ECON8854 - Industrial Organization II

Lecture Notes from Charlie Murry and Michael Grubb's lectures

Paul Anthony Sarkis
Boston College

Contents

Chapter 1

Production Function Estimation

1.1 Introduction

The firm-level production function can be written as:

$$Y_{it} = L_{it}^{\beta_l} K_{it}^{\beta_k} U_{it}$$

where Y is output, L is labor, K is capital and U is the TFP (usually unobserved by the researcher). The goal is to estimate the output elasticities β!

But estimation is not straightforward, some issues arise from the very nature of production. First of all, the production process is dynamic since it uses capital, which can grow or depreciate endogenously (it is a decision based on production). Thus, state variables will be very important (R&D, entry/exit, etc.). Then, we have a problem of simultaneity: variables that are unobserved to the econometrician might be determined by other equations. To see that, consider taking the logs of the production function, decomposing the unobserved TFP into a structural error ω and a random error ε:

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

It might be that l_{it} or k_{it} are chosen with knowledge of ω_{it}, even though it is typically safe to assume that k_{it} is chosen at $t-1$ and is thus free from endogeneity. Finally, we have issues of selction bias since using panel data will give extra weight to successful firms that "survive" throughout the dataset. However, it might be the case that the firms that do not survive have low productivity draws on average: selection bias!

1.2 Olley and Pakes (1996)

1.2.1 Intuition

This paper suggests a way to go around the simultaneity issue (as well as selection bias) by providing a "proxy" for the structural term in the TFP. This proxy is current investment (denoted i_{it}). Intuitively, the argument relies on the fact that while investment is directly correlated with productivity shocks, it will not affect labor or capital until the next period!

This method works only under a set of assumptions:

1. Information: the firm's information set includes only current and past realizations of ω and ε is exogenous ($\mathrm{E}\left[\varepsilon_t | I_t\right] = 0$).
2. First-order Markov processes: the TFP shock follows a first-order Markov process such that $\Pr\left[\omega_{i,t+1} | I_{it}\right] = \Pr\left[\omega_{i,t+1} | \omega_{it}\right]$.
3. Timing of investment: firms accumulate capital according to the law $k_{it} = \kappa(k_{i,t-1}, i_{i,t-1})$. Labor is not dynamic.
4. Scalar unobservable: firms' investment decisions are functions of current capital stock and productivity shock ($i_{it} = f_t(k_{it}, \omega_{it})$).
5. Strict monotonicity: $f_t(\cdot)$ is strictly increasing in ω_{it}.

In particular, assumptions 2 and 5 are crucial to allow for "inverting" investment into TFP, meaning that we can write ω_{it} as a function of k_{it} and i_{it}:

$$\omega_{it} = h(k_{it}, i_{it})$$

Then, by plugging it back into the production function we get:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + h(k_{it}, i_{it}) + \varepsilon_{it}$$

and since h is unknown, we cannot differentiate between the use of k within h and outside of h, thus we effectively use:

$$y_{it} = \beta_l l_{it} + \Phi(k_{it}, i_{it}) + \varepsilon_{it}$$

where Φ is unknown (to be estimated).

1.2.2 Estimation

The econometric procedure relies on two stages: in the first stage, the goal is to recover β_l and Φ, the former is already identified but the latter is not an object of interest per se, thus we only need to be most flexible possible. Then, in the second step, we recover β_k by using our first-stage estimates.

First stage

The first-stage regression is defined as:

$$y_{it} = \beta_l l_{it} + \Phi(k_{it}, i_{it}) + \varepsilon_{it}$$

where the goal is to estimate β_l and Φ as flexibly as possible. To do this, there are two main ways: (1) by "parameterizing" $\Phi(\cdot)$ as a polynomial of k and i or (2) using a semi-parametric approach as in Robinson (1988) where β_l is estimated parametrically and Φ nonparametrically.

Second stage

We can decompose ω into its expected value plus an innovation term ξ_{it}:

$$\omega_{it} = \mathrm{E}\left[\omega_{it} | I_{i,t-1}\right] + \xi_{it} = \mathrm{E}\left[\omega_{it} | \omega_{i,t-1}\right] + \xi_{it} = g(\omega_{i,t-1}) + \xi_{it}$$

With $\hat{\Phi}(k_{it}, i_{it})$ and $\hat{\beta}_l$ in hand, we can rewrite $\omega_{i,t-1}$ as a function of estimated parameters and thus, given a guess for β_k and a parametric form for $g(\cdot)$, we have:

$$\xi_{it}(\beta_k) = \phi_{it} - \beta_k k_{it} - g(\phi_{it-1} - \beta_k k_{it})$$

which first conditional moment can be used to recover β_k. Usually, the moment condition used is $\mathrm{E}\left[\xi k\right] = 0$.

