JOURNAL OF APPLIED ECONOMETRICS, VOL. 11, 343–361 (1996)

THE INCONSISTENCY OF COMMON SCALE ESTIMATORS WHEN OUTPUT PRICES ARE UNOBSERVED AND ENDOGENOUS

TOR JAKOB KLETTE
Statistics Norway, Kongens gt. 6, PB 8131 Dep, N-0033 Oslo, Norway

AND

ZVI GRILICHES
Department of Economics, Harvard University, 125 Littauer Centre, Cambridge, MA 02138, USA and National Bureau of Economic Research

SUMMARY

This paper explores the inconsistency of common scale estimators when output is proxied by deflated sales, based on a common output deflator across firms. The problem arises when firms operate in an imperfectly competitive environment and prices differ between them. In particular, we show that this problem reveals itself as a downward bias in the scale estimates obtained from production function regressions, under a variety of assumptions about the pattern of technology, demand and factor price shocks. The result also holds for scale estimates obtained from cost functions. The analysis is carried one step further by adding a model of product demand. Within this augmented model we examine the probability limit of the scale estimate obtained from an ordinary production function regression. This analysis reveals that the OLS estimate will be biased towards a value below one, and how this bias is affected by the magnitude of the parameters and the amount of variation in the various shocks. We have included an empirical section which illustrates the issues. The empirical analysis presents a tentative approach to solve the problem discussed in the theoretical part of this paper.

1. INTRODUCTION

In many, if not most branches of economics there is a renewed interest in questions related to scale economies and imperfect competition. Despite its long history, empirical estimation of scale economies remains a controversial and unsettled research topic. In particular, scale estimates obtained by estimating production functions on the basis of firm-level data tend to suggest substantial decreasing returns to scale.[1] On the other hand, it is well known that estimates of factor demand equations give results which imply increasing returns to scale.[2] This

[1] This proposition summarizes the studies that apply panel data; see Griliches and Mairesse (1995) for a survey of the literature. Cross-sectional studies of production functions typically suggest increasing returns to scale; see e.g. Griliches and Ringstad (1971) and Ringstad (1974). There is a widely held view that scale estimates from cross-sectional studies are upward biased as these studies do not account for persistent differences in efficiency between firms. This is an old issue which is much commented on in the literature. Discussions of the questions involved in comparing cross-sectional and panel data studies of production functions are provided by e.g. Ringstad (1971), Mundlak (1978), Mairesse (1990) and Griliches and Mairesse (1995).

[2] Griliches and Hausman (1986) and Biørn and Klette (1994) examine to what extent increasing returns to scale in the labour demand equation can be interpreted as errors in variables in the output variable.

CCC 0883–7252/96/040343–19

Received September 1993
Revised August 1995

paper points out a common problem with scale estimates obtained from both production function and cost function regressions. In particular, we suggest that the coefficients usually interpreted as the scale elasticity in such regressions more generally should be considered a mixture of both the scale elasticity and demand-side parameters.

The large amount of heterogeneity between firms, even within narrowly defined industries, is one of the clear facts which have emerged from the many studies of firm behaviour based on micro-data sources. For instance, significant dispersion in output prices seems to exist between firms in several industries (see studies cited below). This paper explores one implication of neglecting such price dispersion. The theoretical part of the paper shows that the practice of using deflated sales as a proxy for real output will, *ceteris paribus*, tend to create a downward bias in the scale estimate obtained from production function regressions. This is so under a variety of conditions if the firms face an imperfectly competitive environment. When estimating cost functions—which is closely related to estimating factor demand equations—we show that there will be a similar bias. The basic issue we are addressing in this paper was clearly recognized already by Marschak and Andrews (1944), but seems to have been neglected in much of the later literature on empirical production analysis.

The analysis is carried one step further by adding a model of product demand to the model of producer behaviour. Within this augmented model we examine the bias in the scale estimate obtained from an ordinary OLS regression of the production function. This analysis consider both the bias caused by replacing output by deflated sales as well as that caused by the correlation between input growth and productivity shocks (often termed the 'transmission bias'[3]). On the basis of this augmented analysis, we show that the scale estimator is biased towards a value below one.

The empirical part of this paper is meant to illustrate the issues pointed out in the theoretical sections. In particular, we present a tentative approach to consistent estimation of price–cost margins and scale economies when deflated sales is the available proxy for real output. One of the main points of this analysis is to show how the parameters in 'production function regressions' are reduced-form parameters, i.e. mixtures of supply- and demand-side coefficients. The analysis presents assumptions sufficient to permit us to recover the elasticities of scale and demand from estimation of production and cost functions, after we have added changes in industry output as an additional regressor. After reinterpreting the estimated parameters as reduced-form coefficients, we find estimates of the scale elasticity to variable factors alone of the order $1{\cdot}06$–$1{\cdot}10$, and estimated demand elasticities which vary from -6 to -12.

The empirical analysis is related to a more general issue. Our regressions reveal that total factor productivity at the plant level is highly correlated with the (industry-wide) changes in average sales, in regressions where we do not impose assumptions about constant returns to scale or perfect competition. Hall (1990) has provided an extensive discussion of the possible explanations for the pro-cyclical behaviour of total factor productivity. Two explanations examined in great detail by Hall are market power and scale economies. Our results suggest that, at the micro level, the pro-cyclical movements in (measured) productivity cannot be fully accounted for by incorporating imperfect competition and scale economies into the model.[4]

[3] The literature on the 'transmission bias' dates back to Marschak and Andrews (1944). This simultaneity problem has been discussed by Mundlak and Hoch (1965) and Zellner, Kmenta and Dreze (1966), among others.

[4] The empirical model presented in the second half of this paper is very similar to a model examined by Bartelsman, Caballero and Lyons (1991). Their main finding is that productivity at the industry level is highly correlated across industries, and in particular with changes in output aggregated across industries. The current paper identifies a similar pattern at the plant level. However, the interpretation we offer differ entirely from the interpretation of Bartelsman *et al.*, who claim their finding suggests some kind of external economies. We interpret our finding as a result of the econometric problem created by replacing the unobserved movements in output with changes in deflated sales.

Abbott (1991) has presented results supporting the perspective of the current paper. He had access to price data for individual firms (which we do not have). His analysis shows that prices differ significantly within the hydraulic cement industry in the USA, also after adjustment is made for differences in output mix. Using individual deflators rather than industry-wide deflators gives different and, in Abbott's terms, more plausible estimates of production function parameters and productivity changes. Dunne and Roberts (1992) also emphasize the importance of price dispersion between firms within narrowly defined industries. Having access to firm-level price data, they examine to what extent price differences can be explained by cost differences as well as variables capturing the firm's competitive environment.

This paper is organized as follows. Section 2 provides a general theoretical analysis of the problem created by using a common deflator when prices differ between firms, in terms of the omitted variable framework. It begins by examining the production function case, and then provides a discussion of the cost function case. Section 3 carries the analysis a step further in the production function case by studying the bias of scale estimates within an explicit, complete model of supply and demand. The analytical expression for the asymptotic bias is derived assuming orthogonality between the idiosyncratic productivity, mark-up, factor price and demand shocks. The empirical analysis is presented in Section 4. Some final comments are added in Section 5.

