Productivity in a Differentiated Products Market Equilibrium*

Jim Levinsohn
Department of Economics and
Ford school of Public Policy
University of Michigan
and
NBER

Marc J. Melitz
Department of Economics
Harvard University
and
NBER

PRELIMINARY AND INCOMPLETE DRAFT
COMMENTS AND SUGGESTIONS ARE WELCOMED

1 Introduction

This paper is about estimating productivity in the presence of product differentiation and in the absence of ideal data. By appealing to standard notions of market equilibrium and imposing structure on consumer demand, we show how one can estimate productivity when: 1) plant-level information about the physical attributes of output are not available; 2) the available measure of output is plant sales; and 3) plants produce differentiated products and may produce more than one product variety. In our experience, this set-up is very common. Even very detailed manufacturing censuses do not typically record information about the physical attributes of the plant's output(s). Further, output is typically measured not in physical units but rather in terms of dollars– i.e. sales. Lastly, we cannot identify an industry in which every plant produces exactly the same product. Rather, plants produce differentiated products that, for administrative reasons, are classified within a given industry code and this holds true at even the most disaggregated industry definitions.

The literatures on estimating productivity and on estimating demand in differentiated products markets have progressed in recent years but without much reference to one another. On the productivity side, a line of research initiated by Olley and Pakes (1996) has made significant

*This research is in progress. When we are further along, it will qualify as "Preliminary." Please direct comments to Levinsohn at jamesl@umich.edu or to Melitz at mmelitz@harvard.edu.

contributions while on the demand side several authors have estimated new models of demand for differentiated products.[1] The link between the productivity and demand literatures arises because of lousy data. In an ideal world, researchers would observe plant-level measures of physical output and the physical attributes of this output. In such a world, one could measure productivity of even differentiated products without any reference to demand. In reality, one usually observes the value of sales at the plant-level and sales depends on price. Price, in turn, is determined in the market equilibrium, and hence the estimation of productivity, typically thought of as a supply-side phenomenon, depends quite crucially on the demand side of the market.

Our goal in this paper is to put some very simple structure on the demand side of the market, show how this structure can yield an estimate of productivity in differentiated products markets, and then take this methodology to the data. The end-product is a set of plant-level estimates of productivity in a differentiated products industry. We investigate the sorts of questions for which accounting for product differentiation seems to matter. That is, for many questions, the standard assumption of plants producing a homogeneous product may be innocuous. For other questions, the assumption may be particularly objectionable. Finally, showing how one can obtain a plant-level productivity measure in a differentiated products industry helps bring the empirical literature a bit closer to the theoretical literature on industry evolution. In this study, we will interpret and measure productivity in a very broad sense, as encompassing both productive efficiency and product quality (our modeling assumptions described in the following section will make this definition more explicit). In fact, our modeling approach will preclude any further decomposition of this plant-level "productivity" concept. Other recent studies by Kraay, Soloaga and Tybout (2001) and Petropoulos (2000) start with a similar framework and further decompose plant level productivity[2] by imposing further restrictions on the production technology (see Kraay et al. (2001)) or the dynamic properties of the productivity shocks (see Petropoulos (2000)) . Nevertheless, we see all these papers as closely related.[3]

Before a more detailed explanation of just how these goals are (hopefully) accomplished, we provide a non-technical discussion of the remainder of the paper. We begin, in Section 2, with a discussion of our methodology. Absent any information on physical product characteristics, we impose some rather severe restrictions on the structure of the demand system. Our approach is to

[1]See for example Berry, Levinsohn and Pakes (1995), Goldberg (1995), and Nevo (2001). Note, though, that all of these studies include data on product characteristics.

[2]In both cases, plant level productivity is decomposed into a productive efficiency component and another reflecting product quality.

[3]We have also strongly benefited from discussions with the authors of both papers.

impose, as a first pass, the simplest of demand structures. In particular, we assume all products are symmetrically differentiated and that consumer utility takes the form of CES preferences. We are not the first to go down this path. Griliches and Klette (1996) adopted a similar strategy when investigating returns to scale in a differentiated products industry. Our approach simply builds on their foundation. Instead of focusing on industry returns to scale, we are more concerned with plant-level productivity. As was the case for Griliches and Klette (1996), our estimation strategy yields, as a by-product, a measure of the elasticity of substitution between varieties in each industry. Given some additional assumptions on market conduct, these elasticities, in turn, yield information on the average markups across sectors. While these parameters are not our principle focus, their estimated magnitude conveys economic meaning and as such they can provide a benchmark for the reasonableness (or not) of our results. The details of our methodology are presented in section 2.

