Dynamic Discrete Choice Structural Models in Empirical IO

Lecture 5: Dealing with Unobserved Heterogeneity: Fixed Effects Approach

Victor Aguirregabiria (University of Toronto)

Carlos III, Madrid
June 30, 2017

Introduction

- In any Dynamic Panel Data (DPD) model (either reduced form or structural), a key econometric issue is distinguishing between **"true dynamics"** (or "true state dependence") and **"spurious dynamics"** due to serially correlated unobservables.

- With short panels, the (reduced form) literature has concentrated on **time-invariant unobserved heterogeneity**.

$$\mathbb{E}\left(\mathbf{x}_{it}\ \eta_i\right) \neq 0$$

- There are two main approaches to deal with this problem:

  (1) the **Fixed effects** approach;
  (2) the **Correlated Random Effects** approach.

Fixed effects (FE) models / methods

- This approach **does not impose any restriction on the joint distribution** of $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ and η_i.

- $CDF(\eta_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ is completely unrestricted. In this sense, the FE model is **nonparametric** with respect the distribution $CDF(\eta_i \mid \mathbf{x}_i)$.

- Typically, fixed effects methods are based on some transformation of the model that eliminates the individual effects, or that make them redundant in a conditional likelihood function.

Correlated Random Effects (CRE) models / methods

- The CRE model imposes some restrictions on the distribution $CDF(\alpha_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$.

- The stronger restriction is that η_i is independent of $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ and $iid(0, \sigma_\eta^2)$. Some textbooks define RE in this restrictive way.
- However, there are more general RE models. For instance, Chamberlain's CRE model:

$$\eta_i = \lambda_0 + \mathbf{x}_{i1}' \ \lambda_1 + ... + \mathbf{x}_{iT}' \ \lambda_T + e_i$$

  where e_i is independent of $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$.
- Based on this assumption, we estimate the parameters β and $\lambda' s$. It is a **parametric approach** because it depends on a parametric assumption on the distribution of $\{x_{i1}, x_{i2}, ..., x_{iT}\}$ and η_i.

Advantages and limitations of FE and CRE models

- (a) FE is more robust because it does not depend on additional assumptions. If the assumption of the CRE is not correct the CRE estimator may be inconsistent.

- (b) The FE transformation may eliminate sample variability of the regressors that is exogenous and useful to estimate the model. Therefore, the FE estimator may be less precise or efficient than the CRE estimator (provided the CRE assumption is consistent).

- (c) For some models (e.g., some nonlinear dynamic models) there is not a root-N consistent FE method, e.g., Chamberlain (ECMA, 2010 on dynamic probit models).

FE and CRE in Dynamic DC Structural Models

- All the literature on DDC structural models with unobserved heterogeneity has focused on CRE models.
- The "common wisdom" is that the FE approach does not work in DDC structural models.
- In these models, unobserved heterogeneity η_i appears nonlinearly in the value function and interacting with observable state variable.
- It seems impossible to "transform" the model or obtain sufficient statistics that control for the unobserved heterogeneity η_i.
- Here I will present results from a recent working paper with my colleagues Jiaying Gu and Yao Luo where we propose FE estimators for some DDC structural models.

Outline

[1] Review of FE estimators in Non-structural DDC Panel Data Models

[2] DDC Structural Models: Assumptions

[3] Identification Results

[4] Estimation

1. Review of FE estimators in Non-structural DDC

Review of FE estimators in Non-structural DDC

- Consider the PD Binary Choice Model:

$$Y_{it} = 1\{\ X_{it}\ \beta + \eta_i + \varepsilon_{it} \leq 0\ \}$$

  N is large and T small.

- X_{it} and η_i can be correlated and we do not impose any restriction on the joint distribution of these variables (FE model).

- Estimators that ignore the correlation between X_{it} and η_i **are inconsistent**.

- The MLE that controls for η_i by including individual-dummies and jointly estimates β and $\eta = (\eta_1, \eta_2, ..., \eta_N)$ [Dummy variables - MLE] **is inconsistent**.

