
Volume 29, issue 3, May 2011
ISSN 0167-7187
ELSEVIER
INTERNATIONAL JOURNAL OF
INDUSTRIAL ORGANIZATION
Editors: B. Caillaud, N. Gandal, J. Hastings
Advisory Board: J. Levinsohn, F. Malerba, M. Riordan, M. Slade, M. Waterson
Co-Editors: J.P. Choi, U. Doraszelski, J. Gans, G.Z. Jin, B. Jullien, J. Holland Mortimer, H. Paarsch, A. Petrin, Y. Spiegel
Associate Editors: V. Aguirregabiria, R. Amir, J. Asker, H. Bar-Isaac, S. Clerides, G.S. Crawford, J. Dana, S. Fisher Ellison, J. Fox, K. Graddy, K. Ho, Y.-f. Fong, A. Goldfarb, T.D. Jeitschko, J. Kastl, F. Khalil, T. Kretschmer, K. Krishna, J. Lerner, S. Markovich, D. McAdams, E.J. Miravete, M. Peitz, B.J. Ruffle, S. Ryan, M. Rysman, P. Schmidt-Dengler, A. Shneyerov, O. Shy, C. Snyder, V. Stango, R. Stenbacka, S. Stern, G. Tan, J. Thisse, N. Vettas, R. Veugelers, J. Wright
Editorial Assistant: A. Simantov-Damti
29 (3) 305–360 (2011)
CONTENTS
Special Issue: Selected Papers, European Association for Research in Industrial Economics 37th Annual Conference, Istanbul, Turkey, September 2–4, 2010
Guest Editor: Luis Cabral
Preface — N. Gandal — 305
Does competition raise productivity through improving management quality? — J. Van Reenen — 306
Do physician incentives affect hospital choice? A progress report — K. Ho and A. Pakes — 317
On the role of empirical industrial organization in competition policy — P. Davis — 323
Competition between exchanges: A research agenda — E. Cantillon and P.-L. Yin — 329
On the anti-competitive effects of quantity discounts — G. Calzolari and V. Denicolò — 337
The impact of the Medicare Rural Hospital Flexibility Program on patient choice — G. Gowrisankaran, C. Lucarelli, P. Schmidt-Dengler and R. Town — 342
Recent progress in the empirical analysis of multi-unit auctions — A. Hortaçsu — 345
Recovering markups from production data — J. De Loecker — 350
Ex ante or ex post competition policy? A progress report — M. Ottaviani and A.L. Wickelgren — 356
Available online at
ScienceDirect
www.sciencedirect.com

International Journal of Industrial Organization 29 (2011) 350–355

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Recovering markups from production data ☆

Jan De Loecker *

Economics Department Princeton University Fisher Hall Princeton, NJ 08540, United States
NBER, United States
CEPR, United Kingdom

ARTICLE INFO

Available online 4 March 2011

JEL classification:
L00
D24

Keywords:
Markups
Production data
Demand estimation

ABSTRACT

In this paper, I discuss, what I call, the Production-Approach to recovering markups. In contrast to the most popular approach in empirical IO, which relies on demand estimation, this approach requires standard production data while allowing for various price-setting models and puts no restrictions on underlying consumer demand. Using production data together with standard cost minimization allows a researcher to obtain markups in a flexible way. After presenting a brief and selective overview of the literature I contrast the production approach to that of the more popular demand estimation approach. This discussion makes it clear that both approaches face important trade-offs and at a minimum empirical economist should have both techniques as part of their toolbox. The hope is that the use of both methods will only depend on the data at hand and the relevant institutional knowledge, paired with the actual research question we are trying to answer.

1. Introduction

Estimating markups has a long tradition in industrial organization (IO) and is important to other fields such as international trade, public economics and macro economics. In general economists and policy makers are interested in measuring the effect of competition and policy changes, such as industrial and trade policies, on market power. Market power is typically measured by the ability of firms to charge prices above marginal costs, and therefore directly feeds into welfare calculations.

The current practice in IO is to obtain markup estimates by first estimating a demand system and adding behavioral assumptions on price setting. These empirical methods rely on the availability of very detailed market-level data with information on prices, quantities sold, characteristics of products and are more recently supplemented with consumer-level attributes.[1] Often, both researchers and government agencies cannot rely on such detailed data, but still need an assessment of whether changes in the operating environment of firms had an impact on markups and therefore on consumer surplus more broadly defined.

