Sufficient Statistics for Unobserved Heterogeneity in Dynamic Discrete Choice Models

Victor Aguirregabiria and Jiaying Gu

University of Toronto
ECO 2403. Topics in Econometrics

Febuary 16, 2018

Overview

1) Conditional maximum likelihood estimator for structural parameters:

- One way to deal with incidental parameter problem in linear or non-linear panel data models is to use the conditional maximum likelihood method.
- Review of Sufficient Statistics with example on static binary choice panel data model.
- Conditional likelihood approach may be a lucky coincidence (Chamberlain (2010): only Logit model for discrete choice!).
- CMLE for dynamic (lagged choice dependence) discrete choice model without forward looking (Binary choice and multinomial choice).
- CMLE for discrete choice models accounting for duration dependence. (Chamberlain 1985, Frederiksen et al, 2007).

Overview

2) Partial identification results:

- Honoré and Tamer (2006): initial condition problem: $p(y_{i0}; \boldsymbol{x}_i, \omega_i)$ is not nonparametrically identified $\rightarrow$ characterize bounds.
- Chernozhukov et al (2003): Even in the logit model where we have point identification through CMLE for the structural parameters, marginal effects are not point identified unless stronger assumptions imposed.

3) Bias correction approach: look for estimators that have smaller biases as opposed to no bias at all.

- Modified conditional likelihood approach: Cox and Reid (1987).

4) Efficiency perspective: do we lose any information in using the conditional maximum likelihood approach?

$$P(\boldsymbol{y}_i; \beta, \omega_i) = \underbrace{P(\boldsymbol{y}_i | S(\boldsymbol{y}_i), \boldsymbol{\beta})}_{\text{conditional likelihood}} P(S(\boldsymbol{y}_i); \boldsymbol{\beta}, \omega_i).$$

$$P(\boldsymbol{y}_i; \beta) = \int P(\boldsymbol{y}_i; \beta, \omega) dF(\omega) = P(\boldsymbol{y}_i | S(\boldsymbol{y}_i), \boldsymbol{\beta}) \int P(S(\boldsymbol{y}_i); \boldsymbol{\beta}, \omega) dF(\omega)$$

- What do we think of the full MLE $(\hat{\beta}, \hat{F}(\omega))$ versus $\hat{\beta}_c = \text{argmax} \sum_i \log P(\boldsymbol{y}_i | S(\boldsymbol{y}_i), \boldsymbol{\beta})$.
- Is $\hat{\beta}$ asymptotically equivalent to $\hat{\beta}_c$? (i.e. is $\hat{\beta}_c$ is semiparametrically efficient?).

Sufficient Statistics

- Definition: Let $\{\mathcal{Y}, \mathcal{P} = \{P_\theta, \theta \in \Theta\}\}$ be a parametric model, the statistic S is sufficient for θ if the conditional distribution of Y given S does not depend on θ.
- Factorization Theorem: A necessary and sufficient condition for S to be a sufficient statistic is that the density factorizes

$$f(y;\theta) = \lambda(y)\psi(S(y);\theta) \; = f_1(y \mid S(Y) = S(y)) \; f_2(S(y), \theta)$$

- Sufficient statistic is not unique, consider iid sample $Y_1, Y_2, \ldots, Y_n \sim N(0, \sigma^2)$.
  - $S_1(Y) = (Y_1, Y_2, \ldots, Y_n)$
  - $S_2(Y) = (Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$
  - $S_3(Y) = (Y_1^2 + Y_2^2 + \cdots + Y_m^2, Y_{m+1}^2 + \cdots + Y_n^2)$
  - $S_4(Y) = \sum_i Y_i^2$ (minimum sufficient)
- Partial sufficiency: suppose $\theta = (\alpha, \beta) \in A \times B$, two cases of factorization for S being sufficient for α (nuisance):
  1. $f(y;\theta) = f_1(y \mid S(Y) = S(y), \beta) \; f_2(S(y); \alpha)$: α and β are likelihood orthogonal
  2. $f(y;\theta) = f_1(y \mid S(Y) = S(y), \beta) \; f_2(S(y); \alpha, \beta)$

  1. using f_1 for estimation of β leads to fully efficient estimator.
  2. using f_1 for estimation of β may or may not be fully efficient (in what sense?)

