RECENT METHODS IN THE ECONOMETRICS OF DYNAMIC GAMES

Victor Aguirregabiria (University of Toronto)

CAMP RESOURCES 2008

Introduction (1)

- In Discrete Choice Dynamic Programming (DC-DP) models, agents are forward looking and maximize expected intertemporal payoffs.

Introduction (1)

- In Discrete Choice Dynamic Programming (DC-DP) models, agents are forward looking and maximize expected intertemporal payoffs.
- The parameters to be estimated are *structural* in the sense that they describe agents' preferences, beliefs and technological and institutional constraints.

Introduction (1)

- In Discrete Choice Dynamic Programming (DC-DP) models, agents are forward looking and maximize expected intertemporal payoffs.
- The parameters to be estimated are *structural* in the sense that they describe agents' preferences, beliefs and technological and institutional constraints.
- These parameters are estimated using micro data on individuals' choices and outcomes and the **principle of revealed preference**.

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;
- education subsidies;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;
- education subsidies;
- investment subsidies/taxes;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;
- education subsidies;
- investment subsidies/taxes;
- patent regulation;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;
- education subsidies;
- investment subsidies/taxes;
- patent regulation;
- land use;

Introduction (2)

- An attractive feature of these structural models is that they can be used to predict the effects of **counterfactual policies**.
- They have been applied to evaluate a wide range of policies (with important dynamic aspects) such as:
- unemployment insurance;
- social security and public pension;
- education subsidies;
- investment subsidies/taxes;
- patent regulation;
- land use;
- Air Clean Act

Introduction (3)

- A well known impediment to the development of this literature has been the **computational complexity of solving DP problems.**

Introduction (3)

- A well known impediment to the development of this literature has been the **computational complexity of solving DP problems.**
- Solving a DP problem can be a non-trivial numerical task, and estimation typically requires the **repeated computation of solutions (or approximations) of the DP problem**.

Introduction (3)

- A well known impediment to the development of this literature has been the **computational complexity of solving DP problems.**
- Solving a DP problem can be a non-trivial numerical task, and estimation typically requires the **repeated computation of solutions (or approximations) of the DP problem**.
- In this context, the recent development of **estimation methods that do not required the solution of the DP problem** has expand significantly the range of models we can estimate.

Introduction (3)

- A well known impediment to the development of this literature has been the **computational complexity of solving DP problems.**
- Solving a DP problem can be a non-trivial numerical task, and estimation typically requires the **repeated computation of solutions (or approximations) of the DP problem**.
- In this context, the recent development of **estimation methods that do not required the solution of the DP problem** has expand significantly the range of models we can estimate.
- These new methods can deal also with the **problem of multiple equilibria** in the estimation of dynamic games.

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

1. Model and Econometrics Issues

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

1. Model and Econometrics Issues
2. Estimation Methods

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

1. Model and Econometrics Issues
2. Estimation Methods
   - Two-Step methods (with Moment Equalities and with Moment Inequalities)

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

1. Model and Econometrics Issues
2. Estimation Methods
   - Two-Step methods (with Moment Equalities and with Moment Inequalities)
   - Methods for Models with Permanent Unobserved Heterogeneity

Outline of the Lecture

- This lecture reviews recent developments in the econometrics of Discrete Choice Dynamic Programming (DC-DP) models of individual behavior, **with special attention to Dynamic Games**.

1. Model and Econometrics Issues
2. Estimation Methods
   - Two-Step methods (with Moment Equalities and with Moment Inequalities)
   - Methods for Models with Permanent Unobserved Heterogeneity
3. Counterfactual Experiments

Model: A Dynamic Game of Market Entry/Exit

- To illustrate econometric issues and estimation methods, I will use a simple dynamic game of market entry/exit.

Model: A Dynamic Game of Market Entry/Exit

- To illustrate econometric issues and estimation methods, I will use a simple dynamic game of market entry/exit.
- I will show how this model can be used **to evaluate the effects of policies that affect demand or costs parameters**: e.g., the 1990 Amendments to the Clean Air Act (Ryan, 2006).

Model: A Dynamic Game of Market Entry/Exit

- To illustrate econometric issues and estimation methods, I will use a simple dynamic game of market entry/exit.
- I will show how this model can be used **to evaluate the effects of policies that affect demand or costs parameters**: e.g., the 1990 Amendments to the Clean Air Act (Ryan, 2006).
- Taking into account **firms' forward-looking and strategic behavior** can lead to very different predictions on the effects of these policies.

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

1. **Local competition.** The industry can be divided into isolated/independent local markets. The model is for a local market.

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

1. **Local competition.** The industry can be divided into isolated/independent local markets. The model is for a local market.
2. **Oligopoly competition.** Highly concentrated local markets

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

1. **Local competition.** The industry can be divided into isolated/independent local markets. The model is for a local market.
2. **Oligopoly competition.** Highly concentrated local markets
3. **Entry costs** which cannot be recovered upon exit (i.e., sunk entry costs).

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

1. **Local competition.** The industry can be divided into isolated/independent local markets. The model is for a local market.
2. **Oligopoly competition.** Highly concentrated local markets
3. **Entry costs** which cannot be recovered upon exit (i.e., sunk entry costs).

- For the sake of simplicity in the presentation, I consider that **only 2 firms are potential entrants in the different local markets.**

Model: A Dynamic Game of Market Entry/Exit

- Consider an industry characterized by:

1. **Local competition.** The industry can be divided into isolated/independent local markets. The model is for a local market.
2. **Oligopoly competition.** Highly concentrated local markets
3. **Entry costs** which cannot be recovered upon exit (i.e., sunk entry costs).

- For the sake of simplicity in the presentation, I consider that **only 2 firms are potential entrants in the different local markets.**
- I index firms by i, local markets by m, and time by t.

A Dynamic Game of Market Entry/Exit

- Every period, firms **decide whether to operate or not a plant in the local market.**

A Dynamic Game of Market Entry/Exit

- Every period, firms **decide whether to operate or not a plant in the local market.**

- When firms make this decision, they maximize expected intertemporal profits in that market:

$$E_t \left(\sum_{s=0}^{\infty} \beta^s \ \Pi_{im,t+s} \right)$$

where $\beta \in (0,1)$ is the discount factor.

A Dynamic Game of Market Entry/Exit

- Every period, firms **decide whether to operate or not a plant in the local market.**

- When firms make this decision, they maximize expected intertemporal profits in that market:

$$E_t \left(\sum_{s=0}^{\infty} \beta^s \; \Pi_{im,t+s} \right)$$

  where $\beta \in (0, 1)$ is the discount factor.

- This decision is **forward-looking** (because sunk entry costs) **and strategic** (because future profits depend on the opponent's entry decisions).

Model: A Dynamic Game of Market Entry/Exit

- Current profit of firm i in a local market is equal to variable profits, VP_{imt}, minus fixed costs of operating a plant, FC_{imt}, and minus the entry cost of setting up a plant by first time, EC_{imt}.

$$\Pi_{imt} = VP_{imt} - FC_{imt} - EC_{imt}$$

Model: A Dynamic Game of Market Entry/Exit

- Current profit of firm i in a local market is equal to variable profits, VP_{imt}, minus fixed costs of operating a plant, FC_{imt}, and minus the entry cost of setting up a plant by first time, EC_{imt}.

$$\Pi_{imt} = VP_{imt} - FC_{imt} - EC_{imt}$$

- In this example we will consider that the researcher does not observe firms' prices, quantities or even revenues at the level of local markets.

Model: A Dynamic Game of Market Entry/Exit

- Current profit of firm i in a local market is equal to variable profits, VP_{imt}, minus fixed costs of operating a plant, FC_{imt}, and minus the entry cost of setting up a plant by first time, EC_{imt}.

$$\Pi_{imt} = VP_{imt} - FC_{imt} - EC_{imt}$$

- In this example we will consider that the researcher does not observe firms' prices, quantities or even revenues at the level of local markets.
- Even with these data limitations, information on firms' entry/exit decisions in local markets can identify demand, variable costs, fixed costs and entry costs parameters for each firm.

Model: A Dynamic Game of Market Entry/Exit

- Current profit of firm i in a local market is equal to variable profits, VP_{imt}, minus fixed costs of operating a plant, FC_{imt}, and minus the entry cost of setting up a plant by first time, EC_{imt}.

$$\Pi_{imt} = VP_{imt} - FC_{imt} - EC_{imt}$$

- In this example we will consider that the researcher does not observe firms' prices, quantities or even revenues at the level of local markets.
- Even with these data limitations, information on firms' entry/exit decisions in local markets can identify demand, variable costs, fixed costs and entry costs parameters for each firm.
- However, the specification of demand and variable costs should be relatively simple.

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".
- The variable profit function is:

$$VP_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ S_{mt}\ \theta_i^M & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 0 \\ S_{mt}\ \theta_i^D & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 1 \end{cases}$$

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".
- The variable profit function is:

$$VP_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ S_{mt}\ \theta_i^M & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 0 \\ S_{mt}\ \theta_i^D & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 1 \end{cases}$$

- S_{mt} is a measure of market size

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".
- The variable profit function is:

$$VP_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ S_{mt}\ \theta_i^M & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 0 \\ S_{mt}\ \theta_i^D & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 1 \end{cases}$$

- S_{mt} is a measure of market size
- a_{jmt} = Indicator of firm j (the opponent) has a plant in the market

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".
- The variable profit function is:

$$VP_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ S_{mt}\ \theta_i^M & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 0 \\ S_{mt}\ \theta_i^D & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 1 \end{cases}$$

- S_{mt} is a measure of market size
- a_{jmt} = Indicator of firm j (the opponent) has a plant in the market
- θ_i^M = Variable profits per-customer if firm i is a monopolist

Variable Profits

- a_{imt} = Indicator of the event "firm i has a plant in market m at period t".
- The variable profit function is:

$$VP_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ S_{mt}\ \theta_i^M & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 0 \\ S_{mt}\ \theta_i^D & \text{if} \quad a_{imt} = 1 \text{ AND } a_{jmt} = 1 \end{cases}$$

- S_{mt} is a measure of market size
- a_{jmt} = Indicator of firm j (the opponent) has a plant in the market
- θ_i^M = Variable profits per-customer if firm i is a monopolist
- θ_i^D = Variable profits per-customer if firm i is a duopolist.

Fixed Costs

- The **fixed cost** is paid every year that the firm has a plant in the market.

$$FC_{imt} \quad = \quad a_{imt} \; (FC_i + \varepsilon_{imt})$$

ε_{imt} represents a firm-idiosyncratic shock in firm i's fixed cost that is iid over firms and over time with a distribution $N(0, \sigma_{\varepsilon}^2)$.

Fixed Costs

- The **fixed cost** is paid every year that the firm has a plant in the market.

$$FC_{imt} = a_{imt} \ (FC_i + \varepsilon_{imt})$$

  ε_{imt} represents a firm-idiosyncratic shock in firm i's fixed cost that is iid over firms and over time with a distribution $N(0, \sigma_\varepsilon^2)$.

- We assume that ε_{imt} is private information of firm i. There are two main reasons to incorporate private information shocks.

Fixed Costs

- The **fixed cost** is paid every year that the firm has a plant in the market.

$$FC_{imt} = a_{imt} \left(FC_i + \varepsilon_{imt} \right)$$

  ε_{imt} represents a firm-idiosyncratic shock in firm i's fixed cost that is iid over firms and over time with a distribution $N(0, \sigma_\varepsilon^2)$.

- We assume that ε_{imt} is private information of firm i. There are two main reasons to incorporate private information shocks.

1. Existence of equilibrium

Fixed Costs

- The **fixed cost** is paid every year that the firm has a plant in the market.

$$FC_{imt} = a_{imt} \left(FC_i + \varepsilon_{imt} \right)$$

ε_{imt} represents a firm-idiosyncratic shock in firm i's fixed cost that is iid over firms and over time with a distribution $N(0, \sigma_\varepsilon^2)$.

- We assume that ε_{imt} is private information of firm i. There are two main reasons to incorporate private information shocks.

1. Existence of equilibrium
2. Convenient econometric errors: they can explain observed heterogeneity in the data without generating endogeneity of opponents' actions.

Entry Costs

- The **entry cost, or cost of setting up a new plant**, is paid only if the plant was not currently active (if $a_{im,t-1} = 0$):

$$EC_{imt} = a_{imt} \left(1 - a_{im,t-1}\right) EC_i$$

Entry Costs

- The **entry cost, or cost of setting up a new plant**, is paid only if the plant was not currently active (if $a_{im,t-1} = 0$):

$$EC_{imt} = a_{imt} (1 - a_{im,t-1}) \, EC_i$$

- We might also incorporate a private information shock in entry costs. Here I consider a simpler version.

Markov Perfect Equilibrium (MPE)

- **Assumption (MPE):** Firms' strategies depend only on payoff relevant variables.

Markov Perfect Equilibrium (MPE)

- **Assumption (MPE):** Firms' strategies depend only on payoff relevant variables.
- The payoff-relevant information of firm i in market m at period t is $\{x_{mt}, \varepsilon_{imt}\}$ where x_{mt} is the vector of **common knowledge state variables**:

$$x_{mt} \equiv \{S_{mt},\ a_{1m,t-1},\ a_{2m,t-1}\}$$

Markov Perfect Equilibrium (MPE)

- **Assumption (MPE):** Firms' strategies depend only on payoff relevant variables.

- The payoff-relevant information of firm i in market m at period t is $\{x_{mt}, \varepsilon_{imt}\}$ where x_{mt} is the vector of **common knowledge state variables**:

$$x_{mt} \equiv \{S_{mt}, \ a_{1m,t-1}, \ a_{2m,t-1}\}$$

- Let $\sigma \equiv \{\sigma_i(x_{mt}, \varepsilon_{imt}) : i = 1, 2\}$ be a set of strategy functions, one for each player.

