Sufficient Statistics for Unobserved Heterogeneity in Dynamic Structural Logit Models

Victor Aguirregabiria, Jiaying Gu, Yao Luo

University of Toronto

November 2, 2017

Motivation

- Incorporating Unobserved Heterogeneity (UH) in Dynamic Panel Data (DPD) models
  - Key issue: distinguish between state dependence (**"true dynamics"**) and **"spurious dynamics"** due to persistent UH. [Heckman (1981)]
- Most common approaches to deal with UH in DPD models are **Fixed Effects (FE)** and **Correlated Random Effects (CRE)**.
  - FE-1: brute force estimate UH and then bias correction for structural parameters (large N large T)
  - FE-2: difference out UH (very common for linear DPD models, challenging for non-linear DPD models).
  - CRE: regularize FE; imposes different types of restrictions: parametric, finite support, group structure, restrictions on initial conditions.

FE in structural dynamic discrete choice (DDC) models

- FE-1 approach is not feasible for structural DDC.
- FE-2 approach is very attractive because it does not impose any restriction on the distribution of the UH conditional on observable explanatory variables and the initial conditions (fully nonparametric).
  - *Non-structural (myopic) dynamic logit: Chamberlain (1985), Honoré and Kyriazidou (2000), Magnac (2000)*
- Not all DDC models can be estimated, root-N consistently, using FE-2 estimators. Examples:
  - *Binary choice models other than logit (Chamberlain 2010).*
  - *Logit with UH multiplicative with explanatory variables.*
- **Structural dynamic logit model:** Common wisdom is FE-2 cannot provide a consistent estimator of structural parameters:
  - Even if UH enters additively into the one-period payoff function, the solution of the model implies that UH appears non-additively in the continuation value.
  - All applications of structural DDC models with UH have considered a CRE approach. Examples:
    - Permanent UH (Finite mixture): Keane and Wolpin (1997), Aguirregabiria and Mira (2007), Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011)
    - Permanent/time varying UH (K-mean classification): Bonhomme et al. (2017)

Contributions

- We propose a FE-2 approach to estimate (root-N) consistently the structural parameters of a structural dynamic logit models (i.e. Rust model with permanent UH).
- We build and extend Chamberlain (1985) to structural (forward-looking) DDC logit models with both **choice-state-dependence** and **duration dependence**.
  - Nonparametric identification of choice-state dependence and duration dependence separately.
  - Minimum sufficient statistics for UH in one-period payoff and continuation value.
  - Construct conditional MLE (à la Anderson (1970)) for structural parameters.

Practicality

- Structural model specified covers many important economic applications in the literature.
  - market entry and exit (binary or multinomial)
  - occupational choice
  - machine replacement
  - dynamic demand of differentiated products
- Identification results rely on finding set of sequences such that, once conditioned on, individual likelihood no longer depends on UH, but still depends on structural parameters of interest.
- FE is consistent for finite T, fully robust, easy to compute.
- Application 1: Revisit Rust (1987) to estimate the structural parameter in maintenance cost.
- Application2 : dynamic demand of a differentiated storable product using consumer scanner data.
  - brand-switching cost and inventory-level dependence
  - Scanner dataset used in Erdem, Imai and Keane (2003).
- (In progress): counterfactual experiment requires identification of distribution of UH.

Outline

(1) Model and Assumptions

(2) Identification of structural parameters [find pair of sequences to difference out UH]

(3) Minimum Sufficiency for UH [efficiency]

(4) Estimation and Inference

(5) Some extensions

(6) Empirical application

Model Elements and Assumptions

- Time is discrete and indexed by t. Agent is indexed by i.
- Time horizon of agent's decision problem is ∞-stationary.
- Observed individual choices have finite length: $t = 1, 2, \ldots, T$.
- **Observables**
  - Decision variable: $y_{it} \in \mathcal{Y} = \{0, 1, \ldots, J\}$.
  - Exogenous state variables: discrete and finite support $\boldsymbol{z}_{it}$ follows a Markov process with transition $f_z(\boldsymbol{z}_{i,t+1}|\boldsymbol{z}_{it})$.
  - Endogenous state variables: $\mathbf{x}_{it}$.
- **Unobservables**
  - $\epsilon_{it}(y)$ i.i.d. type I extreme value distributed
  - $\boldsymbol{\eta}_i$: agent's permanent UH in the payoff (can be multidimensional)
  - $\delta_i \in (0, 1)$: agent-specific discount factor.
- Distribution of incidental parameters $\boldsymbol{\theta}_i = (\boldsymbol{\eta}_i, \delta_i)$ conditional on $\{\mathbf{x}_{it}, \mathbf{z}_{it} : t = 1, 2, \ldots\}$ is unrestricted.
- One-period payoff:

$$U_{it}(y) = \alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + \epsilon_{it}(y)$$

- **Key assumption**: $\boldsymbol{\eta}_i$ and $\mathbf{x}_{it}$ are additively separable in the payoff function.

Endogenous State Variables

- Two endogenous state variables that correspond to two types of state dependence: $\mathbf{x}_{it} = (y_{i,t-1}, d_{it})$.
  - dependence on the lagged decision variable $y_{i,t-1} \in \mathcal{Y}$ for $t = 1, 2, \ldots, T$.
  - dependence on duration: d_{it} is the number of periods since the last change in choice.
- Transition rule for d_{it}: for $t = 1, 2, \ldots, T$,

$$d_{i,t+1} = f_d(y_{it}, \mathbf{x}_{it}) = 1\{y_{it} = y_{i,t-1}\} d_{it} + 1$$

- Deterministic transition $f_{\mathbf{x}}(y, \mathbf{x}_{it}) = (y, f_d(y, \mathbf{x}_{it}))$.
  - stochastic transition à la Rust (1987): cumulative mileage as a state variable (in progress)
- Initial condition $\mathbf{x}_{i1} = (y_{i0}, d_{i1})$.
- Structural state dependence $\beta(y, \mathbf{x}_{it})$: distinguishes two types of dependence

$$\beta(y, \mathbf{x}_{it}) = 1\{y \neq y_{i,t-1}\} \underbrace{\beta_y(y, y_{i,t-1})}_{\text{switching cost}} + 1\{y = y_{i,t-1}\} \underbrace{\beta_d(y, d_{it})}_{\text{duration dependence}}$$

- Occupational choice: (Miller 1984, Keane and Wolpin 1997).
  - cost of switching from occupation y to y': $\beta_y(y', y)$.
  - return of experience on worker's earning: $\beta_d(y, d)$.

Assumptions on structural parameters

- $\beta_y(y, y_{-1})$ and $\beta_d(y, d)$ are bounded.
- Zero switching cost if no switching: $\beta_y(y, y) = 0$.
- Limited return to duration: $\exists d^* < \infty$, $\beta_d(y, d) = \beta_d(y, d^*)$ for $d \geq d^*$.
    - For the moment: assume d^* is known to researcher.
    - d^* can be allowed to change for different $y \in \mathcal{Y}$.
    - Extension: identify d^* from the data.

- Plausible additional assumptions: $\beta_d(0, d) = 0$ for all $d \geq 1$: No duration dependence for "outside alternative" $y = 0$.
    - Occupational choice model: outside alternative is "unemployment".
    - Identification result does not rely on this assumption.
    - With this assumption, more sequences have identifying power.

Examples

- Market entry-exit (binary): (Roberts and Tybout 1997, Dunne et al. 2013)
  - stay active ($y = 1$) vs. exit ($y = 0$).
  - entry cost $\beta_y(1, 0)$ vs. exit cost $\beta_y(0, 1)$.
  - market experience on firm profit: $\beta_d(1, d)$.
  - marginal return of experience is zero once $d \geq d^*$.
- Markets entry-exit (multinomial): (Sweeting 2013, Caliendo et al. 2015)
  - cost of switching from market y to y': $\beta_y(y', y)$.
  - return from experience in market y: $\beta_d(y, d)$.
- Machine replacement: (Rust 1987, Das 1992, Kennet 1993, Kasahara 2009)
  - Keep a machine ($y = 1$) vs. replace a machine ($y = 0$).
  - only state variable d: machine's age.
  - Effect of age on firm's profit: $\beta_d(1, d)$.
- Dynamic demand of differentiated storable products: (Erdem, Imai and Keane 2003, Hendel and Nevo 2006)
  - level of inventory: duration d since last purchase captures inventory level.
  - switching cost from brand y to y': $\beta_y(y', y)$.
  - effect of inventory on consumer utility: $\beta_d(y, d)$.

