STRUCTURAL DYNAMIC DISCRETE CHOICE MODELS WITH FIXED EFFECTS

Econometric Society Summer School
in Dynamic Structural Econometrics

Victor Aguirregabiria (University of Toronto)

August 5, 2024

INTRODUCTION

- Disentangling **true dynamics** (causal effect of past decisions) versus **spurious dynamics** (persistent unobserved heterogeneity [UH] is a fundamental problem in the econometrics of dynamic models.

- Challenges with short panels (Heckman, 1981):

  - **Incidental Parameters Problem (IPP):** Treating UH as fixed parameters implies inconsistent estimation of parameters of interest.

  - **Initial Conditions Problem (ICP):** There is no Nonparametric Identification of the distribution of UH and initial conditions.

- Two alternative approaches to deal with the Nonparametric No-Identification from the ICP:
  - **Random Effects (RE)**.
  - **Fixed Effects (FE)**.

RANDOM EFFECTS (RE) vs. FIXED EFFECTS

- **Random Effects (RE):** Integrating out UH.
  - We deal with the ICP by imposing parametric & finite support restrictions on the joint distribution of UH and initial conditions.
  - **Pros:** Full identification of structural parameters & distribution of UH.
  - **Cons:** Misspecification of parametric restrictions on UH can introduce substantial biases in the estimates of "true dynamics".
- **Fixed Effects (FE)**: Differencing out UH.
  - Focus on identification of structural parameters capturing "true dynamics" and not on the identification of the distrribution of UH.
  - **Pros:** NP specification of UH. Robust identification of true dynamics.
  - **Cons:** Distribution of UH is not fully identified. It limits the counterfactuals we can identify.
  - **Cons:** Not all dynamic models have consistent FE estimators.

FIXED EFFECTS IN STRUCTURAL DDC MODELS

- Until recently, all applications of Structural DDC models use RE models to deal with UH.

- The absence of applications using a FE approach was partly because of two common beliefs.

1. Belief that there are not consistent FE estimators in structural models where agents are forward-looking: **problem with continuation values**.

2. Belief that, even if structural parameters are identified, we cannot identify **Average Marginal Effects (AME)** and other Counterfactuals as these depend on the distribution of the UH.

- Recent developments have challenged these beliefs.

BYPRODUCT OF FE APPROACH: COMPUTATIONAL GAINS

- As we will see, the application of the FE approach to dynamic structural models requires **differencing out the continuation value** component of the conditional choice value function.

- This implies that this estimation approach (Conditional MLE) does not require solving any dynamic programming problem, or computing present values, or even one-period forward expectation.

- The computational cost of implementing the Conditional MLE does not depend on the dimension of the state space.

THIS LECTURE

- This lecture presents recent results on Structural DDC - FE Models.
- **Aguirregabiria, Gu, & Luo (Journal of Econometrics, 2021)**
  - Identification & estimation of structural DDC-FE with lagged decision and duration as state variables.
- **Aguirregabiria (Econometrics Journal, 2023)**
  - Application to dynamic demand for differentiated products
- This approach benefits from recent developments on the identification of nonlinear FE models: Bonhomme (ECMA, 2012), Chernozhukov et al. (ECMA, 2013), Honore & Weidner (2020) Aguirregabiria & Carro (2023), Patel & Weidner (2024).

OUTLINE

1. Model
2. Identification of Structural Parameters.
3. Conditional MLE of Structural Parameters.
4. Empirical application – Dynamic Demand for Differentiated Product.

1. MODEL

MODEL: DECISION & STATE VARIABLES

- **Decision variable**: $y_{it} \in \mathcal{Y} = \{0, 1, ..., J\}$.
- Agent maximizes $\mathbb{E}_t \left[\sum_{s=0}^{\infty} \delta_i^s \; U_{i,t+s} \right]$. U_{it} is the utility function.
- U_{it} depends on current choice, y_{it}, and on:
- **Two types of unobservables** for the researcher, $(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}_{it})$;
- **Two types of observable state variables**:

$$\mathbf{s}_{it} = (\mathbf{z}_{it}, \; \mathbf{x}_{it})$$

$\mathbf{z}_{it}$ = strictly exogenous state variables.
$\mathbf{x}_{it}$ = endogenous state variables.