2. THE OMITTED VARIABLE BIAS

In this section we will provide a general analysis of the inconsistency of scale estimates when estimation proceeds by using deflated sales instead of output in production and cost function analysis. The analysis shows that if the (real, unobserved) prices are correlated with the included variables in the model, an omitted variable bias will arise. More specifically, we will argue that in the analysis of production functions, plausible assumptions suggest that commonly applied scale estimators will be downward biased. Also in the case of cost functions a similar bias of the scale elasticity will occur. To focus ideas we will carry out the argument in terms of panel data estimation of production relationships. More specifically, we will consider models in terms of growth rates, but the argument can easily be altered to be of relevance for pure cross-sectional regressions. However, the biases will probably differ, as the *importance* of the various shocks considered below will be different in the cross-sectional versus the time-series dimension.

2.1. The Production Function Case

Let us assume that the true production function relationship can be written $q = X\alpha_0 + u^q$, where q is a $(N \times 1)$ vector of the growth in 'real' output, X is a $(N \times L)$ matrix of the growth in inputs. α_0 is the $(L \times 1)$ vector of the parameters of interest, while u^q is assumed to be an orthogonal error term. N is the number of observations. This model is the familiar Cobb–Douglas production function, and the scale elasticity is defined as $\varepsilon \equiv \sum_{i=1}^{L} \alpha_{i0}$.

The estimated model is a slight modification to the true production function: $r = X\alpha + u^r$, where the left-hand-side variable now is r, which represents changes in deflated sales based on an industry-wide deflator. The OLS estimator of the parameter vector α, assuming orthogonality of u^r, is

$$\hat{\alpha} = (X'X)^{-1}X'r \tag{1}$$

Define the growth in the firm-specific price relative to the deflator as π. Then the relationship between true output (q) and deflated sales is $\pi + q = r$. Focusing on the probability limit of $\hat{\alpha}$ and using this relationship, we obtain

$$\text{plim}_{N\to\infty}(\hat{\alpha}) = \alpha_0 + \text{plim}_{N\to\infty}[(X'X)^{-1}X'\pi)] + \text{plim}_{N\to\infty}[(X'X)^{-1}X'u^q] \tag{2}$$

Both of the last two terms in equation (2) may have non-zero probability limits. The potential non-zero probability limit of the last term in equation (2) is referred to in the literature as the bias from the 'transmission' of productivity shocks (cf. footnote 3). This problem will be neglected for the moment, as it has been discussed extensively in the existing literature. Let us focus on the second term on the right-hand side of equation (2).

To examine the second term in equation (2), notice that it can be expressed as the OLS estimate of the vector of δ's in the auxiliary regression:[5] $\pi = X\delta + u^{\pi}$, where u^{π} is an orthogonal error term. The direction and the size of the bias in $\hat{\alpha}$ will depend on the sign and the magnitude of the δ coefficients in this auxiliary regression. That is, $\text{plim}_{N\to\infty}(\hat{\alpha}_i) = \alpha_{i0} + \delta_i$. Focusing on the bias in the estimated scale elasticity ($\hat{\varepsilon} \equiv \sum_i \hat{\alpha}_i$), we have

$$\begin{aligned} \text{plim}_{N\to\infty}\hat{\varepsilon} - \varepsilon &= \text{plim}_{N\to\infty}\left(\sum_{i=1}^{L}\hat{\alpha}_i\right) - \sum_{i=1}^{L}\alpha_i \\ &= \sum_{i=1}^{L}\delta_i \end{aligned} \tag{3}$$

where ε is the true scale elasticity. The question now is: Will there be a systematic relationship between the changes in the price a firm charges and the growth of the firm in terms of its *inputs* (cf. the auxiliary regression)? A satisfactory analysis of the relationship between the price a firm charges and the size of the firm requires a more complete model which includes the factors which determine the firm's price-setting behaviour and demand. The next section will provide a simple model incorporating these aspects into a complete formal framework. For the moment we will provide a more general, but less formal discussion.

Let us first consider the impact of idiosyncratic changes in (quality-adjusted) factor prices. This case is simple, and suggests that firms that experience higher costs will, *ceteris paribus*, charge a higher price and lose market shares. Hence, idiosyncratic changes in factor prices suggest a negative relationship between firms' price movements and the changes in input levels, i.e. $\sum_{i=1}^{L} \delta_i < 0$.

The next case to consider is the relationship between price and the level of inputs, when there are idiosyncratic productivity shocks. If efficiency levels differ between firms, it seems plausible that the more productive firm will have a larger market share, and also charge a lower price. If some firms experience productivity improvements beyond the average, they will probably, *ceteris paribus*, obtain a larger market share, measured in terms of (quality-adjusted) output. More output does not necessarily imply more inputs when productivity improves; but with elastic demand the increase in output tends to dominate the productivity gain so that the use of inputs increase. That is, we expect $\sum_{i=1}^{L} \delta_i < 0$ also in this case.

The last case we wish to discuss here is the consequences of demand shocks. If there are scale economies, changes in firm size will affect the price a firm charges in an imperfectly competitive environment. With decreasing returns to scale, we would expect a firm that grows faster than the

[5] This idea has been applied by Griliches (1957), and Griliches and Ringstad (1971, appendix C) to discuss other issues of specification bias in the estimation of production functions.

average, to increase its relative price, and vice versa. Hence, this case suggests positive (negative) δ coefficients in the auxiliary regression if there are negative (positive) scale economies. It follows that demand shocks will bias the estimated scale elasticity towards one.

To summarize, we have shown that in most cases the omitted price variable will create a downward bias in the scale estimator. Before we proceed, we should emphasize that the concept of the *omitted price variable* might be misleading as it suggests that the problem can be solved by means of a suitable choice of instrumental variables. This is not so, since variables that are correlated with the changes in inputs (i.e. potentially useful instruments) will also be correlated with movements in the omitted price variable buried in the residual and therefore illegitimate as instruments.[6] We will return to a solution of the problem in Section 4.

2.2. Scale Estimates from Cost Functions

A similar result to the inconsistency pointed out in the production function case applies when estimating cost functions replacing real output with deflated sales. To make this argument as transparent as possible, we will keep to the 'linear-in-variables' case:[7]

$$c = W\gamma_0 + \beta_0 q + u^c \tag{4}$$

is assumed to be the true relationship. c is a $(N \times 1)$ vector representing the growth in costs. W is a $(N \times K)$ matrix expressing the growth in the K factor prices for the N observations. q is growth in real output as above. β_0 is the parameter of interest. It is the inverse of the scale elasticity. Once more, we assume that we do not observe real output growth; q. The estimated model assumes the same relationship replacing output with deflated sales:

$$\begin{aligned} c &= W\gamma + \beta r + \tilde{u}^c \\ &= Z\lambda + \tilde{u}^c \end{aligned} \tag{5}$$

where Z denotes the $(N \times (K+1))$ matrix obtained by adding the (column-) vector r to the W matrix. λ is the $(K+1)$ column vector containing γ and β.