Markov Perfect Equilibrium (MPE)

- **Assumption (MPE):** Firms' strategies depend only on payoff relevant variables.
- The payoff-relevant information of firm i in market m at period t is $\{x_{mt}, \varepsilon_{imt}\}$ where x_{mt} is the vector of **common knowledge state variables**:

$$x_{mt} \equiv \{S_{mt},\ a_{1m,t-1},\ a_{2m,t-1}\}$$

- Let $\sigma \equiv \{\sigma_i(x_{mt}, \varepsilon_{imt}) : i = 1, 2\}$ be a set of strategy functions, one for each player.
- **σ is a MPE** if, for each firm i, the strategy σ_i maximizes the expected value of firm i at every state $(x_{mt}, \varepsilon_{imt})$ and taking as given the opponent's strategy.

Markov Perfect Equilibrium (MPE)

- It is convenient to represent a strategies and MPE in terms of **choice probabilities.**

Markov Perfect Equilibrium (MPE)

- It is convenient to represent a strategies and MPE in terms of **choice probabilities.**
- Let $P_i(x_{mt})$ represents firm $i's$ probability of being active in the market given state x_{mt}.

Markov Perfect Equilibrium (MPE)

- It is convenient to represent a strategies and MPE in terms of **choice probabilities.**
- Let $P_i(x_{mt})$ represents firm $i's$ probability of being active in the market given state x_{mt}.
- $P_i(x_{mt})$ is the integral of the strategy function $\sigma_i(x_{mt}, \varepsilon_{imt})$ over the distribution of ε_{imt}.

Markov Perfect Equilibrium (MPE)

- It is convenient to represent a strategies and MPE in terms of **choice probabilities.**
- Let $P_i(x_{mt})$ represents firm $i's$ probability of being active in the market given state x_{mt}.
- $P_i(x_{mt})$ is the integral of the strategy function $\sigma_i(x_{mt}, \varepsilon_{imt})$ over the distribution of ε_{imt}.
- We can represent a MPE as a set of probabilities $\mathbf{P} \equiv \{P_i(x_{mt}) : i = 1, 2\}$ **such that** the strategy P_i maximizes the expected value of firm i at every state x_{mt} taking as given the opponent's strategy P_j.

Equilibrium Mapping (1)

- The equilibrium mapping in choice probabilities is the key part of the dynamic game. It summarizes all the relevant structure of the model.

Equilibrium Mapping (1)

- The equilibrium mapping in choice probabilities is the key part of the dynamic game. It summarizes all the relevant structure of the model.
- The form of this equilibrium mapping depends on the payoff/profit function, the transition rule of the state variables, and the distribution of the private information shocks ε_{imt}.

Equilibrium Mapping (1)

- The equilibrium mapping in choice probabilities is the key part of the dynamic game. It summarizes all the relevant structure of the model.
- The form of this equilibrium mapping depends on the payoff/profit function, the transition rule of the state variables, and the distribution of the private information shocks ε_{imt}.
- In this model, the one-period expected profit of firm i can be written as:

$$\Pi_{imt}^{P} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ \mathbf{Z}_{imt}^{P} \ \boldsymbol{\theta}_i - \varepsilon_{imt} & \text{if} \quad a_{imt} = 1 \end{cases}$$

Equilibrium Mapping (1)

- The equilibrium mapping in choice probabilities is the key part of the dynamic game. It summarizes all the relevant structure of the model.
- The form of this equilibrium mapping depends on the payoff/profit function, the transition rule of the state variables, and the distribution of the private information shocks ε_{imt}.
- In this model, the one-period expected profit of firm i can be written as:

$$\Pi^P_{imt} = \begin{cases} 0 & \text{if} \quad a_{imt} = 0 \\ \mathbf{Z}^P_{imt} \; \boldsymbol{\theta}_i - \varepsilon_{imt} & \text{if} \quad a_{imt} = 1 \end{cases}$$

- where

$$\mathbf{Z}^P_{imt} \equiv \{ \; (1 - P_j(x_{mt}))\, S_{mt} \; , \; P_j(x_{mt}) S_{mt} \; , \; -1 \; , \; -(1 - a_{im,t-1})$$

$$\boldsymbol{\theta}_i \equiv \left\{ \; \theta^M_i \; , \; \theta^D_i \; , \; FC_i \; , \; EC_i \; \right\}'$$

Equilibrium Mapping (2)

- In a myopic version of the game (with $\beta = 0$), firm i best response is:

$$\{a_{imt} = 1\} \Leftrightarrow \left\{ \mathbf{Z}_{imt}^{P} \, \boldsymbol{\theta}_i - \varepsilon_{imt} \geq 0 \right\}$$

Equilibrium Mapping (2)

- In a myopic version of the game (with $\beta = 0$), firm i best response is:

$$\{a_{imt} = 1\} \Leftrightarrow \left\{\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_i - \varepsilon_{imt} \geq 0\right\}$$

- And in terms of choice probability, firm i best response is:

$$\Pr\left(a_{imt} = 1 \mid x_{mt}\right) = \Pr\left(\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_i - \varepsilon_{imt} \geq 0 \mid x_{mt}\right) = \Phi\left(\mathbf{Z}_{imt}^{P}\ \frac{\boldsymbol{\theta}_i}{\sigma_{\varepsilon_i}}\right)$$

where $\Phi(.)$ is the CDF of the standard normal.

Equilibrium Mapping (2)

- In a myopic version of the game (with $\beta = 0$), firm i best response is:

$$\{a_{imt} = 1\} \Leftrightarrow \left\{ \mathbf{Z}_{imt}^{P} \ \boldsymbol{\theta}_i - \varepsilon_{imt} \geq 0 \right\}$$

- And in terms of choice probability, firm i best response is:

$$\Pr\left(a_{imt} = 1 \mid x_{mt}\right) = \Pr\left(\mathbf{Z}_{imt}^{P} \ \boldsymbol{\theta}_i - \varepsilon_{imt} \geq 0 \mid x_{mt}\right) = \Phi\left(\mathbf{Z}_{imt}^{P} \ \frac{\boldsymbol{\theta}_i}{\sigma_{\varepsilon_i}}\right)$$

  where $\Phi(.)$ is the CDF of the standard normal.

- A **MPE** in this static game is a pair of probabilities, $\{P_1(x_{mt}), P_2(x_{mt})\}$ that solves the system of equations:

$$\begin{aligned} P_1(x_{mt}) &= \Phi\left(\mathbf{Z}_{1mt}^{P} \ \frac{\boldsymbol{\theta}_1}{\sigma_{\varepsilon_1}}\right) \\ P_2(x_{mt}) &= \Phi\left(\mathbf{Z}_{2mt}^{P} \ \frac{\boldsymbol{\theta}_2}{\sigma_{\varepsilon_2}}\right) \end{aligned}$$

Equilibrium Mapping (3)

- Note that for every value of x_{mt} we have a different BNE. That is not the case in the dynamic version of the game (see below).

Equilibrium Mapping (3)

- Note that for every value of x_{mt} we have a different BNE. That is not the case in the dynamic version of the game (see below).
- By Brower's Theorem, an equilibrium exits.

Equilibrium Mapping (3)

- Note that for every value of x_{mt} we have a different BNE. That is not the case in the dynamic version of the game (see below).
- By Brower's Theorem, an equilibrium exits.
- There may be multiple equilibria for some values of (x_{mt}, θ).

Equilibrium Mapping (4)

- For the dynamic game (with $\beta > 0$), it is possible to show that a MPE is a pair of probability functions, $\mathbf{P} \equiv \{P_1(x_{mt}), P_2(x_{mt}) : x_{mt} \in X\}$, such that, for every firm i and every state x_{mt}:

$$P_i(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \frac{\boldsymbol{\theta}_i}{\sigma_{\varepsilon_i}} + \tilde{e}^{\mathbf{P}}_{imt} \right)$$

where

$$\begin{aligned} \tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \equiv \ & \mathbf{Z}^{\mathbf{P}}_{imt} + E\left(\sum_{s=1}^{\infty} \beta^s \ a_{im,t+s} \ \mathbf{Z}^{\mathbf{P}}_{im,t+s} \mid x_{mt}, \ a_{imt} = 1 \right) \\ & - E\left(\sum_{s=1}^{\infty} \beta^s \ a_{im,t+s} \ \mathbf{Z}^{\mathbf{P}}_{im,t+s} \mid x_{mt}, \ a_{imt} = 0 \right) \end{aligned}$$

and

$$\tilde{e}^{\mathbf{P}}_{imt} \equiv E\left(\sum_{s=1}^{\infty} \beta^s \ a_{im,t+s} \ \varepsilon_{im,t+s} \mid x_{mt}, \ a_{imt} = 1 \right)$$

Econometric Issues

- There are several econometric issues we should deal with when doing statistical inference with dynamic games.

Econometric Issues

- There are several econometric issues we should deal with when doing statistical inference with dynamic games.

1. **Multiple Equilibria.** Models with multiple equilibria do not have unique predictions on the distribution of endogenous variables conditional on exogenous variables and parameters. This can generate problems with some estimation methods.

Econometric Issues

- There are several econometric issues we should deal with when doing statistical inference with dynamic games.

1. **Multiple Equilibria.** Models with multiple equilibria do not have unique predictions on the distribution of endogenous variables conditional on exogenous variables and parameters. This can generate problems with some estimation methods.
2. **Computational burden.** The dimension of the state space increases exponentially with the number of heterogeneous firms.

Econometric Issues

- There are several econometric issues we should deal with when doing statistical inference with dynamic games.

1. **Multiple Equilibria.** Models with multiple equilibria do not have unique predictions on the distribution of endogenous variables conditional on exogenous variables and parameters. This can generate problems with some estimation methods.
2. **Computational burden.** The dimension of the state space increases exponentially with the number of heterogeneous firms.
3. **Endogeneity.** Unobserved (for the researcher) market characteristics are correlated with opponents' decisions.

Econometric Issues

- There are several econometric issues we should deal with when doing statistical inference with dynamic games.

1. **Multiple Equilibria.** Models with multiple equilibria do not have unique predictions on the distribution of endogenous variables conditional on exogenous variables and parameters. This can generate problems with some estimation methods.
2. **Computational burden.** The dimension of the state space increases exponentially with the number of heterogeneous firms.
3. **Endogeneity.** Unobserved (for the researcher) market characteristics are correlated with opponents' decisions.
4. **Counterfactual experiment and multiple equilibria.**

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

- I will describe 5 groups of estimators:

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.
- I will describe 5 groups of estimators:

  1. Constrained MLE / Constrained GMM

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

- I will describe 5 groups of estimators:

  1. Constrained MLE / Constrained GMM
  2. Two-step methods (with moment equalities)

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

- I will describe 5 groups of estimators:

  1. Constrained MLE / Constrained GMM
  2. Two-step methods (with moment equalities)
  3. Estimators based on moment inequalities (MI)

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

- I will describe 5 groups of estimators:

  1. Constrained MLE / Constrained GMM
  2. Two-step methods (with moment equalities)
  3. Estimators based on moment inequalities (MI)
  4. Simulation-based estimation

Estimation: Introduction (1)

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

$$Data = \{S_{mt},\ a_{1mt},\ a_{2mt} : m = 1, 2, ..., M;\ t = 1, 2, ..., T\}$$

- We want to use these data to estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Note: for notational simplicity, I use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i / \sigma_{\varepsilon_i}$.

- I will describe 5 groups of estimators:

  1. Constrained MLE / Constrained GMM
  2. Two-step methods (with moment equalities)
  3. Estimators based on moment inequalities (MI)
  4. Simulation-based estimation
  5. Estimators for models with unobserved market heterogeneity.

Estimation: Introduction (2)

- The following assumption is maintained for some estimation methods.

Estimation: Introduction (2)

- The following assumption is maintained for some estimation methods.
- **Assumption: No-unobserved-market-heterogeneity.** The only unobservables from the point of view of the econometrician are the private information shocks ε_{imt}.

Estimation: Introduction (2)

- The following assumption is maintained for some estimation methods.
- **Assumption: No-unobserved-market-heterogeneity.** The only unobservables from the point of view of the econometrician are the private information shocks ε_{imt}.
- The following assumption is common to the different estimation methods that we will examine.

Estimation: Introduction (2)

- The following assumption is maintained for some estimation methods.
- **Assumption: No-unobserved-market-heterogeneity.** The only unobservables from the point of view of the econometrician are the private information shocks ε_{imt}.
- The following assumption is common to the different estimation methods that we will examine.
- **Assumption: One-MPE-in-the-data.** Define $P_{imt}^{0}(x) \equiv \Pr(a_{imt} = 1 | x_{mt} = x)$ and $\mathbf{P}_{mt}^{0} \equiv \{P_{imt}^{0}(x) : i = 1, 2; x \in X\}$. Then, for any (m, t), $\mathbf{P}_{mt}^{0} = \mathbf{P}^{0}$. Though the model has multiple equilibria, in the population the same MPE is selected at every market and every time period.

Estimation: Introduction (3)

- For the description of the different estimators, it is convenient to define the following **Pseudo Likelihood function**:

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \mathbf{P}) \quad = \quad & \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{imt} \; \ln \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}} \; \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}} \right) \\
& + (1 - a_{imt}) \ln \Phi\left(-\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}} \; \boldsymbol{\theta}_i - \tilde{e}_{imt}^{\mathbf{P}} \right)
\end{aligned}
$$

Estimation: Introduction (3)

- For the description of the different estimators, it is convenient to define the following **Pseudo Likelihood function**:

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \mathbf{P}) &= \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{imt} \ \ln \Phi \left(\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \ \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt} \right) \\
&\quad + (1 - a_{imt}) \ln \Phi \left(-\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \ \boldsymbol{\theta}_i - \tilde{e}^{\mathbf{P}}_{imt} \right)
\end{aligned}
$$

- This pseudo likelihood function treats firms' beliefs **P** as parameters to estimate together with $\boldsymbol{\theta}$.

Estimation: Introduction (3)

- For the description of the different estimators, it is convenient to define the following **Pseudo Likelihood function**:

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \mathbf{P}) \quad = \quad & \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{imt} \ \ln \Phi \left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}} \ \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}} \right) \\
& + (1 - a_{imt}) \ln \Phi \left(-\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}} \ \boldsymbol{\theta}_i - \tilde{e}_{imt}^{\mathbf{P}} \right)
\end{aligned}
$$

- This pseudo likelihood function treats firms' beliefs $\mathbf{P}$ as parameters to estimate together with $\boldsymbol{\theta}$.