In this paper, I discuss, what I call, the Production-Approach to recovering markups which follows the seminal work of Hall (1986). In contrast to the most popular approach in empirical IO which relies on demand estimation, this approach requires standard production data while allowing for various price setting models and puts no restrictions on underlying consumer demand. Using production data together with standard cost minimization allows a researcher to obtain markups in a flexible way. The Hall methodology crucially relies on the insight that the markup can be obtained by considering the difference between the share of an input's expenditure in total cost and total sales. Let us consider the case of labor input. A firm i's share of expenditures on labor in total sales is only equal to the share of expenditures in total costs if the price p_i equals the marginal cost c_i. Any departure from the perfect competitive model will drive a wedge between both shares, as measured by the markup, $\mu_i \equiv p_i/c_i$.

In this paper I briefly discuss a more flexible approach that can relax the original Hall framework by relying on standard production theory which tells us that an input's output elasticity equals that input's revenue share times the relevant markup of the producer. More specifically, I show that it is sufficient to consider the conditional cost function – conditional on input choices for inputs with adjustment costs – to solve for markups as a function of revenue shares and output elasticities. Since revenue shares are directly observed in the data, the problem is reduced to estimating the coefficients of a production function needed to compute output elasticities.

The Production-Approach is, however, not part of the toolbox of most empirical IO economists for reasons I will get to below. In particular, the identification strategy and obtaining an average

☆ This paper is based on both published work and work in progress. I want to especially thank Amil Petrin who discussed an earlier version of work that is cited here (De Loecker and Warzynski 2010) and which substantially improved that paper. Part of this work was presented at EARIE 2010 Istanbul and I thank the organizer, Otto Toivianen, for inviting me and in particular I thank Luis Cabral, Allan Collard-Wexler and Jim Levinsohn for their helpful suggestions and comments. I refer interested readers to De Loecker and Warzynski (2010) for more details and in particular, for a more general discussion of the simplified models discussed here.
* Corresponding author. Tel.: +1 609 258 4219.
E-mail address: jdeloeck@princeton.edu.

[1] See Goldberg (1995) for an application of this for instance. More recently Berry et al. (2004) rely on individual consumer-level data with information on the actual choice set to estimate a random coefficients utility model.

0167-7187/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijindorg.2011.02.002

markup for a given industry over a long period of time were simply not satisfactory for most IO problems. However, in this paper I argue that the availability of micro production data, often publicly available, and the recent advances in econometric techniques make the Production-Approach attractive to IO economists, and other applied micro economists. In particular, recent advances in econometric techniques to estimate production functions have generated a renewed interest in the use of production functions. For the purpose of this short paper I assume those can inform us about the relevant output elasticities.[2]

The paper is organized as follows. I start out by briefly reviewing the relevant literature on the Production-Approach in Section 2, without attempting to do justice to all relevant papers in this research paradigm. In addition, I discuss some outstanding challenges of this approach. Section 3 revisits the challenges in the Production-Approach and presents an empirical framework that generates firm-level markups which are robust to underlying models of competition and demand while relying on fairly flexible production technologies and producer behavior. In a final section, I briefly contrast the Production-Approach to the more popular Demand-Approach in empirical IO by relying on demand estimation to compute markups. The point of this discussion is again by no means to provide an up to date overview, but to simply highlight the set of assumptions required to follow this approach, and which data is needed, and more importantly how it differs from the Production-Approach.

2. The production function approach

I briefly lay out the empirical framework that has been used in one form or another in the literature, and I refer the reader to De Loecker and Warzynski (2010) for a more detailed discussion. I will only introduce the necessary ingredients and hereby make some simplifications. An important one is that I immediately assume that data on firms is available, as opposed to the use of aggregate industry-level data as first used by Hall (1986).

