ECO 2901
EMPIRICAL INDUSTRIAL ORGANIZATION

Lecture 7:
Dynamic Games of Oligopoly Competition
Solution and Estimation Methods

Victor Aguirregabiria (University of Toronto)

February 20, 2020

Solution & Estimation Methods

1. Solution Methods
2. Data & Identification
3. Estimation Methods
4. Counterfactual Experiments with Multiple Equilibria

1. Solution Methods

MPE in vector form

- Suppose that $\mathbf{x}_t$ is discrete: $\mathbf{x}_t \in \{x^1, x^2, ..., x^{|X|}\}$.

- The primitives of the model are:

  [1] **Vectors of payoffs:** $\mathbf{\Pi}_i(a_i, a_{-i})$
  with dimension $|X| \times 1$, for every value of (a_i, a_{-i}).

  [2] **Matrices of transition probabilities:** $\mathbf{F}(a_i, a_{-i})$
  with dimension $|X| \times |X|$, for every value of (a_i, a_{-i}).

  [3] The **discount factor** δ.

MPE in vector form [2]

• Let $\mathbf{P} \equiv \{\mathbf{P}_i(a_i) : \text{for any } i \text{ and } a_i\}$ where $\mathbf{P}_i(a_i)$ is a vector of CCPs with dimension $|X| \times 1$.

• We can define the **vectors of expected payoffs** $\mathbf{\Pi}_i^{\mathbf{P}_{-i}}(a_i)$ and the **matrices of expected transition probabilities** $\mathbf{F}_i^{\mathbf{P}_{-i}}(a_i)$:

$$\mathbf{\Pi}_i^{\mathbf{P}_{-i}}(a_i) \equiv \sum_{a_{-i}} \mathbf{P}_i(a_i) * \mathbf{\Pi}_i(a_i, a_{-i})$$

$$\mathbf{F}_i^{\mathbf{P}_{-i}}(a_i) \equiv \sum_{a_{-i}} \mathbf{P}_i(a_i) * \mathbf{F}(a_i, a_{-i})$$

where $*$ represents the "element-by-element" or Hadamard product.

MPE in vector form [3]

- A MPE is a vector $\mathbf{P} \equiv \{\mathbf{P}_i(a_i) : \text{for any } i \text{ and } a_i\}$ such that (Logit):

$$\mathbf{P}_i(a_i) = \frac{\exp\left\{\mathbf{\Pi}_i^{\mathbf{P}^{-i}}(a_i) + \delta\ \mathbf{F}_i^{\mathbf{P}^{-i}}(a_i)\ \mathbf{V}_i^{\mathbf{P}^{-i}}\right\}}{\sum_{a'}\exp\left\{\mathbf{\Pi}_i^{\mathbf{P}^{-i}}(a') + \delta\ \mathbf{F}_i^{\mathbf{P}^{-i}}(a')\ \mathbf{V}_i^{\mathbf{P}^{-i}}\right\}}$$

- $\mathbf{V}_i^{\mathbf{P}^{-i}} = |X| \times 1$ vector solving the Bellman equation in vector form:

$$\mathbf{V}_i^{\mathbf{P}^{-i}} = \ln\left(\sum_{a_i}\exp\left\{\mathbf{\Pi}_i^{\mathbf{P}^{-i}}(a_i) + \delta\ \mathbf{F}_i^{\mathbf{P}^{-i}}(a_i)\ \mathbf{V}_i^{\mathbf{P}^{-i}}\right\}\right)$$

- Or in compact form, if $\mathbf{P}_i \equiv \{\mathbf{P}_i(a_i) : \text{for any } a_i\}$:

$$\mathbf{P}_i = BR_i\left(\mathbf{P}_{-i}\right)$$

where $BR_i\left(.\right)$ is the best response function that gives firm i's vector of CCPs solving the DP problem given $\mathbf{P}_{-i}$.