Example: static binary choice panel data model

- Model: $y_{it} = 1\{\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \omega_i + \epsilon_{it} > 0\}$, where $\epsilon_{it}|\omega_i, x_{it}$ are iid, $P(\epsilon_{it} \leq a) = \frac{\exp(a)}{1+\exp(a)}$
- Earliest example (Rasch 1961): Each individual i is given T items for test, $y_{it} = \{0, 1\}$ (incorrect, correct), x_{it} is dummy of item t, hence $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)$ measures item difficulty, we have
$$P(y_{it} = 1; \omega_i, \beta_t) = \frac{\exp(\omega_i - \beta_t)}{1 + \exp(\omega_i - \beta_t)}.$$
- Individual Likelihood for $\boldsymbol{y}_i = \{y_{i1}, \ldots, y_{iT}\}$:
$L_i(\boldsymbol{\beta}, \omega_i) = P(\boldsymbol{y}_i; \boldsymbol{\beta}, \omega_i) = \prod_{t=1}^{T} f(y_{it}; \boldsymbol{\beta}, \omega_i) = \frac{\exp(\omega_i \sum_t y_{it} - \sum_t y_{it}\beta_t)}{\prod_{t=1}^{T}(1+\exp(\omega_i - \beta_t))}$.
- Minimum Sufficient statistics for ω_i is: $S(\boldsymbol{y}_i) = \sum_t y_{it}$ (total score of individual).
- Factorization: $L_i(\boldsymbol{\beta}, \omega_i) = P(\boldsymbol{y}_i|S(\boldsymbol{y}_i), \boldsymbol{\beta}, \cancel{\omega_i})P(S_i; \boldsymbol{\beta}, \omega_i)$
- Conditional maximum likelihood estimator $\hat{\boldsymbol{\beta}}_c := \underset{\boldsymbol{\beta} \in B}{\operatorname{argmax}} \sum_i \log P(\boldsymbol{y}_i|S(\boldsymbol{y}_i), \boldsymbol{\beta})$.
- Remark: If S is not minimum sufficient, it could contain information about $\boldsymbol{\beta}$.

Example: static binary choice panel data model

- Suppose $T = 2$, and $\beta_2 = 0$, so $\beta_1 \equiv \beta$ is difficulty level of test item 1 relative to test item 2.
- $S(y_{i1}, y_{i2}) = \{0, 1, 2\}$.
  1. $P(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 0; \beta, \omega_i) = 1$
  2. $P(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 1; \beta, \omega_i) = 1\{y_{i1} = 1, y_{i2} = 0\} \frac{\exp(-\beta)}{1+\exp(-\beta)} + 1\{y_{i1} = 0, y_{i2} = 1\} \frac{1}{1+\exp(-\beta)}$
  3. $P(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 2; \beta, \omega_i) = 1$
- Conditional MLE: $\hat{\boldsymbol{\beta}}_c = \underset{\boldsymbol{\beta} \in B}{\text{argmax}} \sum_{i \in I} 1\{y_{i1} = 1, y_{i2} = 0\} \ln \frac{1}{1+\exp(\beta)} + 1\{y_{i1} = 0, y_{i2} = 1\} \ln \frac{\exp(\beta)}{1+\exp(\beta)} = \log \frac{\#\{0,1\}}{\#\{1,0\}}$. where $I = \{i : y_{i1} + y_{i2} = 1\}$.
- If $\#\{0,1\} > \#\{1,0\}$, then $\hat{\boldsymbol{\beta}}_c > 0$ (test item 1 more difficult).
- CMLE $\hat{\boldsymbol{\beta}}_c$ is root-n consistent under suitable regularity conditions [Anderson (1970)].
  - Usual MLE regularity conditions.
  - ω_i could not take a sequence of too extreme values such that $P(y_{i1} + y_{i2} = 1)$ vanishes as $n \to \infty$.

Example: static binary choice panel data model

- Inference can be conducted in the usual way, asymptotic variance is inverse of Fisher information.
- $L = \sum_{i \in I} w_i \ln \frac{\exp(\beta)}{1+\exp(\beta)} + (1 - w_i) \ln \frac{1}{1+\exp(\beta)}$ with $w_i = 1\{y_{i1} = 0, y_{i2} = 1\}$.
- $\frac{\partial^2 L}{\partial \beta^2} = -\sum_{i \in I} \frac{\exp(\beta)}{(1+\exp(\beta))^2}$ and $\mathbb{E}[-\frac{\partial^2 L}{\partial \beta^2}] = \sum_i P_i \frac{\exp(\beta)}{(1+\exp(\beta))^2}$ with $P_i = \mathbb{E}[1\{y_{i1} + y_{i2} = 1\} | \omega_i, \beta]$.
- Since we do not have a consistent estimator for ω_i, it seems hard to construct consistent estimator for the asymptotic variance.
- However, $\frac{1}{N}\left[-\frac{\partial^2 L}{\partial \beta^2}\right] = \frac{\exp(\beta)}{(1+\exp(\beta))^2} \frac{1}{N} \sum_i 1\{y_{i1} + y_{i2} = 1\}$ converges to $\frac{\exp(\beta)}{(1+\exp(\beta))^2} P_i$ almost surely by LLN.

Identification

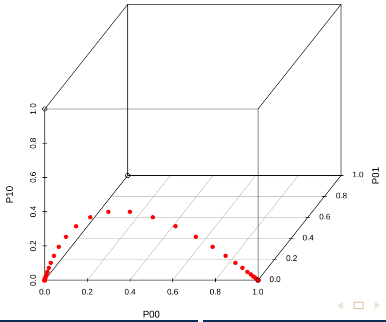
- Not all discrete choice model obtains the root-n consistent CMLE.
- For static binary choice model, Chamberlain (2010) shows that only the logistic model has identification.
- Consider model $P(y_{it} = 1|x_i, \omega_i) = F(\omega_i + x_{it}\beta_0)$, for given x, β, ω_i, the parametric model represents a vector in a $K-1$ ($K = 2^T$) dimensional space, if $T = 2$, it is a 3 dimensional unit simplex.

$$p(x, \beta, \omega) = \begin{pmatrix} P(0,0|x,\omega) \\ P(0,1|x,\omega) \\ P(1,0|x,\omega) \end{pmatrix}$$

- Identification: β_0 is identified if there doesn't exist $\beta' \neq \beta_0$ along with distribution $F'(\omega|x)$ and $F_0(\omega|x)$, such that $\int p(x, \beta_0, \omega) dF_0(\omega|x) = \int p(x, \beta', \omega) dF'(\omega|x)$.

