Applied Economics, 2002, **34**, 1687–1702

Scale effects, time-varying markups, and the cyclical behaviour of primal and dual productivity

SUMRU ALTUG and ALPAY FILIZTEKIN*‡

‡*University of York and CEPR, Heslington, York YO10 5DD, UK* and *Sabanci University, Orhanli, 81474 Tuzla, Turkey*
E-mail: sgal@york.ac.uk and *alpayf@sabanciuniv.edu*

This paper presents estimates of the degree of returns to scale using nonparametric measures of primal and dual productivity for 2-digit US manufacturing industries. As part of the analysis, the cyclical behaviour of primal and dual productivity measures are considered, time-varying markups are allowed for, and the small sample properties of the instrumental variables estimator used to derive the estimates from the primal and dual relations examined. Both the primal and dual estimates indicate the existence of increasing returns to scale for the durable goods industries. The simulation results indicate there is a slight tendency for the dual equation estimates to overestimate the degree of returns to scale. However, small sample bias appears to be most severe for the non-durable goods industries.

I. INTRODUCTION

This study uses nonparametric measures of primal and dual productivity to estimate the degree of returns to scale and the extent of time-variation in markups for 2-digit US manufacturing industries. As part of this analysis, reconciliation of the cyclical properties of the primal and dual productivity residuals is sought. The estimates of the degree of returns to scale reported in the literature are production function estimates that incorporate the implications of firms' primal cost minimization problems. As Shapiro (1987) notes, however, 'prices should provide an independent indication of the source of productivity fluctuations' (p. 119). Under the assumption that the markup of price over marginal cost is a constant, separate estimates of the returns to scale are derived from the primal equation relating output growth to share-weighted input growth, and from the dual equation relating the change in the product price to the share-weighted change in factor prices and output growth. Using the implications of the firm's primal and dual cost minimization problems, tests of the hypothesis that the primal versus dual equation-based returns to scale estimates are equal to each other are also presented. The results are based on the Jorgenson sectoral production data set, which includes information on gross output and primary and intermediate inputs for US manufacturing industries.

In earlier work, Hall (1988, 1990) finds large and significant markups and significant deviations from constant returns to scale using instrumental variables estimation with value-added data. Likewise, Caballero and Lyons (1992) provide evidence for the existence of external effects in industry-wide production functions. However, Basu and Fernald (1995, 1997) have argued that these effects are due to specification error arising from the use of value-added data under nonconstant returns to scale and imperfect competition. Basu and Fernald (1995, 1997a) and Burnside (1996) derive production function estimates of the returns to scale using gross output data for the period 1959–1989. In contrast to Hall (1990), their results imply that the average industry displays constant or even decreasing returns to

* Corresponding author.

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684011011641 4

scale. For example, Burnside (1996) finds that the weighted average of the industry returns to scale estimates is 0.9.

The primal and dual equations are estimated separately for all 21 industries in the study using instrumental variables estimation. Unconstrained returns to scale estimates are presented for all 21 industries as well as estimates that constrain the degree of the returns to scale to be equal across industries. The estimates of the degree of the returns to scale imply that there is considerable heterogeneity across industries. It is found that the non-durable goods industries are characterized by decreasing or constant returns to scale. By contrast, the strongly cyclical durable goods industries display constant or increasing returns to scale. Differences between the primal and dual equation estimates are also found. The mean and weighted mean of the primal equation estimates are all significantly less than 1. By contrast, the mean and weighted mean of the dual equation estimates are greater than 1, implying that the average industry displays constant or increasing returns. Using the equivalence of the primal versus dual versions of the firm's cost minimization problem, a test is derived of the equality of the primal and dual equation returns to scale estimates. It is found that the equality of the primal and dual estimates is rejected for more than half of the industries in the study.

To further investigate the reasons for the differences in the primal versus dual equation estimates, the cyclical behaviour of the primal versus dual Solow residuals is examined. Under the assumption that price-cost margins do not vary over the business cycle, the primal and dual productivity residuals should be equal to each other, irrespective of the degree of returns to scale. A similar equality is used by Shapiro (1987) and Roeger (1995) to examine the sources of cyclical fluctuations.[1] It is found that the primal productivity residual displays marked procyclical behaviour relative to the dual residual, especially for the durable goods industries.

Next, the analysis is extended to allow for the existence of time-varying markups as a way of reconciling the primal and dual equation returns to scale estimates. There is an extensive literature that studies the determinants of markups at the company and industry level. In this study, a simple specification is employed that relates changes in industry-specific markups to changes in aggregate real value-added. It is found that markups for the durable goods are, in general, procyclical, whereas markups tend to move in the opposite direction to aggregate demand conditions for the non-durable goods industries. However, there is little or no change in the number of rejections with respect to the test of the equality of the primal and dual returns to scale estimates.

It is well known that instrumental variables estimators can exhibit poor finite sample performance when the instruments are weakly correlated with the regressors. In Section IV, the small sample properties of the instrumental variables estimator used in this study are examined. For this purpose, diagnostic measures of instrument relevance are provided as well as a Monte Carlo simulation of the model. The simulation results indicate that there is a slight tendency for the dual equation estimates to overestimate to the degree of the returns to scale. However, the dual equation estimates also tend to have larger standard errrors, implying that the incidence of Type I errors is approximately the same for both equations. The simulation results also indicate that small sample bias is most severe for a subset of the non-durables goods industries.

The remainder of this paper is organized as follows. Section II derives expressions for the output-based primal equation and price-based dual equation that are used in estimation. Section III presents single-equation estimates of the degree of the returns to scale and tests of the equality of the primal and dual estimates. It also examines the cyclical behaviour of primal and dual productivity, and allows for time-varying markups. Section IV examines the finite sample properties of the instrumental variables estimator used to estimate the primal and dual equations. Some concluding remarks are in Section V.

II. THE PRIMAL AND DUAL COST MINIMIZATION PROBLEMS

The procyclical behaviour of measured productivity is one of the key issues in the recent macroeconomics literature. Among the various explanations that have been offered, the hypothesis that procyclical movements in productivity reflect endogenous changes in efficiency because the economy operates with increasing returns to scale has far-reaching implications.[2] The issue of the indeterminacy of equilibrium which arises in models of multiple equilibria depends critically on the degree of returns to scale in the

[1] Shapiro argues that the quasi-fixity of capital can be used to reconcile the cyclical behaviour of the primal and dual residuals whereas Roeger (1995) allows for time-invariant markups of price over marginal cost. However, both of these analyses assume constant returns and base their findings on value-added data.

[2] Other well-known explanations for procyclical productivity include exogenous changes in efficiency as stressed by Prescott (1986), unmeasured changes in factor utilization across the business cycle due to labour hoarding or variable input utilization rates as stressed by Abbot *et al.* (1988), Burnside *et al.* (1993) and Basu (1996), and external effects as in Caballero and Lyons (1992).

aggregate economy and the magnitude of the markup parameter.[3]

What follows makes use of the implications of firms' primal and dual cost minimization problems under imperfect competition and nonconstant returns to scale to estimate the degree of the returns to scale and extent of time variation in markups. As part of this exercise, the cyclical behaviour of the primal and dual productivity residuals is also examined.

Implications of the primal problem

First, the primal cost minimization problem is described. For this purpose, a production function for gross output in the ith sector Y_{it} as a function of labour, capital, material and a random technology shock are considered as:

$$Y_{it} = F^i(L_{it}, K_{it}, M_{it}, Z_{it}) \tag{1}$$

where L_{it} denotes man-hours, K_{it} denotes services from capital, M_{it} denotes materials, and Z_{it} is a technology shock. The function F^i is assumed to be homogeneous of degree γ_i in L, K, and M, and homogeneous of degree one in Z. Let P_{it} denote the price of output in the ith sector, P^L_{it} the wage rate, P^K_{it} the rental price of capital, and P^M_{it} the price of materials. To allow for imperfect competition in the product market, the output price is assumed to include a (possibly) time-varying markup over marginal cost as

$$\frac{P_{it}}{MC_{it}} = \mu_{it} \tag{2}$$

where $\mu_{it} \geq 1$. Also define the cost and revenue shares of the inputs by

$$c^J_{it} = \frac{P^J_{it} J_{it}}{P^L_{it} L_{it} + P^K_{it} K_{it} + P^M_{it} M_{it}} \qquad J = L, K, M \tag{3}$$

and

$$s^J_{it} = \frac{P^J_{it} J_{it}}{P_{it} Y_{it}} \qquad J = L, K, M \tag{4}$$

The primal equation is derived by totally differentiating the production function and making use of the first-order conditions for cost minimization given by $P_{Jt} = \lambda_{it} F_J(L_{it}, K_{it}, M_{it}, Z_{it})$, $J = K, L, M$, where λ_{it} is a Lagrange multiplier that has the interpretation of marginal cost and F_J is the derivative of the production function with respect to the Jth input. Using the expression for the markup, it follows that

$$\frac{F_J J_{it}}{Y_{it}} = \mu_{it}\left(\frac{P_{Jt} J_{it}}{P_{it} Y_{it}}\right) = \mu_{it} s^J_{it} \quad J = K, L, M \tag{5}$$

Using the fact that $\gamma_i = \mu_{it} \sum_J s^J_{it}$ together with the definition of the cost shares c^J_{it} yields an expression for the primal equation as

$$\Delta y_{it} = \gamma_i \left[c^L_{it} \Delta l_{it} + c^K_{it} \Delta k_{it} + c^M_{it} \Delta m_{it}\right] + \Delta z_{it} \tag{6}$$

where Δx denotes log-differences of X.The estimates of the degree of the returns to scale reported in the literature have been derived from the output-based primal equation in Equation 6. See, for example, Hall (1988, 1990), Basu and Fernald (1995, 1997a and 1997b), and Burnside (1996).

