Empirical Industrial Organization: Introduction

TSE, MRes

Outline

- Introduction to empirical IO
- An example of structural model

Definition of empirical IO

Quantitative analysis of industrial organization problems

IO problems:

- Microeconomic perspective
- Based on individual behavior of agents
- Agents are typically: firms, consumers, regulator
- Partial equilibrium analysis: one market

Empirical approach

We will review structural models:

- Approach heavily based on theory
- Every parameter has an economic interpretation (even residuals)
- We specify objective functions of agents (utility function for consumers, profit or cost function for firms)

Digression: Structural vs. Reduced-form

Interest of structural approach:

- Estimated parameters are called “primitives" of the model
- They are supposed to be policy and environment invariant
- We can use the model and parameters to simulate counterfactuals: “how does the equilibrium change after a change in the environment"
- Interesting for predictions: ex ante evaluation of a policy, a merger
- Also useful for ex post evaluation: simulation of the “but for” the policy, the merger
- Sometimes it is just that data are too limited, need assumptions to make sense of them...

Digression: Structural vs. Reduced-form

Reduced-form approach:

- We are interested in measuring the causal effect of a variable on an output (e.g. effect of price on sales)
- We must control for all factors to estimate unbiased parameters
- Usually based on instrumental variable ("IV") approach
- Because natural experiments are very rare in IO (prices are not randomly assigned!)
- Controlled experiment also very rare (e.g. impossible to tax randomly consumers)

Interest of reduced-form approach:

- We can do predictions (to some extent...)
- Parameters are typically locally robust but not globally
- Less structure is imposed

But it is pretty rare to have good instruments!

Typical data

Typically, we use data from a single industry, but many different markets (over time, different geographical areas, different groups of consumers...)

Demand, cost functions, and firms' strategies are industry-specific

Typically, we would only observe market outcome data, prices and quantities:

- Cost data would be interesting but very rare to observe or unreliable
- However, we can infer costs from the information/assumption about the equilibrium prices and quantities
- Good news, we can measure price-cost margins without costs!

Typical data

Sample size and sample variability are crucial to obtain precise estimates. We use variations across:

- Firms/products in the industry
- Local/regional markets
- Time periods

It is crucial to think about which variations in the data are needed to estimate the parameters of interest

A cheap but useful comment: we cannot estimate price sensitivity without variation in price!!!

Improvements in empirical work

Improvements in empirical work have come from many directions:

- Better data
- Better computation power
- Advanced estimation techniques
- New general models

Reminder: the generalized method of moments

Structural models are very often non-linear, need to use GMM

GMM relies only on the orthogonality condition between residuals U are some explanatory variables or instruments:

$$\begin{array}{ll} \mathbb{E}(U|X) & \text{if X exogenous} \\ \mathbb{E}(U|Z) & \text{if some X are endogenous, Z are instruments} \end{array}$$

We use empirical counterparts of the expectations: means

Formally, we define the empirical counterpart of the moment condition involving instrument Z^k, $g^k(\theta)$ as:

$$g^k(\theta) = \frac{1}{N}\sum_{i=1}^{N} z_i^k u_i(\theta)$$

Generalized method of moments

We stack the moment conditions:

$$g(\theta) = \begin{pmatrix} g^1(\theta) \\ \vdots \\ g^K(\theta) \end{pmatrix}$$

If more instruments than parameters, $K > dim(\theta)$, we need to weight the different moments. Let W be the weighting matrix, it can be the identity I, $(Z'Z)^{-1}$ or the optimal weighting matrix $(Z'UU'Z)^{-1}$

The GMM estimator $\hat{\theta}$ solves:

$$\min_{\theta} g(\theta)' W g(\theta)$$

This method does not require an assumption on the distribution of the residuals (unlike maximum likelihood)

Identification

Identification is crucial for structural models

Fundamental question of identification: can we have two parameters β_1 and β_2 such that: $M(Y, X, \beta_1) = M(Y, X, \beta_2)$? Where M is the model, X and Y are observable data

If the answer is yes, the model is not identified –> bad news!

It is sometimes difficult to prove formally the identification but need to have a discussion about identification: which restrictions/relationships are important to identify the model

A trivial point: we need at least as many equations as parameters to estimate!

