Sept 22 2021 Class 4

General Discrete Choice Model

- Consider

$$\begin{aligned} U_j &= u_j^*(X) + \varepsilon_j \\ U_J &= \varepsilon_J \\ Y &= \arg\max_{j \in J}\{U_j\} \end{aligned}$$

- Assume distribution of ε_j, $F^*_{\varepsilon_1 \ldots, \varepsilon_J}$, is known (But now, need not be iid, need not be type-I extreme)
- u_j^* identified

General Discrete Choice Model

- why?
- Consider the probability that alternative k is chosen conditional on X:

$$
\begin{aligned}
\Pr(Y = k|X) &= \Pr(u_k + \varepsilon_k = \max_j\{u_j + \varepsilon_j\}) \\
&= \Pr(\{\varepsilon_1 \le u_k - u_1 + \varepsilon_k\} \wedge \ldots \wedge \{\varepsilon_J \le u_k + \varepsilon_k\}) \\
&= \int_{\varepsilon_k=-\infty}^{\varepsilon_k=+\infty} \int_{\varepsilon_1=-\infty}^{\varepsilon_1=u_k-u_1+\varepsilon_k} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_J=u_k+\varepsilon_k} \overbrace{f(\varepsilon_1, \ldots, \varepsilon_J)}^{\frac{\partial^J}{\partial\varepsilon_1\ldots,\partial\varepsilon_J}F_{\varepsilon_1\ldots,\varepsilon_J}} d\varepsilon_1 \cdots d\varepsilon_J \\
&= \int_{\varepsilon_k=-\infty}^{\varepsilon_k=+\infty} \frac{\partial}{\partial\varepsilon_k} F_{\varepsilon_1\ldots,\varepsilon_J}(u_k - u_1 + \varepsilon_k, \cdots, \varepsilon_k, \cdots, u_k + \varepsilon_k) d\varepsilon_k
\end{aligned}
$$

  - Think of this as a mapping Q_k from $\mathbf{u} = (u_1, u_2, \ldots u_{J-1}) \in$ R $^{J-1}$ to $\Pr(Y = k|X)$.
- Define $Q = (Q_1, \ldots, Q_{J-1})$, as a function from $\mathbf{u} = (u_1, u_2, \ldots u_{J-1})$ to $\mathbb{R}^{J-1}$.
  - Q is X dependent
- Want to show that Q is one-to-one (for each X)

General Discrete Choice Model

- Although want to show that Q is one-to-one, first let's show that Q in onto.
- Fix a particular share $\mathbf{s} = (s_1, s_2, \cdots, s_{J-1})$.
- Define $r_k(u_1, ...u_{J-1}, \mathbf{s})$ so that $Q_k(u_1, ..., r_k(u_1, ...u_{J-1}, \mathbf{s}), \cdots, u_{J-1}) = s_k$ for each k.
  - Note that r_k is well-defined.
- Question of onto mapping is equivalent to... "for every $\mathbf{s}$, is there a fixed point for $r(\cdot, \mathbf{s}) = (r_1(\cdot, \mathbf{s}), \cdots, r_{J-1}(\cdot, \mathbf{s}))$?"
- If $F_{\varepsilon_1 ..., \varepsilon_J}$ is smooth, $r(\cdot, \mathbf{s})$ is continuous; can take as mapping from compact set in $\mathbb{R}^{J-1}$ into $\mathbb{R}^{J-1}$.
  - Fill in the gap in prob set.
- Hence, exists fixed point, by Brouer.

General Discrete Choice Model

- Now let's show Q is one-to-one (for each X).
- This is equivalent to saying that $r(\cdot, \mathbf{s})$ has unique fp for every $\mathbf{s}$.
- Sufficient to show that $r(\cdot, \mathbf{s})$ is a contraction: matrix norm of Jacobian r is less than 1.
- Here's a proof: http://terpconnect.umd.edu/˜petersd/466/fixedpoint.pdf

1.3 Proving the Contraction Property

The contraction property is related to the Jacobian $g'(x)$ which is an $n \times n$ matrix for each point $x \in D$. If the matrix norm satisfies $\|g'(x)\| \le q < 1$ then the mapping g must be a contraction:

Theorem 2. *Assume the set $D \subset \mathbb{R}^n$ is convex and the function $g\colon D \to \mathbb{R}^n$ has continuous partial derivatives $\frac{\partial g_j}{\partial k}$ in D. If for $q < 1$ the matrix norm of the Jacobian satisfies*

$$\forall x \in D: \qquad \|g'(x)\| \le q \tag{7}$$

the mapping g is a contraction in D and satisfies (1).