Algorithm: Fixed Point Iterations in BR mapping

- Let $\mathbf{P}^0 \equiv \{\mathbf{P}_i^0 : \text{for any } i\}$ be arbitrary vector of CCPs.
- At iteration n, for any player i:

$$\mathbf{P}_i^n = BR_i\left(\mathbf{P}_{-i}^{n-1}\right)$$

- We check for convergence:

$$\begin{cases} \text{if } \left\|\mathbf{P}^n - \mathbf{P}^{n-1}\right\| \leq \kappa \quad \text{then} \quad \mathbf{P}^n \text{ is a MPE} \\ \text{if } \left\|\mathbf{P}^n - \mathbf{P}^{n-1}\right\| > \kappa \quad \text{then} \quad \text{Proceed to iteration } n+1 \end{cases}$$

where κ is a predetermined constant small positive constant, e.g., $\kappa = 10^{-6}$.

- Convergence? Not guaranteed. More on this later ...

Solving Players' DP Problems

- Evaluation $BR_i(\mathbf{P}_{-i})$ for a value of $\mathbf{P}_{-i}$ involves solving a DP problem for player i.

- I describe here three standard algorithms to solve this DP problem.
  - [1] Value function iterations
  - [2] Newton-Kantorovich iterations
  - [3] Policy iterations

- For notational simplicity, I omit here the subindex i and superindexes $\mathbf{P}_{-i}$, such that the (integrated) Bellman equation is:

$$\mathbf{V} = \Gamma(\mathbf{V})$$

with

$$\Gamma(\mathbf{V}) \equiv \ln\left(\sum_a \exp\left\{\mathbf{\Pi}(a) + \delta\ \mathbf{F}(a)\ \mathbf{V}\right\}\right)$$

Value function iterations

- Let $\mathbf{V}^0$ be an arbitrary vector of values.
- At iteration n:

$$\mathbf{V}^n = \Gamma\left(\mathbf{V}^{n-1}\right)$$

- We check for convergence:

$$\begin{cases} \text{if } \left\|\mathbf{V}^n - \mathbf{V}^{n-1}\right\| \leq \kappa & \text{then} \quad \mathbf{V}^n \text{ is the solution} \\ \text{if } \left\|\mathbf{V}^n - \mathbf{V}^{n-1}\right\| > \kappa & \text{then} \quad \text{Proceed to iteration } n+1 \end{cases}$$

where κ is a predetermined constant small positive constant, e.g., $\kappa = 10^{-6}$.

- Since $\Gamma(.)$ is a contraction mapping, convergence is guaranteed (to the unique fixed point).

Complexity (Computational Cost) of VF iterations

• The computational cost of an algorithm = Cost per iteration * Number of iterations.

• Degree of complexity $C(.)$ = Number of basic operations (sums, products).

• **Cost per iteration =** evaluation of $\Gamma\left(\mathbf{V}^{n-1}\right) =$ $\ln\left(\sum_a \exp\left\{\mathbf{\Pi}(a) + \delta\ \mathbf{F}(a)\ \mathbf{V}^{n-1}\right\}\right)$.

$$\text{Complexity} = O\left(|X|^2\right)$$

• **Number of iterations =** $O\left(\frac{1}{1-\beta}\right)$.

Newton Iterations

- Define the function $f(\mathbf{V}) \equiv \mathbf{V} - \Gamma(\mathbf{V})$.
- Solving for a fixed point of Γ is equivalent to finding a zero (a root) of f.
- We use Newton's method to find a root of f.
- Applying Mean Value Theorem to $f(\mathbf{V})$ around $\mathbf{V}^0$ [$\nabla f(\mathbf{V})$ represents the Jacobian matrix of f]:

$$f(\mathbf{V}) \simeq f(\mathbf{V}^0) + \nabla f(\mathbf{V}^*) \left[\mathbf{V} - \mathbf{V}^0\right]$$

- Making $f(\mathbf{V}) = 0$ and solving for $\mathbf{V}$:

$$\mathbf{V} = \mathbf{V}^0 + \left[\nabla f(\mathbf{V}^*)\right]^{-1} f(\mathbf{V}^0)$$

Newton iterations [2]

- Let $\mathbf{V}^0$ be an arbitrary vector of values.
- At iteration n:

$$\mathbf{V}^n = \mathbf{V}^{n-1} + \left[\nabla f(\mathbf{V}^{n-1})\right]^{-1} f(\mathbf{V}^{n-1})$$

- We check for convergence:

$$\begin{cases} \text{if } \left\|\mathbf{V}^n - \mathbf{V}^{n-1}\right\| \leq \kappa \quad \text{then} \quad \mathbf{V}^n \text{ is the solution} \\ \text{if } \left\|\mathbf{V}^n - \mathbf{V}^{n-1}\right\| > \kappa \quad \text{then} \quad \text{Proceed to iteration } n+1 \end{cases}$$

where κ is a predetermined constant small positive constant, e.g., $\kappa = 10^{-6}$.