- Note that for given $\mathbf{P}$, the function $Q(\boldsymbol{\theta}, \mathbf{P})$ is the likelihood of a Probit model with the parameter of $\tilde{e}_{imt}^{\mathbf{P}}$ restricted to be one.

Estimation: Introduction (4)

- Similarly, for GMM estimation we can define the following **Pseudo GMM Criterion**: (note that there is a minus sign so the GMM estimator maximizes this criterion):

$$Q(\boldsymbol{\theta}, \mathbf{P}) = -c(\boldsymbol{\theta}, \mathbf{P})\ 'A\ c(\boldsymbol{\theta}, \mathbf{P})$$

where

$$c(\boldsymbol{\theta}, \mathbf{P}) = \sum_{m=1}^{M} \left[\begin{array}{c} \dots \\ H(x_{mt})\ \left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}\ \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}} \right) \right\} \\ \dots \text{ for any } (i, t) \end{array} \right]$$

and $H(x_{mt})$ is a vector of functions of x_{mt} (instruments).

Estimation: Introduction (4)

- Similarly, for GMM estimation we can define the following **Pseudo GMM Criterion**: (note that there is a minus sign so the GMM estimator maximizes this criterion):

$$Q(\boldsymbol{\theta}, \mathbf{P}) = -c(\boldsymbol{\theta}, \mathbf{P})\ 'A\ c(\boldsymbol{\theta}, \mathbf{P})$$

where

$$c(\boldsymbol{\theta}, \mathbf{P}) = \sum_{m=1}^{M} \left[\begin{array}{c} \ldots \\ H(x_{mt})\ \left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}\ \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}} \right) \right\} \\ \ldots \text{ for any } (i, t) \end{array} \right]$$

and $H(x_{mt})$ is a vector of functions of x_{mt} (instruments).

- This pseudo GMM criterion function treats firms' beliefs $\mathbf{P}$ as parameters to estimate together with $\boldsymbol{\theta}$.

Constrained MLE / Constrained GMM (1)

- The MLE (the GMM) maximizes $Q(\boldsymbol{\theta}, \mathbf{P})$ subject to the restriction that $\mathbf{P}$ should be an equilibrium associated with $\boldsymbol{\theta}$.

Constrained MLE / Constrained GMM (1)

- The MLE (the GMM) maximizes $Q(\boldsymbol{\theta}, \mathbf{P})$ subject to the restriction that $\mathbf{P}$ should be an equilibrium associated with $\boldsymbol{\theta}$.
- The main issue for the implementation of the MLE (or the GMM) in this model is that for some values of $\boldsymbol{\theta}$ the model has multiple equilibria.

Constrained MLE / Constrained GMM (1)

- The MLE (the GMM) maximizes $Q(\boldsymbol{\theta}, \mathbf{P})$ subject to the restriction that $\mathbf{P}$ should be an equilibrium associated with $\boldsymbol{\theta}$.
- The main issue for the implementation of the MLE (or the GMM) in this model is that for some values of $\boldsymbol{\theta}$ the model has multiple equilibria.
- The MLE (or GMM) is define as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \begin{cases} \max_{\mathbf{P}} \; Q(\boldsymbol{\theta}, \mathbf{P}) \\ \text{subject to: } P_i\left(x_{mt}\right) = \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \; \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt}\right) \text{ for every } (i, \end{cases}$$

Constrained MLE / Constrained GMM (1)

- The MLE (the GMM) maximizes $Q(\boldsymbol{\theta}, \mathbf{P})$ subject to the restriction that $\mathbf{P}$ should be an equilibrium associated with $\boldsymbol{\theta}$.
- The main issue for the implementation of the MLE (or the GMM) in this model is that for some values of $\boldsymbol{\theta}$ the model has multiple equilibria.
- The MLE (or GMM) is define as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \begin{cases} \max_{\mathbf{P}} \; Q(\boldsymbol{\theta}, \mathbf{P}) \\ \text{subject to: } P_i\left(x_{mt}\right) = \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt} \; \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt}\right) \text{ for every } (i, \end{cases}$$

- Since $\mathbf{P}$ is a high-dimension vector, optimization with respect to $\mathbf{P}$ can be very complicated.

Constrained MLE / Constrained GMM (2)

- Note that, without the **one-equilibrium-in-the-data assumption**, the MLE (or the GMM) would be even more complicated to implement.

Constrained MLE / Constrained GMM (2)

- Note that, without the **one-equilibrium-in-the-data assumption**, the MLE (or the GMM) would be even more complicated to implement.
- In that case, the criterion function would be $Q(\boldsymbol{\theta}, \{\mathbf{P}_{mt}\})$, where $\{\mathbf{P}_{mt}\}$ represents the set of MPE, one for each market and time period.

Two-step methods (1)

- Suppose that we knew the equilibrium in the population, $\mathbf{P}^0$.

Two-step methods (1)

- Suppose that we knew the equilibrium in the population, $\mathbf{P}^0$.
- Given $\mathbf{P}^0$ we can construct the variables $\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$ and then obtain a very simple estimator of $\boldsymbol{\theta}^0$.

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

Two-step methods (1)

- Suppose that we knew the equilibrium in the population, $\mathbf{P}^0$.
- Given $\mathbf{P}^0$ we can construct the variables $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$ and then obtain a very simple estimator of $\boldsymbol{\theta}^0$.

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

- 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.

Two-step methods (1)

- Suppose that we knew the equilibrium in the population, $\mathbf{P}^0$.
- Given $\mathbf{P}^0$ we can construct the variables $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$ and then obtain a very simple estimator of $\boldsymbol{\theta}^0$.

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

- 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.
- While equilibrium probabilities are not unique functions of structural parameters, the best response probabilities that appear in $Q(\boldsymbol{\theta}, \mathbf{P})$ are unique functions of structural parameters and players' beliefs.

Two-step methods (2)

- The previous method is infeasible because $\mathbf{P}^0$ is unknown.

Two-step methods (2)

- The previous method is infeasible because $\mathbf{P}^0$ is unknown.
- However, under the Assumptions **"No-unobserved-market-heterogeneity"** and **"One-MPE-in-the-data"** we can estimate $\mathbf{P}^0$ consistently and at with a convergence rate such that the two-step estimator $\hat{\theta}$ is root-M consistent and asymptotically normal.

Two-step methods (2)

- The previous method is infeasible because $\mathbf{P}^0$ is unknown.
- However, under the Assumptions **"No-unobserved-market-heterogeneity"** and **"One-MPE-in-the-data"** we can estimate $\mathbf{P}^0$ consistently and at with a convergence rate such that the two-step estimator $\hat{\boldsymbol{\theta}}$ is root-M consistent and asymptotically normal.
- For instance, a kernel estimator of $\mathbf{P}^0$ is:

$$\hat{P}_i^0(x) = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T} a_{imt} \ K\left(\frac{x_{mt} - x}{b}\right)}{\sum_{m=1}^{M} \sum_{t=1}^{T} K\left(\frac{x_{mt} - x}{b}\right)}$$

Two-step methods (3)

- Given $\hat{\mathbf{P}}^0$, we can construct a consistent estimator of $Z_{imt}^{P^0}$:

$$Z_{imt}^{\hat{\mathbf{P}}^0} = \left\{ \left(1 - \hat{P}_j^0(x_{mt})\right) S_{mt} \ , \ \hat{P}_j^0(x_{mt}) S_{mt} \ , \ -1 \ , \ -(1 - a_{im,t-}$$

Two-step methods (3)

- Given $\hat{\mathbf{P}}^0$, we can construct a consistent estimator of $Z_{imt}^{P^0}$:

$$Z_{imt}^{\hat{\mathbf{P}}^0} = \left\{ \left(1 - \hat{P}_j^0(x_{mt})\right) S_{mt} \;,\; \hat{P}_j^0(x_{mt}) S_{mt} \;,\; -1 \;,\; -(1 - a_{im,t-}$$

- And then, consistent estimators of $\tilde{Z}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$:

$$\tilde{Z}_{imt}^{\hat{\mathbf{P}}^0} = Z_{imt}^{\hat{\mathbf{P}}^0} + \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \; \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \right.$$

$$\tilde{e}_{imt}^{\hat{\mathbf{P}}^0} = \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \; \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \mathbf{e}_i^{\hat{\mathbf{P}}^0} \right\}$$

Two-step methods (3)

- Given $\hat{\mathbf{P}}^0$, we can construct a consistent estimator of $Z_{imt}^{P^0}$:

$$Z_{imt}^{\hat{\mathbf{P}}^0} = \left\{ \left(1 - \hat{P}_j^0(x_{mt})\right) S_{mt} \ , \ \hat{P}_j^0(x_{mt}) S_{mt} \ , \ -1 \ , \ -(1 - a_{im,t-}$$

- And then, consistent estimators of $\tilde{Z}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$:

$$\tilde{Z}_{imt}^{\hat{\mathbf{P}}^0} = Z_{imt}^{\hat{\mathbf{P}}^0} + \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \ \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \right.$$

$$\tilde{e}_{imt}^{\hat{\mathbf{P}}^0} = \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \ \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \mathbf{e}_i^{\hat{\mathbf{P}}^0} \right\}$$

- The two-step estimator is the value of $\boldsymbol{\theta}$ that maximizes $Q(\theta, \hat{\mathbf{P}}^0)$.

Two-step methods (3)

- Given $\hat{\mathbf{P}}^0$, we can construct a consistent estimator of $Z_{imt}^{P^0}$:

$$Z_{imt}^{\hat{\mathbf{P}}^0} = \left\{ \left(1 - \hat{P}_j^0(x_{mt})\right) S_{mt} \; , \; \hat{P}_j^0(x_{mt}) S_{mt} \; , \; -1 \; , \; -(1 - a_{im,t-}$$

- And then, consistent estimators of $\tilde{Z}_{imt}^{\mathbf{P}^0}$ and $\tilde{e}_{imt}^{\mathbf{P}^0}$:

$$\tilde{Z}_{imt}^{\hat{\mathbf{P}}^0} = Z_{imt}^{\hat{\mathbf{P}}^0} + \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \; \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \right.$$

$$\tilde{e}_{imt}^{\hat{\mathbf{P}}^0} = \beta \left(F_x^{\hat{\mathbf{P}}^0}(1, x_{mt}) - F_x^{\hat{\mathbf{P}}^0}(0, x_{mt}) \right)' \left\{ \mathbf{I} - \beta \; \mathbf{F}_x^{\hat{\mathbf{P}}^0} \right\}^{-1} \left\{ \hat{\mathbf{P}}_i^0 * \mathbf{e}_i^{\hat{\mathbf{P}}^0} \right\}$$

- The two-step estimator is the value of $\boldsymbol{\theta}$ that maximizes $Q(\theta, \hat{\mathbf{P}}^0)$.
- For instance, the two-step MLE is simply the MLE in a probit model for a_{imt} with explanatory variables $\tilde{Z}_{imt}^{\hat{\mathbf{P}}^0}$ and $\tilde{e}_{imt}^{\hat{\mathbf{P}}^0}$.

Two-step methods (4)

- We will examine the following issues related to this class of two-step estimators.

Two-step methods (4)

- We will examine the following issues related to this class of two-step estimators.

1. Asymptotic variance

Two-step methods (4)

- We will examine the following issues related to this class of two-step estimators.

1. Asymptotic variance
2. 2. Efficient two-step estimator

Two-step methods (4)

- We will examine the following issues related to this class of two-step estimators.

1. Asymptotic variance
2. 2. Efficient two-step estimator
3. Finite sample properties

Two-step methods (4)

- We will examine the following issues related to this class of two-step estimators.

1. Asymptotic variance
2. 2. Efficient two-step estimator
3. Finite sample properties
4. Dealing with unobserved heterogeneity

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).
- The asymptotic variance of the two-step MLE is:

$$V_{2S} = \Omega_{\theta\theta}^{-1} + \Omega_{\theta\theta}^{-1} \; \Omega_{\theta P} \; \Sigma \; \Omega_{\theta P}' \; \Omega_{\theta\theta}^{-1}$$

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).
- The asymptotic variance of the two-step MLE is:

$$V_{2S} = \Omega_{\theta\theta}^{-1} + \Omega_{\theta\theta}^{-1}\ \Omega_{\theta P}\ \Sigma\ \Omega_{\theta P}'\ \Omega_{\theta\theta}^{-1}$$

- Σ is the variance matrix of $\hat{\mathbf{P}}^0$

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).
- The asymptotic variance of the two-step MLE is:

$$V_{2S} = \Omega_{\theta\theta}^{-1} + \Omega_{\theta\theta}^{-1} \ \Omega_{\theta P} \ \Sigma \ \Omega_{\theta P}' \ \Omega_{\theta\theta}^{-1}$$

- Σ is the variance matrix of $\hat{\mathbf{P}}^0$
- $\Omega_{\theta\theta} \equiv E\left(\nabla_\theta s_{mt} \nabla_\theta s_{mt}'\right)$

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).
- The asymptotic variance of the two-step MLE is:

$$V_{2S} = \Omega_{\theta\theta}^{-1} + \Omega_{\theta\theta}^{-1} \; \Omega_{\theta P} \; \Sigma \; \Omega_{\theta P}' \; \Omega_{\theta\theta}^{-1}$$

- Σ is the variance matrix of $\hat{\mathbf{P}}^0$
- $\Omega_{\theta\theta} \equiv E\left(\nabla_\theta s_{mt} \nabla_\theta s_{mt}'\right)$
- $\Omega_{\theta P} \equiv E\left(\nabla_\theta s_{mt} \nabla_P s_{mt}'\right)$

Two-step methods: Asymptotic variance

- Let $s_{mt}(\boldsymbol{\theta}^0, \mathbf{P}^0)$ be the score for observation (m, t).
- The asymptotic variance of the two-step MLE is:

$$V_{2S} = \Omega_{\theta\theta}^{-1} + \Omega_{\theta\theta}^{-1} \ \Omega_{\theta P} \ \Sigma \ \Omega_{\theta P}' \ \Omega_{\theta\theta}^{-1}$$

- Σ is the variance matrix of $\hat{\mathbf{P}}^0$
- $\Omega_{\theta\theta} \equiv E\left(\nabla_\theta s_{mt} \nabla_\theta s_{mt}'\right)$
- $\Omega_{\theta P} \equiv E\left(\nabla_\theta s_{mt} \nabla_P s_{mt}'\right)$
- and ∇_θ and ∇_P represent partial derivatives w.r.t. θ and P, respectively.

