Unobserved Heterogeneity in Structural Dynamic Discrete Choice Models

Victor Aguirregabiria and Jiaying Gu

University of Toronto

ECO 2403. TOPICS IN ECONOMETRICS

Febraury 9, 2018

Introduction

- Key econometric issue Dynamic Panel Data (DPD) models is distinguishing between **"true dynamics"** and **"spurious dynamics"** due to persistent **unobserved heterogeneity (UH).**
- These lectures deal with this problem in the context of **Dynamic Discrete Choice Structural models**.
- In these models, agents are forward-looking and maximize expected and discounted intertemporal utilities.
- UH enters not only in current utilities but also enters [in a complicated and endogenous way] in continuation values, i.e., in the expected value of future utilities.
- This affects properties and implementation of some estimators.

Introduction [2]

- Most common methods to deal with UH in DPD models are **Fixed Effects (FE)** and **Correlated Random Effects (CRE)**.

- CRE models impose different types of restrictions: parametric, finite support, restrictions on the initial conditions problem.

- FE approach is very attractive because it does not impose any restriction on the distribution of the UH conditional on observable explanatory variables.

FE in structural DDC Models

- **[1] "Brute force" dummy variables method**: inconsistent; bias reduction methods can be computationally intensive.

- **[2] "Sufficient statistics/CMLE method:** Not all DPD models can be estimated root-N consistently using FE estimators. Examples:

  - Discrete choice models other than the logit.

  - Models where UH and predetermined var. are not additively separable.

- Structural dynamic logit model: Common wisdom: FE cannot provide a consistent estimator of structural parameters.

- Even if UH enters additively in one-period utility function, the solution of the model implies that UH appears non-additively in the continuation values.

Outline of the four lectures

- Backwards induction ;) ...
- **[Fourth lecture]** In a recent research project, J. Gu. Y. Luo, and myself show that it is possible to obtain sufficient statistics for UH in a class of models that includes many applications in this literature.
- **[Third & Second lectures]** Literature on sufficient statistics / CMLE method in other related models.
- **[Today's lecture]** Current methods [CRE] to deal with UH in structural DDC models.

Outline

[1] Structural DDC models

[2] Finite mixture – Full solution – MLE

[3] Hotz-Miller: Finite Dependence representation

[4] Hotz-Miller + Nonparametric finite mixtures (Kasahara & Shimotsu)

[5] Hotz-Miller + EM algorithm (Arcidiacono & Miller)

1. Structural DDC models

Model

- Decision variable: $y_{it} \in \mathcal{Y} = \{0, 1, ..., J\}$. Every period t, agent i chooses y_{it} to maximize $\mathbb{E}\left(\sum_{s=0}^{T} \beta^{j}\ U_{it}\right)$.
- The one-period utility of choosing y is:

$$U_{it}(y) = u\left(y, \mathbf{x}_{it}, \omega_i\right) + \varepsilon_{it}(y)$$

- $\{\varepsilon_{it}(0), ..., \varepsilon_{it}(J)\}$ unobservables, i.i.d. over (i, t)
- ω_i unobservable: finite mixture: $\omega_i \in \Omega = \{\omega^1, \omega^2, ..., \omega^L\}$.
- $\mathbf{x}_{it}$ = Observable state variables, with transition probabilities: $f\left(\mathbf{x}_{it+1} | y_{it}, \mathbf{x}_{it}, \omega_i\right)$

Model [2]

- Integrated (over $\varepsilon's$) Bellman equation. For every type ω:

$$V_\omega(\mathbf{x}_t) = \int \arg\max_{y\in\mathcal{Y}} \left[v_\omega\left(y,\mathbf{x}_t\right) + \varepsilon_t(y)\right] \; dG(\varepsilon_t)$$

where $v_\omega\left(y,\mathbf{x}_t\right) \equiv u_\omega\left(y,\mathbf{x}_t\right) + \beta \sum_{\mathbf{x}'} V_\omega(\mathbf{x}') \; f_\omega(\mathbf{x}'|y,\mathbf{x}_t)$.