- Convergence is guaranteed (to the unique fixed point).

Newton iterations [3]

- What is the form of the Jacobian matrix $\nabla f(\mathbf{V})$ in this model?

- Simple to prove for the Logit model but can be generalized to any other distribution:

$$\nabla f(\mathbf{V}) = \mathbf{I} - \delta \, \overline{\mathbf{F}}(\mathbf{V})$$

with

$$\overline{\mathbf{F}}(\mathbf{V}) \equiv \sum_{a} \mathbf{P}(a;\mathbf{V}) \; * \; \mathbf{F}(a)$$

and

$$\mathbf{P}(a;\mathbf{V}) = \frac{\exp\left\{\mathbf{\Pi}(a) + \delta \; \mathbf{F}(a) \; \mathbf{V}\right\}}{\sum_{a'} \exp\left\{\mathbf{\Pi}(a') + \delta \; \mathbf{F}(a') \; \mathbf{V}\right\}}$$

Complexity of Newton iterations

$$\mathbf{V}^n = \mathbf{V}^{n-1} + \left[\mathbf{I} - \delta \sum_a \mathbf{P}(a; \mathbf{V}^{n-1}) * \mathbf{F}(a)\right]^{-1} \left[\mathbf{V}^{n-1} - \Gamma\left(\mathbf{V}^{n-1}\right)\right]$$

• **Cost per iteration =** Main cost comes from inversion of matrix $\mathbf{I} - \delta \overline{\mathbf{F}}(\mathbf{V^{n-1}})$

$$\text{Complexity} = O\left(|X|^3\right)$$

• **Number of iterations =** Substantially smaller than for VF iterations (faster convergence).

• For moderate dimension of $|X|$ – a few thousand cells – Newton's is faster than VF.

• But for larger dimensions of $|X|$, the $O\left(|X|^3\right)$ in Newton's dominates the $O\left(|X|^2\right)$ in VF such that VF becomes faster.

Policy Function Iterations

- The PF algorithm consists in iterating in the **"Policy Function"** (PF) which is a fixed point mapping in the space of the vector of CCPs.

- The solution to the DP can be described as:

$$\mathbf{P} = \varphi(\mathbf{P})$$

- $\varphi(\mathbf{P}) \equiv \{\varphi(a;\mathbf{P}) : \text{for any } a\}$ is the PF, defined as:

$$\varphi(a;\mathbf{P}) = \frac{\exp\{\mathbf{\Pi}(a) + \delta\ \mathbf{F}(a)\ v(\mathbf{P})\}}{\sum_{a'} \exp\{\mathbf{\Pi}(a') + \delta\ \mathbf{F}(a')\ v(\mathbf{P})\}}$$

- $v(\mathbf{P})$ is the **Valuation Operator**. It give the vector of present values in the firm behaves according to the vector of CCPs $\mathbf{P}$. It consists of the calculation of present values.

$$v(\mathbf{P}) = \left[\mathbf{I} - \delta \sum_a \mathbf{P}(a) * \mathbf{F}(a)\right]^{-1} \left[\sum_a \mathbf{P}(a) * \text{"}(a)\right]$$

Policy Function Iterations [2]

- Let $\mathbf{P}^0$ be an arbitrary vector of CCPs.
- At iteration n, for any a:

$$\mathbf{P}^n(a) = \frac{\exp\left\{\mathbf{\Pi}(a) + \delta\ \mathbf{F}(a)\ v\left(\mathbf{P}^{n-1}\right)\right\}}{\sum_{a'} \exp\left\{\mathbf{\Pi}(a') + \delta\ \mathbf{F}(a')\ v\left(\mathbf{P}^{n-1}\right)\right\}}$$

- We check for convergence:

$$\begin{cases} \text{if } \left\|\mathbf{P}^n - \mathbf{P}^{n-1}\right\| \leq \kappa \quad \text{then} \quad \mathbf{P}^n \text{ is the solution} \\ \text{if } \left\|\mathbf{P}^n - \mathbf{P}^{n-1}\right\| > \kappa \quad \text{then} \quad \text{Proceed to iteration } n+1 \end{cases}$$

- Convergence is guaranteed (to the unique fixed point).