Two-step methods: Efficient estimator (1)

- These two-step estimators belong to the class of models defined in terms of conditional moment restrictions.

$$
\begin{aligned}
E\left(a_{imt} - P^0(x_{mt}) \mid x_{mt}\right) &= 0 \\
E\left(a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}^0}\ \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}^0}\right) \mid x_{mt}\right) &= 0
\end{aligned}
$$

Two-step methods: Efficient estimator (1)

- These two-step estimators belong to the class of models defined in terms of conditional moment restrictions.

$$
\begin{aligned}
E\left(a_{imt} - P^0(x_{mt}) \mid x_{mt}\right) &= 0 \\
E\left(a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}^0}_{imt}\ \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}^0}_{imt}\right) \mid x_{mt}\right) &= 0
\end{aligned}
$$

- Newey (1990) obtained the form of the optimal instruments (unconditional moment restrictions) for this class of models:

$$
E\left(H^*(x_{mt})\left\{\begin{array}{c} a_{imt} - P^0(x_{mt}) \quad \text{for any } i,t \\ \\ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}^0}_{imt}\ \boldsymbol{\theta}^0_i + \tilde{e}^{\mathbf{P}^0}_{imt}\right) \quad \text{for any } i,t \end{array}\right\}\right) = 0
$$

where $H^*(x_{mt})$ is the matrix of optimal instruments.

Two-step methods: Efficient estimator (2)

- Given that the model has a triangular form (i.e., $\boldsymbol{\theta}^0$ only enters in the second group of conditional moment restrictions), the matrix of optimal instruments is also triangular, such that the efficient estimator can be implemented in two steps.

$$E\left(H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})\left\{\ a_{imt}-\Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \boldsymbol{\theta}^0_i+\tilde{e}^{\hat{\mathbf{P}}^0}_{imt}\right)\quad \text{for any } i,t\ \right\}\right)\ =\ 0$$

has minimum asymptotic variance.

Two-step methods: Efficient estimator (2)

- Given that the model has a triangular form (i.e., $\boldsymbol{\theta}^0$ only enters in the second group of conditional moment restrictions), the matrix of optimal instruments is also triangular, such that the efficient estimator can be implemented in two steps.
- Given $\hat{\mathbf{P}}^0$, there is an optimal matrix of instruments, $H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})$, such that the GMM estimator of $\boldsymbol{\theta}^0$ based on the unconditional moment restrictions

$$E\left(H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})\left\{\ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \boldsymbol{\theta}^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt}\right) \quad \text{for any } i,t\ \right\}\right) = 0$$

has minimum asymptotic variance.

Two-step methods: Efficient estimator (2)

- Given that the model has a triangular form (i.e., $\boldsymbol{\theta}^0$ only enters in the second group of conditional moment restrictions), the matrix of optimal instruments is also triangular, such that the efficient estimator can be implemented in two steps.
- Given $\hat{\mathbf{P}}^0$, there is an optimal matrix of instruments, $H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})$, such that the GMM estimator of $\boldsymbol{\theta}^0$ based on the unconditional moment restrictions

$$E\left(H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})\left\{\ a_{imt}-\Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \boldsymbol{\theta}^0_i+\tilde{e}^{\hat{\mathbf{P}}^0}_{imt}\right)\quad \text{for any } i,t\ \right\}\right)\ =\ 0$$

has minimum asymptotic variance.

- Pesendorfer and Schmidt-Dengler (REStud, 2008) derive the form of $H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})$.

Two-step methods: Efficient estimator (2)

- Given that the model has a triangular form (i.e., $\boldsymbol{\theta}^0$ only enters in the second group of conditional moment restrictions), the matrix of optimal instruments is also triangular, such that the efficient estimator can be implemented in two steps.
- Given $\hat{\mathbf{P}}^0$, there is an optimal matrix of instruments, $H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})$, such that the GMM estimator of $\boldsymbol{\theta}^0$ based on the unconditional moment restrictions

$$E\left(H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})\left\{\; a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\;\theta^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt}\right) \quad \text{for any } i,t \;\right\}\right) \;=\; 0$$

has minimum asymptotic variance.

- Pesendorfer and Schmidt-Dengler (REStud, 2008) derive the form of $H^*_{\boldsymbol{\theta},\hat{\mathbf{P}}^0}(x_{mt})$.
- When $\mathbf{P}^0$ is known (i.e., when we ignore the estimation error in the first-step), the optimal estimator is the MLE based on probs $\Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}\;\theta^0 + \tilde{e}^{\hat{\mathbf{P}}^0}\right)$

Two-step methods: Finite sample properties (1)

- The most attractive feature of two-step methods is their relative simplicity.

Two-step methods: Finite sample properties (1)

- The most attractive feature of two-step methods is their relative simplicity.
- However, they suffer of a potentially important problem of finite sample bias.

Two-step methods: Finite sample properties (1)

- The most attractive feature of two-step methods is their relative simplicity.
- However, they suffer of a potentially important problem of finite sample bias.
- The finite sample bias of the two-step estimator of $\boldsymbol{\theta}^0$ depends very importantly on the properties of the first-step estimator of $\mathbf{P}^0$. In particular, it depends on the rate of convergence and on the variance and bias of $\hat{\mathbf{P}}^0$.

Two-step methods: Finite sample properties (1)

- The most attractive feature of two-step methods is their relative simplicity.
- However, they suffer of a potentially important problem of finite sample bias.
- The finite sample bias of the two-step estimator of $\boldsymbol{\theta}^0$ depends very importantly on the properties of the first-step estimator of $\mathbf{P}^0$. In particular, it depends on the rate of convergence and on the variance and bias of $\hat{\mathbf{P}}^0$.
- It is well-known that there is a **curse of dimensionality in the NP estimation** of a regression function such as $\mathbf{P}^0$.

Two-step methods: Finite sample properties (2)

- In particular, the rate of convergence of $\hat{\mathbf{P}}^0$ declines, and the variance and bias increase, very quickly as the number of conditioning regressors increases.

Two-step methods: Finite sample properties (2)

- In particular, the rate of convergence of $\hat{\mathbf{P}}^0$ declines, and the variance and bias increase, very quickly as the number of conditioning regressors increases.

- In our simple example, the vector x_{mt} contains only three variables: the binary indicators $a_{im,t-1}$ and the (continuous) market size S_{mt}. In this case, the NP estimator of $\mathbf{P}^0$ has a relatively high rate of convergence an its variance and bias can be small even with relatively small sample.

Two-step methods: Finite sample properties (2)

- In particular, the rate of convergence of $\hat{\mathbf{P}}^0$ declines, and the variance and bias increase, very quickly as the number of conditioning regressors increases.

- In our simple example, the vector x_{mt} contains only three variables: the binary indicators $a_{im,t-1}$ and the (continuous) market size S_{mt}. In this case, the NP estimator of $\mathbf{P}^0$ has a relatively high rate of convergence an its variance and bias can be small even with relatively small sample.

- However, there are applications with more than two (heterogeneous) players and where firm size, capital stock or other predetermined continuos firm-specific characteristics are state variables.

Two-step methods: Finite sample properties (2)

- In particular, the rate of convergence of $\hat{\mathbf{P}}^0$ declines, and the variance and bias increase, very quickly as the number of conditioning regressors increases.
- In our simple example, the vector x_{mt} contains only three variables: the binary indicators $a_{im,t-1}$ and the (continuous) market size S_{mt}. In this case, the NP estimator of $\mathbf{P}^0$ has a relatively high rate of convergence an its variance and bias can be small even with relatively small sample.
- However, there are applications with more than two (heterogeneous) players and where firm size, capital stock or other predetermined continuos firm-specific characteristics are state variables.
- Even with binary state variables ($a_{im,t-1}$), when the number of players is relatively large (e.g., more than 10) .....

Two-step methods: Finite sample properties (3)

- In these situations, with a relatively large number of heterogeneous players, even the best feasible estimator of $\mathbf{P}^0$ will be extremely noisy/imprecise: large variance and/or bias.

Two-step methods: Finite sample properties (3)

- In these situations, with a relatively large number of heterogeneous players, even the best feasible estimator of $\mathbf{P}^0$ will be extremely noisy/imprecise: large variance and/or bias.
- Large variance and/or bias of $\hat{\mathbf{P}}^0$ implies a large bias of the two-step estimator of θ^0.

Two-step methods: Finite sample properties (3)

- In these situations, with a relatively large number of heterogeneous players, even the best feasible estimator of $\mathbf{P}^0$ will be extremely noisy/imprecise: large variance and/or bias.
- Large variance and/or bias of $\hat{\mathbf{P}}^0$ implies a large bias of the two-step estimator of $\boldsymbol{\theta}^0$.
- To see this, note that the moment conditions at the true $\mathbf{P}^0$ hold:

$$E\left(H_{\mathbf{P}^0}(x_{mt})\left\{\ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}^0}\ \boldsymbol{\theta}_i^0 + \tilde{e}_{imt}^{\hat{\mathbf{P}}^0}\right) \quad \text{for any } i, t\ \right\}\right) \ = \ 0$$

Two-step methods: Finite sample properties (3)

- In these situations, with a relatively large number of heterogeneous players, even the best feasible estimator of $\mathbf{P}^0$ will be extremely noisy/imprecise: large variance and/or bias.

- Large variance and/or bias of $\hat{\mathbf{P}}^0$ implies a large bias of the two-step estimator of $\boldsymbol{\theta}^0$.

- To see this, note that the moment conditions at the true $\mathbf{P}^0$ hold:

$$E\left(H_{\mathbf{P}^0}(x_{mt}) \left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\mathbf{P}^0}_{imt} \, \boldsymbol{\theta}^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt} \right) \quad \text{for any } i, t \right\} \right) = 0$$

- but the same moment conditions evaluated at $\hat{\mathbf{P}}^0$ do not hold

$$E\left(H_{\hat{\mathbf{P}}^0}(x_{mt}) \left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt} \, \boldsymbol{\theta}^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt} \right) \quad \text{for any } i, t \right\} \right) \neq 0$$

Two-step methods: Finite sample properties (4)

- There are two sources of finite sample bias.

Two-step methods: Finite sample properties (4)

- There are two sources of finite sample bias.

1. Both the matrix of "instruments" $H_{\hat{\mathbf{P}}^0}(x_{mt})$ and the "error" $\left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^0} \, \theta_i^0 + \tilde{e}_{imt}^{\hat{\mathbf{P}}^0} \right) \right\}$ depend on the random vector $\hat{\mathbf{P}}_0$, and this generates a finite sample correlation between $H_{\hat{\mathbf{P}}^0}(x_{mt})$ and $\left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^0} \, \theta_i^0 + \tilde{e}_{imt}^{\hat{\mathbf{P}}^0} \right) \right\}$.

Two-step methods: Finite sample properties (4)

- There are two sources of finite sample bias.

1. Both the matrix of "instruments" $H_{\hat{\mathbf{P}}^0}(x_{mt})$ and the "error" $\left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \theta^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt} \right) \right\}$ depend on the random vector $\hat{\mathbf{P}}_0$, and this generates a finite sample correlation between $H_{\hat{\mathbf{P}}^0}(x_{mt})$ and $\left\{ a_{imt} - \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \theta^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt} \right) \right\}$.

2. Even when the instruments $H(x_{mt})$ do not depend on $\hat{\mathbf{P}}^0$, we have that the choice probabilities $\Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0}_{imt}\ \theta^0_i + \tilde{e}^{\hat{\mathbf{P}}^0}_{imt} \right)$ are nonlinear functions of the random vector $\hat{\mathbf{P}}_0$, and the expected value of a nonlinear function is not equal to the function evaluated at the expected value.

Two-step methods: Finite sample properties (5)

- As argued by Pakes, Ostrovsky and Berry (RAND, 2007), the first source of bias is present in the two-step MLE or in the efficient-two-step estimators, but not in a simpler method of moments estimator where the matrix of instruments does not depend on $\hat{\mathbf{P}}^0$. They advocate this type of two-step estimator.

Two-step methods: Finite sample properties (5)

- As argued by Pakes, Ostrovsky and Berry (RAND, 2007), the first source of bias is present in the two-step MLE or in the efficient-two-step estimators, but not in a simpler method of moments estimator where the matrix of instruments does not depend on $\hat{\mathbf{P}}^0$. They advocate this type of two-step estimator.
- However, the second source of bias appears in all these two-steps estimators and it can be very important.

Two-step methods: Finite sample properties (5)

- As argued by Pakes, Ostrovsky and Berry (RAND, 2007), the first source of bias is present in the two-step MLE or in the efficient-two-step estimators, but not in a simpler method of moments estimator where the matrix of instruments does not depend on $\hat{\mathbf{P}}^0$. They advocate this type of two-step estimator.
- However, the second source of bias appears in all these two-steps estimators and it can be very important.
- Monte Carlo experiments of several papers (see the Monte Carlo experiments in Hotz et al., 1994, Aguirregabiria and Mira, 2002 and 2007, Kasahara and Shimotsu, 2006) illustrate that this bias is very serious even in relatively simple models.

Recursive K-step estimator

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

$$\hat{\boldsymbol{\theta}}^K = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^{K-1}\right)$$

$$\hat{P}_i^K(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^{K-1}} \; \hat{\boldsymbol{\theta}}_i^K + \tilde{e}_{imt}^{\hat{\mathbf{P}}^{K-1}}\right)$$

Recursive K-step estimator

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

$$\hat{\boldsymbol{\theta}}^K = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^{K-1}\right)$$

$$\hat{P}_i^K(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^{K-1}} \; \hat{\boldsymbol{\theta}}_i^K + \tilde{e}_{imt}^{\hat{\mathbf{P}}^{K-1}}\right)$$

- Aguirregabiria and Mira (2002, 2007) present Monte Carlo experiments which illustrate how this recursive estimators can have significantly smaller bias than the two-step estimator.