Implications of the dual problem

The implications of the dual version of the firms' cost-minimization problem can be derived by considering a general cost function that depends on input prices, the level of output, and the technology shock as

$$C_{it} = C(P^L_{it}, P^K_{it}, P^M_{it}, Y_{it}, Z_{it}) \tag{7}$$

From the firm's cost minimization problem, the degree of returns to scale is equal to the ratio of average cost (AC_{it}) and marginal cost (MC_{it}). Thus, we can write

$$MC_{it} = \frac{AC_{it}}{\gamma_i} \tag{8}$$

Totally differentiating this expression, substituting for $MC_{it} = AC_{it}/\gamma_i$, $AC_{it} Y_{it} = C_{it}$, and making use of Shepard's Lemma to replace the derivative of the cost function with respect to each input with the conditional factor demands for that input yields

$$\begin{aligned}\Delta mc_{it} &= \frac{P^L_{it} L_{it}}{C_{it}} \Delta p^L_{it} + \frac{P^K_{it} K_{it}}{C_{it}} \Delta p^K_{it} + \frac{P^M_{it} M_{it}}{C_{it}} \Delta p^M_{it} \\ &\quad + \frac{MC_{it}}{AC_{it}} \Delta y_{it} + \left(\frac{C_Z Z_{it}}{C_{it}}\right) \Delta z_{it} - \Delta y_{it} \\ &= c^L_{it} \Delta p^L_{it} + c^K_{it} \Delta p^K_{it} + c^M_{it} \Delta p^M_{it} \\ &\quad + \left(\frac{1}{\gamma_i} - 1\right) \Delta y_{it} - \frac{1}{\gamma_i} \Delta z_{it}\end{aligned} \tag{9}$$

The second line follows from the definitions of the cost shares and the fact that $C_Z Z_{it}/C_{it} = -1/\gamma_i$.[4] This expression shows that exogenous technological improvement

[3] For example, Farmer and Guo (1994) require a value of the markup equal to 1.75 for the presence of multiple equilibria. Schmitt-Grohe (1994) shows minimum requirements on underlying parameters for various models to generate multiple equilibria. By contrast, multi-sector models such as those studied by Benhabib and Farmer (1994) and Perli (1998) require only a small degree of increasing returns to scale to display multiple equilibria. As another example, Rotemberg and Woodford (1995) argue that a markup parameter of 1.2 suffices to induce real wage increases in response to increases in government demand.

[4] The second result is obtained by noting that $C_Z \frac{Z_{it}}{C_{it}} = -\lambda_{it} F_Z (Z_{it}/Y_{it})(Y_{it}/C_{it}) = -(\mathrm{MC}_{it}/AC_{it}) = -1/\gamma_i$, where $C_Z = -\lambda_{it} F_Z$ by the envelope theorem and $F_Z Z_{it}/Y_{it} = 1$ by assumption.

under increasing returns to scale has a direct cost-reducing effect as captured by the term $-(1/\gamma_i)\Delta z_{it}$, and an indirect effect due to scale effects as captured by the term $(1/\gamma_i - 1)\Delta y_{it}$. The dual equation is derived by noting that the growth rate of marginal cost equals the difference between the growth in the product price and the growth in time-varying markups, or

$$\Delta p_{it} = c_{it}^L \Delta p_{it}^L + c_{it}^K \Delta p_{it}^K + c_{it}^M \Delta p_{it}^M + \left(\frac{1-\gamma_i}{\gamma_i}\right)\Delta y_{it} - \frac{1}{\gamma_i}\Delta z_{it} + \Delta \mu_{it} \tag{10}$$

This expression shows that variations in the product price can arise from a number of sources. Specifically, increases in the price of output reflect increases in the share-weighted price of inputs or increases in price-cost margins. In the absence of scale effects, exogenous technological improvement leads to declines in the product price solely due to exogenous increases in efficiency. When there are increasing returns to scale, increases in output for a given level of inputs can occur due to endogenous increases in efficiency. Such endogenous increases in efficiency can also lead to a reduction in costs for a given level of factor prices. These facts show that it is possible to infer the magnitude of the returns to scale from the output-based primal and the price-based dual equations, respectively.

In the above discussion, the determinants of time-varying markups have not been specified. As Hall (1990) has argued, the extent of market power possessed by firms may be mitigated by scale effects or other factors. Using the definition of the markup, one has:

$$\mu_{it} = \frac{p_{it}}{MC_{it}} = \frac{p_{it}}{AC_{it}}\frac{AC_{it}}{MC_{it}} = \gamma_i(1+\pi_{it}) \tag{11}$$

where π_{it} refers to the profit rate of firm i.[5] This expression shows that large markups of price over marginal cost are consistent with the small average profit rates in US industry only if there are significant scale economies. Put differently, the observed small average profit rates in the USA are inconsistent with the existence of market power under constant returns to scale. In line with this observation, Chirinko and Fazzari (1994) find that Lerner-type indices of market power (based on the percentage difference between price and marginal cost) and estimates of the degree of the returns to scale tend to be positively correlated, implying that the extent of market power tends to be associated with significant increasing returns in the production technology. In contrast to some other analyses in the literature, it is worth noting that the estimation equations allow for departures from both constant returns to scale and imperfect competition.

Morrison (1992, 1994) provides a decomposition of the markup measure to allow for scale effects, variable capacity utilization, and the quasi-fixity of capital and labour. This decomposition can be used to extend the above arguments to examine the cyclical behaviour of markups. Specifically, Morrison shows that the markup μ_{it} can be written as

$$\mu_{it} = \left(\frac{p_{it}}{AC_{it}}\right)\left(\frac{1}{\varepsilon_{CY}}\right) = \left(\frac{p_{it}}{AC_{it}}\right)\left(\frac{1}{\varepsilon_{CY}^L CU_c}\right) \tag{12}$$

where $\varepsilon_{CY}^L = (1/RTS)$ is the long-run elasticity of costs with respect to output and CU_c is a cost-side capacity utilization measure that is obtained by allowing for the quasi-fixity of the inputs. Thus, to be consistent with zero or close to zero average profits, an increase in markups must be accompanied by an increase in excess capacity. If capacity utilization is taken as a measure of cycles, then markups must exhibit countercyclical behaviour. In what follows, estimates are presented of the degree of the returns to scale under the assumptions that markups are constant over time, and also that they vary over time in response to aggregate demand conditions.

III. RETURNS TO SCALE ESTIMATION

In what follows, single-equation estimates are presented of the returns to scale based on the separate estimation of the primal equation described by Equation 6 and the dual equation described by Equation 10, respectively. Tests are also presented of the equality of the estimates derived from the primal and dual equations, respectively. First, the data are described.

Data

This study makes use of the Jorgenson data set on industry-level on gross output and the inputs of labour, capital, energy use, and materials for 21 manufacturing industries. The data are annual for the period 1959-1989.[6] The reasons for using this data set are three-fold. First, one does not encounter the problem of specification error arising from the use of value-added data because the Jorgenson data set contains information on gross output and materials inputs. Second, use of the Jorgenson data allows one to compare results regarding the degree of the returns to scale with the results obtained by Basu and Fernald (1995, 1997a, b) and Burnside (1996), who also use this data set. Finally, the

[5] This result follows from the fact that the ratio of average cost to marginal cost is equal to the degree of the returns to scale, and by noting that $1 + \pi_{it} = \mathrm{TR}_{it}/\mathrm{TC}_{it} = (p_{it}Y_{it})/(\mathrm{AC}_{it}Y_{it})$.
[6] The data are described in detail in Jorgenson *et al.* (1987) and Jorgenson (1990).