Using the expression $r - q = \pi$, we can rewrite equation (4):

$$\begin{aligned} c &= W\gamma_0 + \beta_0 r + \beta_0(q - r) + u^c \\ &= Z\lambda_0 - \beta_0 \pi + u^c \end{aligned} \tag{6}$$

where λ_0' is the true parameter values (γ_0', β_0). It follows that the probability limit of the OLS estimator of λ is given by

$$\operatorname{plim}_{N\to\infty}(\hat{\lambda}) = \lambda_0 - \beta_0 \operatorname{plim}_{N\to\infty}(Z'Z)^{-1}Z'\pi + \operatorname{plim}_{N\to\infty}(Z'Z)^{-1}Z'\tilde{u}^c \tag{7}$$

At this point we again neglect the last term in equation (7). The problem in focus here is the second term on the right-hand side of equation (7). Once more we can obtain some intuition about this term by noting that $\operatorname{plim}_{N\to\infty}(Z'Z)^{-1}Z'\pi$ can be thought of as the coefficient associated with the regression of π on Z. That is, the term $\operatorname{plim}_{N\to\infty}(Z'Z)^{-1}Z'\pi$ is equal to the $\tilde{\delta}$ vector in the auxiliary regression

$$\begin{aligned} \pi &= Z\tilde{\delta} + v^\pi \\ &= W\tilde{\delta}_W + r\tilde{\delta}_r + v^\pi \end{aligned} \tag{8}$$

[6] With constant returns to scale, this is not true. But, of course, if we assume constant returns to scale to begin with, there is no purpose in estimating scale economies.
[7] That is, we consider the cost function corresponding to a Cobb–Douglas production function.

where v^{π} is an orthogonal error term. The question is whether and how the scale coefficient will be biased, which is equivalent to asking about the significance and magnitude of the $\tilde{\delta}_r$ coefficient in the OLS regression of the model in equation (8). Ruling out the unlikely case of inelastic demand, this coefficient will be negative. Notice that if we neglect the last term in equation (7), we have

$$\text{plim}_{N\to\infty}\hat{\beta} = \beta_0(1 - \tilde{\delta}_r) \tag{9}$$

It follows that the cost function estimates of the scale elasticity—equal to the inverse of $\hat{\beta}$—will be biased downwards in the limit.

3. BIAS FORMULAS IN AN EXPLICIT MODEL

In this section we will provide a more formal argument of the inconsistency of the scale estimate in the production function case. Our analytical expression for the bias in the scale coefficient permits us to examine how the bias is affected by the magnitude and combination of the parameters and the variances of the various shocks. In our analysis we consider simultaneously two sources of bias: (1) The 'omitted price bias', discussed in the previous section, and (2) the more widely discussed bias due to the endogeneity of inputs, i.e. the 'transmission bias' (cf. footnote 3).

Our analysis is carried out in a model where the production function is augmented with a simple demand equation and a price-setting rule. To keep the argument as transparent as possible, we have restricted the analysis to the log-linear case. This restrictive case is sufficient to highlight our main points.

Consider the production function with only one input: $Q_{it} = A_{it}X^{\varepsilon}_{it}$, Q_{it}, A_{it} and X_{it} are the output, productivity and input of firm i at time t. The demand function facing the firm can be expressed by $Q_{it} = D_{it}P^{\eta}_{it}$, where D_{it} is a demand shifter. P_{it} is the firm's output price which we will measure relative the general price level in the industry for convenience. η is the demand elasticity.[8] The environment we have in mind is an industry with horizontal product differentiation.[9] A finite η corresponds to product differentiation between the goods produced by the different firms in the industry.

We assume that the firm applies a mark-up rule $P_{it} = \mu_{it}\, \partial C_{it}/\partial Q_{it} = (\mu_{it}C_{it})/(\varepsilon Q_{it})$, where C_{it} is total (deflated) costs, while μ_{it} is the markup.

This system can be linearized by considering all variables in terms of growth rates.[10] One can then solve the endogenous variables q_{it}, x_{it} and p_{it} in terms of the variables which shock the system. We will use the notation that a lower-case letter is the growth rate for the corresponding upper-case variable. The exogenous shocks are assumed to be idiosyncratic changes in productivity, demand and factor prices. We also allow for changes in the mark-up.[11]

---

[8] A similar demand system has been widely examined in the industrial organization and international trade literature under the label 'the Spence–Dixit–Stiglitz-model'. The characteristic feature of this formalization of monopolistic competition is that it leads to price elasticities which are constant and independent of the number of variants available. Tirole (1988, ch. 7.5) provides a discussion and further references to the micro foundations of this demand system. See e.g. Berry (1994) for a discussion and further references to the literature on econometric models of demand in industries with product differentiation.

[9] Vertical product differentiation is hidden in the possibility that Q_{it} can be an index capturing both the quantity and quality of the output.

[10] I.e. taking the logarithmic differences between year 't' and '$t-1$'.

[11] In the case of price-setting firms, with a finite number of firms, the deviations from a fixed mark-up rule will be related to a firm's market share. That is, u^{m}_{it} is positively related to the deviation in the firm's market share from the mean market share, see Klette (1990, ch. 4). In the case of vertical product differentiation, the mark-ups might be correlated with the quality of the firms' output. See Tirole (1988, ch. 7.5.1) for discussion and references. One can argue that the mark-up will change over time depending on the firm's innovative history. For example, an idiosyncratic productivity shock might lead the innovative firm to charge a higher mark-up, as discussed by Arrow (1962) in the competitive case and by Kamien and Schwartz (1982) in the oligopolistic case. The variations in mark-up will be even more complex if the firm takes into consideration intertemporal dependence in demand or production economies (cf. Tirole, 1988, ch. 1.1.2).

Notice that the production function can be rewritten

$$r_{it} = \varepsilon x_{it} + (r_{it} - q_{it}) + u_{it}^q \tag{10}$$

It follows that the probability limit for the OLS estimator of the scale elasticity, based on deflated sales ($r_{it} = q_{it} + p_{it} - p_t$; p_t is the deflator) as a proxy for real output, can expressed as

$$\begin{aligned}\operatorname{plim}(\hat{\varepsilon}) &= \operatorname{plim}\left[\frac{\sum_{i,t} x_{it} r_{it}}{\sum_{j,s} x_{js}^2}\right] \\ &= \varepsilon + \operatorname{plim}\left[\frac{\sum_{i,t} x_{it}(r_{it} - q_{it})}{\sum_{j,s} x_{js}^2}\right] \\ &\quad + \operatorname{plim}\left[\frac{\sum_{i,t} x_{it} u_{it}^q}{\sum_{j,s} x_{js}^2}\right]\end{aligned} \tag{11}$$

where the sums in the numerator and denominator should be carried out over all observations, both across firms and over time. The probability limit is taken with respect to the number of observations.[12] In this expression, let us first focus on the term involving the discrepancy between deflated sales and real output, which we will term the '*omitted price bias*' (since it can be thought of as caused by omitting the individual price from the production function regression). This bias term will be denoted by Δ_1. The last term on the right-hand side of equation (11) is the '*transmission bias*', due to the correlation between the variable input(s) and the productivity shock. This term will be denoted by Δ_2.

After some algebra,[13] one finds that 'the omitted price bias' can be expressed as

$$\begin{aligned}\Delta_1 &\equiv \operatorname{plim}\left[\frac{\sum_{i,t} x_{it}(r_{it} - q_{it})}{\sum_{j,s} x_{js}^2}\right] \\ &= \frac{\varepsilon\eta(\sigma_w^2 + \sigma_m^2) + (\eta + 1)\sigma_q^2 + (1 - \varepsilon)\sigma_d^2}{\eta^2(\sigma_w^2 + \sigma_m^2) + (\eta + 1)^2\sigma_q^2 + \sigma_d^2}\end{aligned} \tag{12}$$

where we have assumed independent changes in factor price, productivity, price setting and demand for each firm. σ_w^2, σ_m^2, σ_q^2 and σ_d^2 are the variances of the movements in factor prices, the mark-ups, productivity and demand.