Policy Function Iterations

• For this model, it is possible to show Newton Iterations and Policy function Iterations are equivalent algorithms.

• Let $\mathbf{V}^0$ be the initial vector for the Newton's algorithm. Suppose that we initialize the PF algorithm with $\mathbf{P}^0$ such that $\mathbf{P}^0(a) = \frac{\exp\{\mathbf{\Pi}(a)+\delta\ \mathbf{F}(a)\ \mathbf{V}^0\}}{\sum_{a'}\exp\{\mathbf{\Pi}(a')+\delta\ \mathbf{F}(a')\ \mathbf{V}^0\}}$. Then, the two algorithms generate the same sequences of $\{\mathbf{V}^n : n \geq 1\}$ and $\{\mathbf{P}^n : n \geq 1\}$.

• Similarly, let $\mathbf{P}^0$ be the initial vector for the PF algorithm. Suppose that we initialize Newton's algorithm with $\mathbf{V}^0$ such that $\mathbf{V}^0 = v(\mathbf{P}^0)$. Then, the two algorithms generate the same sequences of $\{\mathbf{V}^n : n \geq 1\}$ and $\{\mathbf{P}^n : n \geq 1\}$.

Curse of Dimensionality

• With any of these algorithms, the computation time for solving the DP problem increases with $|X|$ at a rate between $|X|^2$ and $|X|^3$.

• This can become very large (months, years) when $|X|$ in the hundred-thousands or millions, which are common dimensions in some empirical applications.

• **Approximate Dynamic Programming (ADP)** provides different techniques to approximate the solution.

[1] Monte Carlo simulation to approximate the valuation operator $v(\mathbf{P})$.

[2] Projection methods to represent **P** or/and **V** as polynomial (or sieve) function and iterate in the smaller dimension space of the coefficients of the polynomial.

2. Data & Identification

Data

• Suppose that we have a random sample of M local markets, indexed by m, over T periods of time, where we observe:

$$Data = \{\mathbf{a}_{mt},\ \mathbf{x}_{mt} : m = 1, 2, ..., M;\ \ t = 1, 2, ..., T\}$$

• We want to use these data to estimate the model parameters in the population that has generated this data: $\boldsymbol{\theta}^0 = \{\boldsymbol{\theta}_i^0 : i \in I\}$.

• For the moment, we consider that the industry and the data are such that:

(a) each firm is observed making decisions in every of the M markets;

(b) the researcher knows all the payoff relevant market characteristics that are common knowledge to the firms, $\mathbf{x}$.

Data [2]

• Under condition (a) we can allow for rich firm heterogeneity that is fixed across markets and time by estimating firm-specific structural parameters, θ_i.

• Condition (b) rules out the existence of unobserved market heterogeneity. Though it is a convenient assumption, it is also unrealistic for most applications in empirical IO. Later I present estimation methods that relax conditions (a) and (b) and deal with unobserved market and firm heterogeneity.

Identification

• A significant part of this literature has considered the following identification assumptions.

Assumption (ID 1): Single equilibrium in the data. Every observation in the sample comes from the same Markov Perfect Equilibrium, i.e., for any observation (m, t), $\mathbf{P}^0_{mt} = \mathbf{P}^0$.

Assumption (ID 2): No unobserved common-knowledge variables. The only unobservables for the econometrician are the private information shocks ε_{imt} and the structural parameters $\boldsymbol{\theta}$.