- For instance, for the MNL model ($\varepsilon's$ type 1 EV):

$$V_\omega(\mathbf{x}_t) = \ln\left[\sum_{y\in\mathcal{Y}} \exp\left\{v_\omega\left(y,\mathbf{x}_t\right)\right\}\right]$$

- The Conditional Choice Probabilities (CCPs) are:

$$P_\omega(y \mid \mathbf{x}_t) = \Pr\left(y = \arg\max_{j\in\mathcal{Y}} \left[v_\omega\left(j,\mathbf{x}_t\right) + \varepsilon_t(j)\right]\right)$$

- For instance, for the MNL model:

$$P_\omega(y \mid \mathbf{x}_t) = \frac{\exp\left\{v_\omega\left(y,\mathbf{x}_t\right)\right\}}{\sum_{j\in\mathcal{Y}} \exp\left\{v_\omega\left(j,\mathbf{x}_t\right)\right\}}$$

Example: Occupational choice model

- J occupations; $y = 0$ represents "not working".
- Utility depends on earnings, disutility of working, and switching costs.
- Two sources of dynamics: (a) experience in an occupation/job has returns; and (b) switching occupation has switching costs.
- Endogenous state variables in $\mathbf{x}_t$: (a) endogenous: y_{t-1} and duration (experience) in current occupation.
- Exogenous state variables: shocks in wages (occupation specific); health status;
- Unobserved ω: Skills, that can be occupation-specific; taste for leisure; unobserved health; ...

Example: Machine replacement model

- A firm decides whether to replace ($y = 1$) or not ($y = 0$) a machine.
- Profit = Variable Profit - Replacement Cost (if $y = 1$) - Maintenance cost (if $y = 0$).
- Dynamics: Machine depreciates with age.
- Endogenous state var: Machine age: : $x_{t+1} = (1 - y_t)\ (x_t + 1)$
- Exogenous state variables: shocks in profits; price of a new machine.
- Unobserved ω: in maintenance and replacement costs.

Example: Market entry-exit

- A firm decides whether to be active ($y = 1$) or not ($y = 0$) in a market.
- Profit = Variable Profit - Entry cost (if new entrant) - Scrap value (if exiting)
- Two sources of dynamics: (a) experience in the market has returns; and (b) entry costs.
- Endogenous state variables in $\mathbf{x}_t$: (a) endogenous: y_{t-1} and duration (experience) in the market.
- Exogenous state variables: shocks in profits (output and input prices).
- Unobserved ω: Firm or market heterogeneity in costs.

2. Full solution–MLE

Full solution–MLE

- Let θ be the vector of parameters of the model. Given the panel dataset $\{y_{it}, \mathbf{x}_{it} : i = 1, ..., N;\ t = 1, ..., T\}$, the log-likelihood of the model is:

$$
\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{N} \ln \Pr(y_{i1}, \mathbf{x}_{i1}, ..., y_{iT}, \mathbf{x}_{iT} \mid \theta) \\
&= \sum_{i=1}^{N} \ln \left(\sum_{\omega \in \Omega} \pi_\omega \ \Pr(y_{i1}, \mathbf{x}_{i1}, ..., y_{iT}, \mathbf{x}_{iT} \mid \omega, \theta) \right) \\
&= \sum_{i=1}^{N} \ln \left(\sum_{\omega \in \Omega} \pi_\omega \ p\left(\mathbf{x}_{i1} | \omega\right) \prod_{t=1}^{T} P_\omega\left(y_{it} | \mathbf{x}_{it}, \theta\right) \prod_{t=1}^{T-1} f_\omega\left(\mathbf{x}_{it+1} | y_{it},
\end{aligned}
$$

- For the endogenous variables in $\mathbf{x}_{i1}$ (e.g., initial occupation and experience), $p\left(\mathbf{x}_{i1} | \omega\right)$ captures the **initial conditions problem**.