2.1. Towards a simple estimating equation

Robert Hall published a series of papers suggesting a simple way to estimate (industry) markups based on an underlying model of firm behavior (Hall, 1986, 1988, 1990). These papers generated an entire literature that was essentially built upon the key insight that industry specific markups can be uncovered from production data with information on firm or industry level usage of inputs and total value of shipments. The literature also spread to international trade where a great interest lies in analyzing the market power effects of changes in trade policy, and more recently in the effects of increased globalization through direct trade (export and import) and fragmentation of the production chains on firm performance. This approach is based on a production function framework and delivers a markup using the notion that only under perfect competition an input's cost share is equal to its revenue share, and the wedge between both serves as a measure of the relevant markup. This approach set the stage for a separate literature where the very same framework was applied to different settings ranging from privatization and deregulation to innovation, and the interaction with labor market imperfections.

The starting point of this approach is to write down a production function of the following form

$$Q_{it} = F(\mathbf{X}_{it})exp(\omega_{it}) \tag{1}$$

where Q_{it} is output produced by firm i at time t using K inputs $\mathbf{X}_{it}$, where $\mathbf{X}_{it} = \{X_{it}^1, \ldots, X_{it}^K\}$ and $k = 1, 2, \ldots, K$, capturing both intermediate inputs, labor and fixed inputs in production such as capital. The output of a firm is also affected by its level of productivity, ω_{it}, which is modeled as a Hicks neutral productivity shock. We can proceed by considering a first order approximation to the production function by taking first differences over time to arrive at an estimating equation where output growth is related to input growth, where each input is weighted by its cost share. A similar equation was obtained by Klette (1999) by considering the difference of each firm to a representative firm (usually the median) and also relying on cost shares to weigh the various inputs. Both approaches therefore eliminate persistent time invariant productivity shocks from the estimating equation in the hope to reduce the simultaneity bias stemming from unobserved productivity shocks. In both cases we end up with the following expression,

$$\Delta q_{it} = \sum_k \beta_{it}^k \Delta x_{it} + \Delta\omega_{it} \tag{2}$$

where lower cases denote the logs of the relevant variables, relating output growth Δq_{it} to a weighted sum of input growth Δx_{it} and productivity growth $\Delta\omega_{it}$ where the weights β_{it}^k are given by the share of input k's expenditure in total costs.

The last step towards obtaining an estimate of the markup is then to simply observe that for each input k, the cost share β_{it}^k can be written as the product of the relevant markup μ_{it} and the share of the input's expenditure in total sales α_{it}^k.[3] However, the estimation has to be further restricted by either considering a common markup across producers in a given market, or by giving up the identification of firm specific markups and obtaining an estimate of the average markup, or the markup for a specified set of producers by interacting the input growth term with a relevant group variable. The final equation that numerous researchers have taken to the data is then simply

$$\Delta q_{it} = \mu \sum_k \alpha_{it}^k \Delta x_{it} + \Delta\omega_{it} \tag{3}$$

where the interest lies in the markup parameter, μ. The equation above only requires data on output produced, inputs, and revenue shares of the relevant inputs of production. This equation then opens the door for various applications trying to identify whether changes in the operating environment of firms changed markups, on average, by simply interacting the weighted input growth term by the relevant, hopefully, exogenous shifters such as privatization, trade liberalization, R&D and participation in international trade.[4]

2.2. Challenges and outstanding issues: identification and measurement

Eq. (3) highlights some of the features associated with this approach that were undesirable for IO economists. The major shortcoming from an IO point of view is that the average markup is often not a meaningful object since IO is concerned with modeling strategic interactions among firms in settings that clearly depart from a common markup model of imperfect competition. In addition, the identification of the markup parameter is problematic given the potential correlation between input growth and productivity growth. The latter was deemed even more problematic when micro data on production became available and (aggregate) instrumental variable approaches suggested in the literature became irrelevant to deal with the endogeneity of inputs. The use of firm-level panel data

[2] I refer to a recent survey by Ackerberg et al. (2007) and De Loecker and Warzynski (2010) for a detailed treatment of production function estimation in the light of this application.

[3] In Section 3 I will relax the implicit assumption of constant marginal cost, and only rely on an input's output elasticity to be equal to the markups times that input's revenue share. The Hall approach was initially developed under constant returns to scale, but subsequent work relaxed this assumption somewhat.