Linear in Parameters Payoff functions

• For the discussion of identification and estimation, it is convenient to focus in models where players' payoff functions are linear in parameters:

$$\pi_i(a_i, a_{-i}, \mathbf{x}) = \mathbf{h}(a_i, a_{-i}, \mathbf{x})\ \boldsymbol{\theta}_i$$

where $\mathbf{h}(a_i, a_{-i}, \mathbf{x})$ is a vector of functions known to the researcher and $\boldsymbol{\theta}_i$ is the vector of unknown structural parameters for the researcher.

• Most empirical applications of dynamic games in IO have this structure.

• In general, any payoff function can be approximated arbitrarily well by a function $\mathbf{h}(a_i, a_{-i}, \mathbf{x})\ \boldsymbol{\theta}_i$, where $\mathbf{h}(a_i, a_{-i}, \mathbf{x})$ is the vector of basis functions, e.g., terms of a polynomial, sieves.

Linear in Parameters Payoff functions [2]

- This form of the payoff function implies that:

$$\Pi_i^{P_{-i}}(a_i, \mathbf{x}) = \mathbf{h}^{P_{-i}}(a_i, \mathbf{x})\ \boldsymbol{\theta}_i$$

with $\mathbf{h}^{P_{-i}}(a_i, \mathbf{x}) \equiv \sum_{a_{-i}} \left[\prod_{j \neq i} P_j(a_j|\mathbf{x})\right]\ \mathbf{h}(a_i, a_{-i}, \mathbf{x})$.

- Such that:

$$\boldsymbol{\Pi}_i^{P_{-i}}(a_i) = \mathbf{H}^{P_{-i}}(a_i)\ \boldsymbol{\theta}_i$$

with $\mathbf{H}^{P_{-i}}(a_i)$ being the matrix with elements $\mathbf{h}^{P_{-i}}(a_i, \mathbf{x})$.

- And the **Valuation Operator** becomes:

$$v_i(\mathbf{P}) = \widetilde{\mathbf{H}}_i(\mathbf{P})\ \boldsymbol{\theta}_i$$

with

$$\widetilde{\mathbf{H}}_i(\mathbf{P}) = \left[\mathbf{I} - \delta \sum_{a_i} \mathbf{P}_i(a_i) * \mathbf{F}^{P_{-i}}(a_i)\right]^{-1} \left[\sum_{a_i} \mathbf{P}(a_i) * \mathbf{H}^{P_{-i}}(a_i)\right]$$

Linear in Parameters Valuation Operator

$$v_i(\mathbf{P}) = v_{\mathbf{H}i}(\mathbf{P})\ \theta_i$$

with

$$v_{\mathbf{H}i}(\mathbf{P}) = \left[\mathbf{I} - \delta \sum_{a_i} \mathbf{P}_i(a_i) * \mathbf{F}^{P-i}(a_i)\right]^{-1} \left[\sum_{a_i} \mathbf{P}(a_i) * \mathbf{H}^{P-i}(a_i)\right]$$

- The linearity in θ_i of the valuation operator is a property that is particularly convenient (computationally) for the estimation of the model.

- The computation of $v_{\mathbf{H}i}(\mathbf{P})$ is costly because it involves the inversion of a large matrix. But we will need to calculate it only once (or a few times).

- For different trial values of θ_i in the search for the estimates, the valuation operator $v_{\mathbf{H}i}(\mathbf{P})\ \theta_i$ is recalculated using a fixed $v_{\mathbf{H}i}(\mathbf{P})(\mathbf{P})$.

Policy Function

- In the MPE, we have that:

$$\mathbf{P}_i = \Psi_i\left(\mathbf{P}, \boldsymbol{\theta}_i\right)$$

- $\Psi_i\left(\mathbf{P}, \boldsymbol{\theta}_i\right) \equiv \left\{\Psi_i\left(a_i; \mathbf{P}, \boldsymbol{\theta}_i\right) : \text{for any } a_i\right\}$ is the PF, defined as:

$$\Psi_i\left(\dot{a}_i; \mathbf{P}, \boldsymbol{\theta}_i\right) = \frac{\exp\left\{\mathbf{H}^{P_{-i}}(a_i)\boldsymbol{\theta}_i + \delta\ \mathbf{F}^{P_{-i}}(a_i)\ v_{\mathbf{H}i}\left(\mathbf{P}\right)\ \boldsymbol{\theta}_i\right\}}{\sum_{a_i'} \exp\left\{\mathbf{H}^{P_{-i}}(a_i')\boldsymbol{\theta}_i + \delta\ \mathbf{F}^{P_{-i}}(a_i')\ v_{\mathbf{H}i}\left(\mathbf{P}\right)\ \boldsymbol{\theta}_i\right\}}$$

