INCAE PhD SUMMER ACADEMY
DYNAMIC GAMES IN EMPIRICAL IO

Lecture 2:
Single-agent dynamic discrete choice: Estimation

Victor Aguirregabiria (University of Toronto)

June 21, 2022

LECTURE 2: Single-agent dynamic discrete choice: Estimation

1. Data

2. Maximum Likelihood Estimation (MLE):
   Nexted Fixed Point (NFXP) Algorithm

3. Two-Step Hotz-Miller Methods and Finite Dependence

1. Data

DATA

- The researcher has panel data of N individuals (e.g., firms) over T periods of time.

- For each individual i and time t, the researcher observes action y_{it} and vector of state variables $\mathbf{x}_{it}$

$$Data \;=\; \{\; y_{it}\;,\; \mathbf{x}_{it} : i = 1, 2, ..., N\;;\; t = 1, 2, ..., T\; \}$$

- In micro-econometric applications of single-agent models, we typically have that N is large (e.g., hundreds or thousands) and T is small, i.e., short panel.

EXAMPLE: MARKET ENTRY-EXIT

- We have an industry, e.g., supermarket industry.
- Firms operate in local markets, e.g., cities, neighborhoods. That is, consumer demand, output prices, and input prices are determined at the local market level.
- In this context, an "individual" i is a combination of a "firm + local market": e.g., Walmart's entry-exit decision in North-East Toronto.
- The dataset consists of $\{y_{it}, \mathbf{x}_{it} : i = 1, 2, ..., N; t = 1, 2, ..., T\}$:
  - y_{it} = indicator that firm-market i is active at period t.
  - $\mathbf{x}_{it}$ = vector of local market characteristics affecting the profit of an active firm, e.g., consumer population, average income, input prices.

2. Maximum Likelihood Estimation: NFXP ALgorithm

ESTIMATION: GENERAL IDEAS

- Given a panel dataset $\{y_{it}, \mathbf{x}_{it}\}$, we are interested in estimating the unknown parameters in the primitives $\{\pi(.), f_x(.), \delta\}$.

- Let $\boldsymbol{\theta}$ be the vector of structural parameters. We distinguish three components in this vector:

$$\boldsymbol{\theta} \;=\; \{\; \boldsymbol{\theta}_{\pi},\; \boldsymbol{\theta}_{f},\; ,\delta \;\}$$

  where:
  $\boldsymbol{\theta}_{\pi} =$ parameters in utility function π
  $\boldsymbol{\theta}_{f} =$ parameters in transition probability of observable state var.

ESTIMATION OF θ_f

- The parameters $\boldsymbol{\theta}_f$ can be estimated separately from parameters $\boldsymbol{\theta}_\pi$.
- More specifically, the estimation of parameters $\boldsymbol{\theta}_f$ is quite standard as it does not require solving the DP problem.
- **Example.** Market size follows an AR(1) process:

$$s_{it} = \theta_{f,0} + \theta_{f,1} \; s_{i,t-1} + e_{it}.$$

- We can estimate $\theta_{f,0}$ and $\theta_{f,1}$ by OLS in this AR(1) regression eq.
- More generally, given the parametric transition probability function $f_x(\mathbf{x}_{i,t+1}|y_{it}, \mathbf{x}_{it}; \boldsymbol{\theta}_f)$, we can estimate $\boldsymbol{\theta}_f$ by Maximum Likelihood:

$$\widehat{\boldsymbol{\theta}}_f \;=\; argmax_{\boldsymbol{\theta}_f} \; \ell_f\left(\boldsymbol{\theta}_f\right) = \sum_{i=1}^{N} \sum_{t=1}^{T-1} \log f_x\left(\mathbf{x}_{i,t+1}|y_{it}, \mathbf{x}_{it}; \boldsymbol{\theta}_f\right)$$

MAXIMUM LIKELIHOOD ESTIMATION OF θ_π

- We can estimate $\boldsymbol{\theta}_\pi$ by maximizing the likelihood of $(y_{i1}, ..., y_{iT})$ conditional on $(\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})$

- This log-likelihood function is:

$$
\begin{aligned}
\ell(\boldsymbol{\theta}_\pi) &= \sum_{i=1}^{N} \log \Pr\left(y_{i1}, \ldots, y_{iT} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}; \boldsymbol{\theta}_\pi\right) \\
&= \sum_{i=1}^{N} \log \Pr\left(\alpha(\mathbf{x}_{i1}, \varepsilon_{i1}; \boldsymbol{\theta}_\pi) = y_{i1}, \ldots, \alpha(\mathbf{x}_{iT}, \varepsilon_{iT}; \boldsymbol{\theta}_\pi) = y_{iT}\right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} \log P(y_{it} | \mathbf{x}_{it}; \boldsymbol{\theta}_\pi)
\end{aligned}
$$

where $P(y_{it}|\mathbf{x}_{it}; \boldsymbol{\theta}_\pi)$ is the CCP function.