[4] For applications of this approach see Levinsohn (1993), Harrison (1994), Klette (1999), Konings et al. (2005) among others.

made way for the use of fixed effects estimators, and consequently first differencing of the data, which did not produce satisfactory results and relied on hard to maintain assumptions that productivity shocks are not time varying while increasing the role of measurement error in the estimation procedure. The latter is a direct consequence of taking first differences of the production function.[5] In sum, the Hall approach as outlined above did not provide IO economist with a flexible methodology to recover markups at the firm-level while resting on a sound, theoretically founded identification strategy.

Conditional on being interested in an average markup, or an average across a set of producers, the main concern in identifying the markup parameter comes from the potential of unobserved factors of production to impact output growth. An obvious candidate in the framework of a production function is productivity (growth). Not controlling for unobserved productivity shocks biases the estimate of the markup as productivity is potentially correlated with the input choice. The increased availability of firm or plant-level datasets further boosted empirical studies using some version of the Hall approach on micro data. However, dealing with unobserved productivity shocks becomes an even bigger concern when applying the Hall method to micro data given the strong degree of heterogeneity across producers. Therefore the set of instruments suggested in the literature which were mostly aggregate demand factors such as military spending, and oil prices, was no longer useful to control for the endogeneity of input choice.

The Hall methodology and further refinements became a popular tool to analyze how changes in the operating environment – such as privatization, trade liberalization, and labor market reforms – impacted market power as measured by the change in markups. In particular the availability of panel data allowed researchers to exploit both the cross section and time series dimension to argue the identification of a change in markups. The cross sectional variation in the output–input relationship would in theory allow to identify markups at each point time, while the time series is then used to track its change. This identification strategy is again often problematic since the correlation between the change in competition and productivity potentially biases the estimates of the change in the markup. Let us take the case of trade liberalization. If opening up to trade impacts firm-level productivity, which has been documented extensively in the literature, it is clear that the change in the markup due to a change in a trade policy is not identified without controlling for the productivity shock and how trade policy impacts productivity.[6]

In addition to the standard challenges in measuring output and inputs of production consistently across a set of producers and time, the Hall approach required measuring the user cost of capital in order to compute the appropriate revenue share of the capital input, which requires total expenditures on capital. This has proven to be a very difficult job and requires ad hoc assumptions on capital markets and how firms depreciate their various, and often heterogeneous, assets. An initial assumption to alleviate this problem was to assume constant returns to scale in order to estimate the markup parameter. The latter is clearly not without problems and will impact the identification strategy since every increase in input is assumed to lead to a proportional increase in output, impacting the variation we use to identify the markup parameter.

I take up these challenges in the next section by briefly describing an approach, due to De Loecker and Warzynski (2010), which provides an empirical framework to obtain firm-level markups while explicitly dealing with the simultaneity problem in identifying the production function coefficients, while relaxing the underlying assumptions on costs of production. After presenting a short version of this recent approach, I briefly discuss the dominant approach in empirical IO to obtain estimates of markups using demand system estimation.

3. Towards a more flexible approach[7]

I briefly describe a more flexible approach to recover firm-level markups while relying on the same production data described before. The approach resurrects the potentially big advantages of relying on production data by not having to assume any particular form of consumer demand and any specific model of price setting, while directly addressing the econometric concerns of identifying production function coefficients.

The approach relies on cost minimization and there being (at least) one variable input of production free from adjustment cost for which the wedge between that input's revenue share and its output elasticity is a direct measure of the firm's markup. In what follows I assume firms are single product producers selling in one market, but the framework can be extended to allow for multi-product producers selling in potentially many different markets.[8]

3.1. An empirical framework

Let a firm produce under the same production technology as described in Eq. (1), where now variable inputs free of adjustment costs, X_{it}^V, are treated differently in the analysis than the factors of production facing adjustment costs or dynamic inputs such as capital, X_{it}^F. Note that the latter is conditioned on when considering the optimal demand for X_{it}^V. For instance, if a firm's capital stock was decided before time t, optimal demand for variable inputs at time t depends on all the state variables of the firm's problem which were optimized at $t-1$, including capital.