In section 3, we get around to discussing why one might actually care about controlling for product differentiation when estimating productivity. The typical approach to dealing with product differentiation in the productivity literature is pretty simple – just ignore it. For some questions, this is probably a sound approach. In these instances, our approach for accounting for product differentiation is guaranteed to give the same answer as the simpler approach. For other questions of economic interest though, accounting for product differentiation may really matter. In the former category are some particular questions that focus on the productivity ranking of plants. It turns out that - in some particular cases - the rank order of plant productivity is not affected by our adjustment to account for product differentiation. In the latter category of issues where our approach may matter, we consider:

1. Whether accounting for product differentiation provides a substantively different view of the role of reallocation in accounting for industry productivity growth. Recent work, both empirical[4] and theoretical[5], has focused on the role that reallocation might play in explaining industry productivity shifts. The idea is that industry productivity (as typically defined) is a sales-weighted average of plant-level productivities. The industry index, then, can increase either because individual plants have become more productive and/or because more productive plants have increased market shares. A relatively large (and relatively confusing) literature has developed on this topic. Recent work includes [...]. We propose a new way

[4] See ,for example, Baily, Hulten and Campbell (1992), Foster, Haltiwanger and Krizan (1998), and Olley and Pakes (1996).

[5] See, for example, Melitz (2000).

of investigating this issue and then examine whether accounting for product differentiation matters.

2. Whether accounting for product differentiation matters when considering the *relative* productivity of exporting versus non-exporting plants. Recent literature has noted that firms that export tend to be more productive. Examples include [...]. We investigate just how much more productive exporting plants are and whether accounting for product differentiation matters here.

2 Methodology – Estimating the Plant Level Productivities

As we mentioned earlier (and is suggested by our title), we specify a demand system and estimate an industry equilibrium relationship between the plant's sales, their input usage, their productivities, and current industry market demand conditions. The modeling of the demand side is necessary since plant level prices endogenously affect the plants revenues sales, which are the only available measure of the plants' output. If plants produced homogeneous goods, then one could control for the endogenous prices using a single price index for the industry. Plant sales – deflated by this index – then yields a "perfect" proxy for the unobserved physical output – adjusted for quality. Most of the productivity literature relies on the use of this output proxy. However, since plants produce goods that are differentiated, plant level prices will fluctuate relative to the industry price index thereby breaking the link between deflated sales and physical output. The associated problems for the estimation of firm level production functions using the deflated sales proxy have 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.

Our approach builds upon the work of Griliches and Klette (1996), who developed methods to address the problems caused by the deflated sales proxy for firm production analysis in differentiated product industries. Whereas Griliches and Klette (1996) mainly focused on the measurement of the degree of returns to scale in production, we focus the current analysis on the obtainment and interpretation of productivity measures. We will show how the concept of productivity can be re-interpreted in a differentiated product industry where plants produce goods with different quality levels (and possibly produce more than a one variety). For expositional purposes, we first describe 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. We 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.

2.1 The Model

Throughout this paper, we assume that plants, and the firms that own them, are small relative to the industry. We do not explicitly model multi-plant firms and assume that managers at each plant maximize profits independently. We thus model plants as individual firms and use the two terms interchangeably.

2.1.1 Firm Level Demand

We 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:

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

where $M(Z)$ represents a numeraire good that depends on aggregate demand shifters Z and $U(.)$ is assumed to be differentiable and quasi-concave. 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. Firms have no power to influence this industry price index $\tilde{P}$, and take it as given.

If the goods were homogeneous (infinite σ), then there can be no variations in quality adjusted prices across firms: $\frac{P_i}{\Lambda_i} = \tilde{P}$ for all firms, and deflated sales would then be a perfect proxy for the unobserved quality adjusted output, $\frac{R_i}{\tilde{P}} = Q_i\Lambda_i$ 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.

2.1.2 Firm Production and Revenues

We assume that the production technology is homogeneous of degree $\gamma > 0$.[6] Given this assumption, an aggregate input index X_i and factor price index W_i can be constructed such that $X_i = f(\mathbf{X}_i)$, $W_i = h(\mathbf{W}_i)$, and $X_i W_i = \mathbf{X}_i \cdot \mathbf{W}_i$ where $\mathbf{X}_i$ and $\mathbf{W}_i$ are the vectors of inputs and prices and $f(.)$ is linearly homogeneous.[7] Production as a function of the aggregate input index then takes the

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

[7] 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.

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 productive efficiency.

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

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

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

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

Note that firm revenue and cost only depend on the sum of the firm's efficiency and quality index, $\phi_{it} + \lambda_{it}$. A one percent efficiency 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_i s increase), they may improve the quality of the VCRs (both Λ_i s 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 $\tilde{P}$.) We refer to $\varphi_{it} = \phi_{it} + \lambda_{it}$ as a broad measure of a firm's productivity that incorporates both productive efficiency (the ϕ_{it}), and product quality (the λ_{it}).