- Or in a more compact form:

$$\Psi_i\left(a_i; \mathbf{P}, \boldsymbol{\theta}_i\right) = \frac{\exp\left\{\widetilde{\mathbf{H}}_i(a_i; \mathbf{P})\ \boldsymbol{\theta}_i\right\}}{\sum_{a_i'} \exp\left\{\widetilde{\mathbf{H}}_i(a_i'; \mathbf{P})\ \boldsymbol{\theta}_i\right\}}$$

with $\widetilde{\mathbf{H}}_i(a_i; \mathbf{P}) \equiv \mathbf{H}^{P_{-i}}(a_i) + \delta\ \mathbf{F}^{P_{-i}}(a_i)\ v_{\mathbf{H}i}\left(\mathbf{P}\right)$.

Identification

- Under assumptions ID-1 & ID-2, the equilibrium that has generated the data, $\mathbf{P}^0$, can be estimated consistently and nonparametrically from the data. For any $(i, a_i, \mathbf{x})$:

$$P_i^0(a_i|\mathbf{x}) = \Pr(a_{imt} = a_i \mid \mathbf{x}_{mt} = \mathbf{x})$$

- Given that $\mathbf{P}^0$ is identified, we have also identification of expected present values $\widetilde{\mathbf{\Pi}}_i^{P^0{}^{-i}}(a_i)$ at the "true" equilibrium in the population.

- We can invert the best response function to have:

$$\ln\left(\frac{\mathbf{P}_i^0(a_i)}{\mathbf{P}_i^0(0)}\right) = \left[\widetilde{\mathbf{H}}_i(a_i;\mathbf{P}^0) - \widetilde{\mathbf{H}}_i(0;\mathbf{P}^0)\right] \ \theta_i^0$$

Identification (2)

$$\ln\left(\frac{\mathbf{P}_i^0(a_i)}{\mathbf{P}_i^0(0)}\right) = \left[\widetilde{\mathbf{H}}_i(a_i;\mathbf{P}^0) - \widetilde{\mathbf{H}}_i(0;\mathbf{P}^0)\right] \ \theta_i^0$$

- These equilibrium conditions identify θ_i^0.

- Matrix $\widetilde{\mathbf{H}}_i(a_i;\mathbf{P}^0) - \widetilde{\mathbf{H}}_i(0;\mathbf{P}^0)$ is full-column rank if and only if $\mathbf{H}_i^{P_{-i}^0}(a_i) - \mathbf{H}_i^{P_{-i}^0}(0)$ which in turn is an implication of the specification of the elements in vector $h_i(a_i, a_{-i}, \mathbf{x})$.

3. Estimation Methods

Estimation

- We consider the following estimators:

  1. *MLE*
  2. *Two-step estimator Pseudo MLE*
  3. *Recursive K-step Pseudo MLE*

Pseudo Likelihood Function

- In models with multiple equilibria, it is convenient to define an **extended or pseudo likelihood function** that depends both on $\boldsymbol{\theta}$ and $\mathbf{P}$.

$$Q(\boldsymbol{\theta}, \mathbf{P}) = \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{T} \ln \Psi_i \left(a_{imt} | \mathbf{x}_{mt}, \mathbf{P}, \boldsymbol{\theta}_i \right)$$

- where $\Psi_i \left(a_{imt} | \mathbf{x}_{mt}, \mathbf{P}, \boldsymbol{\theta}_i \right)$ is the Policy Function of best response function:

$$\Psi_i \left(a_{imt} | \mathbf{x}_{mt}, \mathbf{P}, \boldsymbol{\theta}_i \right) = \frac{\exp \left\{ \widetilde{\mathbf{H}}_i(a_{imt}, \mathbf{x}_{mt}; \mathbf{P}) \ \boldsymbol{\theta}_i \right\}}{\sum_{a_i'} \exp \left\{ \widetilde{\mathbf{H}}_i(a_i', \mathbf{x}_{mt}; \mathbf{P}) \ \boldsymbol{\theta}_i \right\}}$$