Notice first that as the elasticity of substitution (represented by $-\eta$) between the differentiated goods in the industry tends to infinity, the bias from neglecting the price differences will vanish. This situation corresponds to the standard case with no product differentiation. The important

[12] In practice, we might want to allow for a non-trivial time structure of the productivity shocks, by, for example, incorporating time dummies into the equation. In that case, the relevant limit to consider is with respect to an increasing number of firms.

[13] See the working paper version of this paper for details.

point is that if there is no horizontal product differentiation in the industry, there is no scope for differences in *quality-adjusted* prices. In this case, differences in sales corresponds to differences in inputs (costs), so sales is a valid measure for quality-adjusted output.[14] This is essentially a perfectly competitive situation.

Equation (12) shows that if the demand shocks dominate, the bias will be $1-\varepsilon$. That is, pure demand shifts will bias the scale estimate towards one. Below, we will show that this results holds more generally when we take into consideration that productivity shocks are 'transmitted' to inputs.

Only if there are both decreasing returns to scale *and* the demand shocks dominate will there be an *upward* bias in the scale estimate. If there are increasing returns to scale, demand shocks will also contribute to a downward bias in the scale elasticity. Presence of the various supply shocks will all tend to bias the scale elasticity downwards as we argued in Section 2. Notice that with a fairly high elasticity of substitution between the goods in the industry, or a scale elasticity close to one, the magnitude of the inconsistency due to demand shocks will tend to be dwarfed by the supply shocks.

Assuming orthogonality between the various shocks as before, one can derive the following expression for the 'transmission bias':

$$\Delta_2 \equiv \operatorname{plim}\left[\frac{\sum_{it} x_{it} u_{it}^q}{\sum_{ls} x_{ls}^2}\right] = \frac{(\eta+1)^2[\eta/(\eta+1)-\varepsilon]\sigma_q^2}{\eta^2(\sigma_w^2+\sigma_m^2)+(\eta+1)^2\sigma_q^2+\sigma_d^2} \tag{13}$$

By the second-order condition for profit maximization, we know that the expression inside the square brackets in the numerator is always positive. Hence, Δ_2 is always positive; the 'omitted price bias' and and 'transmission bias' will tend to pull in opposite directions. If we insert the expressions for Δ_1 and Δ_2 into equation (11), we find

$$\operatorname{plim}(\hat{\varepsilon}) = 1-\eta(\eta+1)\left(\frac{\eta}{(\eta+1)}-\varepsilon\right)\frac{\sigma_w^2+\sigma_m^2}{\eta^2(\sigma_w^2+\sigma_m^2)+(\eta+1)^2\sigma_q^2+\sigma_d^2} \tag{14}$$

There are at least three implications from this expression which are worth noting. First, keeping in mind the second-order condition for profit maximization, it is clear that $\operatorname{plim}(\hat{\varepsilon}) \leqslant 1$. Whatever the true magnitude of the scale elasticity, the OLS estimate will converge to a value below one. Second, this expression also reveals that if σ_q^2 or σ_d^2 are much larger than $\sigma_w^2+\sigma_m^2$, then $\operatorname{plim}(\hat{\varepsilon})=1$. That is, if the cross-sectional differences in factor price shocks and changes in price cost margins are negligible relative to the differences in the demand or the productivity shocks, the scale estimator $\hat{\varepsilon}$ will have a probability limit equal to one, whatever the value of the true scale elasticity. Lastly, if $\eta \longrightarrow -\infty$ we find that $\operatorname{plim}(\hat{\varepsilon})=(1-\theta)+\theta\varepsilon$, where θ is $[1+\sigma_q^2/(\sigma_w^2+\sigma_m^2)]^{-1}$. In industries with perfectly elastic demand, the ordinary production function regression provides a scale estimate with a probability limit which converges to a weighted average of one and the true scale coefficient.

[14] This argument has been used to justify quality adjustment of an input price on the basis of changes in the cost of producing the input. See the discussion between Gordon and Triplett in Foss (1982, chs 4 and 5).

4. TOWARDS CONSISTENT ESTIMATES OF THE SCALE ELASTICITY

In this section we present one tentative approach to the estimation of the scale elasticity from regressions based on deflated sales, given that real output is not observable. We make two major points. The first is that the parameters in the augmented production function regressions will be reduced-form parameters, which are mixtures of supply and demand parameters. Second, we argue that—under some assumptions—by adding growth in industry output to the production or the cost function for each firm we can identify the structural parameters of interest.

4.1. Common Estimating Equations

We start by stating the estimating equations considered in our empirical analysis. The models considered are two versions of the Cobb–Douglas production function and a (short-run) cost function. All these estimated models provide inconsistent estimates of the scale elasticity when deflated sales are used to replace output.

The estimating equation for what we will call the four-parameter (Cobb–Douglas) production function is

$$q_{it} = \beta_0^q + \beta_1^q(m_{it} - l_{it}) + \beta_2^q(e_{it} - l_{it}) + \beta_3^q l_{it} + \beta_4^q k_{it} + u_{it}^q \tag{15}$$

The variables m_{it}, l_{it}, e_{it} and k_{it} refer to the four factors of production—materials, labour, energy and capital. Let us emphasize that the transformation of the variables in equation (15) imply that β_3^q corresponds to the scale elasticity of all the (short-run) variable factors (labour, energy and materials).

We have also considered a different specification of this model which relates the growth in output to a cost-weighted input measure, growth in capital inputs and a residual:

$$q_{it} = \tilde{\beta}_0^q + \tilde{\beta}_1^q x_{it} + \tilde{\beta}_2^q k_{it} + \tilde{u}_{it}^q \tag{16}$$

where x_{it} is an index of the (short-run) variable factors of production, defined as $\sum_{j\in\{L,M,E\}} s_{it}^j x_{it}^j$ with s_{it}^j as the cost share of factor j. We will refer to this equation as the two-parameter production function.

By reshuffling the left- and the right-hand variables in equation (16), we obtain essentially a short-run cost function with predetermined capital. The estimating equation in this case is[15]

$$x_{it} = \beta_0^x + \beta_1^x q_{it} + \beta_2^x k_{it} + u_{it}^x \tag{17}$$

Clearly, the inverse of β_1^x in equation (17) is the scale elasticity of variable inputs. Models (17) and (16) are equivalent except, possibly, with respect to stochastic assumptions.

One set of our estimates follow common practice and use deflated sales (r_{it}) as our proxy for real output (q_{it}) in equations (15), (16) and (17).

4.2. Reduced-form Models

The models presented above suffer from the fact that when deflated sales replace the unobserved output variable in the estimation the omitted price variable will be buried in the residual. As

[15] Alternatively, we could have estimated a system of restricted factor demand functions for each of the variable inputs. The approach taken here highlights the similarity, as well as the difference in terms of stochastic specification, between the production function and the factor-demand approach to production econometrics. Estimating a system of factor-demand equations introduces a number of additional parameters and requires exogenous factor price variations, which are usually not available (at the plant level).

pointed out earlier, the problem is not easily solved by an appropriate choice of instrumental variables, since variables which are correlated with the inputs or output (i.e. potentially useful instruments) will essentially always be correlated with this omitted price variable.