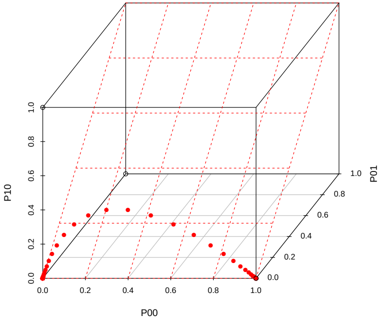
Identification: geometry

- Fix x and β, as ω changes, the model $p(x, \beta, \omega)$ defines a set of points in the 3-d unit simplex
$$C_{x\beta} = \{q \in \mathbb{S}^3 : \exists \omega \text{ such that } q = p(x, \beta, \omega)\}$$
- If F is continuous, then $C_{x\beta}$ traces out a continuous curve in the 3-d unit simplex.
- The observed data frequency $p(x, \beta) = \int p(x, \beta, \omega) dF(\omega|x)$ lives in the convex hull of $H_{x\beta} \equiv coC_{x\beta}$.
- Identification fails if the observed data frequency lies in two different convex hull $H_{x\beta_0}$ and $H_{x\beta'}$.
- Plot the vector $p(x, \beta, \omega)$ with $x_1 = 2, x_2 = 3, \beta = 0, \omega \in [-20, 20]$.



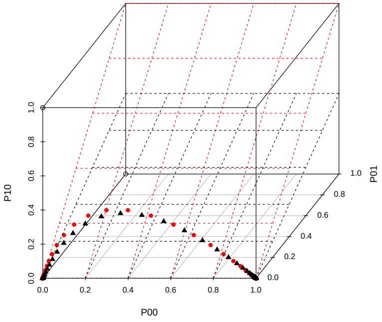
Identification: geometry

- Chamberlain shows that a necessary condition that we have identification of β is when $C_{x\beta}$ lies on a hyperplane of dimension $K-2$. (i.e. $\forall\omega$, $\sum_{j=1}^{K-1} c_j p_j(x, \beta_0, \omega) = c_0$).
- This can only be true if F is the logistic CDF.
- Below I am plotting the regression plane of regressing $P10$ against $P00$ and $P01$, residual is zero, perfect fit!



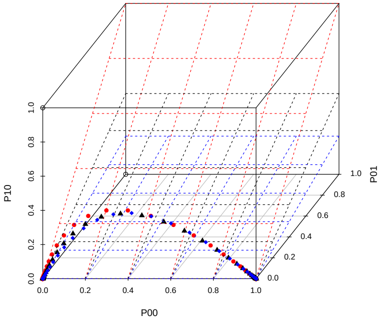
Identification: geometry

- Change β to 0.75 (black).



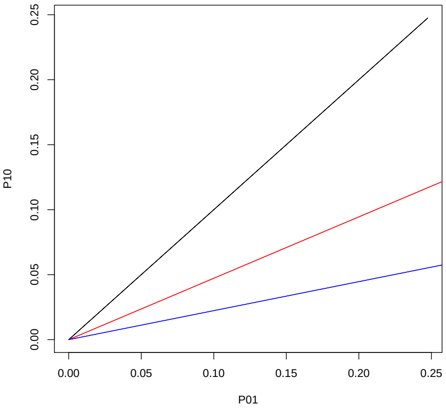
Identification: geometry

- Change β to 1.5 (blue).



Identification: geometry

- For Logit model, in fact $\frac{P10}{P01} = \exp((x_{i1} - x_{i2})\beta)$.

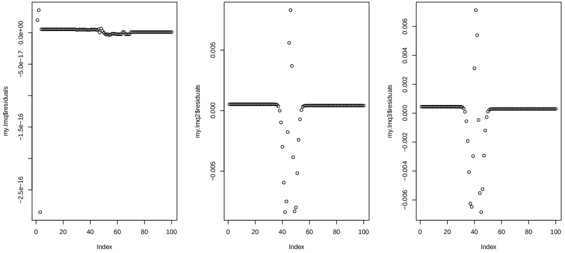


Identification failure: Probit model

- Consider $P(y_{it} = 1|x_i, \omega_i) = \Phi(\omega_i + x_{it}\beta)$: probit model.
- We also have

$$p(x, \beta, \omega) = \begin{pmatrix} P(0,0|x,\omega) \\ P(0,1|x,\omega) \\ P(1,0|x,\omega) \end{pmatrix}$$

- Plot the regression residual of $P10$ regress on $P00$ and $P01$, with $x_{i1} = 2, x_{i2} = 3$
- Left: $\beta = 0$, Middle: $\beta = 0.75$, Right: $\beta = 1.5$.