Maximum Likelihood Estimator

- It is defined as:

$$\left(\widehat{\theta}_{MLE}, \widehat{\mathbf{P}}_{MLE}\right) = \arg\max_{\theta, \mathbf{P}} Q(\theta, \mathbf{P})$$

$$\text{subject to: } \mathbf{P}_i = \Psi_i\left(\mathbf{P}, \theta_i\right) \text{ for any } i$$

- It has all the nice properties of MLE, but it is can be computationally very costly to implement when $|X|$ is large – as large as a few thousand states.

- This is because it requires the computation of the valuation operator $v_{\mathbf{H}i}\left(\mathbf{P}\right)$ for many trial values of $\mathbf{P}$.

Two-step Pseudo MLE

- In a first step, we obtain nonparametric estimates of the CCPs (e.g., frequency estimator, kernel estimator, sieve estimator). Let $\widehat{\mathbf{P}^0}$ be this NP estimator.

- Then, in a second step we estimate $\boldsymbol{\theta}$ as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q(\boldsymbol{\theta}, \widehat{\mathbf{P}^0})$$

- This estimator is root-M consistent and asymptotically normal under the standard regularity conditions. It is not efficient because it does not impose the equilibrium constraints (only asymptotically).

Two-step Pseudo MLE [2]

- The most attractive feature of this two-step method is its **computational simplicity**.
- In the likelihood function $Q(\theta, \widehat{\mathbf{P}^0})$, the probabilities $\Psi_i\left(a_{imt}|\mathbf{x}_{mt}, \mathbf{P}, \theta_i\right)$ have the same structure as standard logit or probits, with indexes that are linear θ_i. Globally concave likelihood in θ_i.
- We need to calculate the valuation operator $v_{\mathbf{H}i}\left(\mathbf{P}\right)$ only once for each player.
- The main limitation of these two-step methods is that they can have **large finite sample bias**.
- The NP estimator of $\mathbf{P}^0$ can be very imprecise (**curse of dimensionality in the NP estimation**) and this implies large finite sample bias in the two-step estimator of θ.

Recursive K-step estimator

- K-step extension of the 2-step estimator. Given an initial consistent (NP) estimator $\widehat{\mathbf{P}^0}$, the sequence of estimators $\{\widehat{\theta^K}, \widehat{\mathbf{P}^K} : K \geq 1\}$ is defined as:

$$\widehat{\theta^{K+1}} = \arg\max_{\theta} Q\left(\theta, \widehat{\mathbf{P}^K}\right)$$

where:

$$\widehat{\mathbf{P}_i^K} = \Psi_i\left(\widehat{\mathbf{P}_i^{K-1}}, \widehat{\theta^K}\right) \quad \text{for any player } i$$

- Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2008) show that this recursive estimator can reduce very significantly the bias of two-step estimator.

- Pesendorfer & Schmidt-Dengler (2010) show that this procedure may not converge to a consistent estimator, when the equilibrium in the DGP is not stable.

4. Counterfactual experiments with multiple equilibria

Counterfactual Experiments (1)

- One of the most attractive features of structural models is that they can be used to predict the effects of new policies or changes in parameters (counterfactuals).
- However, this a challenging exercise in a model with multiple equilibria.
- The data can identify the "factual" equilibrium. However, under the counterfactual scenario, which of the multiple equilibria we should choose?