Solving the model

- Optimal decision rule

$$y_{it} = \underset{y \in \mathcal{Y}}{\text{argmax}} \Big\{ \alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + \epsilon_{it}(y) + \underbrace{\delta_i \mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}} [V(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{i,t+1}, \boldsymbol{\theta}_i)]}_{v(f_x(y,\mathbf{x}_{it}),\mathbf{z}_{it},\boldsymbol{\theta}_i)} \Big\}$$

  with $\epsilon_{it}(y)$ i.i.d. type I extreme value distributed

- Conditional choice probabilities (CCP)

$$\mathbb{P}(y \mid \boldsymbol{\theta}_i, \mathbf{x}_{it}, \mathbf{z}_{it}) = \frac{\exp\{\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + v(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha(j, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(j, \mathbf{x}_{it}) + v(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}$$

- Even though $\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it})$ is separable from state $\mathbf{x}_{it}$,
- $\boldsymbol{\theta}_i = (\boldsymbol{\eta}_i, \delta_i)$ is generally not separable from $\mathbf{x}_{it}$ in the continuation value.
- **Common wisdom**: FE not feasible: It is impossible to difference out $\boldsymbol{\theta}_i$ without also differencing out $\beta(y, \mathbf{x}_{it})$.

Solving the model

- Optimal decision rule

$$y_{it} = \underset{y \in \mathcal{Y}}{\text{argmax}} \Big\{ \alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + \epsilon_{it}(y) + \underbrace{\delta_i \mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}}[V(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{i,t+1}, \boldsymbol{\theta}_i)]}_{v(f_x(y,\mathbf{x}_{it}),\mathbf{z}_{it},\boldsymbol{\theta}_i)} \Big\}$$

  with $\epsilon_{it}(y)$ i.i.d. type I extreme value distributed

- Conditional choice probabilities (CCP)

$$\mathbb{P}(y \mid \boldsymbol{\theta}_i, \mathbf{x}_{it}, \mathbf{z}_{it}) = \frac{\exp\{\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + v(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha(j, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(j, \mathbf{x}_{it}) + v(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}$$

- Even though $\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it})$ is separable from state $\mathbf{x}_{it}$,
- $\boldsymbol{\theta}_i = (\boldsymbol{\eta}_i, \delta_i)$ is generally not separable from $\mathbf{x}_{it}$ in the continuation value.
- **Common wisdom**: FE not feasible: It is impossible to difference out $\boldsymbol{\theta}_i$ without also differencing out $\beta(y, \mathbf{x}_{it})$.

Too pessimistic!

Identification

Some intuition for identification

$$\mathbb{P}(y \mid \boldsymbol{\theta}_i, \mathbf{x}_{it}, \mathbf{z}_{it}) = \frac{\exp\{\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(y, \mathbf{x}_{it}) + v(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha(j, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta(j, \mathbf{x}_{it}) + v(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i)\}}$$

- Structural DDC model **without** duration dependence
  - Only one state variable $x_{it} = y_{i,t-1}$, hence $f_x(y, x_{it}) = y$.
  - Continuation value: $v(f_x(y, x_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i) = v(y, \mathbf{z}_{it}, \boldsymbol{\theta}_i)$, no longer depends on $\mathbf{x}_{it}$!
  - Let $\tilde{\alpha}(y, \mathbf{z}_{it}, \boldsymbol{\theta}_i) \equiv \alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + v(y, \mathbf{z}_{it}, \boldsymbol{\theta}_i)$.
  - Equivalent to a DDC model without continuation value.
  - Use identification strategy similar to Chamberlain (1985) [without $\mathbf{z}_{it}$] and Honoré and Kyriazidou (2000) [set $\mathbf{z}_{it} = \mathbf{z}$ for some periods].

Some intuition for identification

- Structural DDC model **with** duration dependence
  - **Switchers (identification β_y)**: if $y \neq y_{it-1}$

$$\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta_y(y, y_{it-1}) + v(\underbrace{f_x(y, \mathbf{x}_{it})}_{(y,1)}, \mathbf{z}_{it}, \boldsymbol{\theta}_i)$$

  Consider choices $A = \{y_{it-1}, y, y'\}$ vs. $B = \{y_{it-1}, y', y\}$ where $y \neq y' \neq y_{it-1}$

$$\begin{array}{lll} \alpha(y, \boldsymbol{\eta}_i)+ & \beta_y(y, y_{i,t-1}) & +v((y,1), \boldsymbol{\theta}_i) \\ \alpha(y', \boldsymbol{\eta}_i)+ & \beta_y(y', y) & +v((y',1), \boldsymbol{\theta}_i) \end{array} \quad \text{(A)}$$

$$\begin{array}{lll} \alpha(y', \boldsymbol{\eta}_i)+ & \beta_y(y', y_{i,t-1}) & +v((y',1), \boldsymbol{\theta}_i) \\ \alpha(y, \boldsymbol{\eta}_i)+ & \beta_y(y, y') & +v((y,1), \boldsymbol{\theta}_i) \end{array} \quad \text{(B)}$$

  - **Some linear combinations of β_y can be identified** comparing choice histories with switches occurring in different orders.

Some intuition for identification

- **Stayers (identification β_d)** : $y = y_{it-1}$

$$\alpha(y, \boldsymbol{\eta}_i, \mathbf{z}_{it}) + \beta_d(y, d_{it}) + v(\underbrace{f_x(y, \mathbf{x}_{it})}_{(y, \min\{d_{it}+1, d^*\})}, \mathbf{z}_{it}, \boldsymbol{\theta}_i)$$

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

$$\begin{array}{lcr} \alpha(y, \boldsymbol{\eta}_i) + & \beta_d(y, d^* - 1) & + v((y, d^*), \boldsymbol{\theta}_i) \\ \alpha(y, \boldsymbol{\eta}_i) + & \beta_d(y, d^*) & + v((y, d^*), \boldsymbol{\theta}_i) \end{array}$$

- $\beta_d(y, d^*) - \beta_d(y, d^* - 1)$ **can be identified.**

Data and likelihood

- The researcher observes panel data on individual choices over T periods of time.

$$\{y_{it}, \mathbf{z}_{it} : i = 1, 2, \ldots, N; t = 1, 2, \ldots, T\}$$

- N is large, T is small.
- Initial condition $\mathbf{x}_{i1} = (y_{i0}, d_{i1})$ is observed, $\mathbf{x}_{it} = (y_{i,t-1}, d_{it})$.
- Notation:
  - Let $\boldsymbol{\beta}$ collect all $\beta_y(y, y')$ and $\beta_d(y, d)$ for $(y, y') \in \mathcal{Y}^2$ and $d = 1, 2, \ldots, d^*$.
  - Let $\underset{\sim}{\mathbf{y}}_i = \{y_{i1}, \ldots, y_{iT}\}$ and $\underset{\sim}{\mathbf{z}}_i = \{\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}\}$.
  - Let $\tilde{\alpha}(y_{it}, \mathbf{z}_{it}, \boldsymbol{\eta}_i) \equiv \alpha_i(y_{it}, \mathbf{z}_{it})$ and $v(f_x(y_{it}, \mathbf{x}_{it}), \mathbf{z}_{it}, \boldsymbol{\theta}_i) \equiv v_i(f_x(y_{it}, \mathbf{x}_{it}), \mathbf{z}_{it})$.
- Individual likelihood of data conditional on $\mathbf{x}_{i1}$:

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \underset{\sim}{\mathbf{z}}_i, \boldsymbol{\theta}_i) = \prod_{t=1}^{T} \frac{\exp\{\alpha_i(y_{it}, \mathbf{z}_{it}) + \beta(y_{it}, \mathbf{x}_{it}) + v_i(f_x(y_{it}, \mathbf{x}_{it}), \mathbf{z}_{it})\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha_i(j, \mathbf{z}_{it}) + \beta(j, \mathbf{x}_{it}) + v_i(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it})\}}$$

