Estimating Firm-Level Productivity in Differentiated Product Industries

Marc J. Melitz *
Department of Economics
Harvard University
mmelitz@harvard.edu

First version: September 18, 2000
This version: December 15, 2000

THIS DRAFT IS PRELIMINARY
COMMENTS AND SUGGESTIONS ARE WELCOMED

*I would like to thank Susanto Basu for many helpful discussions. As usual, no responsibility for errors or other problems should be inferred.

1 Introduction

Productivity indexes the amount of output a firm can produce with a given set of inputs. Unfortunately, a firm's output is often not directly observable. Most empirical studies of firm level production use firms sales – deflated by a common industry price index – as a proxy for output. If the output produced by firms in an industry is a homogeneous good, then firm deflated sales will yield a "perfect" proxy for output. On the other hand, if the goods produced by different firms are even slightly differentiated, then firm level prices will fluctuate relative to the price index, breaking the link between firm deflated sales and output. The associated problems for the estimation of firm level production functions using the deflated sales proxy has been recognized since the work of Marschak and Andrews (1944). Surprisingly, this problem has largely been sidestepped in the subsequent literature on empirical production analysis. Much of this literature has been devoted to the estimation of firm productivity levels, obtained as residuals from an estimated production function based on the deflated sales proxy. The proxy problem is then either ignored – and the residuals directly interpreted as productivity – or it is mentioned as a disclaimer that the residuals inextricably combine measures of firm productivity and pricing policies. This paper shows that, using only data on firm deflated sales and input use, one can nevertheless obtain meaningful measures of firm productivity that are not tainted by fluctuations in the firm's price relative to the industry index. To achieve this, a restrictive demand structure is imposed. Nevertheless, this structure is less considerably less restrictive than the one implicitly imposed when using the sales proxy.

This paper strongly relies on the work of Klette and Griliches (1996), who develop methods to address the problems caused by the deflated sales proxy for firm production analysis in differentiated product industries. Whereas Klette and Griliches (1996) mainly focus on the measurement of the degree of returns to scale in production, I focus the current analysis on the obtainment and interpretation of productivity measures. I will show how the concept of productivity can be reinterpreted in a differentiated product industry where firms produce goods with different quality levels and where firms may also produce more than a single type of good. I will also show that studies that use deflated sales as the output proxy – without further adjustments – will obtain productivity measures that are spuriously pro-cyclical. I first analyze an estimation method that imposes some strong restrictions on the stochastic nature of productivity, quality and taste shocks in order to focus the discussion on the re-interpretation of the productivity estimates. I then show how these restrictions can be relaxed by adapting some recent econometric methods in order to address issues related to product differentiation, imperfect competition, and multi-product firms.

2 The Model

Firm Level Demand

I initially assume that firms produce a single type of good or variety. These varieties are symmetrically differentiated, with a common elasticity of substitution σ between any two varieties. The demand for each firm's output Q_i is generated by a representative consumer with utility:

$$U\left(\left(\sum_i (\Lambda_i Q_i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, Z\right), \tag{1}$$

where $U(.)$ is assumed to be differentiable and quasi-concave and Z represents aggregate industry demand shifters. Λ_i represents the consumer's valuation of firm i's product quality. Changes in Λ_i over time could come from two effects: the quality "embodied" in the good changes (the actual product is changing) or the consumer's idiosyncratic preferences across varieties change (the product remains unchanged, but the consumer's relative valuations change). By assumption, preference shifts that affect all varieties are captured by Z, so only product quality changes can induce aggregate changes in the Λ_is. Each firm's revenue $R_i = P_i Q_i$ is observable, but not its output Q_i. A price index $\tilde{P}$ that measures aggregate changes in the distribution of firm prices P_i and qualities Λ_i is also available. I will assume throughout this paper that firms are small enough relative to the industry that they have no power to influence the industry price index $\tilde{P}$.

If the goods were perfect substitutes (σ is infinite), then there can be no variations in quality adjusted prices across firms: $\frac{P_i}{\Lambda_i} = \tilde{P}$ for all firms. Deflated sales would then be a perfect proxy for the unobserved quality adjusted output, since then $\frac{R_i}{\tilde{P}} = Q_i \Lambda_i$ for all firms. On the other hand, any finite σ will give firms some flexibility to adjust their price relative to the index and the firms' deflated sales $\frac{R_i}{\tilde{P}}$ no longer yield accurate measures of quality adjusted output. Each firm then faces a downward sloping demand curve for its output that is summarized by the following inverse demand function:

$$Q_i = \Lambda_i^{\sigma-1}\left(\frac{P_i}{\tilde{P}}\right)^{-\sigma}\frac{1}{N}\left(\frac{R}{\tilde{P}}\right), \tag{2}$$

where $\frac{R}{\tilde{P}} = \frac{\sum_{i=1}^{N} R_i}{\tilde{P}}$ represents total industry deflated sales (N indexes the number of firms in the industry). (2) summarizes how, in a differentiated product industry, a firm's output is jointly determined by its product quality, its price relative to the industry index, the number of firms in the industry, and the aggregate industry sales. The price index $\tilde{P}$ that makes (2) the exact demand system for the consumer utility specified in (1) is

$$\tilde{P} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} P_i^{\sigma-1}\right)^{\frac{1}{\sigma-1}}}{\left(\frac{1}{N}\sum_{i=1}^{N} \Lambda_i^{\sigma-1}\right)^{\frac{1}{\sigma-1}}}.$$

It can be shown that a first order approximation for the percentage change in $\tilde{P}$ is obtained by taking a market share weighted average of the percentage changes in firm level prices and qualities. This is essentially the methodology used by the Bureau of Labor Statistics to construct industry price indices.