Recursive K-step estimator

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

$$\hat{\boldsymbol{\theta}}^K = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^{K-1}\right)$$

$$\hat{P}_i^K(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^{K-1}} \; \hat{\boldsymbol{\theta}}_i^K + \tilde{e}_{imt}^{\hat{\mathbf{P}}^{K-1}}\right)$$

- Aguirregabiria and Mira (2002, 2007) present Monte Carlo experiments which illustrate how this recursive estimators can have significantly smaller bias than the two-step estimator.

- Kasahara and Shimotsu (2007) derive a second order approximation to the bias of these K-stage estimators. They show that, if the equilibrium in the population is stable, then this recursive procedure reduces the bias.

Estimation using Moment Inequalities (1)

- Remember that:

$$\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}\ \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt} \equiv \left(\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(1) - \widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(0)\right)\boldsymbol{\theta}_i + \left(e^{\mathbf{P}}_{imt}(1) - e^{\mathbf{P}}_{imt}(0)\right)$$

  where $\mathbf{Z}^{\mathbf{P}}_{imt}(a_i)\boldsymbol{\theta}_i + e^{\mathbf{P}}_{imt}(a_i)$ is the value of choosing alternative a_i today, given that firms behave in the future according to the probabilities in $\mathbf{P}$.

Estimation using Moment Inequalities (1)

- Remember that:

$$\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}} \ \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}} \equiv \left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(1) - \tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(0)\right) \boldsymbol{\theta}_i + \left(e_{imt}^{\mathbf{P}}(1) - e_{imt}^{\mathbf{P}}(0)\right)$$

  where $\mathbf{Z}_{imt}^{\mathbf{P}}(a_i)\boldsymbol{\theta}_i + e_{imt}^{\mathbf{P}}(a_i)$ is the value of choosing alternative a_i today, given that firms behave in the future according to the probabilities in $\mathbf{P}$.

- Then, the value $V_{imt}^{\mathbf{P}}$ is:

$$\begin{aligned} V_{imt}^{\mathbf{P}} &= (1 - P_i(x_{mt})) \left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(0)\boldsymbol{\theta}_i + e_{imt}^{\mathbf{P}}(0)\right) + P_i(x_{mt}) \left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(1)\boldsymbol{\theta}_i + \right. \\ &= W_{imt}^{\mathbf{P}} \begin{pmatrix} \boldsymbol{\theta}_i \\ 1 \end{pmatrix} \end{aligned}$$

Estimation using Moment Inequalities (1)

- Remember that:

$$\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}\ \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt} \equiv \left(\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(1) - \widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(0)\right)\boldsymbol{\theta}_i + \left(e^{\mathbf{P}}_{imt}(1) - e^{\mathbf{P}}_{imt}(0)\right)$$

  where $\mathbf{Z}^{\mathbf{P}}_{imt}(a_i)\boldsymbol{\theta}_i + e^{\mathbf{P}}_{imt}(a_i)$ is the value of choosing alternative a_i today, given that firms behave in the future according to the probabilities in $\mathbf{P}$.

- Then, the value $V^{\mathbf{P}}_{imt}$ is:

$$\begin{aligned} V^{\mathbf{P}}_{imt} &= (1 - P_i(x_{mt}))\left(\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(0)\boldsymbol{\theta}_i + e^{\mathbf{P}}_{imt}(0)\right) + P_i(x_{mt})\left(\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(1)\boldsymbol{\theta}_i \right. - \\ &= W^{\mathbf{P}}_{imt}\begin{pmatrix}\boldsymbol{\theta}_i \\ 1\end{pmatrix}\end{aligned}$$

- where $W^{\mathbf{P}}_{imt}$ is a vector

$$\left(\left(1 - P_i(x_{mt})\right)\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(0) + P_i(x_{mt})\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(1)\ ;\ \left(1 - P_i(x_{mt})\right)e^{\mathbf{P}}_{imt}(0) + P_i\right.$$

Estimation using Moment Inequalities (2)

- Let's split the vector of choice probabilities $\mathbf{P}$ into the sub-vectors $\mathbf{P}_i$ and $\mathbf{P}_{-i}$,

$$\mathbf{P} \equiv (\mathbf{P}_i \ , \ \mathbf{P}_{-i})$$

where $\mathbf{P}_i$ are the probabilities associated to player i and $\mathbf{P}_{-i}$ contains the probabilities of players other than i.

Estimation using Moment Inequalities (2)

- Let's split the vector of choice probabilities $\mathbf{P}$ into the sub-vectors $\mathbf{P}_i$ and $\mathbf{P}_{-i}$,

$$\mathbf{P} \equiv (\mathbf{P}_i \ , \ \mathbf{P}_{-i})$$

  where $\mathbf{P}_i$ are the probabilities associated to player i and $\mathbf{P}_{-i}$ contains the probabilities of players other than i.

- $\mathbf{P}^0$ is an equilibrium associated to θ^0. Therefore, $\mathbf{P}_i^0$ is firm i's best response to $\mathbf{P}_{-i}^0$. This implies that for any $\mathbf{P}_i \neq \mathbf{P}_i^0$ the following inequality should hold:

$$W_{imt}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} \begin{pmatrix} \theta_i^0 \\ 1 \end{pmatrix} \geq W_{imt}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \begin{pmatrix} \theta_i^0 \\ 1 \end{pmatrix}$$

Estimation using Moment Inequalities (2)

- Let's split the vector of choice probabilities $\mathbf{P}$ into the sub-vectors $\mathbf{P}_i$ and $\mathbf{P}_{-i}$,

$$\mathbf{P} \equiv (\mathbf{P}_i \ , \ \mathbf{P}_{-i})$$

  where $\mathbf{P}_i$ are the probabilities associated to player i and $\mathbf{P}_{-i}$ contains the probabilities of players other than i.

- $\mathbf{P}^0$ is an equilibrium associated to θ^0. Therefore, $\mathbf{P}_i^0$ is firm i's best response to $\mathbf{P}_{-i}^0$. This implies that for any $\mathbf{P}_i \neq \mathbf{P}_i^0$ the following inequality should hold:

$$W_{imt}^{\left(\mathbf{P}_i^0,\mathbf{P}_{-i}^0\right)} \begin{pmatrix} \theta_i^0 \\ 1 \end{pmatrix} \geq W_{imt}^{\left(\mathbf{P}_i,\mathbf{P}_{-i}^0\right)} \begin{pmatrix} \theta_i^0 \\ 1 \end{pmatrix}$$

- We can define an estimator of θ^0 based on these (moment) inequalities.

Estimation using Moment Inequalities (3)

- There are infinite alternative policies $\mathbf{P}_i$, and therefore there are infinite moment inequalities. For estimation, we should select a finite set of alternative policies. This is a very important decision for this class of estimators (more below).

Estimation using Moment Inequalities (3)

- There are infinite alternative policies $\mathbf{P}_i$, and therefore there are infinite moment inequalities. For estimation, we should select a finite set of alternative policies. This is a very important decision for this class of estimators (more below).
- Let H be a **(finite) set of alternative policies** for each player.

Estimation using Moment Inequalities (3)

- There are infinite alternative policies $\mathbf{P}_i$, and therefore there are infinite moment inequalities. For estimation, we should select a finite set of alternative policies. This is a very important decision for this class of estimators (more below).
- Let H be a **(finite) set of alternative policies** for each player.
- Define the following criterion function:

$$R\left(\boldsymbol{\theta}, \mathbf{P}^0, H\right) \equiv \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min \left\{ 0 \; ; \; \left[W_{imt}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \right] \left(\begin{array}{c} \boldsymbol{\theta}_i \\ 1 \end{array} \right) \right\} \right.$$

Estimation using Moment Inequalities (3)

- There are infinite alternative policies $\mathbf{P}_i$, and therefore there are infinite moment inequalities. For estimation, we should select a finite set of alternative policies. This is a very important decision for this class of estimators (more below).

- Let H be a **(finite) set of alternative policies** for each player.

- Define the following criterion function:

$$R\left(\boldsymbol{\theta}, \mathbf{P}^0, H\right) \equiv \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min \left\{ 0 \; ; \; \left[W_{imt}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \right] \left(\begin{array}{c} \boldsymbol{\theta}_i \\ 1 \end{array} \right) \right\} \right.$$

- This criterion function penalizes departures from the inequalities.

Estimation using Moment Inequalities (4)

- Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of θ^0 based on moment inequalities (MI):

$$\hat{\theta} = \arg\min_{\theta} \ R\left(\theta, \hat{\mathbf{P}}^0, H\right)$$

or

$$\hat{\theta} = \arg\min_{\theta} \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min\left\{ 0 \ ; \ \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0, \hat{\mathbf{P}}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}_{-i}^0\right)} \right] \left(\begin{array}{c} \theta_i \\ 1 \end{array} \right) \right\} \right)^2$$

Estimation using Moment Inequalities (4)

- Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of $\boldsymbol{\theta}^0$ based on moment inequalities (MI):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; R\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^0, H\right)$$

or

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i,m,t} \sum_{\mathbf{P}\in H} \left(\min\left\{0 \;;\; \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0, \hat{\mathbf{P}}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}_{-i}^0\right)}\right] \begin{pmatrix} \boldsymbol{\theta}_i \\ 1 \end{pmatrix} \right\}\right)^2$$

- There are several relevant comments to make on this MI estimator:

Estimation using Moment Inequalities (4)

- Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of θ^0 based on moment inequalities (MI):

$$\hat{\theta} = \arg\min_{\theta} \; R\left(\theta, \hat{\mathbf{P}}^0, H\right)$$

or

$$\hat{\theta} = \arg\min_{\theta} \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min \left\{ 0 \; ; \; \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0, \hat{\mathbf{P}}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}_{-i}^0\right)} \right] \begin{pmatrix} \theta_i \\ 1 \end{pmatrix} \right\} \right)^2$$

- There are several relevant comments to make on this MI estimator:

1. Point identification / Set identification

Estimation using Moment Inequalities (4)

- Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of $\boldsymbol{\theta}^0$ based on moment inequalities (MI):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; R\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^0, H\right)$$

or

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min\left\{ 0 \; ; \; \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0, \hat{\mathbf{P}}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}_{-i}^0\right)} \right] \begin{pmatrix} \boldsymbol{\theta}_i \\ 1 \end{pmatrix} \right\} \right)^2$$

- There are several relevant comments to make on this MI estimator:

1. Point identification / Set identification
2. Properties (relative to two-step estimators using moment equalities)

Estimation using Moment Inequalities (4)

- Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of $\boldsymbol{\theta}^0$ based on moment inequalities (MI):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; R\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^0, H\right)$$

or

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min \left\{ 0 \; ; \; \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0, \hat{\mathbf{P}}_{-i}^0\right)} - W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}_{-i}^0\right)} \right] \begin{pmatrix} \boldsymbol{\theta}_i \\ 1 \end{pmatrix} \right\} \right)^2$$

- There are several relevant comments to make on this MI estimator:

1. Point identification / Set identification
2. Properties (relative to two-step estimators using moment equalities)
3. Continuous dependent variables

MI Estimator: Point / Set identification

- This estimator is based on exactly the same assumptions as the 2-step moment equalities (ME) estimator. We have seen that $\boldsymbol{\theta}^0$ **is point identified** by the moment equalities of the ME estimators (e.g., by the pseudo likelihood equations).

MI Estimator: Point / Set identification

- This estimator is based on exactly the same assumptions as the 2-step moment equalities (ME) estimator. We have seen that $\boldsymbol{\theta}^0$ **is point identified** by the moment equalities of the ME estimators (e.g., by the pseudo likelihood equations).

- Therefore, if the set H of alternative policies is large enough, then $\boldsymbol{\theta}^0$ is point identified as the unique maximizer of $R\left(\boldsymbol{\theta}, \mathbf{P}^0, H\right)$.

MI Estimator: Point / Set identification

- This estimator is based on exactly the same assumptions as the 2-step moment equalities (ME) estimator. We have seen that $\boldsymbol{\theta}^0$ **is point identified** by the moment equalities of the ME estimators (e.g., by the pseudo likelihood equations).

- Therefore, if the set H of alternative policies is large enough, then $\boldsymbol{\theta}^0$ is point identified as the unique maximizer of $R\left(\boldsymbol{\theta}, \mathbf{P}^0, H\right)$.

- However, it is very costly to consider a set H with many alternative policies. For the type of sets H which are considered in practice, $R\left(\boldsymbol{\theta}, \mathbf{P}^0, H\right)$ does not have a unique maximizes and therefore $\boldsymbol{\theta}^0$ **is set identified.**

Properties of MI estimator (relative to ME) (1)

- **Why to use an estimator that only set-identifies θ^0 when we have alternative estimators which point identified θ^0?** The Moment Inequalities (MI) estimator should have other advantages. Let's start examining which ARE NOT the advantages.

Properties of MI estimator (relative to ME) (1)

- **Why to use an estimator that only set-identifies θ^0 when we have alternative estimators which point identified θ^0?** The Moment Inequalities (MI) estimator should have other advantages. Let's start examining which ARE NOT the advantages.
- **The MI estimator is not more 'robust' than the ME estimator**. Both estimators are based on exactly the same model and assumptions.

Properties of MI estimator (relative to ME) (1)

- **Why to use an estimator that only set-identifies θ^0 when we have alternative estimators which point identified θ^0?** The Moment Inequalities (MI) estimator should have other advantages. Let's start examining which ARE NOT the advantages.
- **The MI estimator is not more 'robust' than the ME estimator**. Both estimators are based on exactly the same model and assumptions.
- **Asymptotically, the MI estimator is less efficient than the ME estimator.** The efficient 2-step Moment Equalities (ME) estimator has lower asymptotic variance than the MI estimator, even as the set H becomes very large.

Properties of MI estimator (relative to ME) (2)

- **Computationally, the MI estimator is more costly than the ME estimator.** The main computational cost of implementing the MI and ME estimators comes from obtaining the vectors of values $\{W_{imt}^{\mathbf{P}}\}$. The 2-step ME estimators has to compute $\{W_{imt}^{\mathbf{P}}\}$ only once: at the estimated $W_{imt}^{\hat{\mathbf{P}}^0}$. Instead, the MI estimator has to calculate also $W_{imt}^{(\mathbf{P}_i, \hat{\mathbf{P}}^0_{-i})}$ at the different alternative policies in H.