(4) will be our main estimating equation that relates the plant's deflated sales to its input usage and productivity. For expositional simplicity, we continue using the aggregate input index construction in this equation; recall that this index can be replaced with any linearly homogeneous function of the various inputs.

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

We initially impose some restrictive assumptions on the stochastic structure of the efficiency 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. We then drop these restrictions, and both describe and estimate a model that allows a richer dynamic structure for the evolution of plant productivity. We thus start by assuming 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 efficiency and quality effects, while ϕ_t and λ_t represent aggregate efficiency and quality levels over time.[8] 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 efficiency and demand shocks.[9] 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:

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

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

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

[9] 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.

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:

$$E[\hat{\alpha}] = \tfrac{\sigma-1}{\sigma}\gamma \tag{6}$$

$$E[\hat{\beta}_i] = \tfrac{\sigma-1}{\sigma}(\phi_i + \lambda_i) \tag{7}$$

$$E[\hat{\beta}_t] = \tfrac{\sigma-1}{\sigma}(\phi_t + \lambda_t) + \tfrac{1}{\sigma}\left[(r_t - \tilde{p}_t) - n_t\right]. \tag{8}$$

As noted earlier, it will be impossible to separately identify the effects of demand shocks and efficiency 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.[10]

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 Griliches and Klette (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

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

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 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 ($\text{cov}(\varphi_t, r_t - \tilde{p}_t) = 0$), the measured aggregate productivity level β_t will be pro-cyclical. In a differentiated product industry, changes over time of the average sales per firm, $(r_t - \tilde{p}_t) - n_t$, will shift the firm-level demand curves (see (2)) and change the relationship between physical output and sales, thus affecting the measured productivities. All firms will therefore appear to be more productive in periods of high demand (and less productive in periods of low demand). In effect, $(r_t - \tilde{p}_t) - n_t$ controls for shifts in market demand conditions. The omission of this market demand regressor has some serious consequences for the interpretation of the productivity residuals:

  - The pro-cyclicality of the plant level productivities will be over-estimated[11]

[11] Unmeasured input utilization could also be driving part of the measured pro-cyclicality of aggregate productivity.

– During consecutive periods of growing or shrinking demand, the entire distribution of plant level productivities will appear to shift over time.[12] A portion of the industry aggregate productivity change will then spuriously be attributed to this notional shift in the distribution of plant level productivity. This also implies that the contribution of market share reallocations to aggregate productivity changes will be under-estimated.[13]

– Since the omitted aggregate sales regressor will be correlated with the plants' input usage, the estimated coefficient on the latter variable will be biased upwards. This will, in turn, bias the productivity of large plants downwards relative to small plants.

The revenue production function (4) shows that it is still possible to uncover 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:

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

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.)

2.3 Multi-Product Firms and Productivity

Up to this point, we have assumed that each firm (plant) produces a single product (or variety). In the absence of adjustment costs to the firm's inputs and persistent markup differences between

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

[12] This, in turn, provides a possible explanation for measured aggregate industry productivity declines: negative industry demand shocks.

[13] For example, increases in export demand for an industry may lead to market share reallocations towards more productive plants. The contribution of this reallocation to aggregate productivity growth will be under-estimated since the demand shift will simultaneously cause a spurious measured increase in all the plants' productivity.

firms, this assumption would entail a perfect correlation between firm size (in terms of revenue or input use) and quality adjusted productivity φ_i. We 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. By assuming the same structure for the production and demand for each variety as was previously developed, it can be shown that the estimating equation (4) can easily be extended to deal with multi-product firms. We now consider the case where each firm i produces M_i varieties. Production of each variety satisfies $Q_{ij} = \Phi_{ij} X_{ij}^{\gamma}$, and its demand satisfies $Q_{ij} = \Lambda_{ij}^{\sigma-1} \left(\frac{P_{ij}}{\tilde{P}}\right)^{-\sigma} \frac{1}{M} \left(\frac{R}{\tilde{P}}\right)$, where Q_{ij}, X_{ij}, Φ_{ij}, and Λ_{ij} represent the physical output, input usage, productive efficiency, and product quality for each variety. $M = \sum_{i=1}^{N} M_i$ now represents the aggregate number of varieties produced.[14]