- To evaluate $\ell(\boldsymbol{\theta}_\pi)$ we need to solve the DP problem for this value $\boldsymbol{\theta}_\pi$.

NESTED FIXED POINT ALGORITHM (NFXP)

- The NFXP algorithm is a gradient iterative search method to obtain the MLE of the structural parameters.
- This algorithm nests a BHHH method (outer algorithm), that searches for a root of the likelihood equations, with a value function iteration method (inner algorithm) that solves the DP problem for each trial value of the structural parameters.
- The algorithm is initialized with an arbitrary vector of structural parameters, say $\boldsymbol{\theta}_{\pi}^{0}$. At BHHH iteration $n \geq 0$:

$$\widehat{\boldsymbol{\theta}}_{\pi}^{n+1} = \widehat{\boldsymbol{\theta}}_{\pi}^{n} + \left(\sum_{i,t} \nabla \log P_{it} \left(\widehat{\boldsymbol{\theta}}_{\pi}^{n} \right) \nabla \log P_{it} \left(\widehat{\boldsymbol{\theta}}_{\pi}^{n} \right)' \right)^{-1} \left(\sum_{i,t} \nabla \log P_{it} \left(\widehat{\boldsymbol{\theta}}_{\pi}^{n} \right)$$

and $\nabla \log P_{it} \left(\boldsymbol{\theta}_{\pi} \right)$ is the gradient of log-CCP with respect to $\boldsymbol{\theta}_{\pi}$.

Nested fixed point (NFXP) algorithm (2/4)

- To illustrate this algorithm in more detail, consider a logit model where the utility function is linear in parameters:

$$\pi(y_{it}, \mathbf{x}_{it}, \boldsymbol{\theta}_\pi) \;=\; \mathbf{z}(y_{it}, \mathbf{x}_{it})' \boldsymbol{\theta}_\pi$$

where $\mathbf{z}(y_{it}, \mathbf{x}_{it})$ is a vector of known functions.

- For this model, the gradient of log-CCP is:

$$\nabla \log P_{it}\left(\boldsymbol{\theta}_\pi\right) \;=\; \mathbf{z}(y_{it}, \mathbf{x}_{it}) - \sum_{j=0}^{J} P(j|\mathbf{x}_{it}, \boldsymbol{\theta}_\pi)\; \mathbf{z}(j, \mathbf{x}_{it})$$

- The CCP function is:

$$P(y|\mathbf{x}, \boldsymbol{\theta}_\pi) = \frac{\exp\left\{\mathbf{z}(y, \mathbf{x})\boldsymbol{\theta}_\pi + \delta\; \mathbf{F}_x(y, \mathbf{x})' \mathbf{V}^\sigma(\boldsymbol{\theta}_\pi)\right\}}{\sum_{j=0}^{J} exp\left\{\mathbf{z}(j, \mathbf{x})\boldsymbol{\theta}_\pi + \delta\; \mathbf{F}_x(j, \mathbf{x})' \mathbf{V}^\sigma(\boldsymbol{\theta}_\pi)\right\}}$$

Nested fixed point (NFXP) algorithm (3/4)

- The vector of values $\mathbf{V}^{\sigma}(\boldsymbol{\theta}_{\pi})$ can be obtained as the unique fixed point of the following Integrated Bellman equation in vector form:

$$\mathbf{V}^{\sigma} = \log\left(\sum_{j=0}^{J} \exp\left\{\mathbf{z}(j)\boldsymbol{\theta}_{\pi} + \delta\ \mathbf{F}_{x}(j)\mathbf{V}^{\sigma}\right\}\right)$$

with $\mathbf{z}(j)$ and $\mathbf{F}_{x}(j)$ are the matrices $\{z(j,x) : x \in X\}$ and $\{\mathbf{F}_{x}(j,x) : x \in X\}$, respectively.

Nested fixed point (NFXP) algorithm (4/4)

- The NFXP algorithm works as follows.

(I) [Inner Algorithm] *Given $\widehat{\boldsymbol{\theta}}_{\pi}^{n}$, we obtain the vector $\mathbf{V}^{\sigma}(\widehat{\boldsymbol{\theta}}_{\pi}^{n})$ by successive iterations in the Integrated Bellman equation.*

(II) *Given $\widehat{\boldsymbol{\theta}}_{\pi}^{n}$ and $\mathbf{V}^{\sigma}(\widehat{\boldsymbol{\theta}}_{\pi}^{n})$, we construct the CCPs $P(y_{it}|\mathbf{x}_{it}, \widehat{\boldsymbol{\theta}}_{\pi}^{n})$ and the gradients of these CCPs using the expression above.*

(III) [Outer iteration] *We apply a BHHH iteration to obtain a new $\widehat{\boldsymbol{\theta}}_{\pi}^{n+1}$.*

** We proceed in this way until the distance between $\widehat{\boldsymbol{\theta}}_{\pi}^{n+1}$ and $\widehat{\boldsymbol{\theta}}_{\pi}^{n}$ is smaller than a pre-specified convergence constant.*

3. Two-Step Hotz-Miller Methods and Finite Dependence

MAIN IDEAS

- The cost of solving some DP problems (the "curse of dimensionality"), limits the range of applications where the NFXP can be applied.