1.2.3 Results

1.3 Levinsohn and Petrin (2003)

1.3.1 Intuition

The key insight behind Levinsohn and Petrin is to use materials instead of investment as a proxy for the TFP. In fact, it seems fair to assume that the current choice of intermediate inputs is correlated with the current productivity shock. Moreover, the authors argue that investment was not a good proxy to begin with because (1) too many zeroes in the data and (2) investment is lumpy and thus might not react to productivity shocks right away. In contrast, intermediate inputs such as materials will be more flexible and have more variation.

To include materials in the regression, it helps to think of it as we did with capital in OP. First, the production function is now:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \varepsilon_{it}$$

Then, under similar assumptions as in OP, we can "invert" the materials function to get $\omega_{it} = h(k_{it}, m_{it}$. Finally, as in OP, we are now able to identify the production function as:

$$y_{it} = \beta_l l_{it} + \Phi(k_{it}, m_{it}) + \varepsilon_{it}$$

1.3.2 Estimation

The estimation procedure follows the procedure in OP, with the addition of a moment condition in the last step. In fact, since we now care about β_k and $\beta_,$, we need an extra moment. Note that we cannot use a similar moment as with capital since current materials are correlated with the productivity shock ξ, thus we need to find an "instrument" in the past realization of materials (m_{it-1}).

1.4 Ackerberg, Caves and Frazer (2015)

1.4.1 Intuition

Ackerberg, Caves and Frazer (2015) raise an important issue within the two previous approaches. In fact, considering how the production function is parametrized, once conditioned on k, ω and m (in LP), the amount of l is completely determined! Essentially, whenever l can be written as a function of only k, m and t, then β_l will not be identified.

To go around this issue, ACF suggests to alter how we think about the timing assumptions. In particular, they assume l is set at time t and k at $t-1$ as usual, but m is now set in between at $t-b$ where $b \in (0,1)$. This implies that at the time when k and m are chosen, there is still a random element that is not known to the firm and will affect production. Moreover, ACF considers the value added production function instead of the gross production. This implies that m can be used to invert productivity but will not end up in the final moment condition.

1.4.2 Estimation

Again, the estimation procedure is very similar to the two previous approaches, with the difference that value added production is preferred to gross, and the additional moment condition is now $\mathrm{E}\left[\xi_{it} l_{it-1}\right]$ in order to recover β_l.

1.5 De Loecker and Warzynski (2012)

1.5.1 From production function to markup

1.5.2 Results

1.5.3 Criticisms

Chapter 2

Endogenous Products

Chapter 3

Common Ownership

Chapter 4

Dynamics

Chapter 5

Nonlinear Pricing

5.1 Two types

This section covers the model of nonlinear pricing where a monopolist offers a menu of two quantity/quality-price pairs to consumers of two types (H, with high value for the good and L with low value).

Without loss of generality, we consider contracts such that each type chooses a contract and has no incentives to deviate. The monopolist chooses a quantity/quality and a price for each contracts, such that its profits are maximized. We get a problem with three elements: (1) an optimization problem, (2) a set of participation constraint (making sure each type buys the contract) and (3) an incentive constraint (making sure no type deviates).

The results of this model tells us that the optimal contracts are designed in such a way that the quantity/quality of the highest type is not distorted, while for the lowest type, they will receive a lower quantity/quality than their first best!

5.2 Continuous types

The continuous types model has the same structure as the one presented in the previous section, however, now types lie on a continuum from lowest to highest. As before, the model is separated in the three same parts and display a similar distrortion where consumers of the highest types get their first best contract, while the lowest types get lower quantity/quality.

5.3 Crawford and Shum (2007)

5.3.1 Summary

The research question of this paper is: "To what extent is quality degradation prevalent in cable TV markets? And what are the effects of regulation on this issue?"

They add quality to the decision of the monopolist (add a dimension). Because of imperfect competition, we might think that quality will also be distorted (as prices are), which would create welfare losses. This is the framework of Mussa-Rosen, which is applied to the setting of cable TV.