Sufficient Statistics

- Build upon Chamberlain (1985) Details, ignore $\mathbf{z}_{it}$ for the moment, the sufficient statistics take the form $\mathbf{S}_i = 1\{\underset{\sim}{\mathbf{y}}_i \in \mathbb{S}(\underset{\sim}{\mathbf{y}}_i)\}$ such that
$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i = 1, \mathbf{x}_{i1}, \boldsymbol{\theta}_i, \boldsymbol{\beta}) = \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i = 1, \mathbf{x}_{i1}, \boldsymbol{\beta})$$
- Individual likelihood of data conditional on $\mathbf{x}_{i1}$ is
$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i) = \prod_{t=1}^{T} \frac{\exp\{\alpha_i(y_{it}) + \beta(y_{it}, \mathbf{x}_{it}) + v_i(f_x(y_{it}, \mathbf{x}_{it}))\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha_i(j) + \beta(j, \mathbf{x}_{it}) + v_i(f_x(j, \mathbf{x}_{it}))\}}$$

Theorem 1: A choice history $\underset{\sim}{\mathbf{y}} \in \mathbb{S}(\underset{\sim}{\mathbf{y}}_i)$ if and only if the following conditions hold:

1. Initial state matches: $\mathbf{x}_1^y = \mathbf{x}_{i1}$.
2. The set of d^*-censored state variables of the two sequences for $t \in [1, T]$ has the same histogram.
3. The set of d^*-censored state variables of the two sequences for $t \in [2, T+1]$ has the same histogram.
4. d^*-censored terminal state matches $\mathbf{x}_{T+1}^{\mathbf{y}} = \mathbf{x}_{i,T+1} = (y_{iT}, d_{i,T+1})$.

$$\underset{\sim}{\mathbf{x}}_i = \mathbf{x}_{i1}, \quad \mathbf{x}_{i2}, \mathbf{x}_{i3}, \ldots, \mathbf{x}_{iT} \quad \mathbf{x}_{iT+1}$$
$$\underset{\sim}{\mathbf{x}}^{\mathbf{y}} = \mathbf{x}_{i1}, \quad \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_T \quad \mathbf{x}_{T+1}$$

- Sufficient statistics for $\boldsymbol{\theta}_i$ are: $\mathbf{x}_{i1}, \mathbf{x}_{iT+1}$ and histogram of $\{\mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\}$.

Corollary: Sufficient Statistics (myopic with duration)

- If there is no forward looking, then individual likelihood of data conditional on $\mathbf{x}_{i1}$

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i) = \prod_{t=1}^{T} \frac{\exp\{\alpha_i(y_{it}) + \beta(y_{it}, \mathbf{x}_{it})\}}{\sum_{j \in \mathcal{Y}} \exp\{\alpha_i(j) + \beta(j, \mathbf{x}_{it})\}}$$

$$\underset{\sim}{\mathbf{x}}_i = \mathbf{x}_{i1}, \quad \mathbf{x}_{i2}, \mathbf{x}_{i3}, \ldots, \mathbf{x}_{iT} \quad , \mathbf{x}_{iT+1}$$
$$\underset{\sim}{\mathbf{x}}^{\mathbf{y}} = \mathbf{x}_{i1}, \quad \mathbf{x}_{2}, \mathbf{x}_{3}, \ldots, \mathbf{x}_{T} \quad , \mathbf{x}_{T+1}$$

- Sufficient statistics for $\boldsymbol{\theta}_i$ is: $\mathbf{x}_{i1}$ and histogram of $\{\mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\}$ and y_{iT}.

Binary Choice dynamic logit model

- Optimal decision rule

$$y_{it} = 1\left\{ \begin{array}{l} \alpha_i(1) - \alpha_i(0) + \beta(1, y_{it-1}, d_{it}) - \beta(0, y_{it-1}, d_{it}) \\ +v_i(f_x(1, y_{it-1}, d_{it})) - v_i(f_x(0, y_{it-1}, d_{it})) + \epsilon_{it}(1) - \epsilon_{it}(0) \geq 0 \end{array} \right\}$$

- Assume no duration dependence for "0": $v_i(0, d) = 0$ and $\beta_d(0, d) = 0$ for any $d \geq 1$.

$$\begin{aligned} \beta(1, y_{it-1}, d_{it}) - \beta(0, y_{it-1}, d_{it}) &= (1 - y_{it-1})\beta_y(1, 0) + y_{it-1}\beta_d(1, d_{it}) - y_{it-1}\beta_y(0, 1) \\ &= \beta_y(0, 1) + \tilde{\beta}_y y_{it-1} + \tilde{\beta}_d(d_{it}) y_{it-1} \end{aligned}$$

with $\tilde{\beta}_y = -\beta_y(1, 0) - \beta_y(0, 1)$ and $\tilde{\beta}_d(d_{it}) = \beta_d(1, d_{it})$.

Myopic BC dynamic logit without duration

- No duration: $\tilde{\beta}_d(d_{it}) = 0$ for all d_{it}.
- The model can be represented as

$$y_{it} = 1\{\alpha_i + \tilde{\beta}_y y_{it-1} + \epsilon_{it} \geq 0\}$$

where

$$\alpha_i \equiv \alpha_i(1) - \alpha_i(0) + \beta_y(1,0)$$
$$\epsilon_{it} \equiv \epsilon_{it}(1) - \epsilon_{it}(0)$$

- Chamberlain (1985): $T = 3$: $A = \{0, 1, y_{i3}\}$ and $B = \{1, 0, y_{i3}\}$ with $\mathbf{x}_{i1} = (y_{i0}, d_{i1})$. The two histories visit the same choice states (with different timing):

$$\frac{\mathbb{P}(A \mid \alpha_i, \mathbf{x}_{i1})}{\mathbb{P}(B \mid \alpha_i, \mathbf{x}_{i1})} = \exp\{\tilde{\beta}_y(y_{i3} - y_{i0})\}$$

- $(y_{i0}, y_{i3}) = (1, 0)$ or $(0, 1)$ (over)-identify $\tilde{\beta}_y$.

Forward-looking BC dynamic logit without duration

The model can be represented as

$$y_{it} = 1\{\alpha_i + \tilde{\beta}_y y_{it-1} + \epsilon_{it} + v_i \geq 0\}$$

where

$$\begin{aligned} \alpha_i &\equiv \alpha_i(1) - \alpha_i(0) + \beta_y(1,0) \\ \beta &\equiv -\beta_y(1,0) - \beta_y(0,1) \\ \epsilon_{it} &\equiv \epsilon_{it}(1) - \epsilon_{it}(0) \\ v_i &\equiv v_i(1) - v_i(0) \end{aligned}$$

- Same fixed effect estimator as before identifies $\tilde{\beta}_y$.

Myopic BC dynamic logit with duration dependence

The model can be represented as

$$y_{it} = 1\{\alpha_i + \tilde{\beta}_y y_{it-1} + \tilde{\beta}_d(d_{it})y_{it-1} + \epsilon_{it} \geq 0\}$$

- $T = 3$: $A = \{0, 1, y_{i3}\}$ and $B = \{1, 0, y_{i3}\}$ with $\mathbf{x}_{i1} = (y_{i0}, d_{i1})$
  - $(y_{i0}, y_{i3}) = (0, 1)$
    $$\frac{\mathbb{P}(A_1 \mid \alpha_i, \mathbf{x}_{i1})}{\mathbb{P}(B_1 \mid \alpha_i, \mathbf{x}_{i1})} = \exp\{\tilde{\beta}_y + \tilde{\beta}_d(1)\}$$
  - $(y_{i0}, y_{i3}) = (1, 0)$ **if** $d^* = 1$.
    $$\frac{\mathbb{P}(A_2 \mid \alpha_i, \mathbf{x}_{i1})}{\mathbb{P}(B_2 \mid \alpha_i, \mathbf{x}_{i1})} = \exp\{\tilde{\beta}_y + \tilde{\beta}_d(1)\}$$
- $T = d + 3$, $A_3 = \{0, 1_{d+2}\}$ and $B_3 = \{1, 0, 1_{d+1}\}$ for $\mathbf{x}_{i1} = (1, d)$,
  $$\frac{\mathbb{P}(A_3 \mid \alpha_i, \mathbf{x}_{i1})}{\mathbb{P}(B_3 \mid \alpha_i, \mathbf{x}_{i1})} = \exp\{\tilde{\beta}_d(d+1) - \tilde{\beta}_d(d)\}$$
- States $\mathbf{x}_{it} = (y_{it-1}, d_{it})$ visited from $t = 1, \ldots, T+1$ for A and B:

$$\begin{array}{lcc|ccccc|c} \mathbf{x}^A = & (1, & d) & (0, \quad 1) & (1, \quad 1) & (1, \quad 2) & \ldots & (1, \quad d+1) & (\mathbf{1}, \quad d+2) \\ \mathbf{x}^B = & (1, & d) & (1, \quad d+1) & (0, \quad 1) & (1, \quad 1) & \ldots & (1, \quad d) & (\mathbf{1}, \quad d+1) \end{array}$$