Let us assume that producers active in a given market are cost minimizing. Consider the associated Lagrangian of the firm's cost minimization problem, where P_{it}^V and P_{it}^F denote a firm's input price for variable and fixed inputs, respectively, and λ_{it} the Lagrangian multiplier. Taking the first order conditions with respect to the variable inputs without adjustment costs, while conditioning on the use of inputs facing adjustment costs, and rearranging terms and multiplying both sides by $\frac{X_{it}^V}{Q_{it}}$, generates the following expression,

$$\frac{\partial F\left(X_{it}^V\right)}{\partial X_{it}^V}\frac{X_{it}^V}{Q_{it}} = \frac{P_{it}^V X_{it}^V}{\lambda_{it} Q_{it}} \tag{4}$$

where λ_{it} measures the marginal cost of production at an output level Q_{it}, where $\frac{\partial L}{\partial Q_{it}} = \lambda_{it}$. Note that this FOC conditions on the use of dynamic inputs of production such as capital, and potentially

[5] I refer to Griliches and Mairesse (1995) for a discussion and related identification issues. In the context of markup estimation, Klette (1999) introduced an improved GMM estimator while relying on additional moments, by applying dynamic panel data techniques.

[6] The same is true in the case where we want to estimate the productivity response to a change in the operating environment such as trade liberalization. See De Loecker (forthcoming) for more on this.

[7] I would like to thank Amil Petrin for his comments and suggestions when discussing an earlier version of De Loecker and Warzynski (2010), on which part of this section rests.

[8] Note that for multi-product firms, the same FOC applies for each product, and in principal product specific markups are obtained in a similar fashion. However, this approach will require product specific input data, and output elasticities, or extra structure. In particular, additional structure is needed on the cost and demand functions of the multi-product firm. In contrast, the demand approach does not require restrictions on the cost function for multi-product firms. I leave the explicit treatment of the multi-product case for future research.

other inputs facing adjustment costs, captured by X^F_{it}. It is the use of this conditional cost function that will allow us to uncover a firm's markup, as cost minimization implies that optimal input demand is satisfied when a firm equalizes the output elasticity of an input X^V_{it} to $\frac{P^V_{it}X^V_{it}}{\lambda_{it}Q_{it}}$.[9]

A final step to obtain an expression for the markup μ_{it} is to simply define it as $\mu_{it} \equiv \frac{P_{it}}{\lambda_{it}}$. This expression is robust to various (static) price setting models, and does not depend on any particular form of consumer demand. The markup will, however, depend on the specific nature of competition among firms. It is important to realize that in this approach only the definition of the markup is used. Using this definition, the FOC relates the output elasticity to the firm's markup and the input's revenue share,

$$\frac{\partial F}{\partial X^V_{it}}\frac{X^V_{it}}{Q_{it}} = \mu_{it}\left(\frac{P^V_{it}X^V_{it}}{P_{it}Q_{it}}\right). \tag{5}$$

An expression for the markup μ_{it} is then simply obtained by multiplying the output elasticity by the input's inverse revenue share. In order to obtain a measure of firm-level markups using production data only estimates of the output elasticity of one (or more) variable input (s) of production and data on expenditure shares are needed.

The output elasticity is obtained by estimating a production function, whereas the revenue shares are directly observed in the data, and therefore we are left with the task of estimating the output elasticities of the underlying production functions, on which a large literature exists. As mentioned before, this literature relied on IV and fixed effect estimators which produced unsatisfactory estimates of the production function coefficients, and for productivity. However, a recent literature initiated by Olley and Pakes (1996) and subsequent work of Levinsohn and Petrin (2003) introduced a full behavioral model to solve for unobserved productivity as a function of observed firm-level decisions (investment and input demand) to deal with the endogeneity of inputs when estimating a production function. De Loecker and Warzynski (2010) discuss the identification and estimation of a flexible production function while explicitly controlling for unobserved productivity shocks, and allowing for an endogenous productivity process. The method relies on a proxy estimator which is consistent with imperfect competition, and produces firm-level output elasticity estimates. In addition, an estimate of unanticipated productivity shocks and measurement error in output is obtained which is important to correctly compute revenue shares from the data, which allows for measured output in the data to be different from the output a firm predicted to produce given its productivity shock and input usage.[10]

For the purpose of this paper it is sufficient to state that obtaining consistent estimates of the output elasticities of the production function follows a careful implementation of those proxy estimators, where knowledge of the institutional details of the industry should govern the estimation routine. Importantly, a great deal of flexibility is available to the applied researcher while not adding additional behavioral assumptions to the already assumed cost minimization to derive an expression of markups. In fact cost minimization gives rise to optimal input demand equations which are used to invert out unobserved productivity shocks and replace productivity by intermediate inputs, i.e. directly applying the insight of Levinsohn and Petrin (2003).