Jorgenson data contain several adjustments that are designed to reduce measurement error in the inputs.[7]

The adjustment for taxes is made by defining the quantity of output in sector i, denoted q_i, as $q_i = (vk_i + vl_i + ve_i + vm_i)/po_i = (vk_i + vl_i + ve_i + vm_i + vt_i)/pi_i$, where vk_i, vl_i, ve_i, vm_i denote the value of capital services, labour inputs, energy inputs and material inputs, respectively, po_i denote the price of output that producers receive, pi_i denotes the price of output that consumers pay, and vt_i is the value of taxes paid by each sector. Under imperfect competition, the payments to capital cannot be defined merely as a residual after all other factors are paid. To define the payments to capital, a series on the user cost of capital r is constructed following Hall and Jorgenson (1967), Hall (1990), and Caballero and Lyons (1992). Thus, the required payment for any type of asset j in industry i, $P_i^j K_i^j$, is then $r_j \pi_i^j K_i^j$, where $\pi_i^j K_i^j$ is the current-dollar value of the stock of type j capital for industry i. For each sector i, data on the current value of 50 types of assets as distinguished by the Bureau of Economic Analysis in constructing the national product accounts, plus land and inventories, are used. Specifically, for capital of type j, the user cost of capital is $r_j = (\rho + \delta_j)(1 - c_j - \tau d_j)/(1 - \tau)$, where ρ is the required rate of return on capital, δ_j is the asset-specific depreciation rate, c_j is the asset-specific investment tax credit, τ is the statutory corporate tax rate, and d_j is asset-specific present value of depreciation allowances.

Estimation results

It is assumed that sectoral markups are constant, and that all factors are variable. An industry-specific constant and a dummy variable that allows a trend break after 1973 are included in each equation. Typically sectoral input use will rise in response to favourable technology shocks. Consequently, the share-weighted change in inputs will be correlated with changes in technology. Likewise, changes in share-weighted input prices will tend to be correlated with sectoral productivity changes. To account for the potential endogeneity problem induced by such correlations, instrumental variables estimation procedures are employed, and intruments are used that are likely to be uncorrelated with sectoral productivity growth for the estimation of both the primal and dual equations. These include the growth rate of real military purchases, the growth rate of the world price of oil, and a dummy variable representing the political party of the president data plus one lagged value of each of these variables. A constant and trend are also included in the instrument set.[8]

Single-equation estimates of the primal and dual equations. Table 1 reports system estimates that are obtained by estimating the primal and dual equations for all 21 manufacturing industries to take into account the correlation in the disturbances across industries. In this table, γ_i^P refers to the primal equation estimate of the degree of the returns to scale for industry i while γ_i^D refers to the dual equation estimate for industry i. We provide system instrumental variables estimates of the primal and dual equations using the instrument set described earlier. We assume that the disturbances are conditionally homoscedastic. It is well known that instrumental variables estimators can exhibit poor finite sample performance when the instruments are weakly correlated with the regressors. These include Nelson and Startz (1990), Staiger and Stock (1994), and Hall *et al.* (1996), amongst others. Section IV, presents a simulation analysis of the instrumental variables estimator used in this study. In this section, following a suggestion by Ligeralde and Brown (1995), test statistics are calculated for the null hypothesis of constant returns to scale by using both unrestricted and restricted residuals to evaluate coefficient standard errors.[9]

The primal equation estimates reported in Table 1 show that the hypothesis of constant returns to scale can be rejected in favour of decreasing returns for around half of the non-durable goods industries and two durable goods industries. By contrast, there are only two industries for which the hypothesis of constant returns can be rejected in favour of increasing returns. While the marginal significance levels obtained by using standard errors based on restricted residuals tend to be larger than those based on unrestricted residuals, these results hold regardless of whether tests of deviations from constant returns to scale are calculated using standard errors based on the unrestricted or restricted residuals.[10] The median, mean and weighted mean of both sets of unrestricted estimates are

[7] In this respect, the series on the labour input is constructed using information from both the household and establishment surveys. (See Jorgenson *et al*. 1987, Ch. 3.) Thus, it accounts for the criticism raised by Prescott (1986) and Evans (1992) that the hours data obtained from household surveys typically differ from hours data based on establishment surveys. The Jorgenson data are also constructed by weighing the hours worked by different types of workers (distinguished by various demographic and occupational characteristics) by their relative wage rates. Thus, the labour input rises either because the number of hours worked rises, or because the 'quality' of this work increases. Similar adjustments are made to the capital input.

[8] The validity and relevance of this instrumentation are discussed at later points in the paper.

[9] These authors examine models with serial correlation and conditional heteroscedasticity, and show that the problem of excessive rejections of the null hypothesis can be reduced when coefficient standard errors are calculated using restricted residuals.

[10] The standard errors and test statistics using the restricted residuals are calculated by restricting the degree of the returns to scale to equal 1 for which the test is being conducted, and equal to its estimated value for the remaining industries.

Table 1. *Single-equation primal and dual estimates*

Industry	γ_i^P	p-value for $H_0:\gamma_i^P=1$*	p-value for $H_0:\gamma_i^P=1$**	γ_i^D	p-value for $H_0:\gamma_i^D=1$*	p-value for $H_0:\gamma_i^D=1$**
Non-durables						
Food	0.599	0.020	0.097	1.249	0.430	0.693
Tobacco	0.684	0.000	0.013	3.051	0.296	0.113
Textiles	0.843	0.073	0.068	0.795	0.011	0.001
Apparel	0.960	0.674	0.665	1.054	0.575	0.455
Paper	0.529	0.000	0.004	0.987	0.857	0.834
Printing	0.886	0.219	0.267	1.178	0.097	0.077
Chemicals	0.374	0.000	0.000	1.090	0.621	0.621
Petroleum products	−0.058	0.000	0.000	0.880	0.725	0.540
Rubber	0.884	0.184	0.224	1.085	0.167	0.093
Leather	1.456	0.071	0.085	1.059	0.604	0.337
Durables						
Lumber and wood	0.874	0.150	0.166	0.819	0.075	0.000
Furniture	1.001	0.983	0.982	1.057	0.099	0.167
Stone, clay and glass	0.897	0.057	0.065	1.023	0.441	0.259
Primary metal	0.968	0.589	0.599	1.147	0.021	0.061
Fabricated metal	1.473	0.000	0.000	1.184	0.001	0.001
Non-electrical machinery	1.096	0.120	0.141	1.284	0.000	0.000
Electrical machinery	1.076	0.250	0.225	0.976	0.443	0.222
Motor vehicles	1.064	0.177	0.231	1.064	0.033	0.000
Transportation equipment	1.116	0.050	0.037	1.106	0.027	0.000
Instruments	0.722	0.000	0.046	0.844	0.006	0.007
Miscellaneous manufacturing	0.133	0.000	0.000	1.357	0.021	0.000
Summary statistics						
γ^{med}	0.886	–	–	1.064	–	–
$\bar{\gamma}$	0.837	–	–	1.157	–	–
	(0.016)	–	–	(0.103)	–	–
$\bar{\gamma}^w$	0.888	–	–	1.113	–	–
	(0.019)	–	–	(0.039)	–	–
σ_γ	0.366	–	–	0.447	–	–
σ_γ^w	0.319	–	–	0.221	–	–
Restricted estimate	0.939	0.000	–	0.999	0.949	
P-value for						
restriction test	–	0.043	–	–	0.836	–
Unrestricted model	–	0.274	–	–	0.116	–

Notes:
(a) Sample period: 1959–1989.
(b) γ_i^P and γ_i^D refer to the primal and dual equation estimates of the degree of returns to scale for industry i.
(c) The definition of the variables and instruments is provided in Section II (Implications of the dual problem) of the text.
(d) γ^{med}, $\bar{\gamma}$, $\bar{\gamma}^w$ denote the median, mean, and weighted mean of the estimates, where $\bar{\gamma}=\sum_{i=1}^N \gamma_i/N$, $\bar{\gamma}^w=\sum_{i=1}^N s_i\gamma_i/N$, and s_i is the average share of industry i in manufacturing value-added.
(f) σ_γ and σ_w are measures of dispersion, with $\sigma_\gamma^2=\sum_{i=1}^N(\gamma_i-\bar{\gamma})^2/N$ and $\sigma_w^2=\sum_{i=1}^N s_i(\gamma_i-\bar{\gamma}^w)^2/N$.
* Standard errors calculated using unrestricted residuals.
** Standard errors calculated using residuals restricted under $H_0:\gamma_i=1$.

all less than 1, implying that the average industry displays decreasing returns to scale. The restricted estimate of γ_i reported in Table 1 refers to the estimate that is obtained by constraining the returns to scale to be equal across all 21 manufacturing industries for the primal and dual equations separately. The restricted estimates for the primal equation show that the hypothesis of constant returns to scale can be rejected in favour of decreasing returns to scale at the 5% level. The over-identifying restrictions associated with constraining the degree of the returns to scale to be equal across industries are rejected at the 5% level. The value of the relevant test statistic, which is distributed as χ^2 (20), is 49.15, and the associated marginal significance levels or p-value is 0.000. These results are similar to the findings reported by Basu and Fernald (1995, 1997a, b) and Burnside (1996).