Firm Production and Revenues

I assume that the production technology is homogeneous of degree $\gamma > 0$.[1] Given this assumption, an aggregate input index X_i and factor price index W_i can be constructed such that $X_i = f(\boldsymbol{X_i})$, $W_i = h(\boldsymbol{W_i})$, and $X_i W_i = \boldsymbol{X_i} \cdot \boldsymbol{W_i}$ where $\boldsymbol{X_i}$ and $\boldsymbol{W_i}$ are the vectors of inputs and prices and $f(.)$ is linearly homogeneous.[2] Production as a function of the aggregate input index then takes the form:

$$Q_i = \Phi_i X_i^{\gamma}, \tag{3}$$

where γ indexes the degree of returns to scale and Φ_i represents firm level productivity.

Writing the demand and production functions (2) and (3) in logs (represented with lower case variables) and adding time subscripts yields:

$$q_{it} = \gamma x_{it} + \phi_{it}$$
$$q_{it} = -\sigma(p_{it} - \tilde{p}_t) + (\sigma - 1)\lambda_{it} + (r_t - \tilde{p}_t) - n_t.$$

[1] Most of the results developed in this paper remain valid with more general production technologies at the expense of a greater expositional burden.

[2] The factor price index is written with a firm subscript as it can be firm specific. Even with an integrated and competitive factor market, adjustment costs will induce differences in the shadow price of the input index across firms based on differences in current levels of the quasi-fixed factors.

Combining the demand function with the production function yields a "revenue production function" that does not contain the unobserved output variable:

$$
\begin{aligned}
r_{it} - \tilde{p}_t &= q_{it} + p_{it} - \tilde{p}_t \\
&= \tfrac{\sigma-1}{\sigma}\gamma x_{it} + \tfrac{1}{\sigma}\big[(r_t - \tilde{p}_t) - n_t\big] + \tfrac{\sigma-1}{\sigma}(\phi_{it} + \lambda_{it}).
\end{aligned} \tag{4}
$$

Note that firm revenue and cost only depend on the sum of the firm's productivity and quality index, $\phi_{it} + \lambda_{it}$, and do not depend on the decomposition of this term into a separate productivity and quality component. A one percent productivity gain affects a firm - in terms of its revenue and cost - in exactly the same way as a one percent quality gain. If a firm's output can not be observed, then it will be impossible to separately identify the effects of these two types of gains. Consider the following example of two firms competing in the video player (VCR) industry. One produces a VCR that plays VHS tapes while the other produces a VCR that plays DVD disks. Over time, firms may find ways of producing more VCRs with a given input bundle (both Φ_is increase), they may improve the quality of the VCRs (both Λ_is increase), or consumers may increasingly prefer DVD players as the availability of DVD rentals increases (the Λ_i for the DVD firm increases, matched by a proportional decrease in the Λ_i of the VCR firm). Assuming that these changes do not affect aggregate variables differently, the DVD firm will be indifferent about the source of its "productivity" gain. Ideally, one would like to obtain a measure of productivity that excludes the effect of the taste shock induced by the increasing availability of DVD rentals. This would require firm level data on both quality and price changes (or quality adjusted price changes) that is seldom available. Of course, note that industry-wide preference shocks (the price of movie theater tickets rises), captured by Z_t, do not affect the productivity measures (so long as the effect of the change in Z_t on average prices is accurately reflected in the price index $\tilde{P}$.)

3 Production Function Regressions Using the Deflated Sales Proxy: A Simple Fixed-Effect Estimation

Assume that production data is available for a panel of firms over time – this data consists of information on firm sales and input use, along with an industry price index for each period. Further assume the functional forms for demand and production specified in (2) and (3). The estimating equations are written in terms of the aggregate input index for simplicity; recall that this index can be replaced with any linearly homogeneous function of the various inputs. I initially impose some restrictive assumptions on the stochastic structure of the productivity and quality indices. These assumptions are made in order to validate a simpler estimation method that nevertheless captures the main issues surrounding the use of the deflated sales proxy. I thus assume that the firm level indices ϕ_{it} and λ_{it} can be decomposed into

$$\phi_{it} = \phi_i + \phi_t + \epsilon_{it}$$
$$\lambda_{it} = \lambda_i + \lambda_t + \eta_{it},$$

where ϕ_i and λ_i represent "fixed" firm productivity and quality effects, while ϕ_t and λ_t represent aggregate productivity and quality levels over time.[3] The ϵ_{it}s and η_{it}s are assumed to be *iid* disturbance terms with zero means and are assumed to be unobserved by the firms. They represent, respectively, idiosyncratic firm-level productivity and demand shocks.[4] Consider running a regression of firm deflated sales $r_{it} - \tilde{p}_t$ on the aggregate index x_{it}, along with firm and time indicator variables χ_i and χ_t:

$$r_{it} - \tilde{p}_t = \alpha x_{it} + \beta_i \chi_i + \beta_t \chi_t + u_{it}. \tag{5}$$

where $u_{it} = \epsilon_{it} + \eta_{it}$ is a zero mean *iid* disturbance term.