MODEL: UTILITY FUNCTION

- The current payoff of choosing alternative j:

$$U_{it}(j) \quad = \quad \alpha_i(j) + \varepsilon_{it}(j) + \beta\left(j, \boldsymbol{s}_{it}\right)$$

- Payoff function $\beta\left(j, \boldsymbol{s}_{it}\right)$ is unrestricted.

- **Unobservables**:

  - Both types of onobservables are additively separable.

  - $\varepsilon_{it}(j)$ i.i.d. type I extreme value distributed;

  - **FE model**: $p(\alpha_i(0), ..., \alpha_i(J) \mid \boldsymbol{x}_{i1},\ \boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iT}\}$ is unrestricted.

OPTIMAL DECISION & CCPs (conditional on α_i)

- The optimal decision is:

$$y_{it} = \arg\max_{j \in \mathcal{Y}} \left\{ \alpha_i(j) + \varepsilon_{it}(j) + \beta(j, \mathbf{s}_{it}) + cv(j, \mathbf{s}_{it}, \boldsymbol{\alpha}_i) \right\}$$

- where $cv(j, \mathbf{s}_{it}, \boldsymbol{\alpha}_i)$ is the **continuation value function**:

$$cv(j, \mathbf{s}_{it}, \boldsymbol{\alpha}_i) \equiv \delta_i \int V(\mathbf{s}_{i,t+1}, \boldsymbol{\alpha}_i) \; f(\mathbf{s}_{i,t+1} \mid j, \mathbf{s}_{it}) \; d\mathbf{s}_{i,t+1}$$

- The extreme value type 1 distribution of the unobservables ε, implies the **conditional choice probability (CCP)** function:

$$P(j | \mathbf{s}_{it}, \boldsymbol{\alpha}_i) = \frac{\exp\left\{ \alpha_i(j) + \beta(j, \mathbf{s}_{it}) + cv(j, \mathbf{s}_{it}, \boldsymbol{\alpha}_i) \right\}}{\sum_{k \in \mathcal{Y}} \exp\left\{ \alpha_i(k) + \beta(k, \mathbf{s}_{it}) + cv(k, \mathbf{s}_{it}, \boldsymbol{\alpha}_i) \right\}}$$

2. IDENTIFICATION OF STRUCTURAL PARAMETERS

A RESTRICTED VERSION OF THE MODEL

- For simplicity, in this lecture I focus on identification results for a version of the model that imposes two additional restrictions.

- R1: No exogenous state variables $\boldsymbol{z}_{it}$.

- R2: Endogenous state variables follow a deterministic transition rule:

$$\boldsymbol{x}_{i,t+1} = f\left(y_{it}, \boldsymbol{x}_{it}\right)$$

- These restrictions have two important implications.

  1. The initial condition + choice path $\widetilde{\boldsymbol{y}}_i = \left\{\boldsymbol{x}_{i1}, y_{i1}, y_{i2}, ..., y_{iT}\right\}$ contains all the information on the path of choices and states.

  2. For two pairs of choices and states, $(j, \boldsymbol{x})$ and $(j', \boldsymbol{x}')$, with $f\left(j, \boldsymbol{x}\right) = f\left(j', \boldsymbol{x}'\right)$, their continuation values are also the same.