For the dual equation system estimates, the evidence against constant returns to scale for non-durable goods industries is much less pronounced compared to the primal equation estimates: there is only one non-durable goods industry for which constant returns can be rejected against the alternative of decreasing returns. By contrast, there is more evidence against constant returns for the durable goods industries, with six out of the 11 durable goods industries displaying evidence for significant increasing returns. That the durable goods industries display significant scale effects is consistent with the capital-intensive nature of production in these industries. In this case, the mean median and weighted mean of the unrestricted estimates are all greater than one while the restricted estimate is nearly equal to one. Unlike the primal equation, however, one cannot reject the hypothesis of constant returns for the restricted estimate. Nor can one reject the over-identifying restrictions associated with setting the returns to scale parameter to be equal across industries. The value of the relevant test statistic is equal to 11.129, implying a marginal significance level of 0.943.

Table 1 also reports the marginal significance levels for the *J*-test of the over-identifying restrictions of the model. Since the *J*-test is a test of the orthogonality of the residuals and the instruments, it can be interpreted as a test of instrument validity.[11] For the system estimates, the *J*-statistic is distributed as a chi-square random variable with degrees of freedom equal to the number of orthogonality conditions minus the number of parameters. The marginal significance levels reported in Table 1 imply that the over-identifying restrictions associated with neither the primal nor the dual version of the firm's cost minimization problem can be rejected based on the unrestricted estimates. Thus, it can be concluded that the demand-side instruments used in this paper are uncorrelated with changes in sectoral productivity, and hence, are valid instruments.[12]

Summarizing, the results in Table 1 suggest that the inference that can be drawn about the magnitude of the degree of the returns to scale differs when one considers the primal versus dual equation. First, the incidence of decreasing returns to scale is greater based on the estimates of the primal equation than the estimates of the dual equation, especially for the non-durable goods industries. By contrast, the dual equation estimates imply that returns to scale are constant or increasing. While the restricted estimate based on the dual equation is consistent with constant returns to scale, the restricted estimate from the primal equation is not. If the findings based on the primal equation are accepted as the basis for the estimates of returns to scale in manufacturing, then there is the problem of decreasing returns to scale for a number of non-durable goods industries, which, if taken literally, would imply that firms are operating, on average, above efficient scale.

Basu and Fernald (1997b) have argued that aggregation or re-allocation effects across firms or industries may be used to justify findings of decreasing returns at higher levels of aggregation. If there are differences in scale economies across firms in a given industry (or across industries within a more broadly defined sector), then such re-allocation effects can lead to omitted variables bias in production function-type regressions. Since firms (or industries) that have higher than average returns to scale also have higher than average input growth, the omitted variable will be positively correlated with observed input use, leading to an underestimate of the degree of the returns to scale. However, this analysis is more useful for reconciling returns to scale estimates at the level of the aggregate economy relative to the industry level. While such aggregation or re-allocation effects can be defined at the industry level, their impact cannot be assessed empirically in the absence of firm-level data on outputs and inputs. Furthermore, is seems difficult to rationalize the findings of significant decreasing returns for such industries such as Tobacco or Petroleum Products, which are comprised of a small number of firms that have access to similar technologies.[13]

It is worth noting that the dual equation estimates are largely consistent with the recent evidence presented by Paul and Siegel (1999a,b). These authors employ a generalized dynamic cost function approach to estimate scale effects. Their results are based on data for the period 1959-1989 for total manufacturing, two-digit SIC level, and four-digit SIC level industries. They find evidence in favour of internal scale effects at all levels of aggregation and for the majority of the industries considered in their study, both in the short- and long-run. Furthermore, such scale economies exist even after accounting for the effects of labour and capital fixity.

The single-equation estimates suggest that there are differences in nature of the estimates of scale effects based on the primal versus dual relations. However, separate estimation of these primal relations does not allow one to determine whether the primal and dual equation estimates are significantly different from each other. In the next section, tests are presented of the equality of the primal and dual equation returns to scale estimates.

Testing for the equality of the primal and dual equation estimates. A simple approach is now described for testing the equality of the returns to scale estimates derived from the primal versus dual versions of the firm's cost minimi-

[11] More accurately, it is an omnibus test of model specification.
[12] This is similar to the results in Burnside (1995) regarding the instrument set used in this paper.
[13] On the other hand, heterogeneity of products and the associated production technology may be one reason for the findings of decreasing returns for an industrial classification such as Miscellaneous Manufacturing.

zation problems. For this purpose, let γ_i^P denote the estimate derived from the primal equation, and let γ_i^D denote the estimate derived from the dual equation. It is initially assumed that markups are constant. Under the assumption that the model is correctly specified, one can solve Δz_{it} from the primal equation defined by Equation 6 and the dual equation defined by Equation 10 separately, and equate these expressions to obtain

$$\Delta y_{it} = \frac{\gamma_i^P}{\gamma_i^D}[c_{it}^L \Delta l_{it} + c_{it}^K \Delta k_{it} + c_{it}^M \Delta m_{it}] + [c_{it}^L \Delta p_{it}^L + c_{it}^K \Delta p_{it}^K + c_{it}^M \Delta p_{it}^M - \Delta p_{it}] + \frac{1}{\gamma_i^D}\varepsilon_{it} \tag{13}$$

Notice that the relation in Equation 13 cannot be used to identify γ_i^P and γ_i^D separately. Nevertheless, it allows for a simple test of model specification based on whether the ratio γ_i^P/γ_i^D is significantly different from 1. Since the relation in Equation 13 holds as an identity, one can estimate the following relation by OLS:

$$\Delta y_{it} = \alpha_i^0 + \alpha_i^1\left[c_{it}^L \Delta l_{it} + c_{it}^K \Delta k_{it} + c_{it}^M \Delta m_{it}\right] + \alpha_i^2\left[c_{it}^L \Delta p_{it}^L + c_{it}^K \Delta p_{it}^K + c_{it}^M \Delta p_{it}^M - \Delta p_{it}\right] + \bar{\varepsilon}_{it} \tag{14}$$

and test whether $\alpha_i^1 = 1$ and $\alpha_i^2 = 1$ individually and jointly.

Table 2 reports the results of this estimation for all 21 industries jointly. Specifically, it shows that constraining the degree of the returns to scale to be equal across the primal and dual equations is not supported by the data. One can reject the equality of the primal and dual equation estimates for 11 of the industries considered in the study, with slightly more rejections for the durable goods industries. Thus, the finding of differences in the value of the returns to scale from the output-based primal versus the price-based dual equation turns out to be statistically significant for over half of the industries in the sample. As a further specification test, whether the coefficient $\alpha_i^2 = 1$ is significantly different from unity is considered separately, and jointly with $\alpha_i^1 = 1$. Table 2 shows that this hypothesis is rejected for 14 industries in the former case, and for 16 industries in the latter.[14]

Cyclical behaviour of primal and dual productivity

An equivalent way to study the implications of the firm's primal versus dual cost minimization problems is to examine the cyclical behaviour of the primal and dual productivity residuals. Solow (1958) introduced the concept of the primal productivity residual as a measure of exogenous technical change under constant returns to scale and perfect competition. Following Hall (1988, 1990), it is straightforward to show that primal productivity residual under imperfect competition and nonconstant returns to scale is given by

$$\mathrm{SC}_{it} = \Delta y_{it} - c_{it}^L \Delta l_{it} - c_{it}^K \Delta k_{it} - c_{it}^M \Delta m_{it} = \left(1 - \frac{1}{\gamma_i}\right)\Delta y_{it} + \frac{1}{\gamma_i}\Delta z_{it} \tag{15}$$

Likewise, Ohta (1975), Hulten (1986) and others have shown that the change in total factor productivity can be calculated as a cost-side measure using data on factor and output prices under the same assumptions that Solow made. Under our assumptions, the dual productivity residual is given by

$$\mathrm{SPC}_{it} = c_{it}^L \Delta p_{it}^L + c_{it}^K \Delta p_{it}^K + c_{it}^M \Delta p_{it}^M - \Delta p_{it} = \left(1 - \frac{1}{\gamma_i}\right)\Delta y_{it} + \frac{1}{\gamma_i}\Delta z_{it} - \Delta\mu_{it} \tag{16}$$

The expressions in Equations 15 and 16 show that if markups are constant and there are no other unobserved variables, then the primal and dual (cost-based) productivity residuals should be identically equal to each other, irrespective of the degree of the returns to scale. As Hall (1988, 1990) and others have argued, if there are increasing returns to scale, then endogenous changes in efficiency will induce procyclical behaviour in the primal residual even in the absence of procyclical technology shocks. In this case, the dual residual will also display procyclical behaviour, provided the effect of endogenous changes in efficiency is not offset by procyclical movements in markups.

Table 3 reports the correlations of the primal and dual productivity residuals SC_{it} and SPC_{it} with changes in aggregate real value-added Δv_t.[15] These correlations show that the primal productivity residual SC_{it} is procycli-

[14] We also estimated the relation in Equation 13 using instrumental variables estimation to account for potential measurement error in the variables. Although there was a slight reduction in the number of rejections, the results were similar to those reported in the text.