Review of FE estimators in Non-structural DDC

- **Manski´s Maximum Score estimator** of β is consistent when X_{it} contains only strictly exogenous variables, but it is **inconsistent** when X_{it} includes **pre-determined endogenous variables** (i.e., in dynamic models).

- **Chamberlain´s Conditional MLE**: For the **Logit model with strictly exogenous** X_{it}, we have that there is a sufficient statistic for the individual effect η_i.

- Let $\widetilde{Y}_i \equiv (Y_{i1}, ..., Y_{iT})$, $\widetilde{X}_i \equiv (X_{i1}, ..., X_{iT})$, and $S_i \equiv \sum_{t=1}^{T} Y_{it}$. Then:

$$\Pr\left(\widetilde{Y}_i \mid \widetilde{X}_i, S_i, \eta_i, \beta\right) = \Pr\left(\widetilde{Y}_i \mid \widetilde{X}_i, S_i, \beta\right)$$

- This result implies that we can estimate consistently β by using an MLE based on the probabilities $\Pr\left(\widetilde{Y}_i \mid \widetilde{X}_i, S_i, \beta\right)$.

Chamberlain CMLE in DDC models

- Unfortunately, when X_{it} includes **pre-determined endogenous variables** (i.e., in dynamic models), S_i is no longer a sufficient statistic for η_i, and the CMLE described above is **inconsistent**.
- However, Chamberlain (1985) shows that for a simple AR(1) PD Logit model, it is possible to obtain other sufficient statistic for η_i and construct a consistent CMLE.
- This result is in the same spirit as the approach we use for structural model, so I spend some slides here describing it in some detail.

Chamberlain CMLE in DDC models (2)

- Consider the dynamic panel data logit model

$$Y_{it} = 1\{\ \beta\ Y_{i,t-1} + \eta_i + \varepsilon_{it} \leq 0\ \}$$

where u_{it} has a logistic distribution.

- We need $T \geq 4$. Suppose that $T = 4$ and let $\widetilde{Y}_i = \{y_{i1}, y_{i2}, y_{i3}, y_{i4}\}$ be the choice history for individual i.

- Conditional on y_1 and y_4, we can distinguish four sets of choice histories:

$$\begin{aligned} A &= \{y_1, 1, 0, y_4\} \\ B &= \{y_1, 0, 1, y_4\} \\ C &= \{y_1, 1, 1, y_4\} \\ D &= \{y_1, 0, 0, y_4\} \end{aligned}$$

Chamberlain CMLE in DDC models (3)

- Define the statistic $S_i = \{y_1, y_4, y_{i2} + y_{i3} = 1\}$.

- It is possible to show that:

$$\Pr\left(\widetilde{Y}_i \mid S_i, \eta_i, \beta\right) = \Pr\left(\widetilde{Y}_i \mid S_i, \beta\right)$$

i.e., S_i is a sufficient statistic for η_i, and [very importantly] $\Pr\left(\widetilde{Y}_i \mid S_i, \beta\right)$ still depends on β.

- More specifically,

$$\Pr\left(\widetilde{Y}_i = A \mid S_i, \beta\right) = \frac{\exp\left(\beta\ [y_1 - y_4]\right)}{1 + \exp\left(\beta\ [y_1 - y_4]\right)} = \Lambda(\beta\ [y_1 - y_4])$$

Chamberlain CMLE in DDC models (4)

- The CMLE is the value of β that maximizes the Conditional log-likelihood function:

$$\begin{aligned} l^{C}(\beta) &= \textstyle\sum_i 1\{y_{i2}=1, y_{i3}=0\} \ \ln\Lambda(\beta \ [y_{1i}-y_{4i}]) \\ &+ 1\{y_{i2}=0, y_{i3}=1\} \ln\Lambda(-\beta\,[y_{1i}-y_{4i}]) \end{aligned}$$

where $\Lambda(.)$ is the logistic function.

- This likelihood is globally concave in β.

