
469	International Dimension of European Monetary Union: Implications For The Dollar	Karen H. Johnson
468	European Monetary Arrangements: Implications for the Dollar, Exchange Rate Variability and Credibility	Hali J. Edison Linda S. Kole
467	Fiscal Policy Coordination and Flexibility Under European Monetary Union: Implications for Macroeconomic Stabilization	Jay H. Bryson
466	The Federal Funds Rate and the Implementation of Monetary Policy: Estimating the Federal Reserve's Reaction Function	Allan D. Brunner
465	Understanding the Empirical Literature on Purchasing Power Parity: The Post-Bretton Woods Era	Hali J. Edison Joseph E. Gagnon William R. Melick
464	Inflation, Inflation Risk, and Stock Returns	John Ammer
463	Are Apparent Productive Spillovers a Figment of Specification Error?	Susanto Basu John S. Fernald
462	When do long-run identifying restrictions give reliable results?	Jon Faust Eric M. Leeper
	1993	
461	Fluctuating Confidence and Stock-Market Returns	Alexander David
460	Dollarization in Argentina	Steven B. Kamin Neil R. Ericsson
459	Union Behavior, Industry Rents, and Optimal Policies	Phillip Swagel
458	A Comparison of Some Basic Monetary Policy Regimes: Implications of Different Degrees of Instrument Adjustment and Wage Persistence	Dale W. Henderson Warwick J McKibbin
457	Cointegration, Seasonality, Encompassing, and the Demand for Money in the United Kingdom	Neil R. Ericsson David F. Hendry Hong-Anh Tran