FE – SUFFICIENT STATISTICS APPROACH

- Let $\widetilde{\boldsymbol{y}} = \{\boldsymbol{x}_1, y_1, y_2, ..., y_T\}$ be an individual's observed history

$$\mathbb{P}\left(\widetilde{\boldsymbol{y}}|\boldsymbol{\alpha}\right) = \prod_{t=1}^{T} \frac{\exp\left\{\alpha\left(y_t\right) + \beta\left(y_t, \boldsymbol{x}_t\right) + v\left(f(y_t, \boldsymbol{x}_t), \boldsymbol{\alpha}\right)\right\}}{\sum\limits_{j \in \mathcal{Y}} \exp\left\{\ \alpha\left(j\right) + \beta\left(j, \boldsymbol{x}_t\right) + v\left(f(j, \boldsymbol{x}_t), \boldsymbol{\alpha}\right)\ \right\}} p(\boldsymbol{x}_1|\boldsymbol{\alpha})$$

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

$$\ln \mathbb{P}\left(\widetilde{\boldsymbol{y}}|\boldsymbol{\alpha}\right) = S(\widetilde{\boldsymbol{y}})'\ g(\boldsymbol{\alpha}) + C(\widetilde{\boldsymbol{y}})'\ \boldsymbol{\beta}$$

where $S(\widetilde{\boldsymbol{y}})$ and $C(\widetilde{\boldsymbol{y}})$ are vectors of statistics.

- For instance:
  - $\sum_{t=1}^{T} 1\{y_t = j\}$ is in $S(\widetilde{\boldsymbol{y}})$.
  - $\sum_{t=2}^{T} 1\{y_{t-1} = k \textit{ and } y_t = j\}$ is in $C(\widetilde{\boldsymbol{y}})$.

FE – SUFFICIENT STATISTICS APPROACH (2)

- This structure has several important implications.

$$\ln \mathbb{P}\left(\widetilde{\boldsymbol{y}}|\boldsymbol{\alpha}\right) = S(\widetilde{\boldsymbol{y}})'\, g(\boldsymbol{\alpha}) + C(\widetilde{\boldsymbol{y}})'\, \boldsymbol{\beta}$$

1. $S(\widetilde{\boldsymbol{y}})$ is a sufficient statistic for $\boldsymbol{\alpha}$.

$$\mathbb{P}\left(\widetilde{\boldsymbol{y}} \mid \boldsymbol{\alpha}, S(\widetilde{\boldsymbol{y}})\right) = \mathbb{P}\left(\widetilde{\boldsymbol{y}} \mid S(\widetilde{\boldsymbol{y}})\right)$$

2. $\boldsymbol{\beta}$ is identified if conditional on $S(\widetilde{\boldsymbol{y}})$ the matrix $C(\widetilde{\boldsymbol{y}})'$ for every $\boldsymbol{y}$ is full-column rank.

A MORE INTUITIVE DESCRIPTION OF IDENTIFICATION

- Suppose that there are two choice histories, say A and B For every parameter in the vector $\boldsymbol{\beta}$, say β_k, there exist two choice histories, say $\widetilde{\boldsymbol{y}} = A$ and $\widetilde{\boldsymbol{y}} = B$ such that:

  - $S(A) = S(B)$

  - $C(A) - C(B)$ is a vector where all the elements are zero except for the element associated with β_k, which is $C_k \neq 0$.

- Under these conditions, we have that:

$$\beta_k \;=\; \frac{\log \mathbb{P}(A) - \log \mathbb{P}(B)}{C_k}$$

- Parameter β_k is identified from the log-odds-ratio of histories A & B.

THE CHALLENGE OF THE CONTINUATION VALUES

- The question is whether such histories A & B exist, or on the contrary, $S(A) = S(B)$ implies that there is no variation left in $C(\widetilde{\boldsymbol{y}})$.

- The continuation value $cv(f(y_t, \boldsymbol{x}_t), \boldsymbol{\alpha}_i)$ depends on $\boldsymbol{\alpha}_i$ in a nonlinear (and unknown) form.

- To difference out/control for $\boldsymbol{\alpha}_i$, we need to difference out the whole continuation value.