Chamberlain CMLE in DDC models (5)

- The approach can be extended to $T > 4$ and it is still straightforward to implement.

- Let $S(\widetilde{Y}_i) = \{y_{i1}\ y_{iT}, \sum_{t=2}^{T-1} y_{it}\}$. Then,

$$\Pr\left(\widetilde{Y}_i \mid \eta_i, S(\widetilde{Y}_i)\right) = \frac{\exp\left(\beta\ \sum_{t=2}^{T-1} y_{it}\ y_{it-1}\right)}{\sum_{\mathbf{d}:\ S(\mathbf{d})=S(\widetilde{Y}_i)} \exp\left(\beta\ \sum_{t=2}^{T-1} d_t\ d_{t-1}\right)}$$

where, for $\mathbf{d} = (d_1, d_2, ..., d_T) \in \{0, 1\}^T$, we have that

$$S(\mathbf{d}) = \left\{d_1,\ d_T,\ \sum_{t=2}^{T-1} d_t\right\}$$

Honore and Kyriadzidou (ECMA, 2000)

- Consider the dynamic panel data logit model

$$Y_{it} = 1\left\{ \beta \; Y_{i,t-1} + X_{it}' \, \delta + \eta_i + \varepsilon_{it} \leq 0 \right\}$$

where ε_{it} is $i.i.d.$ and has a logistic distribution, and X_{it} is a vector of strictly exogenous regressors with respect to ε_{it}.

- For $T = 4$, they show that $S_i = (y_{i1}, y_{i4}, y_{i2} + y_{i3})$ is a sufficient statistic for η_i only if we condition on $x_{i3} = x_{i4}$.

- Using this approach we can identify β but not β and δ.

- They propose a modified version of the CMLE that incorporates kernel weights that depend on the distance $\|x_{i3} - x_{i4}\|$.

2. DDC Structural models

DDC Structural models: Framework

- Decision variable: $Y_{it} \in \mathcal{Y} = \{0, 1, ..., J\}$.

- Expected & discounted intertemporal payoff
$\mathbb{E}_t \left[\sum_{j=0}^{T-t} \delta_i^j \; U_{i,t+j}(Y_{i,t+j}) \right]$

- One-period payoff:

$$U_{it}(y) = \alpha\left(y, \boldsymbol{\eta}_i, \mathbf{Z}_{it}\right) + \beta\left(y, \mathbf{X}_{it}\right) + \varepsilon_{it}(y).$$

  $\mathbf{Z}_{it}$ and $\mathbf{X}_{it}$ are observable; ε_{it} and $\boldsymbol{\eta}_i$ are unobservable.

- $\mathbf{Z}_{it}$ = exogenous state var. with Markov process $f_{\mathbf{z}}(\mathbf{Z}_{i,t+1} | \mathbf{Z}_{it})$.

- $\mathbf{X}_{it}$ is a vector of endogenous state variables.

DDC Structural models: Framework (2)

- The unobservable variables $\{\varepsilon_{it}(y) : y \in \mathcal{Y}\}$ are $i.i.d.$ over (i, t, y) with a extreme value type I distribution.
- Variable(s) $\boldsymbol{\eta}_i$ represents unobserved heterogeneity from the point of view of the researcher.
- This unobserved heterogeneity can be related with the observable state variables $\mathbf{Z}_{it}$ and $\mathbf{X}_{it}$ in an unrestricted way, and has a distribution that is nonparametrically specified, i.e., fixed effects model.

DDC Structural models: Framework (3)

- $\alpha(.,.,.)$ and $\beta(.,.)$ are functions that are nonparametrically specified and bounded.
- For choice alternative $y = 0$, the functions $\alpha(0, \boldsymbol{\eta}_i, \mathbf{Z}_{it})$ and $\beta(0, \mathbf{X}_{it})$ are normalized to zero.