3.2. Revisiting Hall's approach

The framework suggested by De Loecker and Warzynski (2010) produces firm and time specific estimates of markups. I illustrate their approach by discussing a simple and special case, which allows me to revisit the standard approach discussed in Sections 1 and 2. In particular, I consider the special case where markups are constant across firms and time, $\mu_{it} = \mu$. In this way I can retrieve the original Hall specification while explicitly correcting for the endogeneity of inputs, and therefore providing consistent estimates of the (constant) markup. Furthermore, let me assume there are no adjustment costs in labor, and let production be given by a Cobb–Douglas value added production function, which in logs is given by

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it} \tag{6}$$

where y_{it} is the log of measured output in the data. The difference between measured output in the data and in the theoretical framework is due to unanticipated shocks to productivity, measurement and prices.

A final step in recovering the Hall framework is to directly impose the first order conditions from cost minimization. However, as emphasized above, the FOC only applies to variable factors of production and therefore I recover the corrected version of Hall that does not rely on the FOC on capital. The Cobb–Douglas technology assumption implies that the output elasticity of labor β_l does not vary across firms, which together with the assumption that markups are constant then implies that revenue shares have to be constant across firms, since $\beta_l = \mu\alpha^L$. This implies that estimating the following equation is sufficient to get an estimate for μ from the data,

$$y_{it} = \mu\left(\alpha^L l_{it}\right) + \beta_k k_{it} + \omega_{it} \tag{7}$$

where α^L is the wage bill over sales ratio. Under the assumptions of Cobb–Douglas technology and a constant markup, there cannot be any meaningful variation in the revenue share, which is at odds with the micro data. At a minimum this suggests that empirical work should allow for either variation in output elasticities by deviating from Cobb–Douglas, or by considering models of imperfect competition with markup heterogeneity. The framework introduced by De Loecker and Warzynski (2010) allows for both.

It is worth emphasizing that the constant markup condition can either be imposed through economic theory, such as considering a constant elasticity of demand model, or by restricting the goal of the estimation routine to estimate the average markup. Both constraints lead to the same estimating equation but the identification of the parameter μ is quite different. Eq. (7) further highlights that capital is no longer assumed to be a variable input since the first order condition is not used. However, the estimated coefficient on the capital stock could in principal inform us about the extent to which firms face costs in adjusting their capital stock by relying on the estimated markup obtained from the FOC on labor.

In order to obtain a consistent estimate of the markup the same proxy methods can be used to control for unobserved productivity

[9] Note that in the special case where marginal cost is constant across all levels of output, the output elasticity is, only then, equal to the input's cost share. The constant marginal cost assumption was implicitly introduced in the original Hall article by working under constant returns to scale in production. Under that assumption, the markup could in theory be measured by directly comparing revenue and cost shares.

[10] I suppress a discussion on the problems introduced by using revenue data to proxy for output in the context of production function estimation, and refer to De Loecker (forthcoming) and references therein. However, for the purpose of this paper it is important to note that price errors will not impact the ability to identify markups and more importantly their relationship with firm-level characteristics and market-level variables. In fact for a Cobb–Douglas production function, no variation in output elasticities exists, and therefore all that matters for the variation in markups over time and producers is the variation in revenue shares.

which will imply that moments have to be formed on the shocks of the productivity process (ξ_{it}). In the case of an exogenous productivity process where $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$, the following moments are used to identify the markup

$$E\left(\xi_{it}(\mu, \beta_k)\begin{pmatrix} l^*_{it-1} \\ k_{it} \end{pmatrix}\right) = 0 \tag{8}$$

where $l^*_{it} = \alpha^L l_{it}$ and relies on the fact that labor is a variable input into production and therefore lagged labor should not be correlated to current shocks in productivity. The capital stock was formed prior to time t and therefore it should not be correlated with the productivity shock, ξ_{it}, at time t.

Although a simplified case, this example shows that the identification strategy simply restates which factors of production face adjustment costs. Therefore, a great deal of flexibility is available to the applied researcher by considering production functions with various variable inputs such as energy use, intermediate inputs and in some cases labor, and test for overidentifying restrictions.