- But the continuation value also depends on the state variables. So, it seems that differencing out continuation values implies controlling for all the variation in the state variables: **there is no variation left to identify the structural parameters β.**

- **Or there is?**

DIFFERENCING OUT CONTINUATION VALUES

- It turns out that there is a broad and important class of dynamic models where we can difference out continuation values leaving variation in the state variables to identify structural parameters

- Remember that:

$$v(j, \boldsymbol{x}_t, \boldsymbol{\alpha}) \;=\; \alpha(j) + \beta(j, \boldsymbol{x}_t) + cv(f(j, \boldsymbol{x}_t), \boldsymbol{\alpha})$$

- Suppose that the transition rule $f(.)$ is such that there exist two combinations of choice-state $(y_t, \boldsymbol{x}_t)$ such that $\boldsymbol{x}_{t+1}$ is the same:

$$f(j, \boldsymbol{x}) = f(j', \boldsymbol{x}')$$

- Then, it is clear that:

$$v(j, \boldsymbol{x}, \boldsymbol{\alpha}) - v(j', \boldsymbol{x}', \boldsymbol{\alpha}) \;=\; \beta(j, \boldsymbol{x}) - \beta(j', \boldsymbol{x}')$$

- Under this condition, we can identify structural parameters β using a FE – Sufficient Statistics method.

EXAMPLE 1: MULTI-ARMED BANDIT MODELS

- In these models $\boldsymbol{x}_t = y_{t-1}$ such that:

$$\boldsymbol{x}_{t+1} = f(y_t, \boldsymbol{x}_t) = f(y_t, y_{t-1}) = y_t$$

- Therefore, $cv(f(j, y_{t-1}), \boldsymbol{\alpha})$ does not depend on y_{t-1}.

$$v(j, y_{t-1}, \boldsymbol{\alpha}) \ = \ \alpha(j) + \beta(j, y_{t-1}) + cv(j, \boldsymbol{\alpha})$$

- The continuation values $cv(j, \boldsymbol{\alpha})$ are similar as the terms $\alpha(j)$ in the current utility: they do not interact with the state variable y_{t-1}.

- Switching cost parameters, $\beta(y_t, y_{t-1})$ are identified if $T \geq 4$.

- For instance, given choice histories $A = (j, k, j, k)$ and $B = (j, j, k, k)$, we have that:

$$\beta(j, k) = \log \mathbb{P}(A) - \log \mathbb{P}(B)$$

EXAMPLE 2: INVESTMENT MODELS

- In these models $y_t \in \{0, 1, 2, ...\}$ is the investment decision and x_t is the capital stock variable:

$$x_{t+1} = f(y_t, x_t) = x_t + y_t$$

- For any two values of the state, say x and x', we have that:

$$\left[v(1, x, \boldsymbol{\alpha}) - v(0, x+1, \boldsymbol{\alpha})\right] - \left[v(1, x', \boldsymbol{\alpha}) - v(0, x'+1, \boldsymbol{\alpha})\right]$$

$$= [\beta(1, x) - \beta(0, x+1)] - [\beta(1, x') - \beta(0, x'+1)]$$

- Taking into account this structure, it is possible to construct pairs of choice histories, A and B, that identify parameters in β

3. ESTIMATION OF STRUCTURAL PARAMETERS

CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATOR

- Remember that the probability of a choice history $\widetilde{\boldsymbol{y}}_i$ has the following structure:

$$\ln \mathbb{P}\left(\widetilde{\boldsymbol{y}}_i | \boldsymbol{\alpha}_i\right) = S(\widetilde{\boldsymbol{y}}_i)' \ g(\boldsymbol{\alpha}_i) + C(\widetilde{\boldsymbol{y}}_i)' \ \boldsymbol{\beta}$$

and that $S(\widetilde{\boldsymbol{y}}_i)$ is a sufficient statistic for $\boldsymbol{\alpha}_i$.

- We estimate $\boldsymbol{\beta}$ by maximizing the Conditional Likelihood function:

$$\ell^C\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{N} \log \mathbb{P}\left(\widetilde{\boldsymbol{y}}_i \mid S(\widetilde{\boldsymbol{y}}_i), \boldsymbol{\beta}\right)$$

which has the following form:

$$\ell^C\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{N} C(\widetilde{\boldsymbol{y}}_i)'\boldsymbol{\beta} - \sum_{i=1}^{N} \ln\left[\sum_{\widetilde{\boldsymbol{y}} \ S(\widetilde{\boldsymbol{y}})=S(\widetilde{\boldsymbol{y}}_i)} \exp\left\{C(\widetilde{\boldsymbol{y}})'\boldsymbol{\beta}\right\}\right]$$

CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATOR (2)

$$\ell^{C}(\boldsymbol{\beta}) = \sum_{i=1}^{N} C(\widetilde{\mathbf{y}}_i)'\boldsymbol{\beta} - \sum_{i=1}^{N} \ln \left[\sum_{\widetilde{\mathbf{y}}\ S(\widetilde{\mathbf{y}})=S(\widetilde{\mathbf{y}}_i)} \exp\left\{ C(\widetilde{\mathbf{y}})'\boldsymbol{\beta} \right\} \right]$$

- This Conditional Likelihood Function has several important properties:

1. It does not depend on the incidental parameters $\boldsymbol{\alpha}$.

2. It is globally concave in $\boldsymbol{\beta}$.

3. The continuation values enter only in $g(\boldsymbol{\alpha}_i)$. Controlling for $\mathbf{S}$ implies removing the continuation values.

4. Therefore, the computational cost of the Conditional MLE does not depend on the dimension of the state space.

4. EMPIRICAL APPLICATION

Dynamic Demand for Differentiated Product

Laundry Detergent

DATA

- NIELSEN scanner data from Chicago-Kilts center.
- Period 2006-2019. Current estimates using only years 2017-2018.
- More than 40k participating households all over US.
- Rich demographics ($\boldsymbol{w}_i$): ZIP code, income, age, education, occupation, race, family size, family composition, type of residence,
- Data on every shopping trip.
- Product: Laundry detergent

ESTIMATION OF DEMAND PARAMETERS

Fixed Effects provide precise enough estimates ($N = 19,776$).

Estimates of Structural Parameters

Parameter	FE Kernel W. CML Estimate	FE Kernel W. CML (s.e.)	RE (2 types) + $\mathbf{w}_i'\alpha(j)$ Estimate	RE (2 types) + $\mathbf{w}_i'\alpha(j)$ (s.e.)
γ *Price*	1.7392	(**0.3018**)	1.155	(**0.1221**)
β^{sc}(*habits*) *Brand 1*	0.3804	(**0.0290**)	0.7551	(**0.0101**)
β^{sc}(*habits*) *Brand 2*	0.2556	(**0.0573**)	0.6695	(**0.0110**)
β^{sc}(*habits*) *Brand 3*	0.2388	(**0.0591**)	0.7360	(**0.0162**)
β^{dep}(*linear*) *Brand 1*	0.0597	(**0.0112**)	-0.0089	(**0.0040**)
β^{dep}(*linear*) *Brand 2*	0.0611	(**0.0118**)	-0.0161	(**0.0046**)
β^{dep}(*linear*) *Brand 3*	0.0692	(**0.0172**)	-0.0208	(**0.0072**)
Hausman test (p-value)	0.0000			

ESTIMATION OF DEMAND PARAMETERS

Hausman test clearly rejects the Random Effects model.

Estimates of Structural Parameters

	FE Kernel W. CML		**RE (2 types) + $\mathbf{w}_i'\alpha(j)$**	
Parameter	*Estimate*	*(s.e.)*	*Estimate*	*(s.e.)*
γ *Price*	1.7392	(0.3018)	1.155	(0.1221)
$\beta^{sc}(habits)$ *Brand 1*	0.3804	(0.0290)	0.7551	(0.0101)
$\beta^{sc}(habits)$ *Brand 2*	0.2556	(0.0573)	0.6695	(0.0110)
$\beta^{sc}(habits)$ *Brand 3*	0.2388	(0.0591)	0.7360	(0.0162)
$\beta^{dep}(linear)$ *Brand 1*	0.0597	(0.0112)	−0.0089	(0.0040)
$\beta^{dep}(linear)$ *Brand 2*	0.0611	(0.0118)	−0.0161	(0.0046)
$\beta^{dep}(linear)$ *Brand 3*	0.0692	(0.0172)	−0.0208	(0.0072)
Hausman test (p-val)	**0.0000**			

ESTIMATION OF STRUCTURAL PARAMETERS

Random Effects model over-estimates habits parameters.