If output were substituted for deflated sales on the left-hand side, then the interpretation of the coefficients (given the simplifying assumptions) would be straightforward: α measures the degree of returns to scale, the β_is measure the fixed firm productivity levels, and the β_ts measure the aggregate industry productivity levels. As shown by (4), the substitution of output with deflated sales is not innocuous when firms sell differentiated goods. Preserving the simplifying assumptions, the OLS estimates of $\alpha, \beta_i,$ and β_t will now measure:

$$E[\hat{\alpha}] = \tfrac{\sigma-1}{\sigma}\gamma \tag{6}$$
$$E[\hat{\beta}_i] = \tfrac{\sigma-1}{\sigma}(\phi_i + \lambda_i) \tag{7}$$
$$E[\hat{\beta}_t] = \tfrac{\sigma-1}{\sigma}(\phi_t + \lambda_t) + \tfrac{1}{\sigma}\big[(r_t - \tilde{p}_t) - n_t\big]. \tag{8}$$

As noted earlier, it will be impossible to separately identify the effects of demand shocks and

[3] Changes in λ_t will also be reflected in the price index $\tilde{p}_t$.

[4] The η_{it}s thus do not capture changes in product quality (which would be known to the firms) but rather idiosyncratic consumer preference shocks.

productivity shocks. Both will be incorporated in the residual u_{it}. Similarly, only a quality adjusted productivity index $\varphi_i = \phi_i + \lambda_i$ can be identified for each firm. At the aggregate level, the per-period quality adjusted productivity index $\varphi_t = \phi_t + \lambda_t$ *can* be decomposed into a separate quality and productivity component, so long as the decomposition of the aggregate price index into these components is available.[5]

Furthermore, note that the use of the deflated sales proxy significantly changes the interpretation of the regression coefficients and the properties of the residual. Following is a list of these changes:

- As shown by Klette and Griliches (1996), the coefficient α on the aggregate input variable x_{it} will be less than the true degree of returns to scale γ. The intuition for this is as follows: Assume that, in a given time period, two firms with the same quality index λ_i have a one percent relative output difference. In a homogeneous good industry, these two firms would also have a one percent relative revenue difference. In a differentiated product industry, these two firms would only have a $\frac{\sigma-1}{\sigma}$ percent revenue difference between them: the firm with the higher output must have reduced its price relative to the other by $\frac{1}{\sigma}$ percent in order to increase its relative output by one percent. Its relative revenue is thus only $1 - \frac{1}{\sigma} = \frac{\sigma-1}{\sigma}$ percent higher. Within a time period, relative revenue differences thus understate relative (quality adjusted) output differences by $\frac{\sigma-1}{\sigma}$ percent.

  One should also not be surprised to measure higher returns to scale parameters when aggregating up from firms to industries as the coefficient on inputs would no longer be biased down by $\frac{\sigma-1}{\sigma}$. This explanation for the finding of returns to scale estimates rising with the aggregation level is very different than the one developed by Basu and Fernald (1997) who explain why estimates of returns to scale increase with the aggregation from industries to the entire manufacturing sector. Although the reasons differ with the level of aggregation, they both show that measured increases in returns to scale with the level of aggregation do not imply the presence of production externalities.

- Differences in the β_i coefficients across firms will similarly understate true productivity differences: $\Delta\beta_i = \frac{\sigma-1}{\sigma}\Delta\varphi_i$. As was mentioned earlier, revenue differences across firms will understate output differences. Since the definition of productivity is based on output differences, these true productivity differences will be under-estimated. Note that the amount of

[5] One would need to know what proportion of the change in $\tilde{p}_t$ was due to an aggregate quality change.

this measurement bias is linked to the elasticity of substitution: the greater the level of product differentiation (lower σ), the greater the under-measurement bias of the true productivity differences. In particular, caution must be taken when comparing distributions of firm productivity levels across industries. This could also affect the comparison of productivity levels within an industry when firms are partitioned by export status. The exporting firms will be selling a portion of their output on an international market with a different level of product differentiation than the domestic market. If the elasticity of substitution is higher in the international market (which seems likely), then measured productivity differences between exporting and non-exporting firms (in the $\beta_i s$) will be downward biased.