4. Alternative approach: demand estimation

Over the last two decades industrial organization has made substantial progress on the estimation of demand systems (see Ackerberg, et al., 2007 for a specific overview). The current generation of empirical models allows for consumer and product specific elasticities of demand while relying on a random utility framework. An alternative approach to estimating markups presents itself using the estimates of the demand system (own and cross price elasticities). Data on prices, quantities and characteristics of the product and consumers, together with behavioral assumptions on price setting of firms delivers estimates of markups.

In contrast to the production approach, this approach relies on the ability to observe very detailed data and requires different modeling assumptions.

4.1. Data requirements

The literature on demand estimation has moved towards modeling demand in characteristics space (for instance Berry et al., 1995) in order to obtain more realistic price elasticities, and hence implied markups and welfare measures. Although desirable from a modeling perspective, this approach requires a lot of data. In addition to prices and quantities consumed, this approach relies on observing the relevant characteristics of products as well as consumer characteristics, such as income and other relevant demographics. This data constraint implies that our analysis is often constrained to a specific market in a specific location at a given point in time. The production approach on the other hand has the potential to be informative about market power across a wide range of markets, industries and time.

4.2. Model assumptions

Demand estimation relies explicitly on rational consumers making utility maximizing choices while observing the entire relevant choice set of products.[11] With this in place we can learn about the price sensitivity of demand from relating product market shares to prices, while controlling for product and consumer characteristics, where price endogeneity is obviously a major concern in demand estimation, just like the simultaneity bias is in production function estimation. This generates a set of demand elasticities which are used to compute markups. An important additional constraint is that in order to compute markups a specific price setting model has to be introduced. For instance, if firms compete in a Bertrand product differentiation fashion, the relevant markup is given by $\left(1 + \frac{1}{\eta_i}\right)^{-1}$, where η_i is the firm's residual demand elasticity, which we can calculate from the set of estimated demand coefficients. This highlights that we can only recover markups from demand estimation by making explicit assumptions on price setting behavior. This is in contrast to the production approach where we only have to rely on the definition of the markup as the price over marginal cost.

Finally, the demand estimation approach relies on consumers maximizing utility in a static setting, i.e. their past and future choices do not impact current purchasing decisions. The introduction of dynamic demand systems in industrial organization has been the topic of recent research (Gowrisankaran and Rysman, 2009; Hendel and Nevo, 2006, among others). Incorporating the dynamic choice of consumers turns out to be quite challenging and often additional assumptions are required to alleviate some of the computational problems. In sharp contrast, the production approach does not take a stand on how consumers make their purchasing decisions. As discussed above, it does rely on a different static setting, i.e. we require the firm to have control over at least one variable input which is free of adjustment cost in order for an input's revenue share to be informative about the firm's markup.

4.3. Objectives

An important advantage of the demand approach, however, is that researchers can rely on the estimated demand coefficients to run counterfactual exercises for a variety of important policy evaluations, such as merger analysis, product introduction, and trade policy among others.[12] It is also important to note that the obtained markups from both approaches are typically not directly comparable. The markups obtained from demand estimation are informative about margins at the final consumption level, whereas the production approach delivers markups at the level of production and therefore do not include the additional margins that are added further down the distribution chain. However, more recently demand system estimation is used to discipline the analysis of contracting between agents in a bargaining setting, and in vertical relationships along the production chain. In these settings, it is expected that both approaches generate comparable estimates as both approaches are valid and data availability should help selecting among them.

I want to stress that both approaches are inherently complementary and the use of both approaches should depend on the data at hand, and the research question. This paper does highlight the value of using production data in shedding light on markups, and allows us to analyze how markups vary with changes in the operating environment of firms, or with firm-level decision variables such as innovation, exporting, and product introduction, to name but a few.

5. Concluding remarks

This paper briefly discussed, what I feel is, an often overlooked approach in empirical IO to obtain markup estimates by relying on standard micro production data. After presenting a brief and selective overview of the literature I contrasted the production approach to that of the more popular demand estimation approach. This discussion made it clear that both approaches face important trade-offs and at a minimum empirical economist should have both techniques as part of their toolbox. The production approach frees up the data requirements while not having to take a stand on consumer preferences and the exact model of price setting of market participants. The latter is in

[11] In addition distributional assumptions are made on the idiosyncratic errors of the random utility model.