Our approach to identification uses the idea that with an appropriate specification of the demand system, the omitted price variable can be expressed in terms of the firm's output growth relative to industry output, and eventually in terms of observables and parameters already present in the production (or cost) function. With such an expression we can substitute out the omitted price variable.

We will now make this idea more precise. Assume that the demand facing the individual firm is given by $Q_{it} = Q_{It}(P_{it}/P_{It})^{\eta} e^{\tilde{u}^d_{it}}$. That is, demand is given by industry demand and the firm's market share which is determined by the substitution effects across products within the industry, captured by the term $(P_{it}/P_{It})^{\eta}$. We have also allowed for an idiosyncratic part $(\tilde{u}^d_{it})$.[16] P_{it} is the firm-specific price, while P_{It} is the average price in the industry. This demand system in growth terms can be written[17]

$$q_{it} = q_{It} + \eta(p_{it} - p_{It}) + u^d_{it} \tag{18}$$

Using $r_{it} = q_{it} + p_{it} - p_{It}$, it follows that $r_{it} = q_{It} + (\eta + 1)(p_{it} - p_{It}) + u^d_{it}$. Now assume that we can find weights (ν_{it}) such that $\sum_{i\in I} \nu_{it} p_{it} = p_{It}$. That is, since the growth in the deflator (p_{It}) is constructed on the basis of the growth in the individual prices (p_{it}), the task is to identify the weights used to construct the deflator. In the empirical analysis we assume that these weights are given by the plants' market shares. It follows that $\sum_{i\in I} \nu_{it} r_{it} = q_{It}$, if we drop the disturbance term. The important point is that growth in industry output can be estimated by the (weighted) average growth in deflated revenues.

Using equation (18), we have that

$$p_{it} - p_{It} = \frac{1}{\eta + 1}(r_{it} - q_{It} - u^d_{it}) \tag{19}$$

The left-hand side of this expression is the 'omitted price variable' which caused the problems we have discussed above, while the variables on the right-hand side are observable except the disturbance term. With this expression, we can rewrite equations (15), (16) and (17) with deflated sales as the dependent variable and now without the problem due to the omitted price variable.

Combining equations (19) and (15), we get

$$r_{it} = \beta^q_0 + \frac{\eta + 1}{\eta}[\beta^q_1(m_{it} - l_{it}) + \beta^q_2(e_{it} - l_{it}) + \beta^q_3 l_{it} + \beta^q_4 k_{it}] - \frac{1}{\eta} q_{It} + v^q_{it} \tag{20}$$

v^q_{it} is an error term capturing both idiosyncratic demand and productivity shocks. Similarly, for equation (16):

$$r_{it} = \tilde{\beta}^q_0 + \frac{\eta + 1}{\eta}(\tilde{\beta}^q_1 x_{it} + \tilde{\beta}^q_2 k_{it}) - \frac{1}{\eta} q_{It} + \tilde{v}^q_{it} \tag{21}$$

Lastly, combining equations (17) and (19):

$$x_{it} = \beta^x_0 + \frac{\eta\beta^x_1}{\eta + 1} r_{it} + \beta^x_2 k_{it} + \frac{\beta^x_1}{\eta + 1} q_{It} + v^x_{it} \tag{22}$$

[16] There is no problem in allowing for a fixed effect in u^d_{it}, as our estimating equations are in growth rates.
[17] With $u^d_{it} \equiv \tilde{u}^d_{it} - \tilde{u}^d_{i,t-1}$.

In the models presented in equations (20)–(22) growth in *industry* output enters the equation at firm level.[18] Adding this industry variable to the estimating equations ensures (in principle) identification of both the demand elasticity and the production (cost) function parameters of interest. Notice that these equations were derived without assuming profit maximization. Neither did we put constraints on the form of the *industry* demand function, only on the market-sharing rule. The essential and disputable assumption in the derivation of equations (20)–(22) is the symmetric market sharing rule—that the substitution effects are captured by the average price and not the prices of individual competitors.

4.3. The Data Sources and Variable Construction

The data source used in this analysis is the annual census carried out by Statistics Norway. Aggregate numbers and definitions for the census are reported in *Manufacturing Statistics* published annually by Statistics Norway (see also Halvorsen, Jensen and Foyn, 1991). Our unit of observation is a plant. We have employed an unbalanced sample of annual observations for the period 1983–9 (inclusive) for four industry sectors: 'Metal products, except machinery and equipment' (ISIC 381), 'Machinery' (ISIC 382), 'Electrical apparatus and supplies' (ISIC 383), and 'Transport equipment' (ISIC 384). The sample includes only establishments with at least five employees. Plants with incomplete reports for the variables needed in the estimation have been eliminated, but no other cleaning has been carried out.

If we consider, for example the year 1986, the number of plants vary from 196 in 'Electrical equipments' (ISIC 383) to 663 in 'Metal products' (ISIC 381), while the corresponding number of firms are 177 and 629 respectively. That is, almost all plants belong to separate firms. The average number of observations per plant is about four in all four industries. Summary statistics for a number of variables in the employed sample are reported in Table I.

All variables are expressed in terms of growth rates (more precisely, in first differences of logarithms of the variables). In all models, we have applied a Tornquist index for growth in the variable inputs, i.e. the shares are constructed as the average share for the two years used to construct the growth rates.[19] Our 'output' measure is nominal output deflated by the output deflator for the industry. As discussed at length above, this is not a measure of real output and it is the source of the biases we have emphasized throughout this paper. Nominal output is adjusted for duties and subsidies, and (in principle) also for changes in inventories in our data. Price deflators for gross production (at seller prices), materials, energy and capital (at buyer prices) are taken from the Norwegian National Accounts. Labour inputs are represented by man-hours. Wage payments comprise salaries and wages in cash and kind, other benefits for the employees, taxes and social expenses levied by law. Energy use and material inputs are reported as separate variables in our data set.

The capital input variable employed is based on investment figures and the total reported fire insurance value for buildings and machinery. The annual movements are obtained by assuming geometric depreciation at a 3 per cent annual rate, and that investment takes about a year to become productive.[20] An examination of the fire insurance values and a comparison with the investment figures reveal much noise in the fire insurance values. We have therefore constructed

[18] See Pakes (1983) for an analysis of macro variables in micro relations in a different context.

[19] Cf. the shares (s_{it}^{j}) used to construct x_{it} in equations (16), (17), (21) and (22).

[20] This last assumption is imposed also to reduce biases caused by the possibility that investments respond to contemporaneous productivity shocks. The timing of the investment turned out to be unimportant for the estimates reported below.

Table I. Summary statistics for the four industry groups in the applied sample. Period 1983–9

Industry	381		382		383		384	
Variable	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Output[a]	0·005	0·326	0·015	0·363	0·013	0·340	−0·001	0·393
TFP[a]	−0·010	0·149	0·004	0·180	−0·003	0·142	0·003	0·218
Labour[a]	−0·004	0·316	−0·009	0·331	0·004	0·271	−0·010	0·365
Capital[a]	0·036	0·351	0·023	0·296	0·038	0·182	0·016	0·138
Revenue/variable cost	1·124	0·142	1·102	0·173	1·104	0·135	1·092	0·115
Labour share[b]	0·431	0·141	0·441	0·165	0·393	0·131	0·387	0·165
Energy share[b]	0·025	0·022	0·019	0·018	0·013	0·011	0·019	0·017
#Empl.	30·8	50·4	73·4	195·0	76·0	175·0	58·5	109·9
#Obs.	3966		2560		1182		2454	

Notes:
Industry 381: Manufacture of metal products, except machinery and equipment.
Industry 382: Manufacture of machinery.
Industry 383: Manufacture of electrical apparatus and supplies.
Industry 384: Manufacture of transport equipment.