- Hotz and Miller (REStud, 1993) observed that, under standard assumptions in this model, it is not necessary to solve the DP problem, even once, to estimate the structural parameters.

- Hotz-Miller approach is based on **two main ideas**:

1. **CCPs can be estimated nonparametrically** in a first-step, and these estimates can be used to contruct agent's present discounted values without solving the DP problem.

2. A large class of models have a **Finite Dependence** property. This property implies moment conditions that involve CCPs and utilities at only a small number of time periods (sometimes as small as 2).

MAIN IDEAS (2/2)

- Here I present the Finite Dependence version of Hotz-Miller approach, based on Arcidiacono & Miller (2011).
- I present this approach in the context of a dynamic multinomial logit model.

MODEL

- **Integrated Bellman equation**:

$$V^{\sigma}(\mathbf{x}_t) \;=\; \log\left[\sum_{y=0}^{J} \exp\left\{v\left(y, \mathbf{x}_t\right)\right\}\right]$$

- where $v\left(y, \mathbf{x}_t\right)$ are the **Conditional Choice value functions**:

$$v\left(y, \mathbf{x}_t\right) \;\equiv\; \pi\left(y, \mathbf{x}_t\right) + \delta \sum_{\mathbf{x}_{t+1}} V^{\sigma}(\mathbf{x}_{t+1})\; f_x(\mathbf{x}_{t+1} | y, \mathbf{x}_t)$$

- The **Conditional Choice Probabilities (CCPs)** are:

$$P(y \mid \mathbf{x}_t) \;=\; \frac{\exp\left\{v\left(y, \mathbf{x}_t\right)\right\}}{\sum_{j=0}^{J} \exp\left\{v\left(j, \mathbf{x}_t\right)\right\}}$$

FINITE DEPENDENCE PROPERTY

- **Main idea**: Under some conditions, optimal behavior implies that there is a known function that relates CCPs and utility function at periods t and $t+1$ [more generally, at t, $t+1$, ..., $t+s$ where s is finite].

$$\mathbb{E}_t \left[G \left(\pi(y_t, \mathbf{x}_t; \boldsymbol{\theta}_\pi), \ P(y_t|\mathbf{x}_t), \ \pi(y_{t+1}, \mathbf{x}_{t+1}; \boldsymbol{\theta}_\pi), \ P(y_{t+1}|\mathbf{x}_{t+1}) \right) \right] = 0$$

  where $G(.)$ is known.
- This equation has the same flavor as an Euler equation.
- Suppose that we can estimate the CCPs $P(y_t|\mathbf{x}_t)$ directly from the data, as reduced-form probabilities, without solving the model.
- Then, we can estimate the structural parameters in ; $\boldsymbol{\theta}_\pi$ by GMM without having to solve the model even once, and without having to compute any present value.

FINITE DEPENDENCE REPRESENTATION (1/4)

- Given the structure of the Logit CCPs, we have that:

$$\log P(0 \mid \mathbf{x}_t) = -\log \left[1 + \sum_{j=1}^{J} \exp \left\{ v(j, \mathbf{x}_t) - v(0, \mathbf{x}_t) \right\} \right]$$

- This implies the following expression for the integrated value function:

$$\begin{aligned} V^{\sigma}(\mathbf{x}_t) &= \log \left[\sum_{j=0}^{J} \exp \left\{ v\left(j, \mathbf{x}_t\right) \right\} \right] \\ &= v\left(0, \mathbf{x}_t\right) + \log \left[1 + \sum_{j=1}^{J} \exp \left\{ v\left(j, \mathbf{x}_t\right) - v\left(0, \mathbf{x}_t\right) \right\} \right] \\ &= v\left(0, \mathbf{x}_t\right) - \ln P\left(0, \mathbf{x}_t\right) \end{aligned}$$

FINITE DEPENDENCE REPRESENTATION (2/4)

- Second, for any two choice alternatives, say j and k, we have that:

$$\log P(j \mid \mathbf{x}_t) - \log P(k \mid \mathbf{x}_t) \;=\; v(j, \mathbf{x}_t) - v(0, \mathbf{x}_t)$$