Full-Solution MLE. Issue 1: Initial conditions problem

- How to specify $p(\mathbf{x}_{i1}|\omega)$ [or $\pi_\omega \; p(\mathbf{x}_{i1}|\omega) = p(\mathbf{x}_{i1}, \omega)$] in a way that is:

  (a) identified; (b) consistent with rest of the model.

- In general, the probability $p(\mathbf{x}_{i1}, \omega)$ in **NOT nonparametrically identified** in this max. likelihood problem. This is the **initial conditions problem**.

- We need to impose restrictions on $p(\mathbf{x}_{i1}, \omega)$. These restrictions could be wrong, and even incompatible with the rest of the model [but we do not know this without knowing the solution of the model].

- Note that in a FE approach [if feasible !!!], we do not need to make any assumption on $p(\mathbf{x}_{i1}, \omega)$.

Full-Sol. MLE. Issue 2: Computational complexity

- Nested Fixed point algorithm.
- For each trial value of the parameters θ, the algorithm solves the Dynamic Programming (DP) problem. This introduces a substantial computational burden, especially for models with large state spaces [curse of dimensionality].
- This problem is more severe for model with UH ω because:
  (a) The DP should be solved for each type ω;
  (b) In these models, the likelihood has many local maxima; optimization is quite complex.
  *** **Important**: EM algorithm, by itself, is not a solution; many EM iterations implies solving DP problem many times.

3. Hotz-Miller Finite Dependence

Hotz-Miller & Finite Dependence

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

$$\mathbb{E}_t \left[F \left(u_\omega(y_t, \mathbf{x}_t), \; P_\omega(y_t | \mathbf{x}_t), \; u_\omega(y_{t+1}, \mathbf{x}_{t+1}), \; P_\omega(y_{t+1} | \mathbf{x}_{t+1}) \right) \right] = 0$$

where $F(.)$ is known. This is the same flavor [and in fact it can be derived] as an Euler equation.

- Suppose that we can estimate the CCPs $P_\omega(y|\mathbf{x})$ directly from the data, as reduced-form probabilities, without solving the model.

- Then, we can estimate the structural parameters in $u_\omega(y_t, \mathbf{x}_t)$ by GMM without having to solve the model even once, and without having to compute any present value.

Hotz-Miller & Finite Dependence: Some Details

- There is one-to-one relationship between **conditional choice value differences (CCVD)**, $\widetilde{v}_\omega(y, \mathbf{x}) \equiv v_\omega(y, \mathbf{x}) - v_\omega(0, \mathbf{x})$, **conditional choice probabilities (CCP)**, $P_\omega(y|\mathbf{x})$.

- This mapping depends only on distribution of ε and it has a simple closed-form expression for some distributions. Logit model:

$$P_\omega(y \mid \mathbf{x}_t) = \frac{\exp\left\{\widetilde{v}_\omega(y, \mathbf{x})\right\}}{\sum_{j \in \mathcal{Y}} \exp\left\{\widetilde{v}_\omega(j, \mathbf{x})\right\}}$$

- And the inverse mapping is:

$$\widetilde{v}_\omega(y, \mathbf{x}) = \ln P_\omega(y \mid \mathbf{x}_t) - \ln P_\omega(0 \mid \mathbf{x}_t)$$

Hotz-Miller & Finite Dependence [2]

- This implies that the value function $V_\omega(y, \mathbf{x})$ can be written in terms of CCPs and a "baseline" CCV, $v_\omega(0, \mathbf{x})$. For instance, for the logit model:

$$\begin{aligned} V_\omega(\mathbf{x}) &= \ln\left[\sum_{y\in\mathcal{Y}} \exp\left\{v_\omega\left(y, \mathbf{x}\right)\right\}\right] \\ &= v_\omega\left(0, \mathbf{x}\right) + \ln\left[1 + \sum_{y=1}^{J} \exp\left\{\widetilde{v}_\omega\left(y, \mathbf{x}\right)\right\}\right] \\ &= v_\omega\left(0, \mathbf{x}\right) - \ln P_\omega\left(0, \mathbf{x}\right) \end{aligned}$$