Endogenous State Variables

- Two types of endogenous state variables that correspond to two different types of state dependence: $\mathbf{X}_{it} = (Y_{i,t-1}, D_{it})$.
- (a) dependence on the lagged decision variable, $Y_{i,t-1} \in \mathcal{Y}$.
- (b) *duration dependence*, $D_{it} \in \{0, 1, ..., d^*\}$, where D_{it} is the number of periods since the last change in choice.
- Duration variable is right-censored at the positive integer value $d^* > 0$.

Transition of Endogenous State Variables

- We use function $\mathbf{X}_{i,t+1} = \mathbf{x}(y, \mathbf{X}_{it})$ to represent in a compact form the transition rule of the two endogenous state variables when the choice at period t is $Y_{it} = y$.

$$\mathbf{X}_{i,t+1} = \mathbf{x}(y, \mathbf{X}_{it}) = \begin{bmatrix} & y \\ 1\{y = Y_{i,t-1}\} & \min\{D_{it} + 1,\ d^*\} \end{bmatrix}$$

- A key feature of this transition rule is that when the choice is $y \neq Y_{i,t-1}$, the process of the endogenous state variables losses its "memory" and is re-initialized, i.e.,

$$\mathbf{x}(y \neq Y_{i,t-1}, \mathbf{X}_{it}) = (y, 0)$$

that does not depend on $\mathbf{X}_{it}$.

Structural State Dependence

- The term $\beta\left(y, \mathbf{X}_{it}\right)$ in the payoff function captures the dynamics of the model, i.e., structural state dependence.

- We distinguish two additive components in this function:

$$\beta\left(y, \mathbf{X}_{it}\right) = 1\{y = Y_{i,t-1}\}\ \beta^{d}\left(y, D_{it}\right) + 1\{y \neq Y_{i,t-1}\}\ \beta^{y}\left(y, Y_{i,t-1}\right)$$

- $\beta^{d}\left(y, D_{it}\right)$ captures duration dependence, e.g., the effect of experience in occupation.

- $\beta^{y}\left(y, Y_{i,t-1}\right)$ represents switching costs, e.g., the cost of switching from occupation $Y_{i,t-1}$ to occupation y.

- We can set $\beta^{y}(y, y) = 0$ and $\beta^{d}\left(y, 0\right) = 0$ for any y.

Examples

- (1) *Market entry-exit models.* Binary choice. Parameter $\beta^{y}(1,0)$ represents the entry cost. $\beta^{d}(1,d)$ represents the effect of market experience on the firm's profit. The entry-exit model can be extended to J markets.
- (2) *Machine replacement models.* Replacing a machine or not. The only endogenous state variable is the number of periods since the last replacement, D_{it}.
- (3) *Occupational choice models.* A worker chooses between J occupations. There are costs of switching occupations. There is also learning that increases productivity in the current occupation.
- (4) *Dynamic demand of differentiated products.*

Optimal Decision Rule

- Agent i chooses Y_{it} to maximize its expected and discounted intertemporal payoff. The optimal choice at period t is:

$$Y_{it} = \arg\max_{y \in \mathcal{Y}} \left\{ \begin{array}{c} \alpha\left(y, \eta_i, \mathbf{Z}_{it}\right) + \beta\left(y, \mathbf{X}_{it}\right) + \varepsilon_{it}(y) \\ \\ +\delta_i \ \mathbb{E}_{\mathbf{Z}_{i,t+1}|\mathbf{Z}_{it}}\left[V_{i,t+1}\left(\mathbf{x}(y, \mathbf{X}_{it}), \mathbf{Z}_{i,t+1}\right)\right] \end{array} \right\}$$