Proof. Let $x, y \in D$. Then the points on the straight line from x to y are given by $x + t(y-x)$ for $t \in [0,1]$. As D is convex all these points are contained in D. Let $G(t) := g\big(x + t(y-x)\big)$, then by the chain rule we have $G'(t) = g'\big(x + t(y-x)\big)\,(y-x)$ and

$$g(y) - g(x) = G(1) - G(0) = \int_0^1 G'(t)dt = \int_0^1 g'\big(x + t(y-x)\big)\,(y-x)\,dt$$

As an integral of a continuous function is a limit of Riemann sums the triangle inequality implies $\left\| \int_a^b F(t)dt \right\| \le \int_a^b \|F(t)\|\,dt$:

$$\|g(y) - g(x)\| \le \int_0^1 \left\|g'\big(x + t(y-x)\big)\,(y-x)dt\right\| \le \int_0^1 \underbrace{\left\|g'\big(x + t(y-x)\big)\right\|}_{\le q} \|y-x\|\,dt \le q\,\|y-x\|$$

$\square$

This is usually the easiest method to prove that a given mapping g is a contraction, see the examples in sections 1.5, 1.6.

General Discrete Choice Model

- If we consider l^{∞} norm, then matrix norm of A, $\|A\|_{\infty}$is given by
$$\|A\|_{\infty} = \max_{1 \leq i \leq J-1} \sum_k |a_{ik}|.$$
- Hence, sufficient condition for r to be a contraction is $\sum_k \left| \frac{\partial}{\partial u_k} r_i \right| < 1$ for all i.
- Recall: $Q_i(u_1, ..., r_i(u_1, ...u_{J-1}, \mathbf{s}), \cdots, u_{J-1}) = s_i$.
- Hence, $\frac{d}{du_k} Q_i(u_1, ..., r_i(u_1, ...u_{J-1}, \mathbf{s}), \cdots, u_{J-1}) = \frac{\partial}{\partial u_k} Q_i + \frac{\partial}{\partial u_k} Q_i \frac{\partial}{\partial u_k} r_i = 0$.
    - Hence, $\frac{\partial}{\partial u_k} r_i = -\frac{\partial}{\partial u_k} Q_i \left(\frac{\partial}{\partial u_k} Q_i \right)^{-1}$.
- Show $\sum_j \left| \frac{\partial}{\partial u_j} r_k \right| < 1$ for all k (prob set).

Dynamic Discrete Choice

- Consider discrete choice with inifinte horizon.
- Agent has $\{1, \cdots, J\}$ choices at each point in time.
- Let X be the observable state variable, and let us denote its transition by $F_j(X'|X)$.
  - Transition of X depends on choice (indexed by j).
- In each period, agent gets period utility $u_j(X) + \varepsilon_j$, where $\varepsilon_j \perp X$. Denote distribution of $\varepsilon_j \sim F_{\varepsilon_1 \dots, \varepsilon_J}$. Assume **known**.
- Discounts future by $\beta < 1$, and β **known.**
- Then agent's value fn can be written recursively:

$$V(X) = \mathbf{E}_\varepsilon \left[\max_{j \in J} u_j(X) + \varepsilon_j + \beta \int V(X') dF_j(X'|X) \right]$$

Dynamic Discrete Choice

- Define "choice-specific" value function:

$$v_j(X) = \int V(X')dF_j(X'|X).$$

- Then

$$V(X) = \mathbf{E}_\varepsilon\left[\max_{j \in J} u_j(X) + \varepsilon_j + \beta v_j\left(X\right)\right]$$

  and

$$\begin{aligned}&\Pr(j|X)\\ &\quad = \Pr(u_j(X) + \beta v_j\left(X\right) + \varepsilon_j \geq \max_{j'}\{u_{j'}(X) + \beta v_{j'}\left(X\right) + \varepsilon_{j'}\}).\end{aligned}$$

- Note that for any j and j' and X, $u_j(X) + \beta v_j(X) - (u_{j'}(X) + \beta v_{j'}(X))$ is identified (why?)