- Remember that $v_\omega\left(y, \mathbf{x}\right) = u_\omega\left(y, \mathbf{x}\right) + \beta \sum_{\mathbf{x}'} V_\omega(\mathbf{x}')\ f_\omega(\mathbf{x}'|y, \mathbf{x})$. Therefore:

$$v_\omega\left(y, \mathbf{x}\right) = u_\omega\left(y, \mathbf{x}\right) + \beta \sum_{\mathbf{x}'} \left[v_\omega\left(0, \mathbf{x}'\right) - \ln P_\omega\left(0, \mathbf{x}'\right)\right]\ f_\omega(\mathbf{x}'|y, \mathbf{x})$$

Hotz-Miller & Finite Dependence [3]

- If we take any pair of actions j and k, we have that $v_\omega(j, \mathbf{x}) - v_\omega(k, \mathbf{x}) = \ln P_\omega(j|\mathbf{x}) - \ln P_\omega(k|\mathbf{x})$, and:

$$v_\omega(j, \mathbf{x}) - v_\omega(k, \mathbf{x}) = u_\omega(j, \mathbf{x}) - u_\omega(k, \mathbf{x})$$

$$-\beta \sum_{\mathbf{x}'} \ln P_\omega(0, \mathbf{x}') \quad [f_\omega(\mathbf{x}'|j, \mathbf{x}) - f_\omega(\mathbf{x}'|k, \mathbf{x})]$$

$$+\beta \sum_{\mathbf{x}'} u_\omega(0, \mathbf{x}') \quad [f_\omega(\mathbf{x}'|j, \mathbf{x}) - f_\omega(\mathbf{x}'|k, \mathbf{x})]$$

$$+\beta^2 \sum_{\mathbf{x}'} \left[\sum_{\mathbf{x}''} V_\omega(\mathbf{x}'') f_\omega(\mathbf{x}''|0, \mathbf{x}') \right] [f_\omega(\mathbf{x}'|j, \mathbf{x}) - f_\omega(\mathbf{x}'|k, \mathbf{x})]$$

Hotz-Miller & Finite Dependence [4]

- The term

$$\sum_{\mathbf{x}_{t+1}} \left[\sum_{\mathbf{x}_{t+2}} V_\omega(\mathbf{x}_{t+2}) \; f_\omega(\mathbf{x}_{t+2}|0, \mathbf{x}_{t+1}) \right] [f_\omega(\mathbf{x}_{t+1}|j, \mathbf{x}) - f_\omega(\mathbf{x}_{t+1}|k, \mathbf{x})]$$

represents the **difference between the continuation values after** $t+1$, of two choice paths:

- choice path: $\{y_t = j \text{ and } y_{t+1} = 0\}$
- choice path: $\{y_t = k \text{ and } y_{t+1} = 0\}$

- 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; inventory; demand of storable products; etc.

Hotz-Miller & Finite Dependence [5]

- Under one-period finite dependence:

$$\ln P_{\omega}\left(j|\mathbf{x}_{t}\right)-\ln P_{\omega}\left(k|\mathbf{x}_{t}\right)=u_{\omega}\left(j,\mathbf{x}_{t}\right)-u_{\omega}\left(k,\mathbf{x}_{t}\right)$$

$$-\beta\sum_{\mathbf{x}_{t+1}}\ln P_{\omega}\left(0,\mathbf{x}_{t+1}\right)\ \left[f_{\omega}(\mathbf{x}_{t+1}|j,\mathbf{x}_{t})-f_{\omega}(\mathbf{x}_{t+1}|k,\mathbf{x}_{t})\right]$$

$$+\beta\sum_{\mathbf{x}_{t+1}}u_{\omega}\left(0,\mathbf{x}_{t+1}\right)\ \left[f_{\omega}(\mathbf{x}_{t+1}|j,\mathbf{x}_{t})-f_{\omega}(\mathbf{x}_{t+1}|k,\mathbf{x}_{t})\right]$$

- If a NP estimator of the reduced-form CCPs $P_{\omega}\left(j|\mathbf{x}\right)$ exists, then we can estimate structural parameters using a simple two-step GMM estimator.