Forward-looking BC dynamic logit with duration dependence

- The model can be represented as

$$y_{it} = 1\{\alpha_i + \tilde{\beta}_y y_{it-1} + \tilde{\beta}_d(d_{it}) y_{it-1} + \epsilon_{it} + v_i(f_x(1, y_{it-1}, d_{it})) - v_i(f_x(0, y_{it-1}, d_{it})) \geq 0\}$$

If $d^* > 1$

- Suppose $T \geq d^* + 2$. Consider initial condition $\mathbf{x}_1 = (1, d_1)$.

$$A = \{0, 1_{d^*+1}\}$$
$$B = \{1, 0, 1_{d^*}\}$$

- When $d_1 = d^* - 1$, **we can identify**

$$\ln \frac{\mathbb{P}(A \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i)}{\mathbb{P}(B \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i)} = \tilde{\beta}_d(d^*) - \tilde{\beta}_d(d^* - 1)$$

- States $\mathbf{x}_{it} = (y_{it-1}, d_{it})$ visited from $t = 1, \ldots, T+1$ for A and B

$$\begin{array}{ll|ccccc|c} \mathbf{x}^A = & (1, \ d^*-1) & (0, \ 1) & (1, \ 1) & (1, \ 2) & \ldots & (1, \ d^*) & (1, \ d^*) \\ \mathbf{x}^B = & (1, \ d^*-1) & (1, \ d^*) & (0, \ 1) & (1, \ 1) & \ldots & (1, \ d^*-1) & (1, \ d^*) \end{array}$$

Forward-looking BC dynamic logit with duration dependence (2)

If $d^* = 1$

- The model can be represented as

$$y_{it} = 1\{\alpha_i + \tilde{\beta}_y y_{it-1} + \tilde{\beta}_d(1) y_{it-1} + \epsilon_{it} + v_i(1,1) - v_i(0,1) \geq 0\}$$

- Use Chamberlain's estimator we can identify $\tilde{\beta}_y + \tilde{\beta}_d(1)$.

- Can we generalize from Binary choice to multinomial?
- Can we generalize for any T to find all relevant sequences that has identification power?

$$\underset{\sim}{\mathbf{x}}_i = \mathbf{x}_{i1}, \quad \mathbf{x}_{i2}, \mathbf{x}_{i3}, \ldots, \mathbf{x}_{iT} \quad \mathbf{x}_{iT+1}$$
$$\underset{\sim}{\mathbf{x}}^{\mathbf{y}} = \mathbf{x}_{i1}, \quad \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_T \quad \mathbf{x}_{T+1}$$

- Theorem 1 is hard to use for several reasons:
  - We can not permute $\{\mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\}$ freely due to transition rule for d_{it}.
  - It is silent about identification of β.
- We need a better representation of the likelihood: Map choice history into choice runs.

Sufficient Statistics

Definition: A *choice run* is defined as a sequence of periods in which the same choice state is visited consecutively. It can be represented by two values $\begin{pmatrix} y \\ n \end{pmatrix}$ where y is the choice alternative and n is the number of periods of the run.

A choice history $\{y_{i0}, \underset{\sim}{\mathbf{y}}_i\}$ can be represented as a sequence of R_i runs,

$$\mathbf{H}_i = \left\{ \begin{pmatrix} y_i^{(r)} \\ n_i^{(r)} \end{pmatrix}, r = 1, \ldots, R_i \right\}$$

Modify $n_i^{(1)} \to n_i^{(1)} + d_{i1} - 1$ to record initial condition.

Example: $\{y_{i0}, \underset{\sim}{\mathbf{y}}_i\} = \{1, 1, 0, 1, 1, 1, 1\}$ with $d_{i1} = 1$ is represented by 3 runs:

$$\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix} \right\}$$

- **Instead of permuting state variable $\{\mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\}$, it is easier to permute runs.**

Likelihood representation using choice runs

Individual likelihood of data conditional on $\mathbf{x}_{i1}$

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i) = \prod_{t=1}^{T} \frac{\exp\{\alpha_i(y_{it}) + \beta(y_{it}, \mathbf{x}_{it}) + v_i(f_x(y_{it}, \mathbf{x}_{it}))\}}{\sum_{j\in\mathcal{Y}} \exp\{\alpha_i(j) + \beta(j, \mathbf{x}_{it}) + v_i(f_x(j, \mathbf{x}_{it}))\}}$$

$$= \frac{\exp\{\sum_{t=1}^{T} \alpha_i(y_{it}) + \beta(y_{it}, y_{it-1}, d_{it}) + v_i(y_{it}, f_d(y_{it}, y_{it-1}, d_{it}))\}}{\exp\{\sum_{t=1}^{T} \sigma_i(y_{it-1}, d_{it})\}}$$

where $\sigma_i(y_{it-1}, d_{it}) = \ln\left[\sum_{j\in\mathcal{Y}} \exp\{\alpha_i(j) + \beta(j, y_{it-1}, d_{it}) + v_i(j, f_d(j, y_{it-1}, d_{it}))\}\right]$.

The individual log-likelihood can be represented as Details

$$\begin{aligned}
\ln \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \boldsymbol{\beta}, \boldsymbol{\theta}_i) = &\sum_{r=2}^{R_i} \beta_y(y_i^{(r)}, y_i^{(r-1)}) \\
&+ \sum_{r=1}^{R_i} \beta_d^R(y_i^{(r)}, n_i^{(r)}) - 1\{d_{i1} \geq 2\} \sum_{d=1}^{d_{i1}-1} \beta_d(y_i^{(1)}, d) \\
&+ \mathbf{T}_i'\boldsymbol{\alpha}_i + \sum_{r=1}^{R_i} v_i^R(y_i^{(r)}, n_i^{(r)}) - \sum_{d=1}^{d_{i1}} v_i(y_i^{(1)}, d) \\
&- \sum_{r=1}^{R_i} \sigma_i^R(y_i^{(r)}, n_i^{(r)}) + 1\{d_{i1} \geq 2\} \sum_{d=1}^{d_{i1}-1} \sigma_i(y_i^{(1)}, d) + \sigma_i(y_i^{(R_i)}, n_i^{(R_i)})
\end{aligned}$$

Identification of β_y

- Sufficient statistics for $\boldsymbol{\theta}_i$ is the last two rows of the likelihood representation.

$$\mathbf{T}_i'\boldsymbol{\alpha}_i + \sum_{r=1}^{R_i} v_i^R(y_i^{(r)}, n_i^{(r)}) - \sum_{d=1}^{d_{i1}} v_i(y_i^{(1)}, d)$$
$$- \sum_{r=1}^{R_i} \sigma_i^R(y_i^{(r)}, n_i^{(r)}) + 1\{d_{i1} \geq 2\} \sum_{d=1}^{d_{i1}-1} \sigma_i(y_i^{(1)}, d) + \sigma_i(y_i^{(R_i)}, n_i^{(R_i)})$$

where $\boldsymbol{\alpha}_i = (\alpha_i(0), \ldots, \alpha_i(J))'$.

  - Entries in the vector $\mathbf{T}_i$: $\sum_{r=1}^{R_i} 1\{y_i^{(r)} = j\} n_i^{(r)}$. # times j-choice is visited.
  - $v_i^R(y, n) = \sum_{d=1}^{n} v_i(y, d)$ and $\sigma_i^R(y, n) = \sum_{d=1}^{n} \sigma_i(y, d)$.