Dynamic Discrete Choice

Corollary

Consider a dynamic discrete choice problem as above. Furthermore, assume $u_J(X) = 0$ for all X. Then $u_j(X)$ are all identified. (Note that we are **not** *setting $u_J(X) + \beta v_J(X) = 0$ - then would be trivial)*

Dynamic Discrete Choice

- First, note that $\Pr(j|X)$ and $F_j(X'|X)$ are identified.
- Define $\overline{v}_j$ as follows.

$$\overline{v}_j = u_j(X) + \beta v_j(X).$$

- For $j \neq J$,

$$\begin{aligned}
\overline{v}_j(X) &= u_j(X) + \beta v_j(X) \text{ (by def)} \\
&= u_j(X) + \beta \int \mathbf{E}_{\varepsilon'} \left[\max_{j' \in J} \overline{v}_{j'}(X') + \varepsilon_{j'} \right] dF_j(X'|X) \\
&= u_j(X) + \beta \int \mathbf{E}_{\varepsilon'} \max_{j' \in J} \left[\overbrace{\overline{v}_{j'}(X') - \overline{v}_J(X')}^{\Delta\overline{v}_{j'}(X')} + \varepsilon_{j'} \right] + \overline{v}_J(X') dF_j(X'|X).
\end{aligned}$$

- Recall that $\Delta\overline{v}_j = [\overline{v}_j(X) - \overline{v}_J(X)]$ is a known object – a function, say H_j, of choice probabilities:

$$\Delta\overline{v}_j = H_j(\{\Pr(j'|X)\}).$$

Dynamic Discrete Choice

- Note that if we identify $\overline{v}_J(X)$, then we identify $\overline{v}_j(X)$ (why?).
- This, in turn, means that $u_j(X)$ is identified from

$$\overline{v}_j(X) = u_j(X) + \beta \int \mathbf{E}_{\varepsilon'} \left[\max_{j' \in J} \overline{v}_{j'}(X') + \varepsilon_{j'} \right] dF_j(X'|X).$$

- Hence, remains to show that $\overline{v}_J(X)$ is identified.
- Note that

$$\begin{aligned} \overline{v}_J(X) &= \beta \int \mathbf{E}_{\varepsilon'} \left[\max_{j' \in J} \overline{v}_{j'}(X') + \varepsilon_{j'} \right] dF_j(X'|X) \\ &= \beta \int \mathbf{E}_{\varepsilon'} \left[\max_{j' \in J} H_{j'}(\{\Pr(\cdot|X')\}) + \overline{v}_J(X') + \varepsilon_{j'} \right] dF_j(X'|X). \end{aligned}$$

- Note that $u_J(X) = 0$.

Dynamic Discrete Choice

- Consider a mapping T from continuous functions to continuous functions ($C_0 : X \to \mathbb{R}$):

$$T \circ f(X) = \beta \int \mathbf{E}_{\varepsilon'} \left[\max_{j' \in J} H_{j'}(\{\Pr(\cdot|X')\}) + f(X')) + \varepsilon_{j'} \right] dF_j(X'|X).$$

- Function T is identified.
  - Can define T if we know $F_j(X)$ and $\Pr(j|X)$.
  - H_j is determined by $F_{\varepsilon_1 \cdots \varepsilon_J}$.
- Note that $\overline{v}_J$ is the fixed point of T.
- There is a unique FP of T (why?)
- $\overline{v}_J$ is identified.
- Hence $u_j(X)$ are all identified.