Hotz-Miller & Finite Dependence [6]

- It is convenient to write the FD representation as a **"best response" probability function**.

- Let's use the more compact notation $C_\omega(j, k, \mathbf{x}_t, P_\omega;\theta)$ to represent the RHS FD representation. Then, it is simple to show that we can re-write this equation as:

$$P_\omega\left(j|\mathbf{x}_t\right) = \frac{\exp\left\{C_\omega(j, 0, \mathbf{x}_t, P_\omega; \theta)\right\}}{\sum_{k=0}^{J} \exp\left\{C_\omega(j, 0, \mathbf{x}_t, P_\omega; \theta)\right\}}$$

  The RHS can be interpreted as a best response probability function: given then CCPs at $t+1$, what are the optimal CCPs at t.

- We can define a log-likelihood function $\ell(P_\omega; \theta)$ in terms of the choice probabilities $\frac{\exp\{C_\omega(j,0,\mathbf{x}_t,P_\omega;\theta)\}}{\sum_{k=0}^{J}\exp\{C_\omega(j,0,\mathbf{x}_t,P_\omega;\theta)\}}$. Two-step Pseudo-MLE.

4. Hotz-Miller + NP Finite Mixtures (Kasahara & Shimotsu)

Hotz-Miller & NP Finite Mixtures

- For many years since publication of Hotz-Miller (1993) paper, the common wisdom was that this method was feasible only for models with *i.i.d.* unobservables because CCPs $P_\omega(j|\mathbf{x})$ with permanent UH were not NO identified.

- In this context, the recent developments in the literature of NP Finite Mixtures have been very important: Hall & Zhou (AS, 2003); Allman et al. (AS, 2009); Bonhomme et al. (AS, 2016).

- ... and especially Kasahara and Shimotsu (ECTA, 2009) because it deals with NPFM in Markov Discrete Choice models.

- They show that $P_\omega(j|\mathbf{x})$ are NP identified under relatively standard conditions.

5. Hotz-Miller + EM algorithm (Arcidiacono & Miller)

Arcidiacono & Miller (ECTA, 2011)

- They adapt the **EM algorithm** to incorporate UH into CCP estimators with Finite Dependence.

- Remember the FD representation in term of the "best response probabilities", and define:

$$\Psi_\omega(j \mid \mathbf{x}_t, P_\omega; \theta) \equiv \frac{\exp\{C_\omega(j, \mathbf{x}_t, P_\omega; \theta)\}}{\sum_{k=0}^{J} \exp\{C_\omega(j, \mathbf{x}_t, P_\omega; \theta)\}}$$

- Define the log-likelihood function:

$$\ell(P_\omega, \pi, \theta) = \sum_{i=1}^{N} \ln \left(\sum_{\omega \in \Omega} \pi_\omega \prod_{t=1}^{T} \Psi_\omega(y_{it} \mid \mathbf{x}_{it}, P_\omega; \theta) \right)$$

- AM method consists in the application of the EM algorithm to this max likelihood problem.