Generalize this idea: Bonhomme (2011)

- The idea is to find a moment condition for β that does not depend on ω_i.
- Let $[L_{\beta,x}F](y) = \int p(x,\beta,\omega)dF(\omega)$: maps distribution function F to the data frequency.
- Suppose we can find a function $\varphi(\cdot,x,\beta)$ such that for every $F(\omega)$:

$$\int_{\mathcal{Y}} \varphi(y,x,\beta)[L_{\beta,x}F](y)dy = 0$$

- Then we have the moment condition which doesnt depend on ω: $\mathbb{E}[\varphi(y_i,x_i,\beta_0)|x_i = x] = \int_{\mathcal{Y}} \varphi(y,x,\beta_0)[L_{\beta_0,x}F_0](y)dy = 0$.
- In the Chamberlain example, it boils down to find for all x and ω:

$$\sum_{y\in\{0,1\}^T} \varphi(y,x,\beta)P(\boldsymbol{y}_i = y|x,\omega,\beta) = 0.$$

- which leads to

$$\sum_{y\in\{0,1\}^T} 1\{\sum_t y_t = s\}\varphi(y,x,\beta)\exp(\sum_t y_t x_t \beta) = 0$$

  if T=2, the solution for φ is

$$\varphi(\boldsymbol{y} = (0,0),x,\beta) = \varphi(\boldsymbol{y} = (1,1),x,\beta) = 0$$
$$\varphi(\boldsymbol{y} = (1,0),x,\beta)\exp(x_{i1}\beta) + \varphi(\boldsymbol{y} = (0,1),x,\beta)\exp(x_{i2}\beta) = 0$$

Dynamic binary choice model

Econometrics Model: $y_{it} = 1\{\theta y_{i,t-1} + \omega_i + v_{it} \geq 0\}$

- Recall the example of a firm's entry-exit (1-0) decision without forward looking behavior (myopic).
- Endogenous state variable $x_{it} = y_{i,t-1}$.
- $U_{it}(y) = u(y, x_{it}, \omega_i) + \epsilon_{it}(y) = \omega_i(y) + \beta(y, y_{it-1}) + \epsilon_{it}(y)$, with unobservable $\epsilon_{it}(0), \epsilon_{it}(1)$ iid over (i,t),
- Dynamics is captured by $\beta(y, y_{it-1}) = 1\{y \neq y_{it-1}\}\beta(y, y_{it-1})$.
- Observed decision $y_{it} = 1\{\omega_i(1) - \omega_i(0) + \beta(1, x_{it}) - \beta(0, x_{it}) + \epsilon_{it}(1) - \epsilon_{it}(0) \geq 0\}$,
- Since $\beta(1, y_{it-1}) - \beta(0, y_{it-1}) = \beta(1,0)(1 - y_{it-1}) - \beta(0,1)y_{it-1} = \beta(1,0) + y_{it-1}(-\beta(1,0) - \beta(0,1))$.
- Let $\omega_i \equiv \omega_i(1) - \omega_i(0) + \beta(1,0)$ and $v_{it} \equiv \epsilon_{it}(1) - \epsilon_{it}(0)$ and $\theta = -\beta(1,0) - \beta(0,1)$ gives the econometrics model, and θ is the sunk cost of a firm (cost of entry and cost of exit).
- Chamberlain (1985) shows the CMLE of θ is root-n consistent.

Chamberlain (1985): CMLE for dynamic binary choice logit model

- Suppose $\epsilon_{it}(y)$ follows extreme value Type I distribution, and suppose $T = 3$, then
- Individual likelihood for $\boldsymbol{y}_i = \{y_{i0}, y_{i1}, y_{i2}, y_{i3}\}$:

$$
\begin{aligned}
& L_i(\theta, \omega_i) \\
&= P(y_{i0}|\theta, \omega_i) \frac{e^{(y_{i1}+y_{i2}+y_{i3})\omega_i} e^{\theta \sum_{t=1}^{3} y_{it} y_{it-1}}}{(1+e^{(\theta y_{i0}+\omega_i)})(1+e^{(\theta y_{i1}+\omega_i)})(1+e^{(\theta y_{i2}+\omega_i)})} \\
&= \frac{P(y_{i0}|\theta, \omega_i)}{(1+e^{\theta y_{i0}+\omega_i})} \\
&\times \frac{e^{(y_{i1}+y_{i2})\omega_i}}{((1+e^{\theta+\omega_i})^2)^{1\{y_{i1}+y_{i2}=2\}}((1+e^{\theta+\omega_i})(1+e^{\omega_i}))^{1\{y_{i1}+y_{i2}=1\}}((1+e^{\omega_i})^2)^{1\{y_{i1}+y_{i2}=0\}}} \\
&\times e^{y_{i3}\omega_i} \\
&\times e^{\theta \sum_{t=1}^{3} y_{it} y_{it-1}}
\end{aligned}
$$

- Minimum Sufficient Statistics for ω_i: $S(\boldsymbol{y}_i) = \{y_{i0}, y_{i1} + y_{i2}, y_{i3}\}$.
  1. $P(\boldsymbol{y}_i | y_{i0}, y_{i1} + y_{i2} = 0, y_{i3}; \theta, \omega_i) = 1.$
  2. $P(\boldsymbol{y}_i | y_{i0}, y_{i1} + y_{i2} = 2, y_{i3}; \theta, \omega_i) = 1.$
  3. $P(\boldsymbol{y}_i | y_{i0}, y_{i1} + y_{i2} = 1, y_{i3}; \theta, \omega_i) = 1\{y_{i1} = 1, y_{i2} = 0\} \frac{\exp(\theta(y_{i0}-y_{i3}))}{1+\exp(\theta(y_{i0}-y_{i3}))} + 1\{y_{i1} = 0, y_{i2} = 1\} \frac{1}{1+\exp(\theta(y_{i0}-y_{i3}))}$
- CMLE $\hat{\beta}_c = \log\left(\frac{\#\{1,1,0,0\}+\#\{0,0,1,1\}}{\#\{0,1,0,1\}+\#\{1,0,1,0\}}\right)$ (Minimum T =3).