- **To identify** $\boldsymbol{\beta}_y$: permuting "runs in between"

$$\left\{ \begin{pmatrix} y_i^{(1)} \\ n_i^{(1)} \end{pmatrix}, \boxed{\begin{pmatrix} y_i^{(2)} \\ n_i^{(2)} \end{pmatrix} \ldots, \begin{pmatrix} y_i^{(R_i-1)} \\ n_i^{(R_i-1)} \end{pmatrix}}, \begin{pmatrix} y_i^{(R_i)} \\ n_i^{(R_i)} \end{pmatrix} \right\}.$$

- Provided after permuting, it is still a well-defined runs (i.e. $y^{(r)} \neq y^{(r+1)}$ for all $1 \leq r \leq R_i - 1$).

Identification of β_d

$$\left\{ \begin{pmatrix} y_i^{(1)} \\ n_i^{(1)} \end{pmatrix}, \begin{pmatrix} j \\ d^* - 1 \end{pmatrix} \ldots, \begin{pmatrix} j \\ d^* + 1 \end{pmatrix}, \begin{pmatrix} y_i^{(R_i)} \\ n_i^{(R_i)} \end{pmatrix} \right\}.$$

$$\left\{ \begin{pmatrix} y_i^{(1)} \\ n_i^{(1)} \end{pmatrix}, \begin{pmatrix} j \\ d^* \end{pmatrix} \ldots, \begin{pmatrix} j \\ d^* \end{pmatrix}, \begin{pmatrix} y_i^{(R_i)} \\ n_i^{(R_i)} \end{pmatrix} \right\}.$$

- $v_i^R(j, d^* - 1) + v_i^R(j, d^* + 1) = v_i^R(j, d^*) + v_i^R(j, d^*)$
- $\sigma_i^R(j, d^* - 1) + \sigma_i^R(j, d^* + 1) = \sigma_i^R(j, d^*) + \sigma_i^R(j, d^*)$
- $\beta_d^R(j, d^* - 1) + \beta_d^R(j, d^* + 1) \neq \beta_d^R(j, d^*) + \beta_d^R(j, d^*)$
- $\{\beta_d^R(j, d^* - 1) + \beta_d^R(j, d^* + 1)\} - \{\beta_d^R(j, d^*) + \beta_d^R(j, d^*)\} = \beta_d(j, d^*) - \beta_d(j, d^* - 1).$

Identification of β_d

Identification of β_d: Parameter $\beta_d(j, d^*) - \beta_d(j, d^* - 1)$ are identified for all j such that there exists at least one pair of (r, r') that

1. $y_i^{(r)} = y_i^{(r')} = j$
2. $\min\{n_i^{(r)}, n_i^{(r')}\} = d^* - 1$
3. $n_i^{(r)} + n_i^{(r')} \geq 2d^*$.

Choice run examples

Example 1 (Multinomial)

- $A = \{1, 2, 0\}$ and $B = \{2, 1, 0\}$ and $\mathbf{x}_{i1} = (0, d_{i1})$

$$\mathbf{H}_i^A = \begin{pmatrix} 0 & 1 & 2 & 0 \\ d_{i1} & 1 & 1 & 1 \end{pmatrix} \quad \mathbf{H}_i^B = \begin{pmatrix} 0 & 2 & 1 & 0 \\ d_{i1} & 1 & 1 & 1 \end{pmatrix}$$

- Same first run.
- Same last run.
- Permuting two runs in between.

Example 2 (Multinomial)

- $A = \{y', y \times 1_{d^*+1}\}$ and $B = \{y, y', y \times 1_{d^*}\}$ and $\mathbf{x}_{i1} = (y, d^* - 1)$. Let $y = 1$ and $y' = 0$

$$\mathbf{H}_i^A = \begin{pmatrix} 0 & 1 & 0 \\ d^* - 1 & 1 & d^* + 1 \end{pmatrix} \quad \mathbf{H}_i^B = \begin{pmatrix} 0 & 1 & 0 \\ d^* & 1 & d^* \end{pmatrix}$$

- Same sequences of runs.
- $\exists$ two runs of same choice: minimum run length $= d^* - 1$, run length sum $\geq 2d^*$.

Condition MLE

- Let the set $\mathbb{S}(\mathbf{y}_i)$ collects all sequences that have identification power.
- Let γ collects all parameters that can be identified.
- The conditional log-likelihood function for γ

$$\sum_{i=1}^{N} \ln \frac{\exp\{\mathrm{t}_y(\mathbf{y}_i)^\top \gamma\}}{\sum_{\lambda \in \mathbb{S}(\mathbf{y}_i)} \exp\{\mathrm{t}_y(\lambda)^\top \gamma\}}$$

  where $\mathrm{t}_y(\mathbf{y}_i)$ picks out the corresponding elements in γ and is easy to be programmed.
- Duration dependence parameter that can be identified takes the form $\beta_d(y, d^*) - \beta_d(y, d^* - 1)$.
- Switching cost parameters that can be identified takes the form $\delta_y(y, y') = \beta_y(y, y') - \beta_y(y, 0) - \beta_y(0, y')$.
- Interpretation: for $y \neq y' \neq 0$,

$$\delta_y(y, y') = \ln \frac{\mathbb{P}(y_{it} = y | \mathbf{x}_{it} = (y', d))}{\mathbb{P}(y_{it} = 0 | \mathbf{x}_{it} = (y', d))} - \ln \frac{\mathbb{P}(y_{it} = y' | \mathbf{x}_{it} = (0, d))}{\mathbb{P}(y_{it} = 0 | \mathbf{x}_{it} = (0, d))}$$

- Comparison between the switching cost from $y \to y'$ versus $y \to 0 \to y'$.

Some extension and development in progress

Identification of d^*

- Take K as the hypothetical duration censoring point, find CMLE for $\beta_d(y, K) - \beta_d(y, K-1)$.
  If $K > d^*$ then $\beta_d(y, K) - \beta_d(y, K-1) = 0$.
- If $K = d^*$, then $\beta_d(y, d^*) - \beta_d(y, d^* - 1) \neq 0$.
- d^* can be identified as

$$d^* = \max\{K : \beta_d(y, K) - \beta_d(y, K-1) \neq 0\}$$

In progress: counterfactual

Counterfactual: Once we get a robust CMLE for $\boldsymbol{\beta}$, we should come back to estimate distribution of UH in order to do counterfactual analysis.

- Consider **Binary choice model without duration dependence**:

$$y_{it} = 1\{\underbrace{\tilde{\alpha}_i + \tilde{v}_i}_{\mu_i} + \beta y_{it-1} + \epsilon_{it} \geq 0\}$$

1. Get CMLE for β.
2. Nonparametric mixture model with parametric base distribution:

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid y_{i0}) = \int \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid y_{i0}, \mu_i, \beta) dF(\mu_i \mid y_{i0})$$

   Solve for $F(\mu_i \mid y_{i0} = 1)$ and $F(\mu_i \mid y_{i0} = 0)$.
3. The model implies:

$$\tilde{v}_i = \delta \ln \frac{1 + \exp(\tilde{\alpha}_i + \tilde{v}_i + \beta)}{1 + \exp(\tilde{\alpha}_i + \tilde{v}_i)} = \delta \ln \frac{1 + \exp(\mu_i + \beta)}{1 + \exp(\mu_i)} \equiv h(\mu_i, \beta, \delta)$$

   hence $\mu_i = \tilde{\alpha}_i + \tilde{v}_i = \tilde{\alpha}_i + h(\mu_i, \beta, \delta)$ gives distribution for $\tilde{\alpha}_i$, with which we are ready for counterfactual analysis.

- Computationally very simple!
- With duration dependence (under development).

In progress

- We've ignored so far the exogenous state variable $\mathbf{z}_{it}$.
- Extend the results to kernel-weighted CMLE in Honoré and Kiriazidou 2000.
- If $\mathbf{z}_{it}$ is continuous, estimator will no longer be root-N rate. Pairwise approach may be attractive to get better rate.
- Stochastic evolution for d_{it}.