- Remember that: $v(j, \mathbf{x}) = \pi(y, \mathbf{x}) + \delta \sum_{\mathbf{x}_{t+1}} V^{\sigma}(\mathbf{x}_{t+1}) f_x(\mathbf{x}_{t+1}|y, \mathbf{x})$. Therefore:

$$\log P(j|\mathbf{x}_t) - \log P(k|\mathbf{x}_t) = \pi(j, \mathbf{x}_t) - \pi(k, \mathbf{x}_t) +$$

$$+\delta \sum_{\mathbf{x}_{t+1}} \left[v(0, \mathbf{x}_{t+1}) - \log P(0|\mathbf{x}_{t+1}) \right] \left[f_x(\mathbf{x}_{t+1}|j, \mathbf{x}_t) - f_x(\mathbf{x}_{t+1}|k, \mathbf{x}_t) \right]$$

FINITE DEPENDENCE REPRESENTATION (3/4)

- Since $v(0, \mathbf{x}_{t+1}) = \pi(0, \mathbf{x}_{t+1}) + \delta \sum_{\mathbf{x}_{t+2}} V^{\sigma}(\mathbf{x}_{t+2}) f_x(\mathbf{x}_{t+2}|0, \mathbf{x}_{t+1})$, we have that:

$$\log P(j|\mathbf{x}_t) - \log P(k|\mathbf{x}_t) = \pi(j, \mathbf{x}_t) - \pi(k, \mathbf{x}_t) +$$

$$+\delta \sum_{\mathbf{x}_{t+1}} \left[\pi(0, \mathbf{x}_{t+1}) - \log P(0|\mathbf{x}_{t+1})\right] \left[f_x(\mathbf{x}_{t+1}|j, \mathbf{x}_t) - f_x(\mathbf{x}_{t+1}|k, \mathbf{x}_t)\right]$$

$$+\delta^2 \sum_{\mathbf{x}_{t+1}} \sum_{\mathbf{x}_{t+2}} V^{\sigma}(\mathbf{x}_{t+2}) \; f_x(\mathbf{x}_{t+2}|0, \mathbf{x}_{t+1}) \; \left[f_x(\mathbf{x}_{t+1}|j, \mathbf{x}_t) - f_x(\mathbf{x}_{t+1}|k, \mathbf{x}_t)\right]$$

- There is a general class of dynamic models **[one-period finite dependence]** where this term is zero.
  e.g., occupational choice; market entry-exit; machine replacement; dynamic demand of differentiated products; etc.

FINITE DEPENDENCE REPRESENTATION (4/4)

- Under one-period finite dependence, the following condition holds at every-period t, any $\mathbf{x}_t$, and any pair (j, k):

$$\log P\left(j|\mathbf{x}_t\right) - \log P\left(k|\mathbf{x}_t\right) = \pi\left(j, \mathbf{x}_t; \boldsymbol{\theta}_\pi\right) - \pi\left(k, \mathbf{x}_t; \boldsymbol{\theta}_\pi\right)$$

$$-\delta \sum_{\mathbf{x}_{t+1}} \log P\left(0|\mathbf{x}_{t+1}\right) \quad \left[f_x(\mathbf{x}_{t+1}|j, \mathbf{x}_t) - f_x(\mathbf{x}_{t+1}|k, \mathbf{x}_t)\right]$$

$$+\delta \sum_{\mathbf{x}_{t+1}} \pi\left(0, \mathbf{x}_{t+1}; \boldsymbol{\theta}_\pi\right) \quad \left[f_x(\mathbf{x}_{t+1}|j, \mathbf{x}_t) - f\left(\mathbf{x}_{t+1}|k, \mathbf{x}_t\right)\right]$$

- This equation provides moment conditions that can be used to estimate consistently the vector of parameters $\boldsymbol{\theta}_\pi$.

- Given a Nonparametric estimator of the reduced-form CCPs $P\left(j|\mathbf{x}_t\right)$, we can estimate structural parameters using a simple two-step GMM estimator.

EXAMPLE: MARKET ENTRY & EXIT

- $\mathbf{x}_t = (y_{t-1}, s_t)$; $\pi(0, \mathbf{x}_t) = 0$; $\pi(1, \mathbf{x}_t) = \theta_1 + \theta_2 s_t + \theta_3 (1 - y_{t-1})$.

- This model has the one-period dependence property.

- We have:

$$\log P\left(1|\mathbf{x}_t\right) - \log P\left(0|\mathbf{x}_t\right) = \theta_1 + \theta_2 \; s_t + \theta_3 (1 - y_{t-1})$$

$$-\delta \sum_{s_{t+1}} \left[\log P\left(0|1, s_{t+1}\right) - \log P\left(0|0, s_{t+1}\right) \right] \; f_s(s_{t+1}|s_t)$$