[a] Annual growth rates.
[b] Shares in variable costs.

a simple filter to pool the two sources of information about movements in the capital stock. We pooled all the information in the fire insurance values and the investment figures, using the perpetual inventory method, to estimate the level of the capital stock for the first observation for each plant.[21] From this initial level, the capital stocks in the subsequent years are estimated by the perpetual inventory method.

The growth in industry output is estimated by $\sum_{i\in I} \nu_{it}r_{it}$, according to the formula presented in Section 4.2. We have used the share of industry sales as weights (ν_{it}).[22] This seems to be a reasonable approximation to the weights used to estimate the output deflator in the National Accounts, given that it is not possible to identify the actual weights.

4.4. Other Econometric Issues

For both the cost functions (cf. equations (17) and (22)) and the two-parameter production functions (cf. equations (16) and (21)) we have both carried out OLS and instrumental variable regressions. Changes in the number of employees and the capital stock are used as our instrumental variables.[23] The motivation for using an IV approach is twofold. First, it seems likely that the number of employees is less responsive to short-term changes in productivity and demand, as compared to man-hours, materials and energy in the production function. Similarly, in the cost

[21] Extreme fire insurance values have been eliminated.
[22] More precisely, when estimating the growth in industry output from year $t-1$ to t, we have weighted together the growth in deflated sales for all plants in the industry, with the average output shares for the plant in the two years as weights.
[23] We could also have used lagged variables as instruments, as suggested by one referee. This approach has been extensively explored in Klette (1994). The problem with this approach is that there is not much identifying power in past changes for current changes. And as is well known, with weak instruments the IV estimator is very vulnerable to even weak correlation between the error term and the instruments.

function case we believe that the number of employees is less affected by temporary shifts in productivity and demand as compared to deflated sales. If these claims are true, the orthogonality assumption with the error term is a better approximation when using changes in employees as an instrument. Notice that permanent differences in productivity between firms will not cause any simultaneity bias, as all variables are expressed in terms of (annual) growth rates.[24]

Second, measurement errors in the input index will cause a bias in the OLS regression. Measurement errors might be introduced if we are using weights which randomly deviate from the correct shares when constructing the input index, as well as for a variety of other reasons. If these measurement errors are uncorrelated with changes in employment, using the IV approach will remove the bias due to errors of measurement in the input index.

4.5. The Empirical Results

Table II reports the results obtained from estimating the production function in equation (15), as well as a model where the production function is augmented by adding industry output as an explanatory variable. The main finding in Table II is that industry output is a highly significant variable. However, rather than elaborating on these estimates, we will notice that the two-parameter production function spelled out in equation (16) performs uniformly better in terms of the RMSE[25] (see Tables III–VI). Also, the scale estimates obtained by the two-parameter model are likely to be more consistent, as we are able to instrument the variable inputs (man-hours, energy, materials) which are liable to be correlated with high-frequency movements in productivity and demand buried in the residual. In what follows we will concentrate on the two-parameter model, instead of the four-parameter production function model. When we refer to the production function below, we mean the two-parameter version (cf. equations (16) and (21)).

Tables III–VI summarize the main results for each industry group separately. The first two columns in each of the four tables refer to estimation of the production function (16). Columns 3 and 4 present results obtained from the cost function (cf. equation (17)), which is essentially the model as in columns 1 and 2, but with a reshuffling of the left-hand side and right-hand side variables. The last two columns provide results from models augmented by including the industry output variable into the regression. One should notice that the reported RMSE values refer to the root mean square prediction error of deflated sales, rather than the prediction error of the different dependent variables. This is done in order to facilitate comparisons between the various (non-nested) models.

If we consider the first column in Table III we find an estimate of the scale elasticity suggesting significant decreasing returns (0·907). Similar estimates are reported in Tables IV–VI for the other industries. Such decreasing returns to variable inputs (and often stronger decreasing returns) are regularly found in panel data estimation of production functions, as we discussed in the introduction. Column 2 in Table III reports the IV estimates, which are quite similar to the OLS estimates in column 1. Column 3, however, suggests a quite different view of short-run scale economies; recall that the coefficient for deflated sales (the measure of 'output') is equal to the inverse of the short-run scale elasticity if we neglect the omitted price bias for the moment. Our estimate in column 3 suggests that the short-run scale elasticity is 1·13 ($=1/0{\cdot}883$). In column 4, where output is treated as endogenous, the result is entirely different, and, not surprisingly, very similar to the estimates from the production

[24] Furthermore, productivity movements which are not predictable before the inputs are chosen will not cause any bias. Cf. Zellner, Kmenta and Dreze (1966).

[25] Notice that the two models (equations (15) and (16)) are not nested, so traditional testing procedures are not appropriate.

function in columns 1 and 2. As suggested in Section 4, there are two reasons why we will place greater trust in the (higher) IV estimate of the coefficient on deflated sales. First, random measurement errors in deflated sales will cause a downward bias in the estimated sales coefficient (and thereby an upward bias in the scale elasticity).[26] Second, a similar bias will arise due to the correlation between sales and the cost function error term, which incorporates the idiosyncratic shifts in productivity and demand. As discussed above, our instruments are likely to remove at least some of these biases.

We have argued in the previous sections that the coefficients reported in columns 1 to 4 are inconsistent estimates of the (short-run) scale elasticity since they are based on deflated sales rather than real output. In particular, from our discussion in Sections 2 and 3 we would expect the scale elasticity to be larger than our 'preferred estimates' (around 0·93). We have used quotation marks to indicate that we consider these estimates to be inconsistent, as they do not account for the bias due to the omitted price variable.

In Section 4.2 we showed how the true scale elasticity can be identified by adding growth in industry output to the models in columns 1–4, which we have done in the models presented in columns 5 and 6. The industry growth variable is highly significant when introduced into the regressions, suggesting that the firms do not face perfectly elastic demand curves. The estimated coefficients for the other variables in the model are almost unaffected by the introduction of the

Table II. Production function estimates with and without industry output included in the regression. Cf. equations (15) and (20). All industries

Industry	381		382		383		384	
Labour	0·854 (0.009)	0·843 (0·009)	0·814 (0·014)	0·800 (0·014)	0·934 (0·018)	0·925 (0·018)	0·845 (0·011)	0·838 (0·011)
Materials per unit of labour	0·470 (0·006)	0·465 (0·006)	0·400 (0·008)	0·394 (0·008)	0·502 (0·012)	0·500 (0·012)	0·554 (0·007)	0·550 (0·007)
Energy per unit of labour	0·060 (0·005)	0·056 (0·005)	0·070 (0·009)	0·065 (0·009)	0·024 (0·011)	0·021 (0·011)	0·085 (0·008)	0·083 (0·008)
Capital	−0·005 (0·005)	−0·004 (0·005)	−0·013 (0·008)	−0·011 (0·008)	−0·022 (0·022)	−0·020 (0·022)	0·004 (0·004)	0·004 (0·003)
Industry output		0·233 (0·030)		0·282 (0·040)		0·112 (0·034)		0·172 (0·035)
Intercept	−0·009 (0·003)	−0·009 (0·003)	0·007 (0·004)	−0·003 (0·004)	−0·002 (0·005)	−0·002 (0·005)	−0·002 (0·004)	−0·005 (0·004)
R^2	0·75	0·75	0·68	0·69	0·78	0·78	0·80	0·80
RMSE	0·163	0·162	0·206	0·204	0·161	0·160	0·178	0·177

Notes:
Standard errors in parentheses.
Industry 381: Manufacture of metal products, except machinery and equipment.
Industry 382: Manufacture of machinery.
Industry 383: Manufacture of electrical apparatus and supplies.
Industry 384: Manufacture of transport equipment.