Properties of MI estimator (relative to ME) (2)

- **Computationally, the MI estimator is more costly than the ME estimator.** The main computational cost of implementing the MI and ME estimators comes from obtaining the vectors of values $\{W_{imt}^{\mathbf{P}}\}$. The 2-step ME estimators has to compute $\{W_{imt}^{\mathbf{P}}\}$ only once: at the estimated $W_{imt}^{\hat{\mathbf{P}}^0}$. Instead, the MI estimator has to calculate also $W_{imt}^{\left(\mathbf{P}_i, \hat{\mathbf{P}}^0_{-i}\right)}$ at the different alternative policies in H.
- Furthermore, the MI estimator needs an algorithm for set optimization.

Properties of MI estimator (relative to ME) (3)

- There are two potential advantages of MI against ME estimator:

Properties of MI estimator (relative to ME) (3)

- There are two potential advantages of MI against ME estimator:

1. finite sample bias;

Properties of MI estimator (relative to ME) (3)

- There are two potential advantages of MI against ME estimator:

1. finite sample bias;
2. estimation of models with continuous decision variables.

Properties of MI estimator (relative to ME) (3)

- There are two potential advantages of MI against ME estimator:

1. finite sample bias;
2. estimation of models with continuous decision variables.

- In terms of **finite sample bias**, the MI estimator may have lower bias than ME estimators. Exploiting the information in the alternative policies might be useful to reduce the bias. However, there is very little evidence on this point.

Properties of MI estimator (relative to ME) (3)

- There are two potential advantages of MI against ME estimator:

1. finite sample bias;
2. estimation of models with continuous decision variables.

- In terms of **finite sample bias**, the MI estimator may have lower bias than ME estimators. Exploiting the information in the alternative policies might be useful to reduce the bias. However, there is very little evidence on this point.

- This may dependent very much on the choice of the set H. Typically, H contains only a few alternative policies (e.g., 5, 10, 20). The selection of these alternative policies should be very careful, and there should be some intuition of how the inequalities associated with an alternative policy can help to identify a particular parameter or group of parameters.

Properties of MI estimator (relative to ME) (4)

- BBL show that, when combined with simulation techniques to approximate the values $\{W^{\mathbf{P}}_{imt}\}$, this method can be easily extended to the **estimation of dynamic games with continuous decision variables.**

Properties of MI estimator (relative to ME) (4)

- BBL show that, when combined with simulation techniques to approximate the values $\{W_{imt}^{\mathbf{P}}\}$, this method can be easily extended to the **estimation of dynamic games with continuous decision variables.**
- In fact, the BBL estimator of a model with continuous decision variable is basically the same as with a discrete decision variable.

Properties of MI estimator (relative to ME) (4)

- BBL show that, when combined with simulation techniques to approximate the values $\{W_{imt}^{\mathbf{P}}\}$, this method can be easily extended to the **estimation of dynamic games with continuous decision variables.**
- In fact, the BBL estimator of a model with continuous decision variable is basically the same as with a discrete decision variable.
- The ME estimator of models with continuous decision variable may be more complicated.

Simulation-Based Estimation (1)

- Though two-step methods (with either ME or MI) are computationally much cheaper than full solution-estimation methods, they are still impractical for applications where the dimension of the state space X is very large, e.g., a discrete state space with millions of points or a model in which some of the observable state variables are continuous.

Simulation-Based Estimation (1)

- Though two-step methods (with either ME or MI) are computationally much cheaper than full solution-estimation methods, they are still impractical for applications where the dimension of the state space X is very large, e.g., a discrete state space with millions of points or a model in which some of the observable state variables are continuous.

- To deal with this problem, Hotz, Miller, Sanders and Smith (REStud, 1994) proposed an estimator that uses simulation techniques to approximate the values $\mathbf{Z}_{imt}^{\mathbf{P}}(a_i)$ and $e_{imt}^{\mathbf{P}}(a_i)$, or similarly the vector of values $W_{imt}^{\mathbf{P}}$.

Simulation-Based Estimation (1)

- Though two-step methods (with either ME or MI) are computationally much cheaper than full solution-estimation methods, they are still impractical for applications where the dimension of the state space X is very large, e.g., a discrete state space with millions of points or a model in which some of the observable state variables are continuous.
- To deal with this problem, Hotz, Miller, Sanders and Smith (REStud, 1994) proposed an estimator that uses simulation techniques to approximate the values $\mathbf{Z}_{imt}^{\mathbf{P}}(a_i)$ and $e_{imt}^{\mathbf{P}}(a_i)$, or similarly the vector of values $W_{imt}^{\mathbf{P}}$.
- In the context of dynamic games, Bajari, Benkard and Levin (BBL) have proposed to used this simulation and have extended it to models with continuous decision variables.

Simulation-Based Estimation (2)

- Remember that:

$$\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(a_i) \equiv a_i \ \mathbf{Z}^{\mathbf{P}}_{imt} + E\left(\sum_{s=1}^{\infty} \beta^s \ a_{im,t+s} \ \mathbf{Z}^{\mathbf{P}}_{im,t+s} \mid x_{mt}, \ a_{imt} = a_i \right)$$

Simulation-Based Estimation (2)

- Remember that:

$$\widetilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(a_i) \equiv a_i \; \mathbf{Z}^{\mathbf{P}}_{imt} + E\left(\sum_{s=1}^{\infty} \beta^s \; a_{im,t+s} \; \mathbf{Z}^{\mathbf{P}}_{im,t+s} \mid x_{mt}, \; a_{imt} = a_i \right)$$

- and

$$\tilde{e}^{\mathbf{P}}_{imt}(a_i) \equiv E\left(\sum_{s=1}^{\infty} \beta^s \; a_{im,t+s} \; \varepsilon_{im,t+s} \mid x_{mt}, \; a_{imt} = a_i \right)$$

Simulation-Based Estimation (2)

- Remember that:

$$\tilde{\mathbf{Z}}^{\mathbf{P}}_{imt}(a_i) \equiv a_i \, \mathbf{Z}^{\mathbf{P}}_{imt} + E\left(\sum_{s=1}^{\infty} \beta^s \, a_{im,t+s} \, \mathbf{Z}^{\mathbf{P}}_{im,t+s} \mid x_{mt}, \; a_{imt} = a_i \right)$$

- and

$$\tilde{e}^{\mathbf{P}}_{imt}(a_i) \equiv E\left(\sum_{s=1}^{\infty} \beta^s \, a_{im,t+s} \, \varepsilon_{im,t+s} \mid x_{mt}, \; a_{imt} = a_i \right)$$

- The expectations $E(.)$ are taken over all the possible future paths of actions and state variables conditions on $(x_{mt}, a_{imt} = a_i)$ and conditional on future behavior **P**.

Simulation-Based Estimation (2)

- Remember that:

$$\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i) \equiv \; a_i \; \mathbf{Z}_{imt}^{\mathbf{P}} \; + \; E\left(\sum_{s=1}^{\infty} \beta^s \; a_{im,t+s} \; \mathbf{Z}_{im,t+s}^{\mathbf{P}} \mid x_{mt}, \; a_{imt} = a_i \right)$$

- and

$$\tilde{e}_{imt}^{\mathbf{P}}(a_i) \; \equiv \; E\left(\sum_{s=1}^{\infty} \beta^s \; a_{im,t+s} \; \varepsilon_{im,t+s} \mid x_{mt}, \; a_{imt} = a_i \right)$$

- The expectations $E(.)$ are taken over all the possible future paths of actions and state variables conditions on $(x_{mt}, a_{imt} = a_i)$ and conditional on future behavior **P**.
- The simulators of $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i)$ and $\tilde{e}_{imt}^{\mathbf{P}}(a_i)$ are obtained by replacing the true expectations $E(.)$ by Monte Carlo approximations to these expectations.

Simulation-Based Estimation (3)

- For every value of x_{mt} in the sample and every choice alternative a_i (in the sample or not), we consider (a_i, x_{mt}) as the initial state for player i and then we use the probabilities in $\mathbf{P}$, and the transition probabilities of x, to generate R simulated paths of future actions and state variables from period $t+1$ to $t+T^*$ (i.e., T^* periods ahead).

Simulation-Based Estimation (3)

- For every value of x_{mt} in the sample and every choice alternative a_i (in the sample or not), we consider (a_i, x_{mt}) as the initial state for player i and then we use the probabilities in $\mathbf{P}$, and the transition probabilities of x, to generate R simulated paths of future actions and state variables from period $t+1$ to $t+T^*$ (i.e., T^* periods ahead).
- We index simulated paths by $r \in \{1, 2, ..., R\}$. The $r-th$ simulated path associated with the initial state (a_i, x_{mt}) is

$$\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$$

Simulation-Based Estimation (4)

- A simulated path $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$ is obtained as follows.

Simulation-Based Estimation (4)

- A simulated path $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$ is obtained as follows.
- Given (a_i, x_{mt}), we use the transition probability function $F_x(.|a_i, x_{mt})$ to obtain a random draw $x_{m,t+1}^{(r,a_i)}$.

Simulation-Based Estimation (4)

- A simulated path $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$ is obtained as follows.
- Given (a_i, x_{mt}), we use the transition probability function $F_x(.|a_i, x_{mt})$ to obtain a random draw $x_{m,t+1}^{(r,a_i)}$.
- Given $x_{m,t+1}^{(r,a_i)}$, we use the choice probability $P_i(x_{m,t+1}^{(r,a_i)})$ to obtain a random draw $a_{i,m,t+1}^{(r,a_i)}$.

Simulation-Based Estimation (4)

- A simulated path $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$ is obtained as follows.
- Given (a_i, x_{mt}), we use the transition probability function $F_x(.|a_i, x_{mt})$ to obtain a random draw $x_{m,t+1}^{(r,a_i)}$.
- Given $x_{m,t+1}^{(r,a_i)}$, we use the choice probability $P_i(x_{m,t+1}^{(r,a_i)})$ to obtain a random draw $a_{i,m,t+1}^{(r,a_i)}$.
- Given $(a_{i,m,t+1}^{(r,a_i)}, x_{m,t+1}^{(r,a_i)})$, we use the transition probability function $F_x(.|a_{i,m,t+1}^{(r,a_i)}, x_{m,t+1}^{(r,a_i)})$ to obtain a random draw $x_{m,t+2}^{(r,a_i)}$.

Simulation-Based Estimation (4)

- A simulated path $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$ is obtained as follows.
- Given (a_i, x_{mt}), we use the transition probability function $F_x(.|a_i, x_{mt})$ to obtain a random draw $x_{m,t+1}^{(r,a_i)}$.
- Given $x_{m,t+1}^{(r,a_i)}$, we use the choice probability $P_i(x_{m,t+1}^{(r,a_i)})$ to obtain a random draw $a_{i,m,t+1}^{(r,a_i)}$.
- Given $(a_{i,m,t+1}^{(r,a_i)}, x_{m,t+1}^{(r,a_i)})$, we use the transition probability function $F_x(.|a_{i,m,t+1}^{(r,a_i)}, x_{m,t+1}^{(r,a_i)})$ to obtain a random draw $x_{m,t+2}^{(r,a_i)}$.
- And so on.

Simulation-Based Estimation (5)

- Then, given the simulated paths $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$, we construct the simulator of $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i)$ as:

$$\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(a_i) = a_i \; \mathbf{Z}_{imt}^{\mathbf{P}} + \frac{1}{R} \sum_{r=1}^{R} \left[\sum_{j=1}^{T^*} \beta^j \; a_{im,t+j}^{(r,a_i)} \; \mathbf{Z}_i^{\mathbf{P}} \left(x_{m,t+j}^{(r,a_i)} \right) \right]$$

Simulation-Based Estimation (5)

- Then, given the simulated paths $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$, we construct the simulator of $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i)$ as:

$$\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(a_i) \quad = \quad a_i \; \mathbf{Z}_{imt}^{\mathbf{P}} + \frac{1}{R} \sum_{r=1}^{R} \left[\sum_{j=1}^{T^*} \beta^j \; a_{im,t+j}^{(r,a_i)} \; \mathbf{Z}_i^{\mathbf{P}} \left(x_{m,t+j}^{(r,a_i)} \right) \right]$$

- And similarly for the simulator of $\widetilde{\mathbf{e}}_{imt}^{\mathbf{P}}(a_i)$.

Simulation-Based Estimation (5)

- Then, given the simulated paths $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$, we construct the simulator of $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i)$ as:

$$\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(a_i) = a_i \ \mathbf{Z}_{imt}^{\mathbf{P}} + \frac{1}{R}\sum_{r=1}^{R}\left[\sum_{j=1}^{T^*}\beta^j \ a_{im,t+j}^{(r,a_i)} \ \mathbf{Z}_i^{\mathbf{P}}\left(x_{m,t+j}^{(r,a_i)}\right)\right]$$

- And similarly for the simulator of $\widetilde{\mathbf{e}}_{imt}^{\mathbf{P}}(a_i)$.

- The simulator of $W_{imt}^{\mathbf{P}}$ is $W_{imt}^{\mathbf{P},\mathbf{sim}} =$

$$\left((1 - P_i(x_{mt}))\,\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(0) + P_i(x_{mt})\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(1) \ ; \ (1 - P_i(x_{mt}))\,e_{imt}^{\mathbf{P},\mathbf{sim}}(0)\right.$$

Simulation-Based Estimation (5)

- Then, given the simulated paths $\{a_{im,t+j}^{(r,a_i)}, x_{m,t+j}^{(r,a_i)} : j = 1, 2, ..., T^*\}$, we construct the simulator of $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}}(a_i)$ as:

$$\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(a_i) = a_i \; \mathbf{Z}_{imt}^{\mathbf{P}} + \frac{1}{R} \sum_{r=1}^{R} \left[\sum_{j=1}^{T^*} \beta^j \; a_{im,t+j}^{(r,a_i)} \; \mathbf{Z}_i^{\mathbf{P}} \left(x_{m,t+j}^{(r,a_i)} \right) \right]$$

- And similarly for the simulator of $\widetilde{\mathbf{e}}_{imt}^{\mathbf{P}}(a_i)$.