Empirical Application

Rust bus engine replacement

- Dataset from Rust (1987)
- Rust's model

$$U_{it} = \begin{cases} -c_0 - c_1(m_{it}) + \epsilon_{it}(1) & \text{if } y_{it} = 1; \text{ no replacement} \\ -RC + \epsilon_{it}(0) & \text{if } y_{it} = 1; \text{ replacement} \end{cases}$$

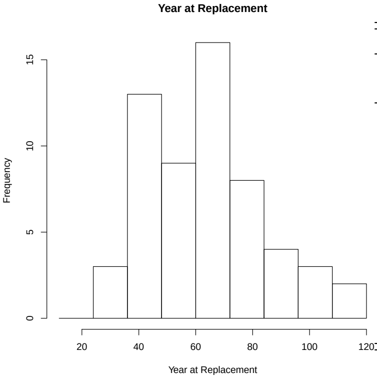
- In our notation: [allow for bus specific replacement cost (RC) and constant maintenance cost (c_0)]

$$U_{it} = \begin{cases} \alpha_i(1) + \beta_d(d_{it}) + \epsilon_{it}(1) & \text{if } y_{it} = 1; \text{ no replacement} \\ \alpha_i(0) + \epsilon_{it}(0) & \text{if } y_{it} = 0; \text{ replacement} \end{cases}$$

  with $\alpha_i(1) = -c_{0i}$ and $\alpha_i(0) = -RC_i$.
- 124 buses [group 1 - 8]. Rust focused on 104 buses [group 1-4].
- 45 buses no replacement, **58 buses one replacement**, 1 bus two replacements.

Rust bus engine replacement

- Assume replacement decision is made every 12 months.



Choice history	Frequency: Absolute	Frequency: %	Frequency: % cumulative
1101111111	3	5.17	5.17
1110111111	11	18.96	24.13
1111011111	9	15.51	39.64
1111101111	18	31.03	70.67
1111110111	7	12.07	82.74
1111111011	5	8.62	91.36
1111111101	3	5.17	96.53
1111111110	2	3.45	100.00

Table: **Empirical Distribution of Choice Histories**

Estimates of d^* and β_d

Estimates of $\beta_d(\mathbf{d}^*) - \beta_d(\mathbf{d}^* - \mathbf{1})$		
Value for d^*	Estimate	s.e.
$d^* = 4$	-0.205	0.295
$d^* = 3$	-1.07^{***}	0.121

Dynamic demand of differentiated storable product

- Consumer scanner data (A.C. Nielsen) on ketchup purchases.
- Same dataset as in Pesendorfer (1998) and Erdem, Imai and Keane (2003).
- 2797 households over 123 weeks.
- Three national brands (Heinz, Hunt's and Del Monte), and one store brand. $\mathcal{Y} = \{1, 2, 3\}$.
- Outside option "0": No purchase.
- Duration since last purchase represents inventory depletion.
- A consumer's choice sequence could look like

$$\{1, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0 \dots\}$$

- A consumer's choice run

$$\mathbf{H} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \dots \right\}$$

- Duration dependence doesn't depend on brand choice: estimate $\beta_d(0, d^*) - \beta_d(0, d^* - 1)$.

Estimation of brand switching costs

Switching Costs parameters (under symmetry)
Estimates (s.e.)

	Heinz	Hunts	Del Monte	Store
Heinz	-	1.052***(0.427)	1.711***(0.421)	2.199**(0.423)
Hunts		-	0.635(0.465)	1.225**(0.465)
Del Monte			-	1.016**(0.472)
Store				-

Estimation of d^*

$\beta_d(0, d^*) - \beta_d(0, d^* - 1)$ (same for all brands) Estimates (s.e.)		
Value for d^*	Estimate	s.e.
...		
$d^* = 16$ weeks	-0.025	0.298
$d^* = 15$ weeks	-0.124	0.295
$d^* = 14$ weeks	-0.287	0.223
$d^* = 13$ weeks	-0.374	0.215
$d^* = 12$ weeks	-0.516^{**}	0.196

Conclusion

- We study identification for fixed effect structural dynamic logit discrete choice model.
- A simple CMLE is proposed for choice-state-dependence and duration dependence parameters.
- FE estimator for structural parameter is consistent and fully robust.
- Results for myopic case fills in the gap for non-structural logit DDC with duration dependence.

Thank you!

Likelihood representation notations:

- $\mathbf{T}_i = (T_i(0), T_i(1), \dots, T_i(J))'$ with $T_i(j) = \sum_{t=1}^{T} 1\{y_{it} = j\}$: one-to-one mapping to row sum of $\mathbf{H}_i$.
- $\boldsymbol{\alpha}_i = (\alpha_i(0), \alpha_i(1), \dots, \alpha_i(J))'$.
- $\beta_d^R(y, n) = 1\{n \geq 2\} \sum_{d=1}^{n-1} \beta_d(y, d)$
- $v_i^R(y, n) = \sum_{d=1}^{n} v_i(y, d)$
- $\sigma_i^R(y, n) = \sum_{d=1}^{n} \sigma_i(y, d)$.

Back to LikRepresent

Chamberlain's CMLE in non-structural DDC logit model

- No forward looking, no duration dependence, no $\mathbf{z}_{it}$, binary outcome

$$y_{it} = 1\{\beta y_{i,t-1} + \boldsymbol{\eta}_i + \epsilon_{it} \geq 0\}$$

- Likelihood conditional on y_{i0}

$$\mathbb{P}(\{y_{i1}, \ldots, y_{iT}\} \mid y_{i0}, \boldsymbol{\eta}_i) = \frac{\exp\{\boldsymbol{\eta}_i \sum_{t=1}^{T} y_{it} + \beta \sum_{t=1}^{T} y_{it} y_{i,t-1}\}}{\prod_{t=1}^{T}(1 + \exp\{\boldsymbol{\eta}_i + \beta y_{i,t-1}\})}$$

- Minimum sufficient statistics for $\boldsymbol{\eta}_i$: $(y_{i0}, y_{iT}, \sum_{t=1}^{T-1} y_{it})$.
- For $T = 3$, $\sum_{t=1}^{T-1} y_{it} \in \{0, 1, 2\}$.
  - $\sum_{t=1}^{T-1} y_{it} = 0$, singleton set of choice paths $\{y_{i0}, 0, 0, y_{i3}\}$.
  - $\sum_{t=1}^{T-1} y_{it} = 2$, singleton set of choice paths $\{y_{i0}, 1, 1, y_{i3}\}$.
  - $\sum_{t=1}^{T-1} y_{it} = 1$, set of choice paths $\{A, B\}$ with $A = \{y_{i0}, 1, 0, y_{i3}\}$ and $B = \{y_{i0}, 0, 1, y_{i3}\}$.
- Identification of β: $\mathbb{P}(\underset{\sim}{\mathbf{y}}_i = A \mid A \cup B, \beta) = \dfrac{\exp\{\beta(y_{i0} - y_{i3})\}}{1 + \exp\{\beta(y_{i0} - y_{i3})\}}$.
- General T: for the observed choice paths, the sufficient statistics picks out a (non-singleton) set of relevant choice paths, denoted $\mathbb{S}(\underset{\sim}{y}_i)$, such that

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid y_{i0}, \{\underset{\sim}{\mathbf{y}}_i \in \mathbb{S}(\underset{\sim}{\mathbf{y}}_i)\}, \boldsymbol{\eta}_i) = \frac{\exp\{\beta \sum_{t=1}^{T-1} y_{it} y_{i,t-1}\}}{\sum_{\mathbf{d}:\mathbf{d} \in \mathbb{S}(\underset{\sim}{\mathbf{y}}_i)} \exp\{\beta \sum_{t=1}^{T-1} d_t d_{t-1}\}}$$

Back to Sufficient Stat

Choice run matrix

There is a one-to-one mapping from the choice runs to a matrix $\mathbf{H}_i$ of dimension $(J+1)\times R_i$ and element

$$\mathbf{H}_i(j,r) = \begin{cases} n_i^{(r)} & \text{if } y_i^{(r)} = j \\ 0 & \text{if } y_i^{(r)} \neq j \end{cases}$$

An augmentation to record initial state d_{i1}, let $n_i^{(1)} = n_{i0} + d_{i1} - 1$. n_{i0} is the number of times y_{i0} is observed consecutively.

- Each column has exactly one entry $\neq 0$.
- In each row: Non-zero entry does not appear in neighbouring position.