In some cases we can only put some bounds on some parameters, typically when the economic model implies some inequalities, this is set identification (e.g. entry of a firm on a market if profit positive)

Different meanings of simulations

1. Monte carlo simulations
    - Useful to analyze the quality of the estimation method, how the precision varies with the sample size, variation in the data
    - Principle: generate **many times** a dataset using the model or DGP (data generating process) and estimate it
    - Standard errors of the estimates is given by the standard deviations of estimates across simulations
2. Counterfactual simulations
    - Use the structural model and estimated parameters to predict the effect of changes in the economic environment, policy...
3. Simulations for numerical approximations
    - Sometimes we need to compute an integral over a distribution which does not have an analytic formula (e.g. $\int_{\varepsilon} x(\epsilon) dF(\epsilon)$)
    - We can approximate this integral using simulations

$$\simeq \frac{1}{ns} \sum_{i=1}^{ns} x_i(\epsilon_i)$$

with ϵ_i are independent draws from $F(.)$

Example of empirical model

How to build a structural model of market equilibrium?

We start from a simple theoretical model

We take the model to the data

Estimate primitives of the model

Use the model to run counterfactual simulations

A model for the cement industry

We want to build an empirical model that describes the cement industry

Objective: estimate the price sensitivity of demand, the cost function (variable and fixed costs), the market power (price-cost margin), consumers surplus and firms profits

Predict equilibrium effects of a new sale tax on cement (including on the number of firms)

Industry background

General information:

- In 2008, more than 2 billion tons cement are produced annually
- Europe accounts for 13% of the world production
- Production very energy intensive (chemical reaction occurs at 1450°C)
- Very polluting (5% of EU greenhouse gas emissions), subject to EU carbon market (EU ETS) from phase IV
- High transportation costs (if distance $>$ 300 km, transportation costs $>$ production costs)

Important features to build the model:

- Homogenous product industry
- Fixed cost of operating a plant
- Marginal costs depend on the installed capacity (i.e. higher marginal cost when firm is close to capacity)
- Oligopoly industry, a concentrated industry

Data

Data on the cement industry in a certain country

Monthly data for this industry over 10 years and 10 localities ($t = 1, ..., 1200$ months×localities)

For each month, we observe:

- the number of plants (=firms) operating in the market: N_t
- the aggregate amount of output produced by all the firms: Q_t
- installed capacity of all the firms (maximum potential output per month): K_t
- the market price: P_t
- some exogenous variables that affect demand and/or costs: population (POP_t), average income (I_t), materials (M_t), and wage (W_t).

$$\text{Data} = \{N_t, Q_t, K_t, P_t, POP_t, I_t, M_t, W_t\}_{t=1,...,120}$$

Empirical model

The model should incorporate:

- A demand equation
- A cost function
- A model of price determination: price maximizes each firm's profit
- A model of entry decision: free entry condition, active firms have positive profits, a potential entrant would make negative profits

Demand equation

Inverse demand function:

$$P_t = A_t - B_t Q_t$$

Assume that both intercept and price sensitivity depend on observable variables : population and income. The intercept also depends on a demand shock which is known by firms but not to the econometrician:

$$\begin{aligned} A_t &= a_0 + a_1 POP_t + a_2 I_t + e_t^D \\ B_t &= b_0 + b_1 POP_t + b_2 I_t \end{aligned}$$

So that the demand equation to estimate is:

$$P_t = a_0 + a_1 POP_t + a_2 I_t + b_0 Q_t + b_1 (POP_t \times Q_t) + b_2 (I_t \times Q_t) + e_t^D$$

Linear equation: dependant variable is price and regressors are $\{intercept, POP_t, I_t, Q_t, POP_t \times Q_t, I_t \times Q_t\}$, demand shock plays the role of residual

Cost function

Assume that all firms have the same cost function (potential entrants too):

$$C_t(q_{it}) = VC_t(q_{it}) + FC_t$$

With VC the variable cost and FC the fixed cost

Assume that marginal cost depends on input prices W (wage) and M (material cost) and how close we are from capacity K and cost shocks:

$$MC_t(q_{it}) = m_0 + m_1 W_t + m_2 M_t + m_3 \frac{q_{it}}{k_{it}} + e_t^{MC}$$

And fixed cost is simply a constant plus a cost shock:

$$FC_t = f_0 + e_t^{FC}$$

Cournot competition

The profit function of a firm i in this industry at month t is:

$$\Pi_{it} = P_t \times q_{it} - VC_t(q_{it}) - FC_t$$

Taking as given the quantity produced by the rest of the firms, a firm chooses q_{it} to maximize its profits. The FOC implies that the marginal revenue is equal to marginal cost, $MR_t = \frac{\partial VC_t}{\partial q_{it}} = MC_{it}$, that implies:

$$A_t - B_t Q_t^{others} - 2B_t q_{it} = MC_{it}$$

Given that $Q_t = Q_t^{others} + q_{it}$, $P_t = A_t - B_t Q_t$, $q_{it} = Q_t/N_t$ and $q_{it}/k_{it} = Q_t/K_t$ (assuming firms are identical), we can write the expression in terms of aggregate outcomes:

$$P_t - MC_t(Q_t) = B_t \frac{Q_t}{N_t}$$

Equilibrium

$$P_t - MC_t = B_t \frac{Q_t}{N_t}$$

If we can estimate demand parameter B_t, then we can recover the price cost margin $P_t - MC_t$, and then the value of the marginal cost, because P_t is observed

We can infer marginal costs using price/quantity data and structural model. This is the power of structural approach!