- The time period coefficient β_t will now measure more than just aggregate quality adjusted productivity levels (recall from (8) that $E[\hat{\beta}_t] = \frac{\sigma-1}{\sigma}\varphi_t + \frac{1}{\sigma}\left[(r_t - \tilde{p}_t) - n_t\right]$). Thus, even if the true aggregate productivity level is a-cyclical ($cov(\varphi_t, r_t - \tilde{p}_t) = 0$), the measured aggregate productivity level β_t will be pro-cyclical. The intuition for this is as follows: a change in output per firm between periods, $(r_t - \tilde{p}_t) - n_t$, will shift the firm-level demand curves (see (2)). Since the firm sales are deflated by an average price index, a firm that increases it's output by the same percentage as the industry average $\Delta\left[(r_t - \tilde{p}_t) - n_t\right]$ will also increase its revenues by the same percentage (this firm must have changed its price by the same percentage as the industry average $\Delta\tilde{p}_t$. The relationship between firm output and revenue is thus different in the cross-section than it is over time. Within the same time period, price differences between firms are not measured and are reflected in revenue differences across firms. On the other hand, average price changes over time *are* captured by the price index and are purged from the deflated revenue proxy.

Cooper and Johri (1997) and Lindstrom (1997) both run firm level production function regressions using the deflated sales output proxy and find that an included aggregate industry sales regressor is strongly significant with a positive coefficient. They interpret this as evidence for "external economies of scale" or "dynamic complementarities". The estimating equation (4) suggests that it would be surprising *not* to find a significant positive coefficient on the aggregate sales regressor $(r_t - \tilde{p}_t)$ and that only the finding of a coefficient *greater* than the inverse elasticity of demand $\frac{1}{\sigma}$ would possibly indicate the presence of external economies.[6]

[6]Unmeasured input utilization could also be driving part of the measured pro-cyclicality of aggregate productivity. This is true for both firm-level and industry-level studies, whereas the bias introduced by the use of sales instead of output only applies to firm-level studies. Basu and Kimball (1997) provide evidence that most of the observed

Unfortunately, average sales per firm can not be added as a regressor in (5) since it would be co-linear with the time indicator variables. Given the present estimation framework, it is thus impossible to separately identify the elasticity of substitution and the true aggregate productivity index φ_t, and test whether the latter is correlated with aggregate industry sales. As I will show in a later section, it is nevertheless possible to separately identify both these parameters (the aggregate productivity level φ_t and the elasticity of substitution σ), once a richer dynamic structure of firm level productivity is introduced.

The measured aggregate productivity β_t will not only fluctuate with the aggregate sales level, but will also be affected by changes in the elasticity of substitution between varieties (as was the case for the firm level measure β_i). If one assumes that increases in the number of varieties produced positively impacts the substitutability between these varieties, then a measured increase in aggregate productivity β_t could partially be driven by an increase in product variety. If the number of varieties then stabilizes as an industry matures, one would then observe a spurious "productivity" slow-down driven by the reduced growth rate of varieties.

- The fixed-effect model for firm productivity is often rejected on the basis of the serial correlation of the error term u_{it}. If the regression (5) is run without individual period indicator variables or only with a time index variable capturing a productivity improvement trend over time, then the presence of serial correlation in u_{it} does not necessarily imply that the fixed effects model is mis-specified: it could just be the consequence of the omitted output per firm regressor $(r_t - \tilde{p}_t) - n_t$, which will almost always be serially correlated.

The revenue production function (4) shows that it is still possible to un-cover productivity differences between firms even though no firm level price information is available. No assumptions concerning profit maximization have been made, so the estimation of (4) does not depend on any particular assumption about firm markups. Some studies have claimed that, without information on firm level prices, it is impossible to identify productivity differences separately from markup differences. This is based on the fact that firm revenue r_{it} can be written as:

$$r_{it} = x_{it} + w_{it} - \log \gamma + \log \mu_{it},$$

pro-cyclicality of productivity in industry level studies can be explained by unmeasured input utilization.

where $\mu_{it} = \frac{P_{it}}{MC_{it}}$ is the firm level markup (and w_{it} is the previously defined factor price index.) It is thus true that differences in revenue per unit of input, $r_{it} - x_{it}$, only capture markup differences between firms (assuming that firms face similar factor shadow prices). This, however, does not imply that the productivity differences can not be estimated (the firm markup μ_{it} will always be inextricably correlated with the input x_{it}: a firm that raises its markup necessarily decreases its output sold, and hence its use of inputs.)

4 Multi-Product Firms and Productivity

Up to this point, I have assumed that each firm produces a single product (or variety). In the absence of adjustment costs to the firm's inputs and persistent markup differences between firms, this assumption would entail a perfect correlation between firm size (in terms of revenue or input use) and quality adjusted productivity φ_i. I now consider an additional factor other than adjustment costs, markups, or productivity that would explain some of the observed dispersion in firm size: differences in the number of varieties produced by firms. I assume the same structure for the production and demand of each variety as was previously developed. Production for each of the M_i varieties produced by firm i still satisfies $Q_{ij} = \Phi_{ij} X_{ij}^{\gamma}$ while the demand for each of these varieties is still given by $Q_{ij} = \Lambda_{ij}^{\sigma-1} \left(\frac{P_{ij}}{\tilde{P}}\right)^{-\sigma} \frac{1}{M} \left(\frac{R}{\tilde{P}}\right)$ where $M = \sum_{i=1}^{N} M_i$ now represents the aggregate number of varieties produced.[7] For each firm, I assume that only the aggregate sales $R_i = \sum_{j=1}^{M_i} R_{ij} = \sum_{j=1}^{M_i} P_{ij} Q_{ij}$ and input use $X_i = \sum_{j=1}^{M_i} X_{ij}$ are observable. I further assume that firms incur a sunk cost (not reflected in their current input use) in order to introduce an additional variety. If there are increasing returns to scale ($\gamma > 1$), then there is another cost of producing additional varieties built into the production function: spreading output over more varieties reduces the total output units a firm can produce with a given input bundle.