General T: CMLE for dynamic binary choice logit model

- For general $T \geq 3$, minimum sufficient statistic $S(\boldsymbol{y}_i) = \{y_{i0}, \sum_{t=1}^{T-1} y_{it}, y_{iT}\}$.
- Conditional likelihood function: $P(\boldsymbol{y}_i \mid S_i, \omega_i, \theta) = \frac{\exp(\theta \sum_{t=1}^{T} y_{it} y_{it-1})}{\sum_{d \in B_i} \exp(\theta \sum_{t=1}^{T} d_t d_{t-1})}$ where

$$B_i = \left\{ (d_0, d_1, \ldots, d_T) \in \{0,1\}^{T+1} : d_0 = y_{i0}, d_T = y_{iT}, \sum_{t=1}^{T-1} d_t = \sum_{t=1}^{T-1} y_{it} y_{it-1} \right\}$$

- Disadvantage: no other observables other than y_{it-1}.
- Honoré and Kyriazidou (2000) considers $y_{it} = 1\{\theta y_{it-1} + x_{it}\beta + \omega_i + v_{it} \geq 0\}$ where x_{it} are exogenous random variables.

Honoré and Kyriazidou (2000), T = 3

- Individual likelihood for $\boldsymbol{y}_i = \{y_{i0}, y_{i1}, y_{i2}, y_{i3}\}$ given $\boldsymbol{x}_i = \{x_{i1}, \dots, x_{iT}\}$:

  $L_i(\theta, \omega_i)$

  $$= P(y_{i0}|\boldsymbol{x}_i, \theta, \omega_i) \frac{e^{(y_{i1}+y_{i2}+y_{i3})\omega_i} e^{\theta \sum_{t=1}^{3} y_{it} y_{it-1} + \beta \sum_t y_{it} x_{it}}}{(1 + e^{(\theta y_{i0} + x_{i1}\beta + \omega_i)})(1 + e^{(\theta y_{i1} + x_{i2}\beta + \omega_i)})(1 + e^{(\theta y_{i2} + x_{i3}\beta + \omega_i)})}$$

  $$\overset{x_{i2}=x_{i3}}{=} \frac{P(y_{i0}|\boldsymbol{x}_i, \theta, \omega_i)}{(1 + e^{\theta y_{i0} + x_{i1}\beta + \omega_i})}$$

  $$\times \frac{e^{(y_{i1}+y_{i2})\omega_i}}{((1 + e^{\theta + x_{i2}\beta + \omega_i})^2)^{1\{y_{i1}+y_{i2}=2\}} ((1 + e^{\theta + x_{i2}\beta + \omega_i})(1 + e^{\omega_i}))^{1\{y_{i1}+y_{i2}=1\}} ((1 + e^{\omega_i})^2)^{1\{y_{i1}+y_{i2}=}}$$

  $$\times e^{y_{i3}\omega_i}$$

  $$\times e^{\theta \sum_{t=1}^{3} y_{it} y_{it-1} + \beta \sum_t y_{it} x_{it}}$$

- Minimum Sufficient Statistics is still $S_i = \{y_{i0}, y_{i1} + y_{i2}, y_{iT}\}$.

Honoré and Kyriazidou (2000), T = 3

- Conditional log Likelihood:
$\sum_i 1\{y_{i1} + y_{i2} = 1\} 1\{x_{i2} - x_{i3} = 0\} \log\left(\frac{\exp((x_{i1}-x_{i2})\beta+\theta(y_{i0}-y_{i3}))^{y_{i1}}}{1+\exp((x_{i1}-x_{i2})\beta+\theta(y_{i0}-y_{i3}))}\right)$
- If all x_{it} are discrete random variable and $P(x_{i2} = x_{i3}) > 0$, then the CMLE $(\hat{\beta}_c, \hat{\theta}_c)$ is root-n consistent.
- If x_{it} are continuous random variables, modify the conditional log likelihood with a kernel weights.

$$\sum_i 1\{y_{i1} + y_{i2} = 1\} K\left(\frac{x_{i2} - x_{i3}}{h_n}\right) \log\left(\frac{\exp((x_{i1} - x_{i2})\beta + \theta(y_{i0} - y_{i3}))^{y_{i1}}}{1 + \exp((x_{i1} - x_{i2})\beta + \theta(y_{i0} - y_{i3}))}\right)$$

- But apparently, age is not allowed as a regressor using this strategy.
- Kernel bandwidth $h_n \to 0$ as $n \to \infty$ and $K(v) \to 0$ as $\|v\| \to \infty$.
- The resulting kernel type CMLE converges to Normal at the rate $\sqrt{nh_n^k}$ where k is column dimension of $\boldsymbol{x}$.