[26] This bias was pointed out by Friedman (1955), and was labeled the 'regression fallacy'. Friedman's discussion was one of the first examinations of the consequences of random measurement errors in the regressors in an econometric context.

Table III. Estimation results for the models in equations (16), (17), (21) and (22). Industry 381: Manufacture of metal products, except machinery and equipment

Model	Basic models				Augmented models	
Dependent variable (eq.)	Deflated sales (16)		Variable inputs (17)		Deflated sales (21)	Variable inputs (22)
Estimation method	OLS	IV	OLS	IV	IV	IV
Variable inputs	0·907 (0·007)	0·930 (0·010)			0·923 (0·010)	
Deflated sales			0·883 (0·007)	1·075 (0·012)		1·083 (0·012)
Capital	−0·005 (0·004)	−0·005 (0·004)	0·007 (0·004)	0·006 (0·005)	−0·005 (0·004)	0·005 (0·005)
Industry output					0·161 (0·027)	−0·174 (0·030)
Intercept	−0·007 (0·002)	−0·007 (0·002)	0·008 (0·002)	0·007 (0·002)	−0·007 (0·002)	0·007 (0·003)
RMSE	0·146	0·146	0·163	0·146	0·145	0·145

Notes:
Numbers in parentheses are standard errors. RMSE refers the root mean square prediction error of deflated sales as implied by the model.

Table IV. Estimation results for the models in equations (16), (17), (21) and (22). Industry 382: Manufacture of machinery

Model	Basic models				Augmented models	
Dependent variable (eq.)	Deflated sales (16)		Variable inputs (17)		Deflated sales (21)	Variable inputs (22)
Estimation method	OLS	IV	OLS	IV	IV	IV
Variable inputs	0·865 (0·010)	0·912 (0·015)			0·905 (0·015)	
Deflated sales			0·871 (0·010)	1·096 (0·018)		1·105 (0·019)
Capital	−0·007 (0·007)	−0·006 (0·007)	0·003 (0·007)	0·007 (0·007)	−0·005 (0·007)	0·006 (0·007)
Industry output					0·167 (0·036)	−0·185 (0·041)
Intercept	0·005 (0·004)	0·005 (0·004)	−0·002 (0·004)	−0·005 (0·004)	−0·005 (0·007)	0·001 (0·004)
RMSE	0·180	0·181	0·207	0·181	0·180	0·180

Notes:
See Table III.

Table V. Estimation results for the models in equations (16), (17), (21) and (22). Industry 383: Manufacture of electrical apparatus and supplies

Model	Basic models				Augmented models	
Dependent variable (eq.)	Deflated sales (16)		Variable inputs (17)		Deflated sales (21)	Variable inputs (22)
Estimation method	OLS	IV	OLS	IV	IV	IV
Variable inputs	0·957 (0·013)	0·989 (0·019)			0·981 (0·019)	
Deflated sales			0·850 (0·012)	1·011 (0·020)		1·020 (0·020)
Capital	−0·023 (0·020)	−0·022 (0·020)	0·015 (0·019)	0·023 (0·020)	−0·021 (0·020)	0·021 (0·020)
Industry output					0·084 (0·031)	−0·085 (0·032)
Intercept	−0·002 (0·004)	−0·003 (0·004)	0·005 (0·004)	0·003 (0·004)	−0·002 (0·004)	0·002 (0·004)
RMSE	0·147	0·148	0·163	0·148	0·147	0·147

Notes:
See Table III.

Table VI. Estimation results for the models in equations (16), (17), (21) and (22). Industry 384: Manufacture of transport equipment

Model	Basic models				Augmented models	
Dependent variable (eq.)	Deflated sales (16)		Variable inputs (17)		Deflated sales (21)	Variable inputs (22)
Estimation method	OLS	IV	OLS	IV	IV	IV
Variable inputs	0·913 (0·009)	0·919 (0·013)			0·912 (0·013)	
Deflated sales			0·903 (0·008)	1·088 (0·015)		1·096 (0·016)
Capital	0·005 (0·003)	0·005 (0·003)	−0·005 (0·003)	−0·005 (0·004)	0·005 (0·003)	−0·006 (0·003)
Industry output					0·139 (0·032)	−0·152 (0·036)
Intercept	0·001 (0·003)	0·001 (0·003)	−0·002 (0·003)	−0·002 (0·004)	−0·001 (0·003)	0·001 (0·004)
RMSE	0·165	0·165	0·182	0·165	0·164	0·164

Notes:
See Table III.

new variable. Our estimates of the (short-run) scale elasticity according to the models in columns 5 and 6 are, however, substantially different from 'our preferred estimates' reported above. The reason is that the interpretations, rather than the estimated values of the coefficients, are different. That is, considering equations (21) and (22), we can identify the demand elasticity from columns 5 and 6 in Table III. Simple calculations show that the two models implies a demand elasticity around −6·2. We can also identify the scale elasticity from the two set of estimates, which turn out to be about 1·10 in for both models.[27]

The results for the other industries are quite similar to the findings discussed above (cf. Tables IV–VI). The growth in industry output is significant in all the four industries we have considered. Our results suggest significant scale economies to variable inputs in all industries. The final point estimates of the scale and demand elasticities are reported in Table VII. The results suggest that the firms face a less than perfectly elastic demand curve, with a demand elasticity around −6 to −12.

A few comments about the other coefficients in these models might also be in place. Consider the intercept terms in Table III; they suggest an average productivity decline around −1 per cent per year (same as the reported TFP figure in Table I). The estimated coefficient for capital growth has the wrong sign, but is not significant. The insignificance of capital is a result which persists throughout our estimations. Our belief is that with the available quality of our capital measures it is not possible to trace the annual changes in capital services. We will add further comments on this problem in the final section of the paper.