Mussa-Rosen

There are three types of consumers (one to allow some consumers to not care about cable TV), and two contracts. The model displays distortion for the lowest types and no distortion at the top. They further go to show that this result holds even if consumer types are continuous if qualities are discrete.

Regulation is set up as a constraint on the optimization problem such that quality cannot go lower than a certain point. This turns out to only restrict the lowest quantity, while the top quantity stays undistorted.

Cable TV industry

Contracts are based on bundles of networks. Basic service is the one that everyone has, then you can buy extended service or premium (maps well to theory presented before, but premium is ignored because horizontal differentiation).

Bundle quality is measured in two ways: number of networks in said bundle (assuming same underlying quality) or through the implied values from consumer distribution (a first-stage problem from Mussa-Rosen).

Empirical model

Results

Quality distortion is present. Regulation mitigates the problem.

5.3.2 Discussion

The theory clearly lacks horizontal differentiation between contracts! Because of this, they ignore some data.

Functional form dictates a lot of the results!

Chapter 6

Price Discrimination

6.1 Dana (1999)

6.1.1 Model

- Consumers on a continuum of types with unit demand.
- Firm has marginal production and capacity cost.
- Firm chooses price before learning demand.
- Proportional rationing imply all types have access and buy to the good proportionately.

6.1.2 Residual Demand

- Residual demand is very important:
  - Start with any price, say $\tilde{p}$: some people buy, some don't, say $\tilde{q}$ have bought.
  - Residual demand at another price, say p is not simply base demand - $\tilde{q}$, because some of those who bought at $\tilde{p}$ would not have bought at p = use proportional rationing to determine residual demand

6.1.3 Perfect Competition

- Competitive market = profit is 0 (but probability!)
- Price dispersion: from $\underline{p}$ to $\bar{p}$

6.1.4 Monopoly

- Market power imply markup: prices support is narrower!

6.1.5 Price Dispersion and Market Structure

- Two results:
  - Support of prices widens with competition.
  - Variance of prices increases with competition (given linear demand).
- In summary: price dispersion increases with PTR (which increases with competition).

6.2 Leslie (2004)

6.2.1 Background

The paper answers the question of welfare effects of price discrimination (consumer = ambiguous; firms = positive). Broadway play where second and third degree price discrimination. Second degree happens because of different seats are offered at different prices based on quality (nonlinear pricing as in Dana (1999)). Third degree is targeted coupons. Finally, discount sales for day-of-performance tickets is damaged goods. Same marginal cost for all seats but capacity costs (again, as in the Dana paper).

6.2.2 Data

Unit of observation is a seat (price, quality, discount, etc). Aggregate discount into two categories (coupons or booth). Aggregate advertising is observed. Competing plays attendance is observed. Finally, consumer variables in the NYC region are observed.

6.2.3 Model

Product space approach.

6.2.4 Assumptions

6.2.5 Results

Main result is that price discrimination improves welfare! But all types of PD are suboptimally designed.

6.2.6 Discussion

Chapter 7

Search

7.1 Introduction

Search is an alternative explanation to observing different market shares, enormous marketing budgets, etc. than simple product differentiation.

Search is defined as looking for another price "quote" for a homogenous product. Firms simultaneously choose prices, then consumers search among a number of prices to find the best one and purchase.

If search is sequential, then consumers would search until finding a price less than or equal to a reservation price r such that the expected benefit of finding r is equal to the search cost.

Diamond (1971) shows that this model is very interesting in that it leads to different results than typical Bertrand competition when $s > 0$, the search cost is greater than 0. However, it does not converge to a Bertrand model as $s \to 0$.

7.2 Other models

7.2.1 Diamond (1971)

- Main result: unique NE is monopoly prices and no search!
  - Independent of the number of firm and search cost (provided $s > 0$).
  - Depends on inability to advertise price cuts.

7.2.2 Stahl (1996)

- Add a portion (μ) of consumers as shopping-lovers $\Rightarrow$ price dispersion!
  - No pure-strategy NE!
  - But very challenging to compute…

7.2.3 Simpler model

- Simplification of Stahl by adding: inelastic demand + fixed search costs.
  - The upper bound of prices ensures that only shoppers search.
  - The lower bound of prices is marginal cost + a markup that depends negatively on the proportion of shoppers and positively the size of the search cost.