Honoré and Kyriazidou (2000), General $T \geq 3$

- For general T, in order for $\{y_{i0}, \sum_{t=1}^{T-1} y_{it}, y_{iT}\}$ to be minimum sufficient statistics, we would have to require $x_{i2} = x_{i3} = \ldots x_{iT}$, this implies the rate of convergence be $\sqrt{nh_n^{(T-2)k}}$ which is too slow!
- Honoré and Kyriazidou (2000) suggests a pair-wise approach to maintain $\sqrt{nh_n^k}$ rate: identification of (θ, β) is based on sequences of histories such that $y_{is} + y_{it} = 1$ for $1 \leq t < s \leq T-1$.
- Consider
  - $A = \{y_{i0} = d_0, \ldots, y_{it-1} = d_{t-1}, y_{it} = 1, y_{it+1} = d_{t+1}, \ldots, y_{is-1} = d_{s-1}, y_{is} = 0, y_{is+1} = d_{s+1}, \ldots, y_{iT} = d_T\}$.
  - $B = \{y_{i0} = d_0, \ldots, y_{it-1} = d_{t-1}, y_{it} = 0, y_{it+1} = d_{t+1}, \ldots, y_{is-1} = d_{s-1}, y_{is} = 1, y_{is+1} = d_{s+1}, \ldots, y_{iT} = d_T\}$
- $P(A|\mathbf{x}_i, \omega_i, A \cup B, x_{it+1} = x_{is+1}) = \frac{\exp\left((x_{it}-x_{is})\beta+\theta(d_{t-1}-d_{s+1})+\theta(d_{t+1}-d_{s-1})1\{s-t>1\}\right)}{1+\exp\left((x_{it}-x_{is})\beta+\theta(d_{t-1}-d_{s+1})+\theta(d_{t+1}-d_{s-1})1\{s-t>1\}\right)}$

  independent of ω_i
- Conditional likelihood:

$$\sum_i \sum_{1 \leq t < s \leq T-1} 1\{y_{is} + y_{it} = 1\} K\left(\frac{x_{it+1} - x_{is+1}}{h_n}\right)$$

$$\times \log\left(\frac{\exp\left((x_{it} - x_{is})\beta + \theta(d_{t-1} - d_{s+1}) + \theta(d_{t+1} - d_{s-1})1\{s - t > 1\}\right)^{y_{it}}}{1 + \exp\left((x_{it} - x_{is})\beta + \theta(d_{t-1} - d_{s+1}) + \theta(d_{t+1} - d_{s-1})1\{s - t > 1\}\right)}\right)$$

Extension to multinomial choice with dynamics

Only covariate: y_{it-1}

- $P(y_{it}=k|y_{it-1}=j,\omega_i,\theta)=\frac{\exp(\omega_{ik}+\theta_{jk})}{\sum_{h=0}^{J}\exp(\omega_{ih}+\theta_{jh})}$.
- Magnac (2000) shows that minimum sufficient statistics for $\boldsymbol{\omega}_i$ is $S_i=\{y_{i0},y_{iT},\sum_{t=1}^{T-1}1\{y_{it}=k\},\forall k\}$ (initial and termination state, and numbers of occurances of all states during period 1 to $T-1$.)
- Conditional likelihood function: $P(\boldsymbol{y}_i|S_i,\boldsymbol{\omega}_i,\boldsymbol{\theta})=\frac{\exp\left(\sum_k\sum_j\left(\sum_{t=1}^{T}1\{y_{it}=k\}1\{y_{it-1}=j\}\theta_{jk}\right)\right)}{\sum_{d\in B_i}\exp\left(\sum_k\sum_j\left(\sum_{t=1}^{T}1\{d_t=k\}1\{d_{t-1}=j\}\theta_{jk}\right)\right)}$ independent of $\boldsymbol{\omega}_i$ where $B_i=\{d=(d_0,\ldots,d_T)\in\{0,\ldots,J\}^T,d_0=y_{i0},d_T=y_{iT},\sum_{t=1}^{T-1}1\{d_t=k\}=\sum_{t=1}^{T-1}1\{y_{it}=k\},\forall k\}$.

Extension to include exogenous covariates x_{jit} (Honoré and Kyriazidou (2000))

- $P(y_{it}=k|y_{it-1}=j,\omega_i,\theta)=\frac{\exp(x_{kit}\beta_k+\omega_{ik}+\theta_{jk})}{\sum_{h=0}^{J}\exp(x_{hit}\beta_h+\omega_{ih}+\theta_{jh})}$.
- Same pair-wise approach as binary choice case: identification via CMLE requires $x_{kit+1}=x_{kis+1}$ for $k=0,\ldots,J$.