- The simulator of $W_{imt}^{\mathbf{P}}$ is $W_{imt}^{\mathbf{P},\mathbf{sim}} =$

$$\Big((1 - P_i(x_{mt})) \, \widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(0) + P_i(x_{mt}) \widetilde{\mathbf{Z}}_{imt}^{\mathbf{P},\mathbf{sim}}(1) \; ; \; (1 - P_i(x_{mt})) \, e_{imt}^{\mathbf{P},\mathbf{sim}}(0)$$

- If the DP problem has finite horizon, or if T^* is large enough such that the approximation error associated with the truncation of paths is negligible, then these simulators are unbiased.

Simulation-Based Estimation (6)

- This simulators can be used either for ME or for MI estimation.

Simulation-Based Estimation (6)

- This simulators can be used either for ME or for MI estimation.
- The **simulation-based ME estimator** is the value of $\boldsymbol{\theta}$ that solves the system of equations:

$$E\left(H(x_{mt})\left\{\ a_{imt}-\Phi\left(\widetilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}}\ \boldsymbol{\theta}_i+\tilde{e}_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}}\right)\quad \text{for any } i,t\ \right\}\right)\ \neq\ 0$$

Simulation-Based Estimation (6)

- This simulators can be used either for ME or for MI estimation.
- The **simulation-based ME estimator** is the value of $\boldsymbol{\theta}$ that solves the system of equations:

$$E\left(H(x_{mt})\left\{\ a_{imt}-\Phi\left(\widetilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}}\ \boldsymbol{\theta}_i+\tilde{e}_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}}\right)\quad \text{for any } i,t\ \right\}\right)\ \neq\ 0$$

- The **simulation-based MI estimator** is the value (or set of values) of $\boldsymbol{\theta}$ that minimizes the criterion function:

$$\sum_{i,m,t}\sum_{\mathbf{P}\in H}\left(\min\left\{0\ ;\ \left[W_{imt}^{\left(\hat{\mathbf{P}}_i^0,\hat{\mathbf{P}}_{-i}^0\right),\mathbf{sim}}-W_{imt}^{\left(\mathbf{P}_i,\hat{\mathbf{P}}_{-i}^0\right),\mathbf{sim}}\right]\left(\begin{array}{c}\boldsymbol{\theta}_i\\ 1\end{array}\right)\right\}\right)^2$$

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.
- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.

- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.
- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$
2. **Simulation error:** because the expectation is not taken over all the possible future histories but only over the simulated paths.

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.
- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$
2. **Simulation error:** because the expectation is not taken over all the possible future histories but only over the simulated paths.
3. **Approximation error:** because $T^* \neq \infty$

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.

- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$
2. **Simulation error:** because the expectation is not taken over all the possible future histories but only over the simulated paths.
3. **Approximation error:** because $T^* \neq \infty$

- (3) can be negligible if T^* is not too small or β too close to one.

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.
- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$
2. **Simulation error:** because the expectation is not taken over all the possible future histories but only over the simulated paths.
3. **Approximation error:** because $T^* \neq \infty$

- (3) can be negligible if T^* is not too small or β too close to one.
- (1) can be very important: curse of dimensionality in NP estimation.

Simulation-Based Estimation (7)

- These estimators are consistent only as the number of simulated paths, R, goes to infinity.
- Note that there are three sources of error in $(W_{imt}^{\hat{\mathbf{P}}^0,\mathbf{sim}} - W_{imt}^{\mathbf{P}^0})$:

1. **Estimation error:** because $\hat{\mathbf{P}}^0 \neq \mathbf{P}^0$
2. **Simulation error:** because the expectation is not taken over all the possible future histories but only over the simulated paths.
3. **Approximation error:** because $T^* \neq \infty$

- (3) can be negligible if T^* is not too small or β too close to one.
- (1) can be very important: curse of dimensionality in NP estimation.
- (2) can be very important. Even with millions of simulated histories we may have a very small proportion of all possible histories.

Estimation of models with unobserved market heterogeneity

- **Unobserved market heterogeneity**. Ignoring persistent market heterogeneity, if present, can generate important biases in the estimation of structural parameters of dynamic oligopoly games.

Estimation of models with unobserved market heterogeneity

- **Unobserved market heterogeneity**. Ignoring persistent market heterogeneity, if present, can generate important biases in the estimation of structural parameters of dynamic oligopoly games.
- Here I describe three approaches which have been proposed to estimate dynamic games where **unobserved market heterogeneity is model as finite mixture model**.

Estimation of models with unobserved market heterogeneity

- **Unobserved market heterogeneity**. Ignoring persistent market heterogeneity, if present, can generate important biases in the estimation of structural parameters of dynamic oligopoly games.
- Here I describe three approaches which have been proposed to estimate dynamic games where **unobserved market heterogeneity is model as finite mixture model**.

1. Two-step method: Kasahara and Shimotsu (2007)

Estimation of models with unobserved market heterogeneity

- **Unobserved market heterogeneity**. Ignoring persistent market heterogeneity, if present, can generate important biases in the estimation of structural parameters of dynamic oligopoly games.
- Here I describe three approaches which have been proposed to estimate dynamic games where **unobserved market heterogeneity is model as finite mixture model**.

1. Two-step method: Kasahara and Shimotsu (2007)
2. Nested Pseudo likelihood (NPL): Aguirregabiria and Mira (2007)

Estimation of models with unobserved market heterogeneity

- **Unobserved market heterogeneity**. Ignoring persistent market heterogeneity, if present, can generate important biases in the estimation of structural parameters of dynamic oligopoly games.
- Here I describe three approaches which have been proposed to estimate dynamic games where **unobserved market heterogeneity is model as finite mixture model**.

1. Two-step method: Kasahara and Shimotsu (2007)
2. Nested Pseudo likelihood (NPL): Aguirregabiria and Mira (2007)
3. NPL + Sequential EM algorithm: Arcidiacono and Miller (2008)

Finite mixture model (1)

- Consider the dynamic game of market entry/exit in local markets (with MD and BK as global players). Remember that the expected profit of firm i is:

$$\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_{i}$$

Finite mixture model (1)

- Consider the dynamic game of market entry/exit in local markets (with MD and BK as global players). Remember that the expected profit of firm i is:

$$\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_i$$

- where

$$\mathbf{Z}_{imt}^{P} \equiv \{\ (1 - P_j(x_{mt}))\, S_{mt}\ ,\ P_j(x_{mt}) S_{mt}\ ,\ -1\ ,\ -(1 - a_{im,t-1})$$

$$\boldsymbol{\theta}_i \equiv \left\{\ \theta_i^M\ ,\ \theta_i^D\ ,\ FC_i\ ,\ EC_i\ \right\}'$$

Finite mixture model (1)

- Consider the dynamic game of market entry/exit in local markets (with MD and BK as global players). Remember that the expected profit of firm i is:

$$\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_i$$

- where

$$\mathbf{Z}_{imt}^{P} \equiv \{\ (1 - P_j(x_{mt}))\, S_{mt}\ ,\ P_j(x_{mt}) S_{mt}\ ,\ -1\ ,\ -(1 - a_{im,t-1})$$

$$\boldsymbol{\theta}_i \equiv \left\{\ \theta_i^M\ ,\ \theta_i^D\ ,\ FC_i\ ,\ EC_i\ \right\}'$$

- Now, suppose that we relax the assumption that $\boldsymbol{\theta}_i$ is invariant across markets. That is, we consider that expected profits are $\mathbf{Z}_{imt}^{P}\ \boldsymbol{\theta}_{im}$, where:

$$\boldsymbol{\theta}_{im} = \boldsymbol{\theta}_i + \Sigma_{\mathbf{i}}\ \boldsymbol{\omega}_m = \begin{pmatrix} \theta_i^M \\ \theta_i^D \\ FC_i \end{pmatrix} + \Sigma_{\mathbf{i}} \begin{pmatrix} \omega_m^M \\ \omega_m^D \\ \omega_m^{FC} \end{pmatrix}$$

Finite mixture model (2)

- In a finite mixture model, we assume that ω_m is a vector of random variables

Finite mixture model (2)

- In a finite mixture model, we assume that ω_m is a vector of random variables
- with discrete and finite support $\Omega = \{\omega^1, \omega^2, \ldots, \omega^L\}$;

Finite mixture model (2)

- In a finite mixture model, we assume that $\boldsymbol{\omega}_m$ is a vector of random variables
- with discrete and finite support $\Omega = \{\boldsymbol{\omega}^1, \boldsymbol{\omega}^2, \ldots, \boldsymbol{\omega}^L\}$;
- i.i.d. across markets with probability mass function $\pi_\ell \equiv \Pr(\boldsymbol{\omega}_m = \boldsymbol{\omega}^\ell)$.

Finite mixture model (3)

- The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only '*One MPE in the data*' to allow for different market types to have different equilibria.

Finite mixture model (3)

- The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only '*One MPE in the data*' to allow for different market types to have different equilibria.
- Let $\mathbf{P}^0_{mt} \equiv \{\Pr(a_{imt} = 1 | x_{mt} = x, m, t) : i = 1, 2; x \in X\}$ be the distributions of a_{imt} conditional on x_{mt} in market m at period t.

Finite mixture model (3)

- The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only '*One MPE in the data*' to allow for different market types to have different equilibria.
- Let $\mathbf{P}^0_{mt} \equiv \{\Pr(a_{imt} = 1 | x_{mt} = x, m, t) : i = 1, 2; x \in X\}$ be the distributions of a_{imt} conditional on x_{mt} in market m at period t.
- We assume that $\mathbf{P}^0_{mt} = \mathbf{P}^0_{\ell}$, where ℓ is the type of market m.

Finite mixture model (3)

- The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only '*One MPE in the data*' to allow for different market types to have different equilibria.
- Let $\mathbf{P}^0_{mt} \equiv \{\Pr(a_{imt} = 1 | x_{mt} = x, m, t) : i = 1, 2; x \in X\}$ be the distributions of a_{imt} conditional on x_{mt} in market m at period t.
- We assume that $\mathbf{P}^0_{mt} = \mathbf{P}^0_{\ell}$, where ℓ is the type of market m.
- Each market type has its own MPE. Though we still assume that only one equilibrium is played in the data *conditional on market type*, the data generating process may correspond to multiple equilibria. Markets which, in term of exogenous characteristics, are observationally equivalent to the econometrician may have different probabilities of entry and exit because the random effect component of profits ω is different.

Nested Pseudo Likelihood (NPL) (1)

- The vector of structural parameters $\boldsymbol{\theta}_i$ now includes Σ_i.

Nested Pseudo Likelihood (NPL) (1)

- The vector of structural parameters $\boldsymbol{\theta}_i$ now includes Σ_i.
- It is convenient to represent firm i's expected profit (under type ℓ) as:

$$\mathbf{Z}_{imt}^{P_\ell}\ \boldsymbol{\theta}_i$$

Nested Pseudo Likelihood (NPL) (1)

- The vector of structural parameters $\boldsymbol{\theta}_i$ now includes Σ_i.
- It is convenient to represent firm i's expected profit (under type ℓ) as:

$$\mathbf{Z}_{imt}^{P_\ell}\ \boldsymbol{\theta}_i$$

- where now:

$$\mathbf{Z}_{imt}^{P_\ell} = \begin{pmatrix} (1 - P_j(x_{mt}))\, S_{mt} \\ P_j(x_{mt}) S_{mt} \\ -1 \\ -(1 - a_{im,t-1}) \\ (1 - P_j(x_{mt}))\, S_{mt}\ \omega_m^M \\ P_j(x_{mt}) S_{mt}\ \omega_m^D \\ -\omega_m^{FC} \\ -(1 - a_{im,t-1})\ \omega_m^{EC} \end{pmatrix} \quad ; \quad \boldsymbol{\theta}_i = \begin{pmatrix} \theta_i^M \\ \theta_i^D \\ FC_i \\ EC_i \\ \sigma_\omega^M \\ \sigma_\omega^D \\ \sigma_\omega^{FC} \\ \sigma_\omega^{EC} \end{pmatrix}$$

Nested Pseudo Likelihood (NPL) (1)

- The vector of structural parameters $\boldsymbol{\theta}_i$ now includes Σ_i.
- It is convenient to represent firm i's expected profit (under type ℓ) as:

$$\mathbf{Z}_{imt}^{P_\ell}\ \boldsymbol{\theta}_i$$

- where now:

$$\mathbf{Z}_{imt}^{P_\ell} = \begin{pmatrix} (1-P_j(x_{mt}))\, S_{mt} \\ P_j(x_{mt}) S_{mt} \\ -1 \\ -(1-a_{im,t-1}) \\ (1-P_j(x_{mt}))\, S_{mt}\ \omega_m^M \\ P_j(x_{mt}) S_{mt}\ \omega_m^D \\ -\omega_m^{FC} \\ -(1-a_{im,t-1})\ \omega_m^{EC} \end{pmatrix} \quad ; \quad \boldsymbol{\theta}_i = \begin{pmatrix} \theta_i^M \\ \theta_i^D \\ FC_i \\ EC_i \\ \sigma_\omega^M \\ \sigma_\omega^D \\ \sigma_\omega^{FC} \\ \sigma_\omega^{EC} \end{pmatrix}$$

- The vector of present values $\widetilde{\mathbf{Z}}_{imt}^{\mathbf{P}_\ell}$ has a similar definition as before, but now it corresponds to the vector $\mathbf{Z}^{P_\ell}$ above.