[15] A measure of aggregate real value-added was constructed as a Divisia index of sectoral real value-added, with the weights defined as the share of nominal value-added in each sector to total nominal value-added over the 34 private industries in the Jorgenson data set as

$$\Delta v_t = \sum_{i=1}^{n} w_{it}\Delta v_{it}$$

Here Δv_t denotes the percentage change in aggregate real value-added, Δv_{it} is the percentage change in sectoral real value-added, and the weights w_{it} are defined as $w_{it} = P_{it}^V V_{it}/\sum_{i=1}^{n} P_{it}^V V_{it}$.

Table 2. *Testing for the equality of the primal and dual equation estimates*

Industry	α_i^1	α_i^2	p-value for $H_0 : \alpha_i^1 = 1$	p-value for $H_0 : \alpha_i^2 = 1$
Non-durables				
Food	0.807	−0.814	0.000	0.000
Tobacco	0.629	−0.497	0.000	0.000
Textiles	1.128	−0.847	0.000	0.016
Apparel	1.013	−0.916	0.479	0.148
Paper	1.095	−0.985	0.014	0.808
Printing	1.019	−0.777	0.609	0.005
Chemicals	0.919	−0.945	0.125	0.431
Petroleum products	0.970	−0.893	0.249	0.001
Rubber	1.017	−1.088	0.285	0.051
Leather	1.151	−1.222	0.001	0.000
Durables				
Lumber and wood	1.036	−0.618	0.322	0.000
Furniture	1.020	−0.879	0.531	0.312
Stone, clay and glass	1.084	−0.352	0.011	0.000
Primary metal	1.126	−0.866	0.000	0.003
Fabricated metal	1.102	−0.603	0.000	0.000
Non-electrical machinery	1.048	−0.919	0.018	0.123
Electrical machinery	1.105	−0.245	0.592	0.000
Motor vehicles	1.179	−0.123	0.000	0.000
Transportation equipment	0.998	−1.089	0.910	0.169
Instruments	0.969	−0.625	0.472	0.001
Miscellaneous manufacturing	0.785	−1.340	0.00	0.001

Notes: (a) Sample period: 1959–1989.

cal for the majority of the durable goods industries, which constitute the most cyclical part of manufacturing industries. By contrast, the primal productivity measure is acyclical or countercyclical for at least half of the non-durable goods industries. When taken together with the results in Table 1, these findings suggest that at least part of the procyclicality of the primal productivity residual for the durable goods industries may be arising from the existence of scale effects.[16]

A second finding that emerges from Table 3 is that the primal productivity residual SC_{it} is more procyclical relative to the dual productivity residual SPC_{it} for 17 out of the 21 industries in the study. The exceptions are three non-durable goods industries, and one durable goods industry (Misc. Manufacturing). Equations 15 and 16 show that in the presence of procyclical markups, the dual productivity will display less procyclical behaviour relative to the primal productivity residual even after controlling for the existence of procyclical technology shocks or endogenous changes in efficiency due to increasing returns to scale. Equation 10 also shows that assuming the existence of time-invariant markups of price over marginal cost induces an omitted-variable bias in the estimates of the dual equation. Thus, the existence of time-varying markups may also help to reconcile the estimates of the degree of the returns to scale obtained from the primal versus dual equations.[17] The next section examines the importance of such factors for reconciling the primal and dual returns to scale estimates.

Time-varying markups

There is an extensive literature that studies the determinants of markups at the company and industry level. Following Hall (1988, 1990), the papers by Domowitz *et al.* (1988), and Haskel *et al.* (1995) estimate industry markups using versions of the primal productivity residual under imperfect competition. The papers by Morrison (1993, 1994), Chirinko and Fazzari (1994) and Galeotti and Schiantarelli (1999) study the cyclical behaviour of industry markups using intertemporal versions of the firm's problem.[18] Unlike Morrison and Chrinko and Fazzari,

[16] The appendix also shows that allowing for the fixity of capital along the lines suggested by Shapiro (1987) is not useful for reconciling the cyclical behaviour of the primal and dual productivity residuals.

[17] Another reason for the difference may be due to cyclical variation in capacity utilization. If the time variation in markups is related to the cyclical variation in cyclical utilization, then allowing for time-varying markups may allow an indirect way of capturing the effect of changes in capacity utilization.

[18] Other approaches to estimating markups involve equating marginal and average variable costs for the firm, and using a static production-theoretic approach. Examples of the former approach include Domowitz *et al.* (1986) and Machin and Van Reenen (1993) while examples of the latter approach are given by Applebaum (1979), among others.

Table 3. *Cyclical properties of the primal and dual residuals*

Industry	Corr. of SC_{it} with Δv_{it}	Corr. of SPC_{it} with Δv_{it}
Non-durables		
Food	0.1268	0.2312
Tobacco	0.0972	0.1421
Textiles	−0.3290	−0.5744
Apparel	0.1211	0.2004
Paper	0.5521	0.2293
Printing	0.5818	0.1985
Chemicals	0.5130	0.3928
Petroleum and coal products	−0.0507	−0.0901
Rubber	0.3821	0.2476
Leather	−0.5643	−0.6508
Durables		
Lumber and wood	−0.6016	−0.7070
Furniture	0.5302	0.4268
Stone, clay and glass	0.6338	−0.2074
Primary metal	0.5657	0.3052
Fabricated metal	0.7005	0.2322
Non-electrical machinery	0.4635	0.1224
Electrical machinery	0.5559	−0.2255
Motor vehicles	0.5859	0.1037
Transportation Equipment	−0.0223	−0.0695
Instruments	0.5325	−0.0361
Miscellaneous manufacturing	0.330	0.4355

Notes:
(a) Sample period: 1959–1989.
(b) SC_{it} and SPC_{it} are defined by Equations 15 and 16.

Galeotti and Sciantarelli assume that sectoral markups vary with state of demand (which they term the level effect), and with the expected changes in future demand (the derivative effect). In what follows, time-varying markups are allowed for by introducing a simple specification that relates the growth rate in industry markups to the growth rate of aggregate real value-added. It is assumed that the percentage change in the markup for sector i is a linear function of aggregate real value-added. Thus, $\Delta\mu_{it}$ is expresses as

$$\Delta\mu_{it} = \psi_i \Delta v_t \tag{17}$$

where v_t is aggregate real value-added.

Proceeding as in Section III (Cyclical behaviour of primal and dual productivity) it can be shown that the equality of the disturbances in the primal and dual equations yields the relation

$$\begin{aligned}\Delta y_{it} = {} & \delta_i^0 + \delta_i^1\left[c_{it}^L \Delta l_{it} + c_{it}^K \Delta k_{it} + c_{it}^M \Delta m_{it}\right] \\ & + \delta_i^2\left[c_{it}^L \Delta p_{it}^L + c_{it}^K \Delta p_{it}^K + c_{it}^M \Delta p_{it}^M - \Delta p_{it}\right] \\ & + \psi_i \Delta v_t + \bar{\varepsilon}_{it}\end{aligned} \tag{18}$$

This relation allows one to test whether $\gamma_i^P = \gamma_i^D$ by testing whether $\delta_i^1 = 1$, and to derive estimates regarding the cyclicality of markups based on the estimates of ψ_i.

The results of estimating the relation in Equation 18 by OLS are reported in Table 4. Compared to the results in Table 2, there are now 16 industries for which one can reject the hypothesis that $\gamma_i^P = \gamma_i^D$. Thus, the inclusion of time-varying markups does not help to reconcile the primal and dual equations estimates. The estimates of δ_i^2 imply that there is additional evidence against the specification in terms of the rejection of $\delta_i^2 = 1$ separately, and jointly with $\delta_i^1 = 1$. In the former case, there are 16 industries for which the hypothesis of $\delta_i^2 = 1$ is rejected separately, and 18 industries for which it is rejected jointly with $\delta_i = 1$.[19]

Turning to the evidence regarding the cyclicality of markups, the estimates of ψ reported in Table 4 imply that markups tend to be more countercyclical for the non-durable goods industries. It is worth noting that the countercyclicality of markups relative to an indicator of aggregate demand for the non-durable goods industries is in line with empirical evidence presented in Galeotti and Schiantarelli (1998) as well as the predictions of many theoretical studies on the topic.[20] By contrast, markups tend to be more procyclical for durable goods industries, which also tend to display more evidence in favour of increasing returns to scale. This is consistent with the evidence in Chirinko and Fazzari (1994), who find that Lerner-type indices of market power tend to be positively correlated with the degree of the returns to scale.

IV. SMALL SAMPLE ISSUES

It is well-known that instrumental variables estimators can exhibit poor finite sample performance when the instruments are weakly correlated with the regressors. In the single regressor–single instrument case, Nelson and Startz (1990) have shown that conventional tests of significance can be badly mis-sized when the instrument has low or

[19] The relation in Equation 18 was also estimated using instrumental variables estimation. There was less evidence against the specification, in terms of both tests of the hypothesis of $\gamma_i^P = \gamma_i^D$ and of $\delta_i^2 = 1$. However, the markup coefficients were imprecisely estimated, implying a potential problem of identifying the individual effects of share-weighted input growth, share-weighted factor price growth, and aggregate real value-added growth using the aggregate instruments used in this study. This is similar to the problem of estimating external effects regressions using the same instrument set that is discussed by Burnside (1996).