Counterfactual Experiments (2)

- Different approaches have been implemented in practice.
- Select the equilibrium to which we converge by iterating in the (counterfactual) equilibrium mapping starting with the factual equilibrium $\mathbf{P}^0$
- Select the equilibrium with maximum total profits (or alternatively, with maximum welfare).
- Homotopy method: Aguirregabiria and Ho (2007)

Counterfactual Experiments: Homotopy method

- Let $\boldsymbol{\theta}$ be the vector of structural parameters in the model. An let $\Psi(\boldsymbol{\theta}, \mathbf{P})$ be the equilibrium mapping such that an equilibrium associated with $\boldsymbol{\theta}$ can be represented as a fixed point:

$$\mathbf{P} = \Psi(\boldsymbol{\theta}, \mathbf{P})$$

- The model could be completed with an equilibrium selection mechanism: i.e., a criterion that selects one and only one equilibrium for each possible $\boldsymbol{\theta}$.
- Suppose that there is a "true" equilibrium selection mechanism in the population under study, but we do not know that mechanism.
- Our approach here (both for the estimation and for counterfactual experiments) is completely agnostic with respect to the equilibrium selection mechanism.

Counterfactual Experiments: Homotopy method

- We only assume that there is such a mechanism, and that it is a smooth function of $\boldsymbol{\theta}$.
- Let $\boldsymbol{\pi}(\boldsymbol{\theta})$ be the (unique) selected equilibrium, for given $\boldsymbol{\theta}$, if we apply the "true" selection mechanism.
- Since we do not know the mechanism, we do not know $\boldsymbol{\pi}(\boldsymbol{\theta})$ for every possible $\boldsymbol{\theta}$.
- However, we DO know $\boldsymbol{\pi}(\boldsymbol{\theta})$ at the true $\boldsymbol{\theta}_0$ because we know that:

$$\mathbf{P}_0 = \boldsymbol{\pi}(\boldsymbol{\theta}_0)$$

and both $\mathbf{P}_0$ and $\boldsymbol{\theta}_0$ are identified.

Counterfactual Experiments: Homotopy method

- Let θ_0 and $\mathbf{P}_0$ be the the population values. Let $(\hat{\theta}_0, \hat{\mathbf{P}}_0)$ be our consistent estimator.
- We do not know the function $\pi(\theta)$. All what we know is that the point $(\hat{\theta}_0, \hat{\mathbf{P}}_0)$ belongs to the graph of this function π.
- Let θ^* be the vector of parameters under a counterfactual scenario.
- We want to know the counterfactual equilibrium $\pi(\theta^*)$.

Counterfactual Experiments: Homotopy method

- A Taylor approximation to $\pi(\theta^*)$ around our estimator $\hat{\theta}_0$ implies that:

$$\begin{aligned}\pi(\theta^*) &= \pi\left(\hat{\theta}_0\right) + \frac{\partial \pi\left(\hat{\theta}_0\right)}{\partial \theta'}\left(\theta^* - \hat{\theta}_0\right) + O\left(\left\|\theta^* - \hat{\theta}_0\right\|^2\right) \\ &= \hat{\mathbf{P}}_0 + \frac{\partial \pi\left(\hat{\theta}_0\right)}{\partial \theta'}\left(\theta^* - \hat{\theta}_0\right) + O\left(\left\|\theta^* - \hat{\theta}_0\right\|^2\right)\end{aligned}$$

- To get a first-order approximation to $\pi(\theta^*)$ we need to know $\frac{\partial \pi\left(\hat{\theta}_0\right)}{\partial \theta'}$.

Counterfactual Experiments: Homotopy method

- We know that $\pi\left(\hat{\theta}_0\right) = \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)$, and this implies that:

$$\frac{\partial \pi\left(\hat{\theta}_0\right)}{\partial \theta'} = \left(I - \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \mathbf{P}'}\right)^{-1} \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \theta'}$$

- Then, $\pi(\theta^*) =$

$$\hat{\mathbf{P}}_0 + \left(I - \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \mathbf{P}'}\right)^{-1} \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \theta'} \left(\theta^* - \hat{\theta}_0\right) + O\left(\left\|\theta^* - \hat{\theta}_0\right\|^2\right)$$

- Therefore, $\hat{\mathbf{P}}_0 + \left(I - \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \mathbf{P}'}\right)^{-1} \frac{\partial \Psi(\hat{\theta}_0, \hat{\mathbf{P}}_0)}{\partial \theta'} \left(\theta^* - \hat{\theta}_0\right)$ is a first-order approximation to the counterfactual equilibrium $\mathbf{P}^*$.