Nested Pseudo Likelihood (NPL) (2)

- The (conditional) pseudo likelihood function has the following finite mixture form:

$$Q\left(\boldsymbol{\theta}, \{\mathbf{P}_\ell\}\right) = \sum_{m=1}^{M} \ln \Pr\left(\text{History of market } m \mid x_{m1}\right)$$

$$= \sum_{m=1}^{M} \ln\left(\sum_{\ell=1}^{L} \Pr\left(\boldsymbol{\omega}_m = \boldsymbol{\omega}^\ell \mid x_{m1}\right) \; \Pr\left(\text{History of market } m \mid \boldsymbol{\omega}_m = \boldsymbol{\omega}\right.\right.$$

$$= \sum_{m=1}^{M} \ln\left(\sum_{\ell=1}^{L} \pi_{\ell|x_{m1}} \left[\prod_{t=1}^{T}\prod_{i=1}^{N} \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_\ell}\boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}}\right)^{a_{imt}} \Phi\left(-\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_\ell}\boldsymbol{\theta}_i - \tilde{e}_{imt}^{\mathbf{P}}\right)\right.\right.$$

Nested Pseudo Likelihood (NPL) (2)

- The (conditional) pseudo likelihood function has the following finite mixture form:

$$Q\left(\boldsymbol{\theta}, \{\mathbf{P}_{\ell}\}\right) = \sum_{m=1}^{M} \ln \Pr\left(\text{History of market } m \mid x_{m1}\right)$$

$$= \sum_{m=1}^{M} \ln \left(\sum_{\ell=1}^{L} \Pr\left(\boldsymbol{\omega}_m = \boldsymbol{\omega}^{\ell} \mid x_{m1}\right) \; \Pr\left(\text{History of market } m \mid \boldsymbol{\omega}_m = \boldsymbol{\omega}\right.\right.$$

$$= \sum_{m=1}^{M} \ln \left(\sum_{\ell=1}^{L} \pi_{\ell|x_{m1}} \left[\prod_{t=1}^{T} \prod_{i=1}^{N} \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_{\ell}} \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}}\right)^{a_{imt}} \Phi\left(-\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_{\ell}} \boldsymbol{\theta}_i - \tilde{e}_{imt}^{\mathbf{P}}\right)\right.\right.$$

- where $\pi_{\ell|x_{m1}}$ is the conditional probability $\Pr(\omega^{\ell}|x_{m1})$.

Nested Pseudo Likelihood (NPL) (2)

- The (conditional) pseudo likelihood function has the following finite mixture form:

$$Q(\boldsymbol{\theta}, \{\mathbf{P}_\ell\}) = \sum_{m=1}^{M} \ln \Pr(\text{History of market } m \mid x_{m1})$$

$$= \sum_{m=1}^{M} \ln \left(\sum_{\ell=1}^{L} \Pr(\boldsymbol{\omega}_m = \boldsymbol{\omega}^\ell \mid x_{m1}) \ \Pr(\text{History of market } m \mid \boldsymbol{\omega}_m = \boldsymbol{\omega}$$

$$= \sum_{m=1}^{M} \ln \left(\sum_{\ell=1}^{L} \pi_{\ell|x_{m1}} \left[\prod_{t=1}^{T} \prod_{i=1}^{N} \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_\ell} \boldsymbol{\theta}_i + \tilde{e}_{imt}^{\mathbf{P}}\right)^{a_{imt}} \Phi\left(-\tilde{\mathbf{Z}}_{imt}^{\mathbf{P}_\ell} \boldsymbol{\theta}_i - \tilde{e}_{imt}^{\mathbf{P}}\right)\right.\right.$$

- where $\pi_{\ell|x_{m1}}$ is the conditional probability $\Pr(\omega^\ell | x_{m1})$.
- $\pi_{\ell|x_{m1}}$ is not equal to the unconditional probability π_ℓ. Incumbent statuses at period 1, which are components of the vector x_{m1}, are not independent of market type, i.e., more profitable markets according to ω_m tend to have more incumbent firms.

Nested Pseudo Likelihood (NPL) (3)

- Under the assumption that x_{m1} is drawn from the stationary distribution induced by the MPE, we have that $\pi_{\ell|x_{m1}}$ depend only on choice probabilities in $\mathbf{P}_{\ell}$ and the (known) unconditional probabilities $\{\pi_{\ell}\}$.

Nested Pseudo Likelihood (NPL) (3)

- Under the assumption that x_{m1} is drawn from the stationary distribution induced by the MPE, we have that $\pi_{\ell|x_{m1}}$ depend only on choice probabilities in $\mathbf{P}_\ell$ and the (known) unconditional probabilities $\{\pi_\ell\}$.
- Let $\mathbf{p}^*(\mathbf{P}_\ell) \equiv \{p^*(x|\mathbf{P}_\ell) : x \in X\}$ be the stationary distribution of x induced by the equilibrium $\mathbf{P}_\ell$ and the transition $f_x(.|.,.)$.

Nested Pseudo Likelihood (NPL) (3)

- Under the assumption that x_{m1} is drawn from the stationary distribution induced by the MPE, we have that $\pi_{\ell|x_{m1}}$ depend only on choice probabilities in $\mathbf{P}_\ell$ and the (known) unconditional probabilities $\{\pi_\ell\}$.
- Let $\mathbf{p}^*(\mathbf{P}_\ell) \equiv \{p^*(x|\mathbf{P}_\ell) : x \in X\}$ be the stationary distribution of x induced by the equilibrium $\mathbf{P}_\ell$ and the transition $f_x(.|.,.)$.
- This stationary distribution can be very simply obtained as the solution to the system of linear equations:

$$p^*(x|\mathbf{P}_\ell) = \sum_{x' \in X} p^*(x'|\mathbf{P}_\ell) \left(\sum_{a \in A^N} \left[\prod_{j=1}^{N} P_{\ell j}(a_j|x) \right] f_x(x'|a,x) \right)$$

Nested Pseudo Likelihood (NPL) (4)

- Then, by Bayes' rule, we have that:

$$\pi_{\ell|x_{m1}} = \frac{\pi_{\ell} \; p^{*}(x_{m1}|\mathbf{P}_{\ell})}{\sum_{\ell'=1}^{L} \pi_{\ell'} \; p^{*}(x_{m1}|\mathbf{P}_{\ell'})}$$

Nested Pseudo Likelihood (NPL) (4)

- Then, by Bayes' rule, we have that:

$$\pi_{\ell|x_{m1}} = \frac{\pi_\ell \; p^*(x_{m1}|\mathbf{P}_\ell)}{\sum_{\ell'=1}^{L} \pi_{\ell'} \; p^*(x_{m1}|\mathbf{P}_{\ell'})}$$

- Therefore, given $\{\mathbf{P}_\ell\}$ (and the known unconditional probabilities π_ℓ), the conditional probabilities $\pi_{\ell|x_{m1}}$ are known.

Nested Pseudo Likelihood (NPL) (4)

- Then, by Bayes' rule, we have that:

$$\pi_{\ell|x_{m1}} = \frac{\pi_{\ell}\ p^{*}(x_{m1}|\mathbf{P}_{\ell})}{\sum_{\ell'=1}^{L} \pi_{\ell'}\ p^{*}(x_{m1}|\mathbf{P}_{\ell'})}$$

- Therefore, given $\{\mathbf{P}_{\ell}\}$ (and the known unconditional probabilities π_{ℓ}), the conditional probabilities $\pi_{\ell|x_{m1}}$ are known.
- Furthermore, for any $\{\mathbf{P}_{\ell}\}$, $Q(.)$ is globally concave in $\boldsymbol{\theta}$. This is a very convenient feature of this model and of the NPL method.

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$(1) \qquad \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$$

$$(2) \qquad \hat{P}_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell}\, \hat{\boldsymbol{\theta}}_i + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell}\right) \quad \text{for any } \ell \text{ and } (i, m, t)$$

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$(1) \qquad \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$$

$$(2) \qquad \hat{P}_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell}\; \hat{\boldsymbol{\theta}}_i + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell}\right) \quad \text{for any } \ell \text{ and } (i, m, t)$$

- A simple procedure to obtain an NPL fixed point is the following. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, ..., L\}$.

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$
\begin{array}{lrcl}
(1) & \hat{\boldsymbol{\theta}} & = & \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right) \\
(2) & \hat{P}_{\ell,i}(x_{mt}) & = & \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell} \; \hat{\boldsymbol{\theta}}_i + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell}\right) \quad \text{for any } \ell \text{ and } (i, m, t)
\end{array}
$$

- A simple procedure to obtain an NPL fixed point is the following. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, ..., L\}$.
  1. Step 1: Obtain the probabilities $\{\pi_{\ell|x_{m1}}\}$.

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$(1) \qquad \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$$

$$(2) \quad \hat{P}_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell} \; \hat{\boldsymbol{\theta}}_i + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell}\right) \quad \text{for any } \ell \text{ and } (i, m, t)$$

- A simple procedure to obtain an NPL fixed point is the following. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, ..., L\}$.
  1. Step 1: Obtain the probabilities $\{\pi_{\ell|x_{m1}}\}$.
  2. Step 2: Obtain $\hat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell^0\}\right)$

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$(1) \qquad \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$$

$$(2) \quad \hat{P}_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell} \; \hat{\boldsymbol{\theta}}_i + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell}\right) \quad \text{for any } \ell \text{ and } (i, m, t)$$

- A simple procedure to obtain an NPL fixed point is the following. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, ..., L\}$.
  1. Step 1: Obtain the probabilities $\{\pi_{\ell|x_{m1}}\}$.
  2. Step 2: Obtain $\hat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell^0\}\right)$
  3. Step 3: Update the vector of players' choice probabilities using the best response probability mapping. That is,
  $\hat{P}_{\ell,i}^1(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}_{imt}^{\hat{\mathbf{P}}_\ell^0} \; \hat{\boldsymbol{\theta}}_i^1 + \tilde{e}_{imt}^{\hat{\mathbf{P}}_\ell^0}\right)$

Nested Pseudo Likelihood (NPL) (4)

- An NPL fixed point is defined as a pair $\left(\hat{\boldsymbol{\theta}}, \{\hat{\mathbf{P}}_\ell\}\right)$ that satisfies two conditions:

$$(1) \qquad \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$$

$$(2) \quad \hat{P}_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}_\ell}_{imt} \; \hat{\boldsymbol{\theta}}_i + \tilde{e}^{\hat{\mathbf{P}}_\ell}_{imt}\right) \quad \text{for any } \ell \text{ and } (i, m, t)$$

- A simple procedure to obtain an NPL fixed point is the following. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}^0_\ell : \ell = 1, 2, ..., L\}$.
  1. Step 1: Obtain the probabilities $\{\pi_{\ell|x_{m1}}\}$.
  2. Step 2: Obtain $\hat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}, \{\hat{\mathbf{P}}^0_\ell\}\right)$
  3. Step 3: Update the vector of players' choice probabilities using the best response probability mapping. That is,
     $\hat{P}^1_{\ell,i}(x_{mt}) = \Phi\left(\tilde{\mathbf{Z}}^{\hat{\mathbf{P}}^0_\ell}_{imt} \; \hat{\boldsymbol{\theta}}^1_i + \tilde{e}^{\hat{\mathbf{P}}^0_\ell}_{imt}\right)$
  4. If, for every type ℓ, $||\hat{\mathbf{P}}^1_\ell - \hat{\mathbf{P}}^0_\ell||$ is smaller than a predetermined small

Nested Pseudo Likelihood (NPL) (5)

- If there is a unique NPL fixed point, then $\left(\hat{\theta}, \{\hat{\mathbf{P}}_{\ell}\}\right)$ is a consistent estimator of $\left(\hat{\theta}^{0}, \{\hat{\mathbf{P}}_{\ell}^{0}\}\right)$.

Nested Pseudo Likelihood (NPL) (5)

- If there is a unique NPL fixed point, then $\left(\hat{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$ is a consistent estimator of $\left(\hat{\theta}^0, \{\hat{\mathbf{P}}_\ell^0\}\right)$.
- Otherwise, if there are multiple NPL fixed points, then the consistent NPL estimator is the NPL fixed point that provides the maximum value of the likelihood $Q\left(\theta, \{\hat{\mathbf{P}}_\ell\}\right)$.

Nested Pseudo Likelihood (NPL) (5)

- If there is a unique NPL fixed point, then $\left(\hat{\theta}, \{\hat{\mathbf{P}}_\ell\}\right)$ is a consistent estimator of $\left(\hat{\theta}^0, \{\hat{\mathbf{P}}_\ell^0\}\right)$.
- Otherwise, if there are multiple NPL fixed points, then the consistent NPL estimator is the NPL fixed point that provides the maximum value of the likelihood $Q\left(\theta, \{\hat{\mathbf{P}}_\ell\}\right)$.
- Therefore, it is important to check for multiple NPL fixed points by applying the recursive procedure to different initial vector of probabilities $\{\hat{\mathbf{P}}_\ell^0\}$.

Counterfactual Experiments with Estimated Model (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).

Counterfactual Experiments with Estimated Model (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.

Counterfactual Experiments with Estimated Model (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 with Estimated Model (2)

- Different approaches have been implemented in practice.

Counterfactual Experiments with Estimated Model (2)

- Different approaches have been implemented in practice.

1. Select the equilibrium to which we converge by iterating in the (counterfactual) equilibrium mapping starting with the factual equilibrium $\mathbf{P}^0$

Counterfactual Experiments with Estimated Model (2)

- Different approaches have been implemented in practice.

1. Select the equilibrium to which we converge by iterating in the (counterfactual) equilibrium mapping starting with the factual equilibrium $\mathbf{P}^0$
2. Select the equilibrium with maximum total profits (or alternatively, with maximum welfare).

Counterfactual Experiments with Estimated Model (2)

- Different approaches have been implemented in practice.

1. Select the equilibrium to which we converge by iterating in the (counterfactual) equilibrium mapping starting with the factual equilibrium $\mathbf{P}^0$
2. Select the equilibrium with maximum total profits (or alternatively, with maximum welfare).
3. Taylor approximation: Aguirregabiria and Ho (2007)

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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:

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

Counterfactual Experiments: Aguirregabiria-Ho (2007)

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

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

- The model could be completed with an equilibrium selection mechanism: i.e., a criterion that selects one and only one equilibrium for each possible θ.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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: Aguirregabiria-Ho (2007)

- We only assume that there is such a mechanism, and that it is a smooth function of $\boldsymbol{\theta}$.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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 $\boldsymbol{\theta}$.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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: Aguirregabiria-Ho (2007)

- Let θ_0 and $\mathbf{P}_0$ be the the population values. Let $(\hat{\theta}_0, \hat{\mathbf{P}}_0)$ be our consistent estimator.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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: Aguirregabiria-Ho (2007)

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

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

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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 $\dfrac{\partial \pi\left(\hat{\theta}_0\right)}{\partial \theta'}$.

Counterfactual Experiments: Aguirregabiria-Ho (2007)

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

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

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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:

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

Counterfactual Experiments: Aguirregabiria-Ho (2007)

- 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:

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

- 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 $\mathbf{P}^*$.