For each firm, an average quality adjusted productivity level $\tilde{\varphi}_i$ can be constructed such that firm i's total sales and input use match those of a hypothetical firm producing the same number of varieties at the *same* quality adjusted level $\tilde{\varphi}_i$. In other words, $\tilde{\varphi}_i$ is the productivity level that converts $\frac{X_i}{M_i}$ units of inputs into sales $\frac{R_i}{M_i}$ according to the revenue production function outlined in

[7]Preserving the same form of production and demand both across varieties and across firms imposes some additional restrictions on the structure of production and demand. These restrictions preclude the possibility of cost synergies within firms across varieties. The restrictions also rule out the possibility that varieties may be less differentiated within firms than across firms.

(4).[8] Thus, for any period t,

$$r_{it} - m_i - \tilde{p}_t = \tfrac{\sigma-1}{\sigma}\gamma(x_{it} - m_i) + \tfrac{1}{\sigma}\left[(r_t - \tilde{p}_t) - m_t\right] + \tfrac{\sigma-1}{\sigma}\tilde{\varphi}_{it}.$$

The relationship between the firm total sales and its total input use and average productivity level is then given by:

$$r_{it} - \tilde{p}_t = \tfrac{\sigma-1}{\sigma}\gamma x_{it} + \tfrac{1}{\sigma}\left[(r_t - \tilde{p}_t) - m_t\right] + \tfrac{\sigma-1}{\sigma}\left[\tilde{\varphi}_{it} + \left(\tfrac{1}{\sigma-1} - (\gamma-1)\right) m_i\right].$$

Consider the same firm level regression outlined in (5). The estimates of α and β_t will have the same interpretation as was described in the previous section ((6) and (8) still hold). On the other hand, the estimate of the firm effect β_i will now include an additional term:

$$E[\hat{\beta}_i] = \tfrac{\sigma-1}{\sigma}\left[\tilde{\varphi}_{it} + \left(\tfrac{1}{\sigma-1} - (\gamma-1)\right) m_i\right].$$

As I will show below, $\frac{1}{\sigma-1} - (\gamma-1)$ must be positive if firms choose to produce more than one variety (if $m_i = \log M_i > 0$). Thus, two firms with the same (quality adjusted) productivity level $\tilde{\varphi}_i$ will have different measured productivity levels β_i if they produce a different number of varieties: the firm producing a greater number of varieties will have a higher β_i. The intuition for this measured productivity difference is as follows: consider first the case of constant returns to scale (γ=1). The measured productivity difference between two firms with identical productivity parameters $\tilde{\varphi}_i$ will then be $\frac{1}{\sigma-1}\Delta m_i$. This is the product variety effect: holding the quality of goods constant, consumers prefer a bundle of goods with more varieties to one with the same number of units spread over a smaller number of varieties.[9] The presence of non-constant returns to scale will also affect the measured productivity linked to product variety: with increasing returns, the measured productivity difference would be smaller than in the constant returns case. This reflects the fact that under increasing returns, a multi-product firm reduces the total number of output units it

[8] Formally,

$$e^{-\frac{1}{\gamma}\tilde{\varphi}_i} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left[\left(\frac{M_i R_{ij}}{R_i}\right)^{\frac{\sigma}{\sigma-1}} e^{-\varphi_{ij}}\right]^{\frac{1}{\gamma}}.$$

$\tilde{\varphi}_i$ is a weighted average of the quality adjusted productivity levels for each variety φ_{ij}. The weights are proportional to the varieties' share of the firm's revenue.

[9] A consumer would be willing to forego exactly $\frac{1}{\sigma-1}$ percent total units of goods consumed in return for spreading those units over a one percent greater number of varieties (holding qualities fixed).

could produce by spreading its output over more than one variety. Firms are willing to incur this efficiency loss because they can sell the reduced output quantities of each variety at higher prices (as long as $\frac{1}{\sigma-1} > \gamma-1$, as will be shown below). With decreasing returns, the measured productivity difference would be larger than the constant returns case: spreading production over fewer varieties increases output efficiency. In both cases, the sunk cost of adding new varieties imposes a trade-off for firms between the cost and benefit of producing additional varieties.

5 A Re-Interpretation of Firm Productivity

The previous section discussed why – holding the productivity of producing any given variety fixed – firms producing more varieties would "appear" to be more productive than firms producing fewer varieties. Adjusting a measure of firm productivity to reflect differences in the "inherent" quality of the goods produced seems quite intuitive. Adjusting the productivity measure to reflect differences in consumer tastes for the varieties is somewhat less intuitive. And lastly, adjusting the productivity measure to reflect differences in the number of varieties produced by firms seems even less intuitive. A natural follow-up question is then to ask whether the productivity measures β_i obtained from the production function regression (5), $\tilde{\varphi}_i + \left(\frac{1}{\sigma-1} - (\gamma-1)\right) m_i$, can somehow be re-interpreted as a "valid" measure of firm productivity. As will be shown below, the answer to this question is affirmative: the use of an intuitive quantity index to measure firm output across varieties and qualities gives the productivity measure $\tilde{\varphi}_i + \left(\frac{1}{\sigma-1} - (\gamma-1)\right) m_i$ the standard interpretation of productivity applied to homogeneous good industries.

