# Identification of Structural Parameters in Dynamic Discrete Choice Games with Fixed Effects Unobserved Heterogeneity

Victor Aguirregabiria, Jiaying Gu, & Pedro Mira

**IAAE CONFERENCE 2019 - NICOSIA**

Jue 26, 2019

# Empirical Dynamic Discrete Games

- Dynamic Discrete Choice games are useful tools for empirical analysis of **strategic and social interactions** that involve dynamic (investment) decisions.
- Important applications:
  - Firms' entry/exit and investment.
  - Dynamic social interactions and peer effects.
  - Electoral competition (Sieg and Yoon, 2017)
  - Labor supply within the family (Eckstein and Lifshitz, 2015)

# Structural Parameters and Unobserved Heterogeneity

- Two types of structural parameters are particularly important:
  - Switching or Adjustment cost parameters (dynamics)
  - Strategic or social interaction parameters (game)

- Identification of these parameters depends crucially on the specification of **Unobserved Heterogeneity (UH)**.

- Misspecifying **Persistent UH** can imply substantial biases in the estimation of parameters capturing dynamics. **Spurious Dynamics**.

- Misspecifying **UH common to players** can imply substantial biases in the estimation of parameters capturing strategic or social interactions between players. **Spurious positive spillover**.

# Fixed Effects - Sufficient Statistics Approach

- We study the identification of these structural parameters in dynamic games where $N$ players are observed playing the game at $M$ markets and $T$ periods of time, where **T is small** and **M is large**.

- We consider a **Fixed Effects model** for the time-invariant Player-Market UH.
  - The joint distribution of the UH and the initial values of endogenous variables is unrestricted.

- We consider a **Sufficient Statistics - Conditional Likelihood approach** (Chamberlain, 1985, Honore & Kyriadzidou, 2000).

- Is there a vector of statistics such that conditional on this vector the probability of the observed history of players' choices does not depend on UH but still depends on structural parameters?

# Two main challenges for Dynamic Games

**[1] Continuation values.** If players are **forward-looking**, their decisions depend on continuation values and these are nonlinear functions of UH and endogenous state variables.

- For single-agent dynamic discrete choice models, Aguirregabiria, Gu, and Luo (2018) show that we can get identification.

**[2] Multiple equilibria.** The model implies only bounds on probabilities of choice histories.

- The standard Sufficient Statistic - Conditional ML approach needs to be extended to deal with this case.

# Outline

1. Model

2. Identification
   2.1. Model with unique predictions
   2.2. Model with multiple equilibria (bounds)

3. Conclusions and Extensions

# 1. Model

# Model: Basic Features

- For today's talk, I will focus on a dynamic game with:
  - Two players: $i \in \{1, 2\}$
  - Binary choice: $y_{imt} \in \{0, 1\}$
  - Logit i.i.d. transitory shocks $\varepsilon_{imt}$
  - The only endogenous state variables are $(y_{1m,t-1}, y_{2m,t-1})$
  - Complete information

- At every period $t$, players choose simultaneously their actions ($y_{1mt}$ and $y_{2mt}$) to maximize: $\mathbb{E}_t \left[ \sum_{s=0}^{\infty} \delta_{im}^{s} \; U_{im,t+s} \right]$.

- Let $\Delta U_{imt} \equiv U_{imt}(1, y_{jmt}) - U_{imt}(0, y_{jmt})$.

$$\Delta U_{imt} = \alpha_{im} + \gamma_i \; y_{jmt} + \beta_{ii} \; y_{imt-1} + \beta_{ij} \; y_{jmt-1} - \varepsilon_{imt}$$

# Model: Best response equations

- The equilibrium of the dynamic game can be described by:

$$y_{1mt} = 1\left\{ \begin{array}{l} \varepsilon_{1mt} \leq \alpha_{1m} + \gamma_1 \ y_{2mt} + \beta_{11} \ y_{1mt-1} + \beta_{12} \ y_{2mt-1} \\ \qquad\quad + v_1(y_{2mt}, \boldsymbol{\alpha}_m) \end{array} \right\}$$