7.3 Salz (2017)

7.3.1 Summary

Background

Search costs in decentralized markets are important and incentivize the role for intermediaries to enter the market. But how do they influence market outcomes? Empirical study of the NYC trade waste industry. Companies generate waste and have to comply by law to finding a waste management service. Brokers might help with this process by awarding contracts on a competitive bidding process. Given that search costs are about 11-35% of expenses, what is the effect of brokers on the rents in that market?

Data

Observation unit is a contract between company and waste management (from regulatory institution). Variables observed include zip code, negotiated price, quantity of waste and whether or not it was brokered. Some markets are very concentrated while others are very competitive.

There is a lot of price dispersion (even with the same observables), implying search intermediation is useful.

Model

Sequential game between customers and carters, where brokers are non-strategic. Customers observe search costs, carters observe their search cost (this is private information). The game plays as follows:

1. Customers learn his search cost and carters learn their costs.
2. Customers decide between delegating search or performing a number of search (not sequential).
3. Carters submit prices to either broker (given a first-price auction) or to customers directly.

Assumptions

- Simultaneous search rather than sequential.
- No broker competition.
- Contract length is fixed to two years: underestimation of search costs.
- Number of bidders is fixed to observed winners: underestimate number of bidders $\Rightarrow$ search costs are underestimated.

- Carters are grouped into types.

Results

Intermediaries create benefits to both customers who use brokers and those who do not.

Chapter 8

Switching Costs

8.1 Effects

8.1.1 Model 1: Investing and Harvesting

- Firms "invest" in consumers by lowering their prices below marginal cost, then "harvest" their loyal base by extracting all surplus!
  - Only if switching cost is high enough and perfect competition!
  - Welfare is intact (average price is still marginal cost)
- No switching in equilibrium.

Model 1a: Add heterogenous values

- Now there is a DWL: consumers with high value of the good lose!
- No switching in equilibrium.

Model 1b: Add heterogenous switching costs

- Identical welfare outcomes but more switching (those who have low draws of switching costs).

8.1.2 Model 2: Hotelling Duopoly

- Distance between consumer taste and firm is key to compute demand.
  - Consumer has $\theta \in [0, 1]$ in both periods, firms are at 0 and 1.
  - Budget is B.
  - Transportation cost is τ.

Model 2a: Taste is known and stable

- As if no switching costs (same equilibrium as pure Hotelling).
- No switching in equilibrium.

Model 2b: Taste in second period is unknown

- Average price is half of previous model!
  - This is because when the decision is made, products are not differentiated (taste unknown), thus firms need to "invest" harder.

Model 2c: Consumer myopia

- Same as model 2b!
  - Same intuition, myopia means that second period does not come into decision, thus attracting consumers is harder.

8.2 Measurement

8.2.1 Measuring switching costs or inertia

- Challenge: is it switching costs or persistent heterogenous preferences?
  - Osborne (2011): look at previous period event's effect on current period (price cut increases probability to buy again? = switching costs).

    - Handel (2013): look at difference between old and new consumers (if old buy more than new = inertia)

8.2.2 Decomposing inertia

- Inertia is: switching costs, habit formation, search costs, learning, inattention, etc.
    - Usually only one form is considered (others assumed away) = bad (Wilson, 2012)

8.2.3 Comments on Wilson (2012)

- Insight: a 1$ search cost has more impact than a 1$ switching cost because:
    - Searching will cause the cost with probability 1, while switching is only incurred if better option available
    - Potential of multiple searches, while only one switch
    - Inversting from firms might help consumers offset cost
- Thus, design a "quick and easy" method for estimating search and switch costs:

8.3 Honka (2014)

8.3.1 Summary

Background

Insurance markets can be inefficient for many reasons, two of them being search costs and switching costs (market frictions). What is their value in the auto insurance market? And do they affect consumer choice and welfare? are the two questions Honka (2014) answers. Auto insurance market is perfect in the sense that it includes both (very high retention rate). Honka puts both types of costs in her model: new thing in the literature‼

Data

the observation unit is a transaction, with contract observables (prices, premium, etc.), consumer observables (previous contract, number of quotes, by whom, and demographics, etc.). Consideration set is very important to identify search costs, previous contract to identify switching costs (but no panel data).

Model

Multinomial logit model with a search + previous insurer consideration set. There a "first-step" search model based on expected utility to decide how many searches.