The quantity index is constructed as follows: pick any one variety as a reference good, then measure a firm's output as the amount (quantity) of that variety required to leave the consumer indifferent between this quantity of the one variety and the actual bundle of goods produced by the firm. Let Q_i^o denote this quantity index for firm i's output. Assuming, only for expositional simplicity, that $\Lambda^o = 1$ for the reference variety, then Q_i^o can be written:

$$Q_i^o = \left(\sum_{j=1}^{M_i} (\Lambda_{ij} Q_{ij})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

If firm i only produces one variety, then its production function with output measured in units of the quantity index is given by:

$$q_i^o = \gamma x_i + \phi_i + \lambda_i = \gamma x_i + \varphi_i.$$

If firm i produces M_i varieties, then the production function becomes:

$$q_i^o = \gamma x_i + \tilde{\varphi}_i + \left[\tfrac{1}{\sigma-1} - (\gamma-1)\right] m_i, \tag{9}$$

where $\tilde{\varphi}_i$ is the average quality adjusted productivity level that was previously defined. (9) shows how $\hat{\varphi}_i = \tilde{\varphi}_i + \left(\tfrac{1}{\sigma-1} - (\gamma-1)\right) m_i$ is an appropriate measure of firm i's relative productivity: given the use of any input bundle x_i, the percentage difference in the quantity index of output produced by any two firms will be given by $\Delta\hat{\varphi}_i$. This is exactly the interpretation of productivity in homogeneous good industries where the construction of a quantity index is unnecessary.

The quantity index production function (9) also demonstrates why $\frac{1}{\sigma-1} - (\gamma-1)$ must be positive if firms choose to produce more than one variety. (9) implies that the amount of the aggregate input required to produce q_i^o units of output spread over M_i varieties is:

$$x(q_i^o, m_i) = \frac{1}{\gamma}\left[q_i^o - \left(\tfrac{1}{\sigma-1} - (\gamma-1)\right) m_i - \tilde{\varphi}_i\right].$$

If $\frac{1}{\sigma-1} - (\gamma-1)$ were negative, then the firm's input requirement to produce q_i^o output units increases with the number of varieties m_i. A firm could produce the same output quantity index q_i^o using fewer inputs by only producing a single variety with the highest φ_i.[10] On the other hand, if $\frac{1}{\sigma-1} > (\gamma-1)$, then firms trade-off the positive benefit of producing more varieties against the sunk cost required to introduce them.[11]

6 Time-Varying Firm Productivity and Quality Levels

In this section, I relax the stochastic assumptions imposed on the firm productivity and quality indices ϕ_{it} and λ_{it}. Although the estimation of (4) using fixed firm effects is no longer valid, I show how the implementation of the Olley and Pakes (1996) estimation method proposed by Levinsohn

[10] Note that since a firm has no control over the industry price index, there is a fixed price at which the firm can sell the output quantity q_i^o.

[11] Even if firms only produce one variety, $\frac{1}{\sigma-1} < (\gamma-1)$ does not seem reasonable as this would imply that any firm earning zero or positive profits could indefinitely increase its profits by increasing production. Large factor adjustment costs would then be the only reason preventing the market structure transition towards monopoly.

and Petrin (2000) – hereafter L-P – can be extended to the case of imperfect competition and product differentiation. I now assume that each firm's productivity and quality index follow a first order Markov process, although I restrict these processes to be identical for both indices. In essence, this forces productivity, quality, and preference innovations to have the same amount of serial correlation so that they can all be combined into a single unobserved state variable.[12] In the fixed effect model, only aggregate quality changes were possible and consumer preference shocks were independent across time. Now, long lasting quality *and* preference shocks are possible at the firm level (the consumer's preference for DVD players over VCR players may be long-lasting). Given the assumption of a common Markov process for all innovations, firms will be indifferent about the source of these innovations and will only care about their quality adjusted productivity $\varphi_i = \phi_i + \lambda_i$. The only other unobserved variable affecting a firm's performance will be the number of varieties it produces. Since the firm does not face any uncertainty concerning this variable, it can also be combined with the unobserved productivity φ_i to form a unique unobserved state variable

$$\hat{\varphi}_i = \varphi_i + \left[\tfrac{1}{\sigma-1} - (\gamma-1)\right] m_i.$$

As was shown in the previous section, this variable appropriately indexes firm level productivity in a differentiated product industry with multi-product firms.