This extension to model multi-product firms is motivated by the common problem concerning the unavailability of variety level data. As is almost always the case with production census data, only plant level aggregate (across varieties) sales $R_i = \sum_{j=1}^{M_i} P_{ij} Q_{ij}$ and input usage $X_i = \sum_{j=1}^{M_i} X_{ij}$ are observable. The estimating equation (4) relating these aggregate plant level variables with productivity then becomes

$$r_{it} - \tilde{p}_t = \tfrac{\sigma-1}{\sigma} \gamma x_{it} + \tfrac{1}{\sigma} \left[(r_t - \tilde{p}_t) - m_t\right] + \tfrac{\sigma-1}{\sigma} \left[\tilde{\varphi}_{it} + \xi \, m_i\right] \tag{9}$$

where $\xi = \frac{1}{\sigma-1} - (\gamma - 1)$ is a positive decreasing function of both σ and γ, $m_i = \log M_i > 0$, and $\tilde{\varphi}_{it}$ is an average of the plant's productivity levels (quality adjusted) across varieties.[15]

The only difference between (9) and (4) is that a plant's productivity now depends - positively - on the number of varieties it produces. With product differentiation, holding the quality of varieties fixed, consumers prefer s bundle of goods with more varieties to one with the same number of units spread over a smaller number of varieties. By splitting their inputs over the production of more varieties, firms can increase the average price of units sold – again, holding quality fixed. Firms weigh this benefit of higher product variety against the sunk cost (not reflected in the overall current input usage) of introducing more varieties. A natural follow-up question is: Does adjusting a firm's productivity for the number of varieties it produces make sense, given the economic definition of

[14] 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.

[15] Melitz (2001) shows that ξ must be positive in any equilibrium where firms choose to produce more than one variety. See sections 4 and 5 for further details.

productivity? 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 $\tilde{\varphi}_{it} + \xi\, m_i$ the standard interpretation of productivity applied to homogeneous good industries[16].

2.4 Time-Varying Firm Productivity and Quality Levels

In this section, we 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, the implementation of the Olley and Pakes (1996) estimation method proposed by Levinsohn and Petrin (2000) – hereafter L-P – can be extended to our case with imperfect competition and product differentiation. We now assume that each firm's productivity and quality index follow a first order Markov process, although we 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.[17] 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

$$\hat{\varphi}_i = \varphi_i + \xi\, m_i.$$

As was previously noted, this variable appropriately indexes firm level productivity in a differentiated product industry with multi-product firms.

Following L-P, we use intermediate inputs as a proxy for the unobserved productivity variable $\hat{\varphi}_i$. We therefore need to ensure that a firm's use of intermediate inputs is a monotonic function of its productivity index $\hat{\varphi}_i$. Melitz (2001) shows that this monotonicity condition will hold so long as more productive firms (with higher $\hat{\varphi}_i$) do not set inordinately higher markups than less productive

[16]See Melitz (2001), section 5, for details.

[17]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.

firms. In this situation, an inordinate markup difference would imply that a productivity increase would lead a firm to increase its markup by such an amount that it would lead to a decrease in the firm's input usage. We consider this to be a quite extraordinary case and rule out this possibility in our empirical work. We also note that the profit maximizing assumptions of monopolistic competition would automatically rule out this special case.

Given this assumption that the monotonicity condition is satisfied, the extension of L-P's estimation strategy to our current framework is straightforward. 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

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

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),[18] the estimating equation becomes:

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

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 $[(r_t - \tilde{p}_t) - n_t]$ (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 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 $\gamma = \frac{\sigma}{\sigma-1}\left(\beta_k + \beta_l + \beta_e\right)$ 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.

[18]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 $\hat{\varphi}_{it}$.

References

Baily, M. N., Hulten, C. and Campbell, D. (1992), 'Productivity dynamics in manufacturing plants', *Brookings Papers on Economic Activity* **0**, 187–249. Microeconomics.

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* .

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **60**(4), 889–917.

Foster, L., Haltiwanger, J. and Krizan, C. (1998), 'Aggregate productivity growth: Lessons from microeconomic evidence', *NBER Working Paper No. 6803* .

Goldberg, P. K. (1995), 'Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry.', *Econometrica* **63**, 891–951.

Griliches, Z. and Klette, T. J. (1996), 'The inconsistency of common scale estimators when output prices are unobserved and endogenous', *Journal of Applied Econometrics* **11**, 343–361.

Kraay, A., Soloaga, I. and Tybout, J. (2001), 'Product quality, productive efficiency, and international technology diffusion', The Pennsylvania State University, mimeo.

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

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

Melitz, M. J. (2000), 'The impact of trade on intra-industry reallocations and aggregate industry productivity', Harvard University, Mimeo.

Melitz, M. J. (2001), 'Estimating firm-level productivity in differentiated produc industries', Harvard University, mimeo.

Nevo, A. (2001), 'Measuring market power in the ready-to-eat cereal industry', *Econometrica* **69**, 307–42.

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.