[20] Some potential reasons for the countercyclicality of markups with respect to the level of demand include the difficulty in sustaining collusive behaviour and the ease of entry and increased competitiveness during periods of high demand, increases in concentration and profit margins due to increases in bankruptcy of credit-constrained firms and increases in markups due to changes in the elasticity of demand during recessions. See Galeotti and Schiantarelli (1998) for a more detailed discussion of the theoretical literature.

Table 4. *Testing the equality of the primal and dual estimates with time-varying markups*

Industry	δ_i^1	δ_i^2	p-value for $H_0: \delta_i^1 = 1$	p-value for $H_0: \delta_i^2 = 1$	ψ_i	p-value for $H_0: \psi i = 0$
Non-durables						
Food	0.791	−0.856	0.000	0.000	−0.038	0.426
Tobacco	0.624	−0.493	0.000	0.000	−0.022	0.853
Textiles	1.094	−0.813	0.025	0.006	0.025	0.809
Apparel	1.042	−0.971	0.046	0.622	−0.075	0.144
Paper	0.980	−0.860	0.602	0.016	0.360	0.000
Printing	0.907	−0.672	0.014	0.000	0.299	0.000
Chemicals	0.710	−0.675	0.000	0.000	0.709	0.000
Petroleum products	0.967	−0.895	0.188	0.000	0.069	0.255
Rubber	0.970	−1.043	0.187	0.371	0.238	0.009
Leather	1.155	−1.142	0.002	0.034	−0.083	0.554
Durables						
Lumber and wood	1.023	−0.620	0.588	0.000	0.060	0.652
Furniture	0.920	−0.762	0.045	0.029	0.357	0.002
Stone, clay and glass	0.886	0.239	0.003	0.000	0.540	0.000
Primary metal	1.114	−0.837	0.000	0.005	0.044	0.680
Fabricated metal	0.861	−0.588	0.000	0.000	0.557	0.000
Non-electrical machinery	0.918	−0.936	0.001	0.225	0.486	0.000
Electrical machinery	0.880	−0.312	0.000	0.000	0.587	0.000
Motor vehicles	1.162	−0.102	0.000	0.000	0.122	0.417
Transportation Equipment	0.993	−1.107	0.731	0.112	0.050	0.562
Intruments	0.885	−0.657	0.003	0.001	0.506	0.000
Miscellaneous Manufacturing	0.752	−1.254	0.000	0.011	0.339	0.038

Notes: (a) Sample period: 1959–1989.

poor relevance. They suggest a measure of relevance based on the R^2 of the first-stage regression. Shea (1997) has argued that in multivariate regressions a partial R^2 measure is more appropriate. Hall *et al*. (1996) examine the performance of this measure in the multiple regressor–multiple instrument case. This section first examines such diagnostic measures of instrument relevance. Then the small sample distribution of the instrumental variables estimator used in the study is derived.

The regressors in the primal and dual equations consist of share-weighted input growth, output growth, and the difference between changes in the product price and share-weighted changes in input prices. Table 5 shows that the industry average of the R^2 and $\bar{R}^2$ measures for the first-stage regressions of the endogenous variabless on the instruments. The R^2 values for share-weighted input growth, output growth and the difference in the change in the product price and share-weighted changes in input prices are given by 0.3360, 0.1339 and 0.2707 while the $\bar{R}^2$ values for these variables are given by 0.3178, 0.1102 and 0.0488. respectively.[21] It is worth noting that the R^2 measures for output growth are only slightly smaller than those for share-weighted input growth. By contrast, the diagnostic measures for the difference in the change in the product price and share-weighted changes in input prices suggest that the instruments correlate less well with this variable.

In the earlier literature, the findings of large increasing returns for the manufacturing industries reported by Hall (1990) have been attributed to the existence of small sample bias in the reverse regressions that Hall estimates. Specifically, instead of using the version of the primal equation described by Equation 6, Hall (1990) regresses share-weighted input growth on output growth, and obtains an estimate of γ_i by inverting the resulting coefficient. If the instruments that Hall uses are more weakly correlated with output growth than with share-weighted input growth, then small sample bias may lead to the findings of spuriously large increasing returns for the manufacturing industries that Hall reports in his study.[22] For the present study, differences in the R^2 measures for output growth and share-weighted input growth, on the one hand, and the growth rates of the product and input prices, on the

[21] These values are similar to those reported by Burnside (1996) for his baseline instrument set, which is identical to the instrument set used in this paper. Out of the five instruments considered by Burnside, this is also one of the instrument sets that provides the best fit in terms of the first-stage regressions.

[22] In Hall's study, there is also the problem of the use of value-added data.

Table 5a. *Diagnostic measures of instrument relevance*

	R^2	$\bar{R}^2$
Industry input growth	0.3360	0.1339
Industry output growth	0.3178	0.1102
Industry net price growth	0.2707	0.0488

Table 5b. *Simulated distribution for* $T = 31$

			True null			False null		
γ	$\bar{\gamma}$	SE (γ)	10%	5%	1%	10%	5%	1%
Primal estimate								
0.8	0.8001	0.0369	12.7	7.4	0.4	93.7	91.3	78.0
1.2	1.2002	0.0369	12.7	7.4	0.4	93.9	91.4	77.7
Dual estimate								
0.8	0.8199	0.0491	12.9	7.0	0.4	83.3	80.1	68.9
1.2	1.2152	0.0577	12.7	7.0	0.4	88.7	84.2	66.9

Notes:
(a) $\bar{\gamma}$ is the estimated returns to scale coefficient, and $SE(\gamma)$ its standard error, averaged over replications and over industries.
(b) Rejection rates for the true null show the sizes of the tests, rejection rates for the false null show the power of the tests.
(c) True null: γ equal to 0.8, or 1.2. False null: γ eqal to 1.0.

other hand, suggest that differences between the primal versus dual equation estimates may be attributed to small sample bias arising from instrument relevance. The remainder of this section derives the small sample distribution of the instrumental variables estimator used in the study to quantitatively assess the nature of such bias.

In what follows, only the sample size of 31 used in the study is considered. The reason is that the interest is in the impact of small sample bias on the primal versus dual equation estimates for a given sample size. The small sample distribution of the relevant estimator is generated from 1000 replications of the variables, and the estimation is done for all 21 industries. The simulation is done to preserve the correlation properties of the variables observed in the data. For this purpose, the vector of instruments from the actual sample is used, and the following procedure is employed. The actual series on share-weighted input growth is regressed against the instrument vector described earlier. We define $\widehat{\Delta x}_{it}$ as the fitted values, $\bar{\varepsilon}^x_{it}$ as the estimated residual, and $\hat{\sigma}^2_x$ as the estimated variance of the residual from this regression. The simulated series on share-weighted input growth is obtained by adding a random error $\bar{\varepsilon}^x_{it} \sim N(0, \hat{\sigma}^2_x)$ to these fitted values as

$$\Delta x^s_{it} = \widehat{\Delta x}_{it} + \Delta\varepsilon^x_{it} \tag{19}$$

The simulated series on the growth rate of sectoral technology shocks are constructed to be uncorrelated with the instrument vector but correlated with share-weighted input growth. Specifically, an estimate of changes in sectoral technology $\widehat{\Delta z}_{it}$ is obtained from the (nonlinear) least squares estimation of the primal and dual equations comprised by Equations 6 and 10. The estimated residual from the primal equation is then regressed on the estimated residual $\hat{\varepsilon}^x_{it}$ from the regression of share-weighted input growth on the instruments. Letting β^x denote the regression coefficient and $\hat{\sigma}^2_z$ denote the estimated variance of the residual from this regression, the simulated series on sectoral technology growth is obtained as

$$\Delta z^s_{it} = \beta^x \varepsilon^x_{it} + \varepsilon^z_{it} \tag{20}$$

where $\varepsilon^z_{it} \sim N(0, \hat{\sigma}^2_z)$. The simulated series on Δy_{it} is obtained by

$$\Delta y^s_{it} = \gamma \Delta x^s_{it} + \Delta z^s_{it} \tag{21}$$

This yields the simulated version of the *primal equation*. Let Δp^x_{it} denote the share-weighted changes in input prices. The *dual equation* is simulated by using the simulated values of output growth and changes in technology plus an idiosyncratic error $\varepsilon^p_{it} - \sim N(0, \hat{\sigma}^2_p)$ as

$$(\Delta p_{it} - \Delta p^x_{it})^s = \frac{\gamma_i - 1}{\gamma_i} \Delta y^s_{it} - \frac{1}{\gamma_i} \Delta z^s_{it} + \varepsilon^p_{it} \tag{22}$$

where $\hat{\sigma}^2_p$ is the estimated variance of the idiosyncratic error in the dual equation.