- The CCP function has the following form:

$$P_{it}\left(y|\mathbf{X}_{it}, \mathbf{Z}_{it}\right) = \frac{\exp\left\{ \begin{array}{c} \alpha\left(y, \eta_i, \mathbf{Z}_{it}\right) + \beta\left(y, \mathbf{X}_{it}\right) \\ +\delta_i \ \mathbb{E}_{\mathbf{Z}_{i,t+1}|\mathbf{Z}_{it}}\left[V_{i,t+1}\left(\mathbf{x}(y, \mathbf{X}_{it}), \mathbf{Z}_{i,t+1}\right)\right] \end{array} \right.}{\sum\limits_{j \in \mathcal{Y}} \exp\left\{ \begin{array}{c} \alpha\left(j, \eta_i, \mathbf{Z}_{it}\right) + \beta\left(j, \mathbf{X}_{it}\right) \\ +\delta_i \ \mathbb{E}_{\mathbf{Z}_{i,t+1}|\mathbf{Z}_{it}}\left[V_{i,t+1}\left(\mathbf{x}(j, \mathbf{X}_{it}), \mathbf{Z}_{i,t+1}\right)\right] \end{array} \right.}$$

CCP: Stationary model

- When the model has infinite horizon ($T = \infty$), and payoff and transition prob functions are time homogeneous, Blackwell's Theorem establishes that optimal decision rules and CCP functions are time-invariant.

- The CCP function of the stationary model is:

$$P_i\left(y|\mathbf{X}_{it}, \mathbf{Z}_{it}\right) = \frac{\exp\left\{ \alpha_i\left(y, \mathbf{Z}_{it}\right) + \beta\left(y, \mathbf{X}_{it}\right) + v_i(y, \mathbf{X}_{it}, \mathbf{Z}_{it}) \right\}}{\sum_{j \in \mathcal{Y}} \exp\left\{ \alpha_i\left(j, \mathbf{Z}_{it}\right) + \beta\left(j, \mathbf{X}_{it}\right) + v_i(j, \mathbf{X}_{it}, \mathbf{Z}_{it}) \right\}}$$

For the sake of notational simplicity, we use $\alpha_i\left(y, \mathbf{Z}_{it}\right)$ to represent $\alpha\left(y, \eta_i, \mathbf{Z}_{it}\right)$ and $v_i(y, \mathbf{X}_{it}, \mathbf{Z}_{it})$ to represent

$$v_i(y, \mathbf{X}_{it}, \mathbf{Z}_{it}) \equiv \delta_i \left\{ \begin{array}{c} \mathbb{E}_{\mathbf{Z}_{i,t+1}|\mathbf{Z}_{it}}\left[V_i\left(\mathbf{x}(y, \mathbf{X}_{it}), \mathbf{Z}_{i,t+1}\right)\right] \\ \\ -\mathbb{E}_{\mathbf{Z}_{i,t+1}|\mathbf{Z}_{it}}\left[V_i\left(\mathbf{x}(0, \mathbf{X}_{it}), \mathbf{Z}_{i,t+1}\right)\right] \end{array} \right\}$$

3. Identification results

Data

- The researcher observes panel data of individuals over several periods of time:

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

N 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 that capture "true dynamics" or "true state dependence", i.e., $\beta^{d}(y, D_{it})$ and $\beta^{y}(y, Y_{i,t-1})$.

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

Identification

- Let $\widetilde{\mathbf{y}}_i = \{y_{i1}, y_{i2}, ..., y_{iT}\}$ and $\widetilde{\mathbf{z}}_i = \{\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT}\}$. Define $\theta_i \equiv (\eta_i, \delta_i)$. The model implies that:

$$\mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \widetilde{\mathbf{z}}_i,\ \theta_i\right) = \prod_{t=1}^{T} \frac{\exp\left\{\ \alpha_i\left(y_{it}, \mathbf{z}_{it}\right) + \beta\left(y_{it}, \mathbf{x}_{it}\right) + v_i\left(y_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\right)\right.}{\sum\limits_{j \in \mathcal{Y}} \exp\left\{\ \alpha_i\left(j, \mathbf{z}_{it}\right) + \beta\left(j, \mathbf{x}_{it}\right) + v_i\left(j, \mathbf{x}_{it}, \mathbf{z}_{it}\right)\right.}$$

- We look for an statistic S_i that is sufficient θ_i but does not completely remove the dependence with respect to $\boldsymbol{\beta}$, such that:

$$\mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \widetilde{\mathbf{z}}_i,\ \theta_i,\ S_i,\ \boldsymbol{\beta}\right) = \mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1},\ \widetilde{\mathbf{z}}_i,\ S_i,\ \boldsymbol{\beta}\right)$$

Identification Results

- Propositions 1 and 2 present our main identification result for single-agent models.
- Proposition 1 deals with the identification of switching costs parameters $\beta^{y}(y, y_0)$.
- Proposition 2 deals with the identification of duration dependence parameters $\beta^{d}(y, d)$.
- Before presenting these propositions, I start presenting identification results for the simpler versions of the model.