Taking into account our specification of the marginal cost, we have:

$$\left(P_t - B_t \frac{Q_t}{N_t}\right) = m_0 + m_1 W_+ m_2 M_t + m_3 \frac{Q_t}{K_t} + e_t^{MC}$$

After estimating B_t, the left-hand side variable is known.

Then, this equation is linear, the dependent variable is $\left(P_t - \hat{B}_t \frac{Q_t}{N_t}\right)$ and regressors are a constant, W_t, M_t and $\frac{Q_t}{K_t}$

Estimate demand and cost

How to estimate the parameters of demand and marginal cost?

We have linear relationships between variables

However, P_t and Q_t simultaneously determined, OLS estimates would be biased

Solution: use IV regression (2SLS). We need instruments

Cost shifters as instruments for demand: affect price but not directly Q_t

Demand shifters as instruments for cost equation: affect Q_t but not directly cost

Free entry condition

The industry does not make negative profits with N_t firms:

$$\Pi_t(N_t) = \frac{1}{B_t}\left(\frac{A_t - MC_t}{N+1}\right)^2 - FC_t \geq 0$$

Additional entry would imply negative profits for the industry:

$$\Pi_t(N_t+1) = \frac{1}{B_t}\left(\frac{A_t - MC_t}{N+2}\right)^2 - FC_t \leq 0$$

So we have bounds for the fixed costs:

$$\frac{1}{B_t}\left(\frac{A_t - MC_t}{N+2}\right)^2 < FC_t < \frac{1}{B_t}\left(\frac{A_t - MC_t}{N+1}\right)^2$$

If we assume that fixed cost is in the middle of the bounds, we can estimate the parameter f_0 of the fixed cost equation: $FC_t = f_0 + e_t^{FC}$.

Simulate the model

Objective = generate data from the model

How to simulate the model?

- Set the parameters (primitives)
- Generate (draw) the random variables
- Solve for the endogenous variables

Why simulating the model?

To check identification strategy and estimation algorithm works before working on real data.

Simulate the model

Generate data following the model and using the following parameters, distribution for the variables:

- Set the random seed (so all of us will generate the same data!): rng(2021)
- 12 month, 10 years, 10 localities: T = 1,200 independent observations
- Demand function: 2 demand shifters, pop and income, a demand shock e_t^d:

$$pop_t \sim U(0,1) \qquad inc_t \sim U(0,4) \qquad e_t^d \sim N(0,1)$$

- Demand parameters:

$$a_0 = 20 \quad a_1 = 4 \quad a_2 = 1 \quad b_0 = 1$$

- We cannot directly draw Q_t and P_t since they are endogenous (simultaneously determined). Also endogenous is the number of firms N_t

Simulate the model

- Supply side: 3 cost shifters: capacity, wage and material; one cost shock:

$$cap_t \sim U(10, 20) \quad wage_t \sim U(0, 2) \quad material_t \sim U(0, 3) \quad e_t^S \sim N(0, 1)$$

- Parameters of the cost function:

$$m_0 = 2 \quad m_1 = 2 \quad m_2 = 1 \quad m_3 = 1$$

- Fixed cost:

$$f_t \sim U(2, 12)$$

Simulate the model

How to solve for Q_t and P_t, N_t?

For given N_t, we use firm FOC and demand function to solve for P_t, Q_t.

N_t s.t. additional entrant would imply negative profits (free entry condition). Se we loop over values for N_t

1. Start with $N_t = 2$, solve for Q_t, P_t
2. If profits > 0, continue; if profits <0 stop $N_t^* = 1$
3. Then try $N_t = 3$, solve for Q_t, P_t and check profits
4. Iterate until profits become negative

Note: here the parameters and distributions were chosen such that we always have at least one firm

Exercise (in class and continue at home)

- Generate one dataset following the parameters values and distributional assumptions
- Estimate the model parameters using OLS and IV. How large is the endogeneity bias? Which parameters are the most biased, why? Is the sign of the bias expected?
- Use the estimated parameters to predict the effect of a green sale tax of 20% on cement
- Do a monte-carlo study of the performances of the OLS and IV estimates. For this, generate 100 synthetic datasets. Compute the mean estimates, the mean bias and the standard deviations of the estimates.
- Replicate the monte carlo study, but this time decrease the variance of the cost and demand shocks (ϵ^D and ϵ^S). What are the consequences on the parameter estimates?