Following L-P, I investigate using intermediate inputs as a proxy for the unobserved productivity variable $\hat{\varphi}_i$. I thus need to find some sufficient conditions under which a firm's use of intermediate inputs is a monotonic function of its productivity index $\hat{\varphi}_i$. The homogeneity assumption for the production technology implies that no factor, and in particular the intermediate input, can be inferior. This ensures that an increase in the aggregate input index implies an increase in the use of the intermediate input. I therefore investigate when higher productivity (as measured by $\hat{\varphi}_i$) leads to the use of a higher aggregate input level. To do this, I convert both the firm level output bundle and the aggregate output bundle using the previously defined quantity index. Let q_i^o and q^o denote these quantity indices. Further let $p_i^o = r_i - q_i^o$ and $p^o = r - q^o$ be the price indices associated with

[12] Petropoulos (2000) develops an estimation procedure that does not impose this restriction and allows demand side innovations to have lower levels of serial correlation than productivity innovations.

these quantities. Firm i's production and demand functions then be take the form:

$$q_i^o = \gamma x_i + \hat{\varphi}_i$$
$$q_i^o - q^o = -\sigma(p_i^o - p^o).$$

Let $\log \mu_i = p_i^o - c_i'$ denote the log of firm i's markup measured in terms of the price and cost of a marginal unit of the quantity index q_i^o. c_i', the log of firm i's marginal cost, is then given by:

$$\begin{aligned} c_i' = \log\left(\frac{\partial W_i X_i}{\partial Q_i^o}\right) &= \log\left(\frac{\partial x_i}{\partial q_i^o}\right) + w_i + x_i - q_i^o \\ &= w_i - \big[(\gamma{-}1)x_i + \hat{\varphi}_i + \log\gamma\big]. \end{aligned}$$

The factor markets for the intermediate input and any other flexible input (whose levels can be adjusted without cost by the firms) are assumed to be competitive. As was previously mentioned, the shadow price w_i of the aggregate input bundle will vary across firms, but will only depend on the firm's current level of the quasi-fixed inputs. Holding the levels of these quasi-fixed inputs fixed, the impact of productivity on the firm's aggregate input use is given by the partial derivative:

$$\frac{\partial x_i}{\partial \hat{\varphi}_i} = \frac{1 - \left(\frac{\sigma}{\sigma-1}\right)\frac{\partial \log \mu_i}{\partial \hat{\varphi}_i}}{\frac{1}{\sigma-1} - (\gamma{-}1)}. \tag{10}$$

As was previously discussed, the denominator in (10) must be positive if firms choose to produce more than one variety, and I continue to assume that this condition is satisfied.[13] The monotonicity condition will therefore be satisfied if the numerator in (10) is positive, or equivalently, if $\frac{\partial \log \mu_i}{\partial \hat{\varphi}_i} < \frac{\sigma-1}{\sigma}$. The elasticity of the markup must therefore be bounded above by $\frac{\sigma-1}{\sigma}$. Assuming that firms can not threaten to produce a variety already produced by another firm (i.e. there are no limit-pricing motives) and that firms have perfect knowledge of their own productivity $\hat{\varphi}_i$, then profit maximization, even with factor adjustment costs, would ensure that the monotonicity condition be satisfied as the markup elasticity $\frac{\partial \log \mu_i}{\partial \hat{\varphi}_i}$ would then be zero (all firms would choose the same markup, $\frac{\sigma}{\sigma-1}$, regardless of their productivity level $\hat{\varphi}_i$). Without imposing these additional assumptions, the monotonicity condition will be satisfied if more productive firms do not

[13]This condition is actually stronger than what is needed to ensure that the monotonicity condition is satisfied. The returns to scale parameter γ in the denominator can be replaced with the returns to scale to *only* the variable inputs, which will be less than γ. (The levels of the quasi-fixed factors are implicitly held fixed).

have "disproportionately" higher markups than less productive ones.[14] The intuition for this is straightforward: more productive firms use fewer inputs to produce the same output as do less productive firms. More productive firms must therefore produce "sufficiently" more output than less productive firms in order to use more inputs. This will only be possible if these more productive firms do not set "disproportionately" higher markups than the less productive firms. Thus, if $\frac{\partial \log \mu_i}{\partial \hat{\varphi}_i} < \frac{\sigma-1}{\sigma}$, the monotonicity condition will be satisfied and intermediate inputs can be used as a proxy for productivity.

I now briefly discuss how this proxy can be used to obtain consistent estimates of the elasticity of substitution σ and the firm level productivity indices $\hat{\varphi}_{it}$. Given the re-interpretation of firm productivity $\hat{\varphi}_{it}$ and the validity of the intermediate input proxy, the estimation method developed by L-P can be applied to the current case of differentiated product industries and multi-product firms. Consider, for simplicity, the case of Cobb-Douglas production involving a variable factor (labor, l), a quasi-fixed factor (capital, k), and the intermediate input (fuel or materials, e). The input index x_i is then given by $x_i = \alpha_k k_i + \alpha_l l_i + \alpha_e e_i$ where the cost shares $\alpha_k, \alpha_l, \alpha_e$ sum to one. Deflated firm sales can then be written

$$r_{it} - \tilde{p}_t = \tfrac{\sigma-1}{\sigma}\gamma\big(\alpha_k k_{it} + \alpha_l l_{it} + \alpha_e e_{it}\big) + \tfrac{1}{\sigma}\big[(r_t - \tilde{p}_t) - m_t\big] + \tfrac{\sigma-1}{\sigma}\hat{\varphi}_{it}.$$