We recognize that our interpretation of these results are not very robust, and that alternative interpretations of our findings can be put forward. The industry output variable can be significant as this variable acts as a proxy for other omitted variables. At least two candidates suggest themselves: factor utilization and year- and industry-specific changes in productivity. Our capital variable is not adjusted for utilization. We believe that this is likely to create an upward bias in the coefficients on variables' inputs in the production function case and a downward bias in the output coefficient in the cost function. Changes in factor utilization might, however, also create an upward bias to the coefficient on growth in industry output. Productivity changes are assumed to be captured by the constant term and the residual in our model. This is not entirely comforting, and biases might be present in the estimated coefficients. Time dummies can be added to the equation to capture productivity shocks, but that leaves little scope for estimating the effect of changes in industry output, as we confirmed by some experimental estimations. In future work we believe that adjustments for changes in factor utilization could be pushed

Table VII. Point estimates of the demand and scale elasticities as implied by the results in Tables III–VI, columns 5 and 6

Model	Semi-parametric model		Factor-demand model	
Industry	Demand	Scale	Demand	Scale
381	−6·2	1·10	−6·2	1·10
382	−6·0	1·09	−6·0	1·09
383	−11·9	1·07	−12·0	1·07
384	−7·2	1·06	−7·2	1·06

[27] In fact, we could have dropped column 5 or 6, as these two models are essentially identical. We have reported both for completeness.

forward by exploiting the information in variables such as energy use and hours per worker. Adjustment for productivity changes based on observables rather than time and industry dummies might be more difficult.

5. ADDITIONAL COMMENTS

This paper has identified some problems of interpretation of the production function parameters when deflated sales is used instead of real output as the dependent variable. The problems arise in situations where the firms compete in an environment with differentiated products and imperfect competition, where prices will reflect idiosyncratic differences in cost. The basic insight is that a firm experiencing a cost improvement beyond the industry average will be inclined to undercut its competitor's price and thereby expand its market share. However, since the firm will have a relative decline in price, the expansion in sales will be less than proportional to the growth in output. It follows that replacing changes in real output by growth in deflated sales will introduce a bias in the standard approaches to estimation of production parameters. The main contribution of this paper is to present an analysis of the nature and consequences of this bias for widely applied estimators of scale economies.

Our paper also provides some empirical results from production and cost function models, augmented by adding growth in industry output to the estimating equations. By adding this variable, we demonstrate how scale economies and demand elasticities can be identified from the augmented estimating equations. Different specifications of the production model are examined. Our findings suggest that the firms in the industries we examine face downward-sloping demand curves with significant but moderate price elasticities. The results also reveal some scale economies.

One problem with our findings is the non-significance of capital in our empirical models. As the equations are in first differences, this result is consistent with the general experience with these kind of data. The usual explanation is simple. The quality of our capital variable is too poor to identify the annual variations in capital services. Two key problems in this respect are variations in capacity utilization and investment lags. These issues raise the question of how we should incorporate capital into our models. Does capital have a positive (shadow) price only when the firm operates at full capacity? And closely related; how does the firm incorporate the shadow price of capital into its pricing decisions? We believe that a more satisfactory solution to these questions requires an explicit dynamic model of investment behaviour, incorporating uncertainty. In our estimates, we suspect that changes in the capital services are picked up by movements in the variable inputs (particularly energy) in our production function regressions, and by the deflated sales variable in the cost function regressions. In this interpretation, our scale estimates are closer to the long run rather than the short-run scale elasticities.

ACKNOWLEDGEMENTS

Comments on an earlier draft from Jacques Mairesse, M. Hashem Pesaran and seminar participants at NBER and the Tel Aviv University are gratefully acknowledged. The paper has also benefited from the comments of two anonymous referees. Klette's research on this project has been financed by grants from the Norwegian Research Council. Griliches' work has been supported by the NSF and the Bradley Foundation. This research project was carried out while Klette was visiting the NBER.

REFERENCES

Abbott, T. A., Z. Griliches and J. A. Hausman (1988), 'Short run movements in productivity: Market power versus capacity utilization', Unpublished paper.

Arrow, K. J. (1962), 'Economic welfare and the allocation of resources for invention', in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton University Press, Princeton, NJ.

Bartelsman, E. J., R. J. Caballero and R. K. Lyons (1991), 'Short and long run externalities', NBER Working Paper No. 3810.

Berry, S. T. (1994), 'Estimating discrete-choice models of product differentiation', *Rand Journal of Economics*, **25**, 242–62.

Biørn, E. and T. J. Klette (1994), 'Errors in variables and panel data: The labour demand response to permanent changes in output', Discussion Paper 125, Statistics Norway.

Dunne, T. and M. J. Roberts (1992), 'Costs, demand, and imperfect competition as determinants of plant-level output prices', Discussion Paper CES 92–5, Center for Economic Studies, US Bureau of the Census.

Friedman, M. (1955), 'Comment on 'Survey of the Empirical Evidence on Economies of Scale' by C. A. Smith', in *Business Concentration and Price Policy, A Report of the National Bureau of Economic Research*, Princeton University Press, Princeton, NJ.

Foss, M. F. (1982), *The U.S. National Income and Product Accounts*, University of Chicago Press, Chicago, IL.

Griliches, Z. (1957), 'Specification bias in estimates of production functions', *Journal of Farm Economics*, **39**, 8–20.

Griliches, Z. and J. A. Hausman (1986), 'Errors in variables in panel data', *Journal of Econometrics*, **18**, 93–118.

Griliches, Z. and J. Mairesse (1995), 'Production functions: The search for identification', NBER Working Paper no. 5067.

Griliches, Z. and V. Ringstad (1971), *Economies of Scale and the Form of the Production Function*, North-Holland, Amsterdam.

Hall, R. E. (1990), 'Invariance properties of Solow's productivity residual', in P. Diamond (ed.), *Growth, Productivity, Unemployment*, MIT Press, Cambridge, MA.

Halvorsen, R., R. Jensen and F. Foyn (1991), 'Dokumentasjon av Industristatistikkens Tidsseriebase (Documentation of the Panel Data Base for Manufacturing)', Mimeo, Statistics Norway, Oslo (in Norwegian).

Kamien, M. I. and N. L. Schwartz (1982), *Market Structure and Innovation*, Cambridge University Press, Cambridge.

Klette, T. J. (1990), *Four Essays on Industrial Policy under Imperfect Competition*, Økonomiske Doktoravhandlinger No. 6, Department of Economics, University of Oslo.

Klette, T. J. (1994), 'Estimating price–cost margins and scale economies from panel data', Discussion Paper, Statistics Norway.

Mairesse, J. (1990), 'Time-series and cross-sectional estimates on panel data: Why are they different and why should they be equal?', in J. Hartog, G. Ridder and J. Theeuwes (eds), *Panel Data and Labour Market Studies*, North-Holland, Amsterdam.

Marschak, J. and W. Andrews (1944), 'Random simultaneous equations and the theory of production', *Econometrica*, **12**, 143–205.

Mundlak, Y. (1978), 'On the pooling of time-series and cross-section data', *Econometrica*, **46**, 49–85.

Mundlak, Y. and I. Hoch (1965), 'Consequences of alternative specifications of Cobb–Douglas production functions', *Econometrica*, **33**, 814–28.

Pakes, A. (1983), 'On group effects and errors in variables in aggregation', *Review of Economics and Statistics*, **64**, 168–73.

Ringstad, V. (1971), *Estimating Production Functions and Technical Change from Micro Data*, Social and Economic Studies 21, Statistics Norway, Oslo.

Ringstad, V. (1974), 'Some empirical evidence on the decreasing scale elasticity', *Econometrica*, **42**, 87–101.

Tirole, J. (1988), *The Theory of Industrial Organization*, MIT Press, Cambridge, MA.

Zellner, A., J. Kmenta and J. Dreze (1966), 'Specification and estimation of Cobb–Douglas production functions', *Econometrica*, **34**, 784–95.