In Table 5b, we report various statistics of the simulated distribution of the instrumental variables estimator that is used to derive the estimates of the primal and dual equations. In part (b), $\bar{\gamma}$ is the estimated returns to scale coefficient, averaged over replications and industries. Likewise $SE(\gamma)$ denotes the estimated standard error of the returns to scale coefficient, averaged over replications and industries. The rejection rates for the true hypothesis show the incidence of Type I errors at the 10, 5 and 1% significance levels. Likewise, the rejection rates for the false hypothesis show the power of the test at the corresponding significance levels. The simulations are done for $\gamma_i = 0.8$ and $\gamma_i = 1.2$. Since γ_i enters the primal equation linearly, the simulated distribution of the estimated coefficient for this equation does not change with different values of γ_i. However, allowing for different values of γ_i enables the impact of the nonlinearity in the dual equation to be assessed with respect to the estimation of the returns to scale coefficient.

Part (b) of Table 5 shows that when all 21 industries are considered, the primal equation estimates tend to approximate the true value of γ_i quite closely. By contrast, the 21-industry average of the dual estimates point to an upward bias in the estimates of γ_i. For example, for $\gamma_i = 0.8$, the average percentage bias in estimating this coefficient is close to zero for the primal equation estimates but around 2.5% for the dual equation estimates. The estimated average standard errors for the dual equation estimates also tend to be larger. Turning to simulated sizes of the tests, it is found that the instrumental variables estimator used in this study tends to reject the null hypothesis too frequently.

For example, at a nominal size of 5%, the actual rejection rates are approximately equal to 7% for both the primal and dual equation estimates. This is similar to other results regarding the small sample behaviour of the instrumental variables estimator, as discussed by Nelson and Startz (1990), Ferson and Foerster (1994), Hall *et al.* (1996) and others and it provides a justification for calculating coefficient standard errors using restricted residuals as suggested by Ligeralde and Brown (1995). However, more important for the purposes of this study, it is found that the actual rejection rates for the true null hypothesis differ very little for the primal versus dual estimates. Thus, while there seems to be some upward bias in the dual equation estimates (arising from the fact that the instruments correlate less well with the some of the regressors in this equation), the rejection rates suggest that one should not observe a significant upward bias in rejecting the true hypothesis of constant returns to scale, say, from the dual equation relative to the primal equation.[23] Part (b) also reports the power of the tests to reject the hypothesis of constant returns to scale. The tests based on both the primal and the dual equation display considerable power to reject the false null hypothesis, although the power of the tests are slightly lower for the dual equation estimates.

When the results for 21 industries are considered separately, it can be shown that the slight upward bias in the dual equation estimates and in the standard errors of the primal equation are primarily due to the small sample bias in the estimates for Food, Chemicals, Petroleum Products, and Miscellaneous Manufacturing. For example, when the true value of γ is 0.8, the simulated average of the dual equation returns to scale coefficients for these industries are given by 0.85156, 0.87225, 0.85333 and 0.90395, respectively. Moreover, that standard errors of the primal equation returns to scale estimates also tend to be significantly larger than average for these industries. For example, the average standard errors of the primal equation estimates for Chemicals, Petroleum Products, and Miscellaneous Manufacturing are given by 0.089, 0.076 and 0.116. For an industrial group such as Food, one reason for these results may be product variety, where different products are characterized by differences in brand loyalty and hence, the nature of their markup behaviour. Taken together, these results suggest that small sample bias is likely to be most severe for at least a subset of the non-durable goods industries.

V. CONCLUSION

This paper has presented estimates of the degree of returns to scale using both production data on inputs and outputs as well as data on product and factor prices for 2-digit US manufacturing industries. As part of the analysis, the cyclical behaviour of primal and dual productivity measures have been considered, allowing for time-varying markups, and examining the small sample properties of the primal versus dual equation estimates. Both the primal and dual equation estimates indicate the existence of increasing returns to scale for the durable goods industries, and evidence is found for differences in the cyclicality of markups for the non-durable versus durable goods industries. Furthermore, the analysis of the small sample properties of the instrumental variables estimator used to derive estimates of the primal and dual equations shows that there is a tendency for the dual equation to slightly overestimate the degree of the returns to scale. Finally, the simulation results indicate that small sample bias is likely to be more severe for some of the non-durable goods industries.

As the discussion in Section II indicates, deviations from nonconstant returns to scale and the extent of market power have important implications for the modelling of cyclical fluctuations. The results indicate such effects are empirically important. Furthermore, the results indicate that heterogeneity across industry groups is also empirically important. In the study, such features as time-varying markups have been incorporated and as a result, an implicit measure of cyclical variation in capacity utilization. Future extensions of this work involve analysing models that include more explicit dynamic features.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the North American Summer meeting of the Econometric Society, Montreal 1998, Econometric Society European meeting, Berlin 1998, Money, Macro, Finance Group, York 2000, and at seminars at the Athens University of Economics and Business, Birkbeck College, the University of Cambridge, and the University of Bristol. The authors are grateful to Rick Ashley, Leslie Godfrey and Peter N. Smith for helpful comments, and to Dale Jorgenson, Susanto Basu and Craig Burnside for providing the data used in this paper.

[23] Explicitly allowing for conditional heteroscedasticity or serial correlation in the disturbances of the primal and dual equations may worsen these results. The authors tried to allow for conditional heteroscedasticity in the actual estimation reported in Section III but the instrumental variables estimator became ill-conditioned under this assumption.

REFERENCES

Abbott III, T., Griliches, Z. and Hausmann, J. (1988) Short run movements in productivity: market power versus capacity utilization, unpublished manuscript.

Applebaum, E. (1979) Testing price taking behaviour, *Journal of Econometrics*, **9**, 283–94.

Basu, S. (1996) Procyclical productivity: overhead inputs or capacity utilization?, *Quarterly Journal of Economics*, **111**, 719–51.

Basu, S. and Fernald, J. (1995) Are apparent spillovers a figment of specification error?, *Journal of Monetary Economics*, **36**, 165–88.

Basu, S. and Fernald, J. (1997a) Returns to scale in US production: estimates and implications, *Journal of Political Economy*, **105**, 249–83.

Basu, S. and Fernald, J. (1997b) Aggregate productivity and aggregate technology, Board of the Governors of the Federal Reserve System, International Finance Discussion Paper No. 593.

Benhabib, J. and Farmer, R. (1994) Indeterminacy and increasing returns, *Journal of Economic Theory*, **63**, 19–41.

Burnside, C. (1996) Production function regressions, returns to scale, and externalities, *Journal of Monetary Economics*, **37**, 177–201.

Burnside, C., Eichenbaum, M. and Rebelo, S. (1993) Labor hoarding and the business cycle, *Journal of Political Economy*, **101**, 245–73.

Caballero, R. and Lyons, R. (1992) External effects in US procyclical productivity, *Journal of Monetary Economics*, **29**, 209–25.

Chirinko, R. and Fazzari, S. (1994) Economic fluctuations, market power, and returns to scale: evidence from firm-level data, *Journal of Applied Econometrics*, **9**, 47–69.

Domowitz, I., Hubbard, G. and Petersen, C. (1986) Business cycles and the relationship between concentration and price cost margins, *Rand Journal of Economics*, **17**, 1–17.

Domowitz, I., Hubbard, G. and Petersen, C. (1988) Market structure and cyclical fluctuations in US manufacturing, *Review of Economics and Statistics*, **70**(1), 55–66.

Evans, C. (1992) Productivity shocks and real business cycles, *Journal of Monetary Economics*, **29**, 191–208.

Farmer, R. and Guo, J. (1994) Real business cycles and the animal spirits hypothesis, *Journal of Economic Theory*, **63**, 42–72.

Ferson, W. and Foerster, S. (1994) Finite sample properties of the generalized method of moments estimator in tests of conditional asset pricing models, *Journal of Financial Economics*, **36**, 29–55.

Galeotti, M. and Schiantarelli, F. (1998) The cyclicality of markups in a model with adjustment costs: econometric evidence for US industry, *Oxford Bulletin of Economics and Statistics*, **60**, 121–42.

Hall, R. (1988) The relation between price and marginal cost in US industry, *Journal of Political Economy*, **96**, 921–47.

Hall, R. (1990) Invariance properties of Solow's productivity residual, In *Growth/Productivity/Employment*, P. Diamond (ed), Cambridge, MA: MIT Press.

Hall, R. and Jorgenson, D. (1967) Tax policy and investment behavior, *American Economic Behavior*, **57**, 391–414.

Hall, A., Rudebusch, G. and Wilcox, D. (1996) Judging instrument relevance in instrumental variables estimation, *International Economic Review*, **37**, 283–98.

Haskel, J., Martin, C. and Small, I. (1995) Price, marginal cost, and the business cycle, *Oxford Bulletin of Economics and Statistics*, **57**, 25–41.

Hulten, C. (1986) Productivity change, capacity utilization, and the sources of efficiency growth, *Journal of Econometrics*, **33**, 31–50.

Jorgenson, D. (1990) Productivity and economic growth, in *Fifty Years of Economic Measurement*, E. R. Berndt and J. E. Triplett (eds.), Chicago, Il: University of Chicago Press.