Duration dependence (Frederiksen et al 2007)

- Single spell data with Grouped fixed effect: groups indexed by i and individuals within groups are indexed by $j = 1, \ldots, J_i$. Number of groups goes to ∞ relative to group size and time periods $t = 1, \ldots, T$.
- $y_{jit} = \{0, 1\}$ (unemployed-employed). Let T_{ji} be the time period in which unemployment spell ends.

$$y_{jit} = 1\{\delta_{S_{jit}} + \omega_i + v_{jit} \geq 0\} \text{for } t = 1, \ldots, T_{ji}$$

  where $S_{jit} = S_{ji1} + t$.
- We observe $\{y_{jit}, T_{ji}\}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, J_i$.
- Suppose $J_i = 2$ for all i, the likelihood function for $\{y_{1i1}, \ldots, y_{1iT_{1i}}, y_{2i1}, \ldots, y_{2iT_{2i}}\}$ is

$$\left(\prod_{s=1}^{T_{1i}-1} \frac{1}{1+\exp(\delta_{S_{i1s}} + \omega_i)}\right) \frac{\exp(\delta_{S_{1iT_{1i}}} + \omega_i)}{1+\exp(\delta_{S_{1iT_{1i}}} + \omega_i)}$$

$$\times \left(\prod_{s=1}^{T_{2i}-1} \frac{1}{1+\exp(\delta_{S_{2is}} + \omega_i)}\right) \frac{\exp(\delta_{S_{2iT_{2i}}} + \omega_i)}{1+\exp(\delta_{S_{2iT_{2i}}} + \omega_i)}$$

$$= \frac{\exp(2\omega_i)\exp(\delta_{S_{1iT_{1i}}} + \delta_{S_{2iT_{2i}}})}{\prod_{s=1}^{T_{1i}}(1+\exp(\delta_{S_{i1s}} + \omega_i))\prod_{s=1}^{T_{2i}}(1+\exp(\delta_{S_{i1s}} + \omega_i))}$$

  Sufficient statistics is (T_{1i}, T_{2i}) with no further reduction, hence no CMLE for δ_t.
- No SS $\neq$ No identification or root-n consistent estimator.

Duration dependence (Frederiksen et al 2007)

- Identification comes from comparing two events:
  - $A = \{T_{1i} = t_1, T_{2i} > t_2\}$ and $B = \{T_{1i} > t_1, T_{2i} = t_2\}$; WLOG $t_1 < t_2$.
  - $P(A|A \cup B) = \frac{a_1}{a_1 + a_2}$ with

$$
\begin{aligned}
a_1 &= P_{t_1}(y_{1it_1} = 1, y_{2it_1} = 0 \mid \{y_{1is} = 0, y_{2is} = 0\}_{s<t_1}) \\
&\quad \times P_{t_2}(y_{2it_2} = 0 \mid \{y_{1is} = 0\}_{s<t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s<t_2}) \\
&= F(\delta_{t_1+S_{1i1}} + \omega_i)(1 - F(\delta_{t_1+S_{2i1}} + \omega_i))(1 - F(\delta_{t_2+S_{2i1}} + \omega_i)) \\
a_2 &= P_{t_1}(y_{1it_1} = 0, y_{2it_1} = 0 \mid \{y_{1is} = 0, y_{2is} = 0\}_{s<t_1}) \\
&\quad \times P_{t_2}(y_{2it_2} = 1 \mid \{y_{1is} = 0\}_{s\leq t_1}, \{y_{2is} = 0\}_{s<t_2}) \\
&= (1 - F(\delta_{t_1+S_{1i1}} + \omega_i))(1 - F(\delta_{t_1+S_{2i1}} + \omega_i))F(\delta_{t_2+S_{2i1}} + \omega_i)
\end{aligned}
$$

  - If F is logistic distribution function, $P(A|A \cup B) = \frac{\exp(\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}})}{1+\exp(\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}})}$.

- Extremum estimator (root-n consistent):

$$
\begin{aligned}
\delta_t = \text{argmax} \sum_i \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} & 1\{T_{1i} = t_1, T_{2i} > t_2\} \log \frac{\exp(\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}})}{1 + \exp(\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}})} \\
& + 1\{T_{1i} > t_2, T_{2i} = t_2\} \log \frac{1}{1 + \exp(\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}})}
\end{aligned}
$$

Duration dependence (Chamberlain 1985)

- If there is no duration dependence, then the choice prior to y_{it-1} should have no effect on the probability of the choice y_{it}.
- Chamberlain (1985) proposes to test $H_0 : \gamma_2 = 0$ from the following model

$$P(y_{it} = 1 | y_{it-1}, y_{it-2}) = \frac{\exp(\omega_i + \gamma_{1i} y_{it-1} + \gamma_2 y_{it-2})}{1 + \exp(\omega_i + \gamma_{1i} y_{it-1} + \gamma_2 y_{it-2})}$$

- Sufficient statistics for (ω_i, γ_{1i}): $S_i = \{y_{i0}, y_{i1}, \sum_{t=2}^{T-2} y_{it}, \sum_{t=2}^{T-1} y_{it} y_{it-1}, y_{iT-1}, y_{iT}\}$.
- Conditional likelihood function:

$$P(\mathbf{y}_i | S_i, \omega_i, \gamma_{1i}, \gamma_2) = \frac{\exp(\gamma_2 \sum_{t=2}^{T} y_{it} y_{it-2})}{\sum_{d \in B_i} \exp(\gamma_2 \sum_{t=2}^{T} d_t d_{t-2})}$$

  with $B_i = \{d = \{d_0, \ldots, d_T\} : d_0 = y_{i0}, d_1 = y_{i1}, \sum_{t=2}^{T-2} d_t = \sum_{t=2}^{T-2} y_{it}, \sum_{t=2}^{T-1} d_t d_{t-1} = \sum_{t=2}^{T-1} y_{it} y_{it-1}, d_{T-1} = y_{iT-1}, d_T = y_{iT}\}$
- Requires at least $T = 5$.