Estimates of Structural Parameters

	FE Kernel W. CML		**RE (2 types) + $\mathbf{w}_i'\alpha(j)$**	
Parameter	*Estimate*	*(s.e.)*	*Estimate*	*(s.e.)*
γ *Price*	1.7392	(0.3018)	1.155	(0.1221)
β^{sc}(*habits*) *Brand 1*	**0.3804**	(0.0290)	**0.7551**	(0.0101)
β^{sc}(*habits*) *Brand 2*	**0.2556**	(0.0573)	**0.6695**	(0.0110)
β^{sc}(*habits*) *Brand 3*	**0.2388**	(0.0591)	**0.7360**	(0.0162)
β^{dep}(*linear*) *Brand 1*	0.0597	(0.0112)	-0.0089	(0.0040)
β^{dep}(*linear*) *Brand 2*	0.0611	(0.0118)	-0.0161	(0.0046)
β^{dep}(*linear*) *Brand 3*	0.0692	(0.0172)	-0.0208	(0.0072)
Hausman test (p-value)	0.0000			

ESTIMATION OF STRUCTURAL PARAMETERS

Random Effects model provides wrong sign for duration dependence.

Estimates of Structural Parameters

	FE Kernel W. CML		RE (2 types) + $\mathbf{w}_i'\alpha(j)$	
Parameter	*Estimate*	*(s.e.)*	*Estimate*	*(s.e.)*
γ *Price*	1.7392	(0.3018)	1.155	(0.1221)
$\beta^{sc}(habits)$ *Brand 1*	0.3804	(0.0290)	0.7551	(0.0101)
$\beta^{sc}(habits)$ *Brand 2*	0.2556	(0.0573)	0.6695	(0.0110)
$\beta^{sc}(habits)$ *Brand 3*	0.2388	(0.0591)	0.7360	(0.0162)
$\beta^{dep}(linear)$ *Brand 1*	**0.0597**	(0.0112)	**-0.0089**	(0.0040)
$\beta^{dep}(linear)$ *Brand 2*	**0.0611**	(0.0118)	**-0.0161**	(0.0046)
$\beta^{dep}(linear)$ *Brand 3*	**0.0692**	(0.0172)	**-0.0208**	(0.0072)
Hausman test (p-value)	0.0000			

ESTIMATION OF DEMAND PARAMETERS

Random Effects model under-estimates price-sensitivity of demand.

Estimates of Structural Parameters

	FE Kernel W. CML		**RE (2 types) + $\mathbf{w}_i'\alpha(j)$**	
Parameter	*Estimate*	*(s.e.)*	*Estimate*	*(s.e.)*
γ *Price*	**1.7392**	(0.3018)	**1.155**	(0.1221)
β^{sc}(*habits*) *Brand 1*	0.3804	(0.0290)	0.7551	(0.0101)
β^{sc}(*habits*) *Brand 2*	0.2556	(0.0573)	0.6695	(0.0110)
β^{sc}(*habits*) *Brand 3*	0.2388	(0.0591)	0.7360	(0.0162)
β^{dep}(*linear*) *Brand 1*	0.0597	(0.0112)	-0.0089	(0.0040)
β^{dep}(*linear*) *Brand 2*	0.0611	(0.0118)	-0.0161	(0.0046)
β^{dep}(*linear*) *Brand 3*	0.0692	(0.0172)	-0.0208	(0.0072)
Hausman test (p-value)	0.0000			

EXTENSIONS

- This paper presents a Fixed Effects dynamic panel data model of demand for different products where consumers are forward looking.

- Some relevant extensions:

1. Identification of aggregate price elasticities following Aguirregabiria & Carro (2023) results on the identification of Average Marginal Effects.

2. Identification of FE Dynamic games in Aguirregabiria, Gu, and Mira (2022).

3. Introducing stochastic transitions in endogenous state variables.

4. Counterfactuals ?????