Estimation

Simulated semi-parametric way of recovering number of searches (analogous to ordered probit?). With consumer beliefs, recover search costs. Finally, switching costs are inside RUM model.

Assumptions

Main model assumptions are: (1) search and purchase are conditional on coverage (no search across coverages); (2) search is used to discover prices only and (3) search is simultaneous rather than sequential.

Other assumptions include: static preferences (utility fixed effect is not correlated with previous choice); switching costs are not exactly identified (confounded with heterogenous persistent preferences.

Results

Search costs are three times higher than switching if made in person, but comparable if made through internet. Search costs are what drives most of the retention.

Both have negative effects on welfare.

Chapter 9

Learning

9.1 Literature Review

9.2 Covert (2015)

9.2.1 Summary

The research question in this paper is: "Do firms learn (in production)?"

Using data on hydraulic fracturing in the Bakken Shale and a model of input choice under technology uncertainty, Covert shows that firms only learned partially, leaving out 40% of profits in the process.

Background

New industry after advances on how to extract shale gas. Almost 1000% growth in 8 years.

Fracking is pumping a mix of sand, water and chemicals in the ground = choice to make on the "recipe" that affects production and costs! But no one knew at the time how to do it = opportunities for learning.

Firms have their own data (private for 6 months) and then get access to other data (after six months).

Evidence for learning

Covert looks for three types of learning: (1) is experience (age of firms) correlated with productivity (oil per well drilled)? Which is estimated using a Benkard type of model. (2) is the choice of inputs more profitable over time? which is estimated using ex ante and ex post profits comparison.

Results

One of the first empirical analyses of learning behavior in production. Find that firms increased the profit capturing rate from 20% to 60%. No experimenting to learn as firms go to more certain input choices. Firms overweight their data compared to other firms' data.

9.2.2 Model

Production function

The output is log-log specified as a function of:

- t: the number of days of operation of the well.
- D: the number of days of production.
- H: length of the well.
- Z: other topologic controls
- ϵ: well-specific shock
- ν: idiosyncratic shock

Profits

Firm's profits depend on the usual stuff: share, market size, price, and cost of inputs.

Preferences

Firms get utility from the mean profit and standard deviation of profits given an input.

Gaussian Process Regressions

9.2.3 Comments

Results

Firms underutilize sand and water but: are costs measured correctly? If costs are convex, then findings corroborate optimal choice? Overestimate return to production?

Public policy

Delayed disclosure of information lowers barriers to entry while leaving rents on the table.

Experimenting

Nice framework but conclusion too hasty? Experimenting is too linked with risk, what if even when experimenting firms would choose safe levels of inputs? Or sub-optimal experimenting?

Other questions

Risk aversion or myopia?

Prior beliefs are correctly specified?

Chapter 10

Insurance

10.1 Cohen and Einav (2007)

10.1.1 Summary

Background

There is little empirical research on the link between risk aversion and individual characteristics (how different agents might have different aversion to risk). But it is possible to recover these by looking at market outcomes in which participants had to make a choice between risky outcomes.

Data

Auto insurance company in Israel, data for 100k new customers, only focus on the first choice (to remove discussion about switching and search costs). Observation unit is thus transaction-level with all customer info that insurance provider got + menu of choice (four policies and premia), but also the length of contract and the actual use of the contract (claims and amounts claimed).

Model

Using Expected Utility Theory to derive utility function (vNM style). Two main factors to the choice: risk aversion and claim risk (tradeoff is given by an equation recoverable from the data). Estimation by ML.

Assumptions

Claims are drawn from a Poisson distribution estimated on consumer characteristics. No moral hazard.

Results

Risk averse is heterogenous (dependent on observables). Claim risk is heterogenous.

10.1.2 Discussion

- Authors add dimensions to the question of adverse selection (how are people risk averse, are they different, etc.)

10.2 Einav, Finkelstein and Cullen (2010)

10.2.1 Summary

Background

Welfare loss in insurance market is not to be proved anymore, but it is still difficult to quantify (because of hidden information mostly). With fewer assumptions, yet enough structure, this paper provides a methodology to evaluate welfare in insurance market.

Model

Given two choices of policies, normalize price and quantities to represent the difference between the two contracts. Then, crossing between demand and MC is the efficient, while crossing with AC is equilibrium. The difference is the deadweight loss.

Data

Employer-provided health insurance at Alcoa.