Sufficiency through run matrix

The individual log-likelihood

$$
\begin{aligned}
\ln \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{x}_{i1}, \boldsymbol{\beta}, \boldsymbol{\theta}_i) = & \sum_{r=2}^{R_i} \beta_y(y_i^{(r)}, y_i^{(r-1)}) \\
& + \sum_{r=1}^{R_i} \beta_d^R(y_i^{(r)}, n_i^{(r)}) - 1\{d_{i1} \geq 2\} \sum_{d=1}^{d_{i1}-1} \beta_d(y_i^{(1)}, d) \\
& + \mathbf{T}_i' \boldsymbol{\alpha}_i + \sum_{r=1}^{R_i} v_i^R(y_i^{(r)}, n_i^{(r)}) - \sum_{d=1}^{d_{i1}} v_i(y_i^{(1)}, d) \\
& - \sum_{r=1}^{R_i} \sigma_i^R(y_i^{(r)}, n_i^{(r)}) + 1\{d_{i1} \geq 2\} \sum_{d=1}^{d_{i1}-1} \sigma_i(y_i^{(1)}, d) + \sigma_i(y_i^{(R_i)}, n_i^{(R_i)})
\end{aligned}
$$

- $\mathbf{T}_i$ records the total number of times each choice is visited $\Rightarrow$ **row sum of the run matrix $\mathbf{H}_i$**.
- $v_i^R(y, n) \equiv \sum_{d=1}^{n} v_i(y, d)$: Each run necessarily contributes $v_i(y_i^{(r)}, 1) \Rightarrow$ **# non-zero entry in each row of $\mathbf{H}_i$**
- Similar for $\sigma_i^R(y, n) \equiv \sum_{d=1}^{n} \sigma_i(y, d)$.
- **Goal**: Find another run matrix $\mathbf{H}'$ that keeps the same row sum and # non-zero entries in each row as $\mathbf{H}_i$.

Sufficient Statistics for $(\boldsymbol{\theta}_i, \beta_d)$

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i^2 = 1, \boldsymbol{\theta}_i, \boldsymbol{\beta}) = \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i^2 = 1, \beta_y) \quad \text{for} \quad \mathbf{S}_i^2 = 1\{\underset{\sim}{\mathbf{y}}_i \in \mathbb{S}^2(\underset{\sim}{\mathbf{y}}_i) \equiv \mathbb{S}_i^2\}$$

Necessary and sufficient conditions for $\mathbb{S}_i^2$: Provided $d^* > 1$, a sequence $\underset{\sim}{\mathbf{y}}_i' \in \mathbb{S}_i^2$ if and only if

1. Initial state and d^*-censored termination state matches: $\mathbf{x}_{i1}^{\mathbf{y}'} = \mathbf{x}_{i1}$; $\bar{\mathbf{x}}_{iT+1}^{\mathbf{y}'} = \bar{\mathbf{x}}_{iT+1}$.
2. The run matrix of $\underset{\sim}{\mathbf{y}}_i'$ has the same row sum as $\mathbf{H}_i$.
3. Total number of non-zero entries in each row matches.
4. The run matrix of $\underset{\sim}{\mathbf{y}}_i'$ is a result of any (or a combination) of the following matrix operations on $\mathbf{H}_i$.
   - (a) *Column Swapping*: say (r, r') columns and $r < r'$ WLOG.
     - (a.1) Columns in between: $1 < r < r' < R_i$
       - (a.1.1) $y_i^{(r)} = y_i^{(r')}$.
       - (a.1.2) $y_i^{(r)} \neq y_i^{(r)}$.
     - (a.2) First column and column in between: $1 = r < r' < R_i$ provided $y_i^{(r)} = y_i^{(r')}$ and $n_i^{(r')} \geq d_{i1}$.
   - (b) *Run length revision*: for any $1 \leq r < r' \leq R_i$, provided $y_i^{(r)} = y_i^{(r')}$ and $\min\{n_i^{(r)}, n_i^{(r')}\} \geq d^*$, revise $(n_i^{(r)}, n_i^{(r')})$ to any element in

$$\mathcal{R}_i^1(r, r') = \{(n^{(r)}, n^{(r')}) : n^{(r)} + n^{(r')} = n_i^{(r)} + n_i^{(r')}, \min\{n^{(r)}, n^{(r')}\} \geq d^*\}$$

Sufficient statistics for $(\boldsymbol{\theta}_i, \beta_y)$

$$\mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i^3 = 1, \boldsymbol{\theta}_i, \boldsymbol{\beta}) = \mathbb{P}(\underset{\sim}{\mathbf{y}}_i \mid \mathbf{S}_i^3 = 1, \beta_d) \quad \text{for} \quad \mathbf{S}_i^3 = 1\{\underset{\sim}{\mathbf{y}}_i \in \mathbb{S}^3(\underset{\sim}{\mathbf{y}}_i) \equiv \mathbb{S}_i^3\}$$

Necessary and sufficient conditions for $\mathbb{S}_i^3$: Provided $d^* > 1$, a sequence $\underset{\sim}{\mathbf{y}}_i' \in \mathbb{S}_i^3$ if and only if

1. Initial state and d^*-censored termination state matches: $\mathbf{x}_{i1}^{\mathbf{y}'} = \mathbf{x}_{i1}$; $\bar{\mathbf{x}}_{iT+1}^{\mathbf{y}'} = \bar{\mathbf{x}}_{iT+1}$.
2. The run matrix of $\underset{\sim}{\mathbf{y}}_i'$ has the same row sum as $\mathbf{H}_i$.
3. Total number of non-zero entries in each row matches.
4. The run matrix of $\underset{\sim}{\mathbf{y}}_i'$ is a result of any (or a combination) of the following matrix operations on $\mathbf{H}_i$.
   - (a) *Column Swapping*: say (r, r') columns and $r < r'$ WLOG.
     - (a.1.1) Columns in between: $1 < r < r' < R_i$ and $y_i^{(r)} = y_i^{(r')}$.
     - (a.2) First column and column in between: $1 = r < r' < R_i$ provided $y_i^{(r)} = y_i^{(r')}$ and $n_i^{(r')} \geq d_{i1}$.
   - (b) *Run length revision*: for any $1 \leq r < r' \leq R_i$, provided $y_i^{(r)} = y_i^{(r')}$ and $(n_i^{(r)}, n_i^{(r')})$ satisfy Condition E*, revise $(n_i^{(r)}, n_i^{(r')})$ to any element in

$$\mathcal{R}_i^3(r, r') = \{(n^{(r)}, n^{(r')}) : n^{(r)} + n^{(r')} = n_i^{(r)} + n_i^{(r')}, (n^{(r)}, n^{(r')}) \text{ satistifies Condition E*}\}$$

Condition E*: (1) $\min\{n^{(r)}, n^{(r')}\} \geq d^* - 1$; (2) $n^{(r)} + n^{(r')} \geq 2d^*$; (3) $n^{(r')} \geq d^*$ if $\max\{r, r'\} = R_i$.

When $d^* = 1$

- Only first year of experience matters, no learning afterwards. $d^* = 1$ and $\beta_d(y, 1) \neq 0$.
- From a likelihood point of view, state variable d_{it} always take value 1, so no need to track it.
- If the only state variable is lagged choice, then we can let $\tilde{\alpha}_i(y) = \alpha_i(y) + v_i(y)$.
- $\mathbb{P}(y \mid \boldsymbol{\theta}_i, y_{it-1}) = \dfrac{\exp\{\tilde{\alpha}_i(y) + \beta(y, y_{it-1})\}}{\sum_{j \in \mathcal{Y}} \exp\{\tilde{\alpha}_i(j) + \beta(j, y_{it-1})\}}$.
- In the binary choice case:
  - No duration dependence for "0" choice: $\beta_d(0, 1) = 0$
  - Duration dependence for "1" choice: $\beta_d(1, 1) = \beta(1, 1) \neq 0$.
  - $y_{it} = 1\{\gamma y_{it-1} + \check{\alpha}_i + \epsilon_{it} \geq 0\}$.
  - $\check{\alpha}_i = \tilde{\alpha}_i(1) - \tilde{\alpha}_i(0) + \beta(1, 0)$.
  - $\gamma = -\beta(1, 0) - \beta(0, 1) + \beta(1, 1)$.