Jorgenson, D., Gollop, F. and Fraumeni, B. (1987) *Productivity and US Economic Growth*, Cambridge, Mass.: Harvard University Press.

Ligeralde, A. and Brown, B. (1995) Band covariance matrix estimation using restricted residuals: a Monte Carlo analysis, *International Economic Review*, **36**, 751–67.

Machin, S. and Van Reenen, J. (1993) Profit margins and the business cycle: evidence for UK manufacturing firms, *Journal of Industrial Economics*, **41**, 29–50.

Morrison, C. (1992) Unravelling the productivity growth slowdown in the United States, Canada, and Japan: the effects of subequilibrium, scale economies, and markups, *Review of Economics and Statistics*, **74**, 381–93.

Morrison, C. (1993) Productive and financial performance in US manufacturing industries: an intregrated structural approach, *Southern Economic Journal*, **60**, 376–92.

Morrison, C. (1994) The cyclical nature of markups in Canadian manufacturing: a production theory approach, *Journal of Applied Econometrics*, **9**, 269–82.

Nelson, C. and Startz, R. (1990) The distribution of the instrumental variables esimator and its *t*-ratio when the instrument is a poor one, *Journal of Business*, **63**, 125–40.

Ohta, M. (1975) A note on the duality between production and cost functions: rate of returns to scale and rate of technical progress, *Economic Studies Quarterly*, **25**, 63–5.

Paul, C. Morrison, and Siegel, D. (1999a) Scale economies and industry agglomeration externallities: a dynamic cost function approach, *American Economic Review*, **89**, 272–90.

Paul, C. Morrison and Siegel, D. (1999b) Estimation of scale economies underlying growth and productivity: the empirical implications of data aggregation, *Southern Economic Journal*, **65**, 730–56.

Perli, R. (1998) Indeterminacy, home production, and the business cycle: a calibrated analysis, *Journal of Monetary Economics*, **41**, 105–25.

Prescott, E. (1986) Theory ahead of business cycle measurement, *Carnegie-Rochester Conference Series on Public Policy*, **25**, 11–44.

Roeger, W. (1995) Can imperfect competition explain the difference between the primal and dual productivity measure? Estimates for US manufacturing, *Journal of Political Economy*, **21**, 316–30.

Rotemberg, J. and Woodford, M. (1995) Dynamic general equilibrium models with imperfectly competitive product markets, in *Frontiers of Business Cycle Research*, Thomas Cooley (ed.), Princeton: Princeton University Press, pp. 243–93.

Schmitt-Grohe, S. (1994) Comparing four models of fluctuations due to self-fulfilling expectations, unpublished manuscript, Federal Reserve Board.

Shapiro, M. (1987) Are cyclical fluctuations in productivity due more to supply shocks or demand shocks?, *American Economic Review Papers and Proceedings*, **77**, 118–24.
Shea, J. (1997) Instrument relevance in multivariate linear models: a simple measure, *Review of Economics and Statistics*, **79**(2), 348–52.
Solow, R. (1958) Technical change and the aggregate production function, *Review of Economics and Statistics*, **39**, 312–20.
Staiger, D. and Stock, J. (1994) Instrumental variables estimation with weak instruments, NBER Working Paper 151.

APPENDIX

In the analysis in the text, it has been assumed that the capital input is a fully variable input. In an earlier paper, Shapiro (1987) argues that the cyclical behaviour of the primal and dual productivity residuals can be reconciled under the assumption that the capital input is fixed in the short-run. The analysis is now extended to allow for the fixity of capital under imperfect competition and nonconstant returns to scale. As in Shapiro (1987), the nonparametric measures of productivity that are used above are modified to account for the fact that capital is fixed and may not be valued by its rental rate in the short run.

The results are derived under the assumption that the production function is of the Cobb-Douglas variety as

$$Y_{it} = z_{it} K_{it}^{\alpha_i^1} M_{it}^{\alpha_i^2} L_{it}^{\alpha_i^3}, \quad \alpha_i^1 + \alpha_i^2 + \alpha_i^3 = \gamma_i \tag{A1}$$

An expression can be derived for the primal equation by logarithmically differentiating Equation A1. However, the resulting expression depends on the parameters α_1, α_i^2, and α_i^3. With the quasi-fixity of capital, one cannot proceed as in the derivation of Equation 6 and rewrite this expression in terms of the cost shares of the inputs and the returns to scale parameter γ_i because capital is not valued at its rental cost. However, it is easy to show that cost minimization with the quasi-fixity of capital implies that $\alpha_i^2 = (\gamma_i - \alpha_i^1) c_{M,it}^{VC}$ and $\alpha_3 = (\gamma_i - \alpha_i^1) c_{L,it}^{VC}$, where $c_{j,it}^{CV}$ denotes the share of input j in total variable cost for $j = L, M$. Using the expressions for α_2 and α_3, it is straightforward to show that the primal equation under the fixity of capital is given by

$$\Delta y_{it} = (\gamma_i - \alpha_i^1)\left[c_{M,it}^{VC} \Delta m_{it} + c_{L,it}^{VC} \Delta l_{it}\right] + \alpha_i^1 \Delta k_{it} + \Delta z_{it} \tag{A2}$$

and the modified primal productivity residual SR_{it} is given by

$$SCV_{it} = \Delta y_{it} - c_{L,it}^{VC} \Delta l_{it} - c_{M,it}^{VC} \Delta m_{it} = \left(1 - \frac{1}{\gamma_i - \alpha_i^1}\right) \Delta y_{it} + \frac{\alpha_i^1}{\gamma_i - \alpha_i^1} \Delta k_{it} + \frac{1}{\gamma_i - \alpha_i^1} \Delta z_{it} \tag{A3}$$

where $c_{j,it}^{VC}$ denotes the share of input j in total variable costs for $j = L, M$ and α_i^1 denotes the elasticity of output with respect to capital.

To derive the implications of the dual problem, it is noted that the marginal cost function is given by

$$MC_{it} = \frac{p_{it}^L L_{it} + p_{it}^M M_{it}}{(\alpha_i^2 + \alpha_i^3) Y_{it}} \tag{A4}$$

Logarithmically differentiating the expression in Equation A4 and using the form of the production function in Equation A1 to substitute for ΔL_{it} in the resulting expression yields

$$\Delta \mathrm{MC}_{it} = c_{L,it}^{VC} \Delta p_{it}^L + c_{M,it}^{VC} \Delta p_{it}^M + \left[c_{M,it}^{VC} - \frac{\alpha_i^2}{\alpha_i^3} c_{L,it}^{VC}\right] \Delta m_{it} + \left[\frac{c_{L,it}^{VC}}{\alpha_i^3} - 1\right] \Delta y_{it} - c_{L,it}^{VC} \left(\frac{\alpha_i^1}{\alpha_i^3}\right) \Delta k_{it} - \frac{c_{L,it}^{VC}}{\alpha_3} \Delta z_{it} \tag{A5}$$

Making use of the fact that $\Delta \mathrm{MC}_{it} = \Delta p_{it} - \Delta \mu_{it}$ together with the expressions for α_i^2, α_i^3, and $\gamma_i - \alpha_i^1$, one can derive an expression for the dual equation

$$\Delta p_{it} = c_{L,it}^{VC} \Delta p_{it}^L + c_{M,it}^{VC} \Delta p_{it}^M + \left(\frac{1 - (\gamma_i - \alpha_i^1)}{\gamma_i - \alpha_i^1}\right) \Delta y_{it} - \frac{\alpha_i^1}{\gamma_i - \alpha_i^1} \Delta k_{it} + \Delta \mu_{it} - \frac{1}{\gamma_i - \alpha_i^1} \Delta z_{it} \tag{A6}$$

and the modified dual productivity residual as

$$\mathrm{SPV}_{it} = c_{L,it}^{VC} \Delta p_{it}^L + c_{M,it}^{VC} \Delta p_{it}^M - \Delta p_{it} = \left(1 - \frac{1}{\gamma_i - \alpha_i^1}\right) \Delta y_{it} + \frac{\alpha_i^1}{\gamma_i - \alpha_i^1} \Delta k_{it} + \frac{1}{\gamma_i - \alpha_i^1} \Delta z_{it} - \Delta \mu_{it} \tag{A7}$$

These expressions show that with the fixity of capital, the primal and dual productivity measures depend on the change in the marginal product of capital evaluated in terms of quantities. Furthermore, if markups are constant, then the expressions in Equations (A3) and (A7) show that scale effects can be identified either from the primal or the dual productivity measure.

As in the case when capital is a fully variable input, one can construct a specification test based on the primal and dual version of the firm's cost minimization problem following an approach used to derive Equations 13 or 18. One can examine the correlations of the modified primal and dual productivity residuals with aggregate real value-added. In the former case, it was found that the equality of the primal and dual equation estimates is rejected for the majority of the industries in the study. It was also found that the modified primal residual is more procyclical relative to the modified dual residual. These results indicate that allowing for the fixity of capital using a simple specification for the production technology is not necessarily useful for reconciling the primal and dual versions of the firm's cost minimization problem.