Duration dependence (Chamberlain 1985)

- $A_1 = \{1,0,1,0,0,0\}, A_2 = \{1,0,0,1,0,0\}$: $P(A_1|A_1 \cup A_2) = \frac{\exp(\gamma_2)}{1+\exp(\gamma_2)}$
- $B_1 = \{0,1,0,1,1,1\}, B_2 = \{0,1,1,0,1,1\}$: $P(B_1|B_1 \cup B_2) = \frac{\exp(\gamma_2)}{1+\exp(\gamma_2)}$
- $C_1 = \{1,1,0,1,1,0\}, C_2 = \{1,1,1,0,1,0\}$: $P(C_1|C_1 \cup C_2) = \frac{1}{1+\exp(\gamma_2)}$.
- $D_1 = \{0,0,1,0,0,1\}, D_2 = \{0,0,0,1,0,1\}$: $P(D_1|D_1 \cup D_2) = \frac{1}{1+\exp(\gamma_2)}$.

Duration dependence (AGL, 2018wp)

- Recall the example of firm's entry-exit (1,0) decision without forward looking behavior.
- Endogenous state variable $x_{it} = \{y_{it-1}, d_{it}\}$ where $d_{it+1} = 1\{y_{it} = y_{it-1}\}d_{it} + 1$.
- $U_{it}(y) = \omega_i(y) + \beta(y, y_{it-1}, d_{it}) + \epsilon_{it}(y)$
- Structural parameter $\beta(y, y_{it-1}, d_{it}) = 1\{y = y_{it-1}\}\beta_d(y, d_{it}) + 1\{y \neq y_{it-1}\}\beta_y(y, y_{it-1})$.
- Suppose $\beta_d(y, d) = \beta_d(y, d^*)$ for $d \geq d^* = 2$ for both $y = 0, 1$.

Duration dependence (AGL, 2018wp)

- Optimal decision rule

$$y_{it} = 1\Big\{ \ \omega_i(1) - \omega_i(0) + \beta(1, y_{it-1}, d_{it}) - \beta(0, y_{it-1}, d_{it}) + \epsilon_{it}(1) - \epsilon_{it}(0) \geq 0 \ \Big\}$$

$$\begin{aligned} \beta(1, y_{it-1}, d_{it}) &= y_{it-1}\beta_d(1, d_{it}) + (1 - y_{it-1})\beta_y(1, 0) \\ &= y_{it-1}y_{it-2}\beta_d(1, 2) + y_{it-1}(1 - y_{it-2})\beta_d(1, 1) + (1 - y_{it-1})\beta_y(1, 0) \\ \beta(0, y_{it-1}, d_{it}) &= (1 - y_{it-1})(1 - y_{it-2})\beta_d(0, 2) + (1 - y_{it-1})y_{it-2}\beta_d(0, 1) + y_{it-1}\beta_y(0, 1) \end{aligned}$$

- Combine terms, we have

$$\begin{aligned} y_{it} = 1\Big\{ \quad & \omega_i + y_{it-1}y_{it-2}(\beta_d(1, 2) - \beta_d(1, 1)) + y_{it-2}(1 - y_{it-1})(\beta_d(0, 2) - \beta_d(0, 1)) \\ & + y_{it-1}(\beta_d(1, 1) + \beta_d(0, 2) - \beta_y(1, 0) - \beta_y(0, 1)) + \epsilon_{it} \geq 0 \Big\} \end{aligned}$$

with $\omega_i = \omega_i(1) - \omega_i(0) + \beta_y(1, 0) - \beta_d(0, 2)$ and $\epsilon_{it} = \epsilon_{it}(1) - \epsilon_{it}(0)$.

Duration dependence (AGL, 2018wp)

$$y_{it} = 1\Big\{ \quad \omega_i + y_{it-1}y_{it-2}(\beta_d(1,2) - \beta_d(1,1)) + y_{it-2}(1 - y_{it-1})(\beta_d(0,2) - \beta_d(0,1)) \\ + y_{it-1}(\beta_d(1,1) + \beta_d(0,2) - \beta_y(1,0) - \beta_y(0,1)) + \epsilon_{it} \geq 0 \Big\}$$

- If $\delta_d \equiv \beta_d(0,2) - \beta_d(0,1) = \beta_d(1,2) - \beta_d(1,1)$, then

$$y_{it} = 1\Big\{\omega_i + y_{it-2}\delta_d + y_{it-1}\underbrace{(\beta_d(1,1) + \beta_d(0,2) - \beta_y(1,0) - \beta_y(0,1))}_{not-identified} + \epsilon_{it} \geq 0\Big\}$$

  Two lags dynamic binary choice model: δ_d is identified.

- If no duration dependence for zero choice, $\beta_d(0,2) = \beta_d(0,1) = 0$, and $\delta_d \equiv \beta_d(1,2) - \beta_d(1,1)$, then

$$y_{it} = 1\Big\{\omega_i + y_{it-1}y_{it-2}\delta_d + y_{it-1}(\beta_d(1,1) - \beta_y(1,0) - \beta_y(0,1)) + \epsilon_{it} \geq 0\Big\}$$

  Cumulative lag: both δ_d and $(\beta_d(1,1) - \beta_y(1,0) - \beta_y(0,1))$ are identified.

- AGL provides a generalization of the Chamberlain results which allows more general identification of duration dependence.