In practice, the number of varieties sold in any period, $M_t = e^{m_t}$, is not observable. Assuming that the *average* number of varieties sold per firm remains constant over time, m_t can be replaced by the log of the number of firms, n_t, plus a constant. Adding an idiosyncratic *iid* "productivity" disturbance shock u_{it} that is unobserved by the firms (when making their pricing decisions),[15] the estimating equation becomes:

$$r_{it} - \tilde{p}_t = \beta_o + \beta_k k_{it} + \beta_l l_{it} + \beta_e e_{it} + \beta_\sigma\big[(r_t - \tilde{p}_t) - n_t\big] + \hat{\varphi}(k_{it}, e_{it}) + u_{it},$$

where $\hat{\varphi}(k_{it}, e_{it})$ is some non-parametric function of k_{it} and e_{it}. This estimating equation is structurally equivalent to the one derived by L-P. The only difference is the additional regressor $\big[(r_t - \tilde{p}_t) - n_t\big]$ (average firm deflated sales) and the re-interpretation of the coefficients $\beta_k, \beta_l, \beta_e$, and of productivity $\hat{\varphi}$. A first stage regression using a non-parametric function of k_{it} and e_{it} will

[14] Of course, the elasticity of demand σ affects the interpretation of "disproportionate".

[15] These shocks need only be unobserved by the firm at the time that prices are set. These shocks could reflect both unexpected cost or firm demand fluctuations. The firm's markup and input use does therefore not respond to these shocks. The firm's markup is chosen based only on information on $\hat{\varphi}_{it}$.

produce consistent estimates for β_σ (and hence for $\sigma = \frac{1}{\beta_\sigma}$) and β_l. β_k and β_e can then be estimated in a second stage (see L-P for details), yielding estimates for $\gamma = \frac{\sigma}{\sigma-1}\,(\beta_k + \beta_l + \beta_e)$ and $\hat{\varphi}_{it} = \frac{\sigma}{\sigma-1}\,\hat{\varphi}(k_{it}, e_{it})$. The $\hat{\varphi}_{it}$s can then be further decomposed into an aggregate component φ_t (common across firms in a given time period) and a firm relative effect. The true pro-cyclicality of productivity can then be assessed by examining the correlation between φ_t and $r_t - \tilde{p}_t$.

7 Conclusion

In order to measure productivity differences between firms or productivity changes over time, one must first find a way of measuring and comparing output levels across firms and over time. If the output produced by different firms is a homogeneous good, then there is a straightforward way to make these comparisons using observed differences in firm revenues. Although this remains an interesting reference case, it is one that is seldom directly applicable in practice. In most (if not all) industries, firms produce goods that are differentiated, and they often produce more than one type of good. In a large number of cases, economists can not observe the quantities and qualities of the varieties produced by a firm – only a firm's revenue sales across these varieties is observable. A common approach to this problem has been to use these revenues as a direct measure of output that is then homogeneous across firms. This paper has shown how the use of this approach can lead to some serious mis-interpretations of the obtained productivity estimates. On the other hand, this paper has also shown that the information on firm revenues can be used to obtain meaningful estimates of productivity differences between firms and over time, *without* requiring additional information on the firms' output quantities and qualities. Thus, firm revenue does aggregate information on a firm's output in a way that is useful for productivity analysis, although these revenues can not directly be used as a measure of output. The paper also shows how the obtained productivity estimates must be interpreted in order to match the definition of productivity commonly applied to the production of homogeneous goods.

In order to use revenues to obtain information on a firm's output, this paper has made some very restrictive assumptions on the structure of consumer demand and the type of product differentiation. When additional information on the varieties produced by firms is available, then a much richer demand structure can be developed – as has been done in a number of industry studies. However, when product characteristic data is unavailable, the default approach has been to assume perfect substitutability between goods – either explicitly or implicitly, through the use of the deflated sales proxy for output. Although the demand side structure imposed in this paper is quite restrictive, it is nevertheless *less* restrictive than this default assumption of a *common* and *infinite* elasticity of substitution between varieties. Furthermore, this possibility is nested within the proposed estimation framework.

References

Basu, S. and Fernald, J.: 1997, Returns to scale in U.S. production: Estimates and implications, *Journal of Political Economy* **105**, 249–283.

Basu, S. and Kimball, M.: 1997, Cyclical productivity with unobserved input variation, *NBER Working Paper No. 5915*.

Cooper, R. and Johri, A.: 1997, Dynamic complementarities: A quantitative analysis, *Journal of Monetary Economics* **40**, 97–119.

Klette, T. J. and Griliches, Z.: 1996, The inconsistency of common scale estimators when output prices are unobserved and endogenous, *Journal of Applied Econometrics* **11**, 343–361.

Levinsohn, J. and Petrin, A.: 2000, Estimating production functions using inputs to control for unobservables, *NBER Working Paper No. 7819*.

Lindstrom, T.: 1997, External economies at the firm level: Evidence from Swedish manufacturing, *Mimeo*.

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

Olley, S. and Pakes, A.: 1996, The dynamics of productivity in the telecommunications equipment industry, *Econometrica* **64**, 1263–97.

Petropoulos, W.: 2000, Productivity in an industry with differentiated products, *University of Michigan, Mimeo*.