Identification of Switching Costs

- Conditional on $y_{i0} = y_0$ and $y_{i3} = y_3$, and given arbitrary values y and y^* with $y \neq y^*$, define the choice histories

$$\begin{aligned} A &= \{y_0,\ y,\ y^*,\ y_3\} \\ B &= \{y_0,\ y^*,\ y,\ y_3\} \end{aligned}$$

- Define S_i^{y,y^*} as:

$$S_i^{y,y^*} = 1\left\{ \begin{array}{c} \mathbf{z}_{it} = \mathbf{z}_i \text{ for } t = 1,2,3;\ y_{i0} = y_0;\ y_{i3} = y_3; \\ \text{and } \widetilde{\mathbf{y}}_i \in A \cup B \end{array} \right\}$$

- Then:

$$\mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \widetilde{\mathbf{z}}_i,\ \theta_i,\ S_i^{y,y^*},\ \boldsymbol{\beta}\right) = \mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1},\ \widetilde{\mathbf{z}}_i,\ S_i^{y,y^*},\ \boldsymbol{\beta}\right)$$

and the whole vector $\boldsymbol{\beta}$ is identified.

Example: Entry-Exit Model

- Consider an entry exit model with switching (entry) cost but without duration dependence.

$$P_i\left(y_{i,t-1}, \mathbf{z}_{it}\right) = \frac{\exp\left\{\ \alpha_i\left(\mathbf{z}_{it}\right) + \beta\ y_{i,t-1} + v_i(\mathbf{z}_{it})\ \right\}}{1 + \exp\left\{\ \alpha_i\left(\mathbf{z}_{it}\right) + \beta\ y_{i,t-1} + v_i(\mathbf{z}_{it})\ \right\}}$$

where remember $v_i(\mathbf{z}_{it}) \equiv v_i(1, \mathbf{z}_{it}) - v_i(0, \mathbf{z}_{it})$.

- A **key property** of this model is that the continuation values $v_i(y_{it}, \mathbf{z}_{it})$ do not depend on $y_{i,t-1}$.

- Therefore, the structure the model is very similar to Honore & Kyriadzidou (2000):

$$P_i\left(y_{i,t-1}, \mathbf{z}_{it}\right) = \frac{\exp\left\{\ \widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) + \beta\ y_{i,t-1}\ \right\}}{1 + \exp\left\{\ \widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) + \beta\ y_{i,t-1}\ \right\}}$$

where $\widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) \equiv \alpha_i\left(\mathbf{z}_{it}\right) + v_i(\mathbf{z}_{it})$.

Example: Multinomial case

- Consider the multinomial case with switching costs but without duration dependence.

$$P_i\left(y \mid y_{i,t-1}, \mathbf{z}_{it}\right) = \frac{\exp\left\{ \alpha_i\left(y, \mathbf{z}_{it}\right) + \beta(y, y_{i,t-1}) + v_i(y, \mathbf{z}_{it}) \right\}}{1 + \sum\limits_{j \neq 0} \exp\left\{ \alpha_i\left(j, \mathbf{z}_{it}\right) + \beta(j, y_{i,t-1}) + v_i(j, \mathbf{z}_{it}) \right.}$$

- Again, a **key property** of this model is that the continuation values $v_i(y_{it}, \mathbf{z}_{it})$ do not depend on $y_{i,t-1}$.