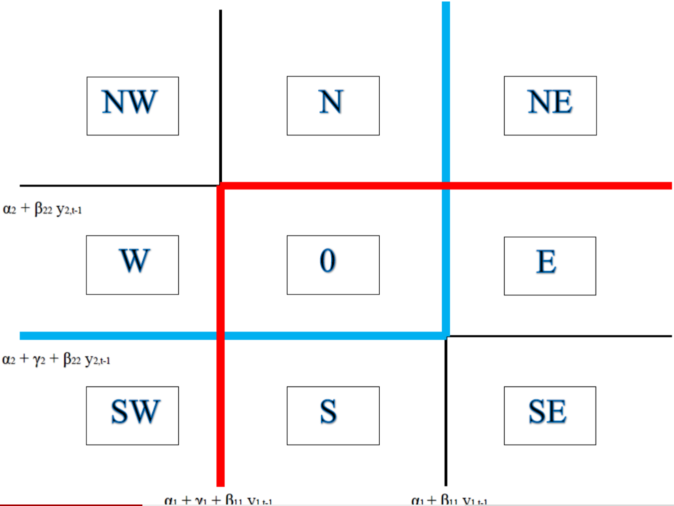
$$y_{2mt} = 1\left\{ \begin{array}{l} \varepsilon_{2mt} \leq \alpha_{2m} + \gamma_2 \ y_{1mt} + \beta_{21} \ y_{1mt-1} + \beta_{22} \ y_{2mt-1} \\ \qquad\quad + v_2(y_{1mt}, \boldsymbol{\alpha}_m) \end{array} \right\}$$

- $v_1(y_{2mt}, \boldsymbol{\alpha}_m)$ and $v_2(y_{1mt}, \boldsymbol{\alpha}_m)$ are the **"differential continuation values":** difference between future values of choosing 1 versus 0.

- $p(\alpha_{1m}, \alpha_{2m}, y_{1m0}, y_{2m0})$ is unrestricted, i.e., FE model.

# Multiple Equilibria

- There is a region in the space of $(\varepsilon_{1mt}, \varepsilon_{2mt})$ with non-unique predictions about the outcome $(y_{1mt}, y_{2mt})$:



# Bounds on Choice Probabilities

- With $\gamma_1 \leq 0$ and $\gamma_2 \leq 0$, the model implies lower and upper bounds for the outcomes $(0,1)$ and $(1,0)$.

$$L(0,1 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m) \leq \mathbb{P}(0,1 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m) \leq U(0,1 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m)$$

$$L(1,0 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m) \leq \mathbb{P}(1,0 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m) \leq U(1,0 \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m)$$

# Stackelberg Dynamic Game

- If $\beta_{12} = \gamma_1 = 0$, then **player 1 is the leader** and its current or future decisions are not affected by player's 2.

$$y_{1mt} \quad = \quad 1\left\{ \; \varepsilon_{1mt} \leq \quad \alpha_{1m} + \beta_{11} \; y_{1mt-1} + v_1(\boldsymbol{\alpha}_m) \; \right\}$$

- **Player 2 is the follower** such that:

$$y_{2mt} \quad = \quad 1\left\{ \begin{array}{ll} \varepsilon_{2mt} \leq & \alpha_{2m} + \gamma_2 \; y_{1mt} + \beta_{21} \; y_{1mt-1} + \beta_{22} \; y_{2mt-1} \\ & + v_2(y_{1mt}, \boldsymbol{\alpha}_m) \end{array} \right\}$$

- This model has **unique predictions** for the choice probabilities $\mathbb{P}(\mathbf{y}_{mt} \mid \mathbf{y}_{mt-1}; \boldsymbol{\alpha}_m)$ of any outcome $\mathbf{y}_{mt}$.

# 2. Identification

# Data

- The researcher observes panel data for the players playing the game at $M$ markets and $T$ time periods:

$$\text{Data} = \{ \; y_{1mt}, \; y_{2mt} : m = 1, 2, ..., M \; ; \; t = 0, 1, ..., T \}$$

$M$ is large and $T$ is small.

- Given these data and the restrictions from the model, the researcher is interested in the estimation of the structural parameters $\beta_{11}$, $\beta_{22}$, $\beta_{12}$, $\beta_{21}$, $\gamma_1$, $\gamma_2$.

- We denote these structural parameters using the vector $\boldsymbol{\beta}$.

# Identification of Stackelberg model

- Let $\widetilde{\mathbf{y}}_m \equiv (y_{1mt}, y_{2mt} : t = 0, 1, .., T)$ be a *market history*:

$$\mathbb{P}\left(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}\right) = p_{\boldsymbol{\alpha}_m}\left(y_{1m0}, y_{2m0}\right)$$

$$\prod_{t=1}^{T} \frac{\exp\left\{ y_{1mt}\left[\alpha_{1m} + \beta_{11}\ y_{1mt-1} + v_1(\boldsymbol{\alpha}_m)\right] \right\}}{1 + \exp\left\{ y_{1mt}\left[\alpha_{1m} + \beta_{11}\ y_{1mt-1} + v_1(\boldsymbol{\alpha}_m)\right] \right\}}$$