Additional restriction on β

- In some applications, it is plausible to assume **no duration dependence in outside alternative**: $\beta_d(0, d) = 0$ for all $d \geq 1$.
- This does not change set $\mathcal{M}_i^3$.
- This may **enlarge** $\mathcal{M}_i^2 \setminus \mathcal{M}_i^1$.
- Revisit trinomial example: $A = \{0, 1, 2\}$ and $B = \{1, 0, 2\}$ with $\mathbf{x}_{i1} = (0, 1)$.

$$\mathbf{x}^A = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

$$\mathbf{x}^B = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

- Associated run matrix

$$\mathbf{H}_i^A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{H}_i^B = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- The column with non-zero entries for **first row** is re-distributed. # non-zero entry in first row is different
- Column sum and # non-zero entries in **other** rows stay the same.
- $\ln \dfrac{\mathbb{P}(A \mid \boldsymbol{\theta}_i, \mathbf{x}_{i1})}{\mathbb{P}(B \mid \boldsymbol{\theta}_i, \mathbf{x}_{i1})}$ identifies $\delta_y(1, 2)$.

Trinomial Example (identification of β_y)

- $\mathcal{Y} = \{0, 1, 2\}$, $T = 3$.
- $A = \{1, 2, 0\}$ vs. $B = \{2, 1, 0\}$ and $\mathbf{x}_{i1} = (0, d_{i1})$.

$$\mathbf{x}^A = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathbf{x}^B = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

- $\ln \dfrac{\mathbb{P}(A|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}{\mathbb{P}(B|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}$ identifies $\Delta_y(0 \to 0; 1, 2)$

$$\Delta_y(0 \to 0; 1, 2) = \{\beta_y(1,0) + \beta_y(2,1) + \beta_y(0,2)\} - \{\beta_y(2,0) + \beta_y(1,2) + \beta_y(0,1)\}$$

Trinomial Example (identification of β_y)

- $\mathcal{Y} = \{0, 1, 2\}$, $T = 3$.
- $A = \{1, 2, 0\}$ vs. $B = \{2, 1, 0\}$ and $\mathbf{x}_{i1} = (0, d_{i1})$.

$$\mathbf{x}^A = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathbf{x}^B = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

- $\ln \dfrac{\mathbb{P}(A|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}{\mathbb{P}(B|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}$ identifies $\Delta_y(0 \to 0; 1, 2)$

$$\Delta_y(0 \to 0; 1, 2) = \{\beta_y(1,0) + \beta_y(2,1) + \beta_y(0,2)\} - \{\beta_y(2,0) + \beta_y(1,2) + \beta_y(0,1)\}$$

- Similarly $\Delta_y(1 \to 1; 2, 0)$ and $\Delta_y(2 \to 2; 1, 0)$ can be identified.

$$\Delta_y(1 \to 1; 2, 0) = \{\beta_y(2,1) + \beta_y(0,2) + \beta_y(1,0)\} - \{\beta_y(0,1) + \beta_y(2,0) + \beta_y(1,2)\}$$
$$\Delta_y(2 \to 2; 1, 0) = \{\beta_y(1,2) + \beta_y(0,1) + \beta_y(2,0)\} - \{\beta_y(0,2) + \beta_y(1,0) + \beta_y(2,1)\}$$

Trinomial Example (identification of β_y)

- $\mathcal{Y} = \{0,1,2\}$, $T = 3$.
- $A = \{1,2,0\}$ vs. $B = \{2,1,0\}$ and $\mathbf{x}_{i1} = (0, d_{i1})$.

$$\mathbf{x}^A = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathbf{x}^B = \left\{ \begin{pmatrix} 0 \\ d_{i1} \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

- $\ln \dfrac{\mathbb{P}(A|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}{\mathbb{P}(B|\boldsymbol{\theta}_i, \mathbf{x}_{i1})}$ identifies $\Delta_y(0 \to 0; 1, 2)$

$$\Delta_y(0 \to 0; 1,2) = \{\beta_y(1,0) + \beta_y(2,1) + \beta_y(0,2)\} - \{\beta_y(2,0) + \beta_y(1,2) + \beta_y(0,1)\}$$

- Similarly $\Delta_y(1 \to 1; 2, 0)$ and $\Delta_y(2 \to 2; 1, 0)$ can be identified.

$$\Delta_y(1 \to 1; 2,0) = \{\beta_y(2,1) + \beta_y(0,2) + \beta_y(1,0)\} - \{\beta_y(0,1) + \beta_y(2,0) + \beta_y(1,2)\}$$
$$\Delta_y(2 \to 2; 1,0) = \{\beta_y(1,2) + \beta_y(0,1) + \beta_y(2,0)\} - \{\beta_y(0,2) + \beta_y(1,0) + \beta_y(2,1)\}$$

- Let $\delta_y(1,2) \equiv \beta_y(1,2) - \beta_y(1,0) - \beta_y(0,2)$ and $\delta_y(2,1) \equiv \beta_y(2,1) - \beta_y(2,0) - \beta_y(0,1)$, we can identify $\delta_y(1,2) - \delta_y(2,1)$.
- If we further assume **no duration dependence for "0" choice**, then $\Delta_y(0 \to 2; 0, 1) = \delta_y(2,1)$ and $\Delta_y(0 \to 1; 0, 2) = \delta_y(1,2)$ are identified.
- Interpretation of $\delta_y(1,2)$

Trinomial Example (identification of β_d)

- Suppose $T \geq d^* + 2$. Consider initial condition $\mathbf{x}_1 = (y, d_1)$ for $y \in \{0, 1, 2\}$.
- Let $y' \neq y$ and let

$$A = \{y', y \times 1_{d^*+1}\}$$
$$B = \{y, y', y \times 1_{d^*}\}$$

- If $d_1 \geq d^* - 1$

$$\ln \frac{\mathbb{P}(A \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i)}{\mathbb{P}(B \mid \mathbf{x}_{i1}, \boldsymbol{\theta}_i)} = \beta_d(y, d_1 + 1) - \beta_d(y, d_1)$$

- When $d_1 = d^* - 1$, parameter $\beta_d(y, d^*) - \beta_d(y, d^* - 1)$ is identified for all $y \in \{0, 1, 2\}$.

$$\mathbf{x}^A = \begin{pmatrix} y \\ d^* - 1 \end{pmatrix} \boxed{\begin{pmatrix} y' \\ 1 \end{pmatrix} \begin{pmatrix} y \\ 1 \end{pmatrix} \begin{pmatrix} y \\ 2 \end{pmatrix} \cdots \begin{pmatrix} y \\ d^* \end{pmatrix}} \begin{pmatrix} y \\ d^* \end{pmatrix}$$

$$\mathbf{x}^B = \begin{pmatrix} y \\ d^* - 1 \end{pmatrix} \boxed{\begin{pmatrix} y \\ d^* \end{pmatrix} \begin{pmatrix} y' \\ 1 \end{pmatrix} \begin{pmatrix} y \\ 1 \end{pmatrix} \cdots \begin{pmatrix} y \\ d^* - 1 \end{pmatrix}} \begin{pmatrix} y \\ d^* \end{pmatrix}$$

Connection to Binary choice DDC with more than one lag

- Binary choice DDC with two lages (Chamberlain 1985)
  $y_{it} = 1\{\alpha_i + \gamma_{i1} y_{it-1} + \gamma_2 y_{it-2} + \epsilon_{it} \geq 0\}$
- Logit: $\mathbb{P}(y_{it} = 1 \mid y_{it-1}, y_{it-2}) = \dfrac{\exp\{\alpha_i + \gamma_{i1} y_{it-1} + \gamma_2 y_{it-2}\}}{1 + \exp\{\alpha_i + \gamma_{i1} y_{it-1} + \gamma_2 y_{it-2}\}}$
- Test for duration dependence: $H_0 : \gamma_2 = 0$
- Use pair of sequences $\{y_{i0}, \ldots, y_{i5}\}$: $A = \{1, 0, 1, 0, 0, 0\}$ and $B = \{1, 0, 0, 1, 0, 0\}$. Suppose $d_{i1} = d \geq 1$.

$$\mathbf{H}^A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ d & 1 & 1 & 3 \end{pmatrix} \quad \mathbf{H}^B = \begin{pmatrix} 1 & 0 & 1 & 0 \\ d & 2 & 1 & 2 \end{pmatrix}$$

- Identification for β_d when $d^* = 2$.
- Test for $\gamma_2 = 0$ is equivalent to test $\beta_d(0, d^*) - \beta_d(0, d^* - 1) = 0$ if $d^* = 2$.
- If $d^* > 2$, we may reject $H_0 : \gamma_2 = 0$ due to forward looking behaviour.
- If $d^* = 1$ and $\beta_d(1, 1) \neq 0$, but we will always get $\gamma_2 = 0$.