- Therefore, the structure the model is:

$$P_i\left(y \mid y_{i,t-1}, \mathbf{z}_{it}\right) = \frac{\exp\left\{ \widetilde{\alpha}_i\left(y, \mathbf{z}_{it}\right) + \beta(y, y_{i,t-1}) \right\}}{1 + \sum\limits_{j \neq 0} \exp\left\{ \widetilde{\alpha}_i\left(j, \mathbf{z}_{it}\right) + \beta(j, y_{i,t-1}) \right\}}$$

where $\widetilde{\alpha}_i\left(y, \mathbf{z}_{it}\right) \equiv \alpha_i\left(y, \mathbf{z}_{it}\right) + v_i\left(y, \mathbf{z}_{it}\right)$.

Identification of Duration Dependence

- Suppose that $T \geq d^* + 2$. Consider that the initial condition is $x_{i1} = (y_0, d^* - 1)$, and define the following two types of choice histories:

$$A = \left\{ y_0,\ y_1^A = 0,\ y_t^A = y \text{ for } t = 2, ..., d^* + 2 \right\}$$

$$B = \left\{ y_0,\ y_1^B = y,\ y_t^B = y \text{ for } t = 2, ..., d^* + 2 \right\}$$

- Define S_i^d as:

$$S_i^d = 1 \left\{ \begin{array}{c} \mathbf{z}_{it} = \mathbf{z}_i \text{ for } t = 1, .., d^* + 2;\ y_{i0} = y_0; \\ \\ \text{and } \widetilde{\mathbf{y}}_i \in A \cup B \end{array} \right\}$$

- Then:

$$\mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \widetilde{\mathbf{z}}_i,\ \theta_i,\ S_i^{y,y^*},\ \boldsymbol{\beta}^d\right) = \mathbb{P}\left(\widetilde{\mathbf{y}}_i \mid \mathbf{x}_{i1},\ \widetilde{\mathbf{z}}_i,\ S_i^{y,y^*},\ \boldsymbol{\beta}^d\right)$$

and the $\beta^d\,(y, d^*) - \beta^d\,(y, d^* - 1)$ is identified.

Example: Machine replacement

- Consider a binary choice machine replacement model

$$P_i\left(d_{it}, \mathbf{z}_{it}\right) = \frac{\exp\left\{ \alpha_i\left(\mathbf{z}_{it}\right) + \beta^d\left(d_{it}\right) + v_i\left(d_{it}, \mathbf{z}_{it}\right) \right\}}{1 + \exp\left\{ \alpha_i\left(\mathbf{z}_{it}\right) + \beta^d\left(d_{it}\right) + v_i\left(d_{it}, \mathbf{z}_{it}\right) \right\}}$$

  where $v_i(d_{it}, \mathbf{z}_{it}) \equiv v_i(0, \mathbf{z}_{it}) - v_i(\min\{d_{it}+1, d^*\}, \mathbf{z}_{it})$.

- A **key property** of this model is that the continuation value $v_i(d_{it}, \mathbf{z}_{it})$ is the same for $d_{it} = d^* - 1$ and $d_{it} = d^*$, i.e., the continuation value does not depend on d_{it}.

- Therefore, for $d_{it} \in \{d^* - 1, d^*\}$:

$$P_i\left(d_{it}, \mathbf{z}_{it}\right) = \frac{\exp\left\{ \widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) + \beta^d\left(d_{it}\right) \right\}}{1 + \exp\left\{ \widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) + \beta^d\left(d_{it}\right) \right\}}$$

  where $\widetilde{\alpha}_i\left(\mathbf{z}_{it}\right) \equiv \alpha_i\left(\mathbf{z}_{it}\right) + v_i(0, \mathbf{z}_{it}) - v_i(d^*, \mathbf{z}_{it})$.

4. Estimation

Fixed Effect Estimation of DDC Structural Models

- Based on the previous identification results, we can use Chamberlain´s or Honore-Kyrazidou´s Conditional MLE to estimate the structural parameters β^y and β^d.
- The implementation of the CMLE is very similar to the one for "non-structural" DDC models.