$$\frac{\exp\left\{ y_{2mt}\left[\alpha_{2m} + \gamma_2\ y_{1mt} + \beta_{21}\ y_{1mt-1} + \beta_{22}\ y_{2mt-1} + v_2(y_{1mt}, \boldsymbol{\alpha}_m)\right]\right.}{1 + \exp\left[\alpha_{2m} + \gamma_2\ y_{1mt} + \beta_{21}\ y_{1mt-1} + \beta_{22}\ y_{2mt-1} + v_2(y_{1mt}, \boldsymbol{\alpha}_m)\right]}$$

where $p_{\boldsymbol{\alpha}_m}\left(y_{1m0}, y_{2m0}\right)$ = probability of initial condition given $\boldsymbol{\alpha}_m$.

# Identification of Stackelberg model [2]

- The log-probability of a market history has the following form:

$$\ln \mathbb{P}\left(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}\right) = \mathbf{s}(\widetilde{\mathbf{y}}_m)'\ g(\boldsymbol{\alpha}_m) + \mathbf{c}(\widetilde{\mathbf{y}}_m)'\ \boldsymbol{\beta}$$

  where $\mathbf{s}(\widetilde{\mathbf{y}}_m)$ and $\mathbf{c}(\widetilde{\mathbf{y}}_m)$ are vectors of statistics.

- This structure implies that $\mathbf{s}(\widetilde{\mathbf{y}}_m)$ is a sufficient statistics for $\boldsymbol{\alpha}_m$.

- If the elements in the vector $[\mathbf{s}(\widetilde{\mathbf{y}}_m)', \mathbf{c}(\widetilde{\mathbf{y}}_m)']$ are linearly independent, then the maximization of the Conditional log-likelihood function implies the identification of the vector of parameters $\boldsymbol{\beta}$.

# Identification of Stackelberg Dynamic Game [3]

- The switching costs parameters $(\beta_{11}, \beta_{22})$ and the lagged strategic effect $\beta_{21}$ are point identified.

- The vectors of sufficient statistics is:

$$\mathbf{s}(\widetilde{\mathbf{y}}) = \left[ y_{10},\ y_{20},\ y_{1T},\ y_{2T},\ T_{(1)},\ T_{(2)},\ T_{(1,2)} \right]$$

where $T_{(1)} \equiv \sum_{t=1}^{T} y_{1t}$; $T_{(2)} \equiv \sum_{t=1}^{T} y_{2t}$; $T_{(1,2)} \equiv \sum_{t=1}^{T} y_{1t} y_{2t}$

- And:

$$\mathbf{c}(\widetilde{\mathbf{y}})'\ \beta = \beta_{11}\ C_{(1,1)} + \beta_{22}\ C_{(2,2)} + \beta_{21}\ C_{(2,1)}$$

where $C_{(i,j)} \equiv \sum_{t=1}^{T} y_{it}\ y_{jt-1}$

# Examples of identifying pair of histories

**Examples of histories and identified parameters with T=3**

$$A = \{\mathbf{y}_0,\ \mathbf{a},\ \mathbf{b},\ \mathbf{y}_3\}; \quad B = \{\mathbf{y}_0,\ \mathbf{b},\ \mathbf{a},\ \mathbf{y}_3\}$$

| | $\mathbf{y}_0$ | $\mathbf{a}$ | $\mathbf{b}$ | $\mathbf{y}_3$ | $\ln \mathbb{P}(A) - \ln \mathbb{P}(B)$ |
|---|---|---|---|---|---|
| Case 1: | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ | $\beta_{11}$ |
| Case 2: | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ | $\beta_{22}$ |
| Case 3: | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ | $\beta_{21}$ |

# Conditional Likelihood

- Consider the Conditional Likelihood (CL) function:

$$\ell^c(\boldsymbol{\beta}) = \sum_{m=1}^{M} \ln \mathbb{P}\left(\widetilde{\mathbf{y}}_m \mid \mathbf{s}(\widetilde{\mathbf{y}}_m),\ \boldsymbol{\beta}\right)$$

- Given the structure described above, the CL function has the form:

$$\ell^c(\boldsymbol{\beta}) = \sum_{m=1}^{M} \mathbf{c}(\widetilde{\mathbf{y}}_m)'\ \boldsymbol{\beta} - \ln \left[ \sum_{\widetilde{\mathbf{y}}:\ \mathbf{s}(\widetilde{\mathbf{y}})=\mathbf{s}(\widetilde{\mathbf{y}}_m)} \exp\left\{\mathbf{c}(\widetilde{\mathbf{y}})'\ \boldsymbol{\beta}\right\} \right]$$

- This log-likelihood function is globally concave in the vector $\beta$ such that it is straightforward to compute the CMLE of $\beta$ using standard gradient methods (Newton, or BHHH).