[12] Applications of this are for instance Nevo (2000), Petrin (2002), Goldberg (1995) and Berry et al. (1999).

particular often impossible to test, and remains a challenge for that literature. On the other side, the production approach relies on cost minimization behavior of producers and generates estimates of firm-level markups, and is not directly suited for counterfactual exercises often carried out in IO, such as merger simulations for instance. The hope is that the use of both methods will only depend on the data at hand and the relevant institutional knowledge, paired with the actual research question we are trying to answer.

Finally, let me end by suggesting some future work in this area. First, as mentioned before, this method can in principal be extended to the case of multi-product producers, selling in multiple markets. However, this requires a treatment of the multi-product cost function and observing input usage of a producer broken down by product or market; or assumptions thereof. Secondly, an important next step in this research program is to compare markup estimates across both methods on data capturing both production data, and the relevant variables for demand estimation such as prices, quantities and product-consumer characteristics. This would allow us to evaluate the different sets of assumptions of both approaches, and obtain robust estimates of price cost margins. To this end, De Loecker and Scott (in progress) started to collect both standard demand (scanner data) and production data (census) for a few selected products across US industries to compare markups across both methods.

References

Ackerberg, D., Benkard, L., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. Handbook of Econometrics 6, 4173–4276.
Berry, S., Levinsohn, L., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.
Berry, S., Levinsohn, L., Pakes, A., 1999. Voluntary export restraints on automobiles: evaluating a strategic trade policy. American Economic Review 89 (3), 400–430.
Berry, S., Levinsohn, L., Pakes, A., 2004. Differentiated products demand systems from a combination of micro and macro data: the new car market. Journal of Political Economy 112, 68–105.
De Loecker, J., forthcoming, Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity, Econometrica.
De Loecker, J. and Scott, P.T., in progress. Comparing Markup Estimates: Scanner Data versus Producer Data, Princeton mimeo.
De Loecker, J. and Warzynski, W., 2010. Markups and Firm-level Export Status, Princeton mimeo (NBER # WP 15198, CEPR WP # 7450).
Goldberg, P., 1995. Product differentiation and oligopoly in international markets: the case of the U.S. automobile industry. Econometrica 63 (4), 891–951.
Gowrisankaran, G. and M. Rysman, 2009. Dynamics of Consumer Demand for New Durable Goods, Arizona University mimeo.
Griliches, Z., Mairesse, J., 1995. Production functions: the search for identification. NBER WP 5067.
Hall, R.E., 1986. Market structure and macroeconomic fluctuations. Brookings Papers on Economic Activity 2, 285–322.
Hall, R.E., 1988. The relation between price and marginal cost in U.S. industry. Journal of Political Economy 96 (5), 921–947.
Hall, R.E., 1990. Invariance properties of solow's productivity residual. In: Diamond, P. (Ed.), Growth, productivity, Unemployment: Essays to Celebrate Bob Solow's Birthday. MIT Press, Cambridge, MA, pp. 71–112.
Harrison, A., 1994. Productivity, imperfect competition, and trade reform: theory and evidence. Journal of International Economics 36, 53–73.
Hendel, I., Nevo, A., 2006. Measuring the implications of sales and consumer inventory behavior. Econometrica 74 (6), 1637–1673.
Klette, T.J., 1999. Market power, scale economies and productivity: estimates from a panel of establishment data. Journal of Industrial Economics 47 (4), 451–476.
Konings, J., Van Cayseele, P., Warzynski, F., 2005. The effect of privatization and competitive pressure on firms' price-cost margins: micro evidence from emerging economies. Review of Economics and Statistics 87 (1), 124–134.
Levinsohn, J., 1993. Testing the imports-as-market-discipline hypothesis. Journal of International Economics 35, 1–21.
Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 317–342 April.
Nevo, A., 2000. Mergers with differentiated products: the case of the ready-to-eat cereal industry. The RAND Journal of Economics 31 (3), 395–421.
Olley, S.G., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–1297.
Petrin, A., 2002. Quantifying the benefits of new products: the case of the minivan. Journal of Political Economy 110, 705–729.