Preliminary notes on EM-algorithm

- Consider a general finite mixture model where $p(y|\pi, \theta) = \sum_{\omega \in \Omega} \pi_\omega$ $p(y \mid \omega, \theta)$, and the log-likelihood is:

$$\ell(\pi, \theta) \quad = \quad \sum_{i=1}^{N} \ln p(y_i \mid \pi, \theta) = \sum_{i=1}^{N} \ln \left[\sum_{\omega \in \Omega} \pi_\omega \; p(y_i \mid \omega, \theta) \right]$$

- Define the posterior probabilities: $q_i(\omega \mid \pi, \theta) \equiv \Pr(\omega \mid y_i, \pi, \theta)$. By Bayes' rule:

$$q_i(\omega|\pi, \theta) = \frac{\pi_\omega \; p(y_i \mid \omega, \theta)}{\sum_{\omega' \in \Omega} \pi_{\omega'} \; p(y_i \mid \omega', \theta)}$$

- The EM algorithm does not maximizes the original log-likelihood but the **extended expected likelihood function**:

$$Q(q(\pi, \theta), \theta) \quad = \quad \sum_{i=1}^{N} \sum_{\omega \in \Omega} q_i(\omega|\pi, \theta) \; \ln p(y_i \mid \omega, \theta)$$

EM-algorithm: Steps

- We start the algorithm with initial values $\{\pi^0, \theta^0\}$. At every iteration $n \geq 1$ we update these parameters by applying two steps.
- **Expectation (E) Step.** We update the posterior probabilities q_i and the $\pi's$ as follows:

$$q_i^n(\omega) = \frac{\pi_\omega^{n-1} \; p(y_i \mid \omega, \theta^{n-1})}{\sum_{\omega' \in \Omega} \pi_{\omega'}^{n-1} \; p(y_i \mid \omega', \theta^{n-1})}$$

$$\pi_\omega^n = \frac{1}{N} \sum_{i=1}^{N} q_i^n(\omega)$$

- **Maximization (M) Step.** We update θ by maximizing the expected log-likelihood $Q(q^n, \theta)$

$$\theta^n = \arg\max_\theta \sum_{i=1}^{N} \sum_{\omega \in \Omega} q_i^n(\omega) \; \ln p(y_i \mid \omega, \theta)$$

CCP + EM-algorithm

- Now we have the log-likelihood function:

$$\ell(P_\omega, \pi, \theta) = \sum_{i=1}^{N} \ln \left(\sum_{\omega \in \Omega} \pi_\omega \prod_{t=1}^{T} \Psi_\omega(y_{it} \mid \mathbf{x}_{it}, P_\omega; \theta) \right)$$

- If the vector of CCPs, P_ω, **were known**, then the application of the EM algorithm to this problem would be very straightforward.

- Simply, we apply the same equations for $q_i(\omega|\pi, \theta)$ and for $Q(q(\pi, \theta), \theta)$ with the only difference that now we have that:

$$p(y_i \mid \omega, \theta) = \prod_{t=1}^{T} \Psi_\omega(y_{it} \mid \mathbf{x}_{it}, P_\omega; \theta)$$

- Since we do not know the CCPs P_ω, we need to nest the EM algorithm (inner algorithm) within an "outer' algorithm that updates the CCPs P_ω.

CCP + EM-algorithm [2]

- We start the algorithm with $\{\pi^0, \theta^0, P_\omega\}$. At every iteration $n \geq 1$ we update these parameters as follows.

- **Inner algorithm: EM.** Taking P_ω^{n-1} as given, we apply the EM algorithm to estimate $\{\pi, \theta\}$. This give us $\{\pi^n, \theta^n\}$.

- **Outer algorithm: Updating of CCPs.** Given $\{\pi^n, \theta^n\}$, we update the CCPs as follows:

$$P_\omega^n(j \mid \mathbf{x}, \theta) = \Psi_\omega(j \mid \mathbf{x}, P_\omega^{n-1}; \theta^n) \equiv \frac{\exp\left\{C_\omega(j, \mathbf{x}, P_\omega^{n-1}; \theta^n)\right\}}{\sum_{k=0}^{J} \exp\left\{C_\omega(j, \mathbf{x}, P_\omega^{n-1}; \theta^n)\right\}}$$

- Alternatively, we could use Kasahara-Shimotsu estimates of P_ω and not iterate. Trade-offs.