# Identification of Model with Multiple Equilibria

- Using the bounds for players' choice probabilities we can construct bounds for the probability of a market history $\mathbb{P}\left(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}\right)$.

$$L(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}) \leq \ln \mathbb{P}\left(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}\right) \leq U(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta})$$

- We can obtain (non sharp) bounds that have a logit structure and with the following form:

$$\begin{aligned} L(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}) &= \mathbf{s}(\widetilde{\mathbf{y}}_m)' \, g(\boldsymbol{\alpha}_m) + \mathbf{c}_L(\widetilde{\mathbf{y}}_m)' \, \boldsymbol{\beta} \\ U(\widetilde{\mathbf{y}}_m \mid \boldsymbol{\alpha}_m, \boldsymbol{\beta}) &= \mathbf{s}(\widetilde{\mathbf{y}}_m)' \, g(\boldsymbol{\alpha}_m) + \mathbf{c}_U(\widetilde{\mathbf{y}}_m)' \, \boldsymbol{\beta} \end{aligned}$$

- Very importantly, these lower and upper bounds depend on the incidental parameters exactly in the same way, $\mathbf{s}(\widetilde{\mathbf{y}}_m)' \, g(\boldsymbol{\alpha}_m)$.

# Identification of Model with Multiple Equilibria [2]

- Let $A$ and $B$ be two market histories such that:

$$\begin{cases} \mathbf{s}(A) = \mathbf{s}(B) \\ \mathbf{c}_L(A) - \mathbf{c}_U(B) \neq 0 \quad \text{OR} \quad \mathbf{c}_U(A) - \mathbf{c}_L(B) \neq 0 \end{cases}$$

- Then, we have that:

$$[\mathbf{c}_L(A) - \mathbf{c}_U(B)]' \boldsymbol{\beta} \leq \ln\left(\frac{\mathbb{P}(A)}{\mathbb{P}(B)}\right) \leq [\mathbf{c}_U(A) - \mathbf{c}_L(B)]' \boldsymbol{\beta}$$

- These inequalities provide set identification of the structural parameters.

- Interestingly, there are pairs of histories such that:

$$\mathbf{c}_L(A) - \mathbf{c}_U(B) = \mathbf{c}_U(A) - \mathbf{c}_L(B) \neq 0$$

such that there is point identification of some structural parameters, or linear combination of structural parameters.

# Dynamic Game with Multiple Equilibria

- The parameters $\beta_{11}$, $\beta_{22}$, $\gamma_1$ and $\gamma_2$ are set identified.

- The vectors of sufficient statistics is:

$$\mathbf{s}(\widetilde{\mathbf{y}}) = \left[ y_{10},\ y_{20},\ y_{1T},\ y_{2T},\ T_{(1)},\ T_{(2)},\ C_{(1,2)},\ C_{(2,1)} \right]$$

with either $C_{(1,2)} = C_{(2,2)}$ or $C_{(2,1)} = C_{(1,1)}$

- And:

$$\begin{aligned}
\mathbf{c}_{L1}(\widetilde{\mathbf{y}})'\ \beta &= \beta_{11}\ C_{(1,1)} + \beta_{22}\ C_{(2,2)} + \gamma_1\ T_{(1)} + \gamma_2\ T_{(1,2)} \\
\mathbf{c}_{L2}(\widetilde{\mathbf{y}})'\ \beta &= \beta_{11}\ C_{(1,1)} + \beta_{22}\ C_{(2,2)} + \gamma_1\ T_{(1,2)} + \gamma_2\ T_{(2)} \\
\mathbf{c}_{U}(\widetilde{\mathbf{y}})'\ \beta &= \beta_{11}\ C_{(1,1)} + \beta_{22}\ C_{(2,2)} + \gamma_1\ T_{(1,2)} + \gamma_2\ T_{(1,2)}
\end{aligned}$$

# Example of identifying pair of histories

**Examples of histories and identified parameters with T=3**

| | Lower | Upper |
|---|---|---|
| $A = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ | $\beta_{11} + \gamma_1 + \gamma_2$ | $\beta_{11} + \gamma_1$ |

# Summary & Conclusions

- We study the identification of dynamic discrete choice games with a FE specification of UH and using short panels.
- Relative to reduced form panel data models, the identification should control for the continuation values and deal with multiple equilibria.
- We extend the Sufficient Statistics approach to deal with bounds on the probabilities of market histories.
- We find positive (and some negative) identification results.
- Things to do: Inference; Empirical application; Relax the logit assumption.