Dynamic Discrete Choice

- How to estimate $u_j(X;\theta)$?
- 1. Likelihood based (Nested FP) - 1st gen.
- Estimate $F_j(X'|X)$: transition matrix
- Fix θ, compute value function $V(X,\theta)$ and policy function, $d_j(X,\theta)$.
- Compute likelihood for θ, $L(\theta)$.
- Maximize $L(\theta)$ over θ.

Dynamic Discrete Choice

- 2. Conditional Choice Probability (CCP, HM) - 2nd gen.
- Estimate $F_j(X'|X)$: transition matrix
- Estimate $\Pr(j|X)$: choice probability
- Construct an estimate of $\widehat{T}$
- Solve for the fixed point of $\widehat{T}$.
  - Sometimes this is easier if absorbing state (can easily compute v_J)
- c.f. next example

Dynamic Discrete Choice

- 2'. Conditional Choice Probability (CCP, HM) - 2nd gen
- Estimate $F_j(X'|X)$: transition matrix
- Estimate $\Pr(j|X)$: choice probability
  - Note that we can estimate the policy function $d(X, \varepsilon) \in \{1, \cdots, J\}$
- Fix θ, compute $\widehat{v}_j(X;\theta)$ by simulating draws of $\{\varepsilon_t, \varepsilon_{t+1}, \cdots\}$ and $\{X = X_t, X_{t+1}, \cdots\}$ using the policy function.
- Estimate $\Pr(j|X;\theta)$ given by

$$\Pr\widehat{(j|X;\theta)} = \Pr(\widehat{v}_j(X;\theta) + \varepsilon_j \geq \max_{j'}\{\widehat{v}_j(X;\theta) + \varepsilon_{j'}).$$

- Minimize $\|\Pr(j|X;\theta) - \Pr(j|X)\|$
  - Similar to ideas in Pakes-Ostrovsky-Berry.
  - No need to solve for fixed point

Dynamic Discrete Choice

- 2". Conditional Choice Probability (CCP, HM) - 2nd gen
- Estimate $F_j(X'|X)$: transition matrix
- Estimate $\Pr(j|X)$: choice probability
  - Note that we can estimate the policy function $d(X, \varepsilon) \in \{1, \cdots, J\}$
- Take policy functions that are not being used, $d'(X, \varepsilon)$
- Fix θ, compute $\widehat{\bar{v}}_j(X; d, \theta)$ and $\widehat{\bar{v}}_j(X; d', \theta)$ by simulating draws of $\{\varepsilon_t, \varepsilon_{t+1}, \cdots\}$ and $\{X_t, X_{t+1}, \cdots\}$ using the policy function $d(X, \varepsilon)$ and $d'(X, \varepsilon)$
- Estimate θ by minimizing

$$\left\| \widehat{\bar{v}}_j(X; d, \theta) - \widehat{\bar{v}}_j(X; d', \theta) \right\|^-$$

- In practice, use many d' by minimizing

$$\sum_{d'} \left\| \widehat{\bar{v}}_j(X; d, \theta) - \widehat{\bar{v}}_j(X; d', \theta) \right\|^-$$

  - Similar to ideas in Bajari Benkard Levin
  - No need to solve for fixed point

Example: Entry & Exit

- Firm decides whether to operate or not.
  $j \in \{1, 2\} (= \{in, out\})$
  - $j = 1$ is be in market, $j = 2$ is to be out of market.
- Macro conditions S (observed) which affect profitability of being in the market.
- Let $\pi(s) + \varepsilon_1$ be the period profit from being in the market when $S = s$.
- Firm gets profit $0 + \varepsilon_2$, when out of market.
  - ε_1, ε_2 i.i.d. EV.
- If S is finite, (say $|S| = 2$) there is a transition matrix: Π,

$$\Pi = \begin{pmatrix} \Pr(g|g) & \Pr(b|g) \\ \Pr(g|b) & \Pr(b|b) \end{pmatrix} = \begin{pmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{pmatrix}.$$

  Otherwise, there is a transition measure, $f(s'|s)$.
- Entry is costly, pay κ if previously out of market and enter.
- What are the state variables?
- Assume actions and states are observed. Want to know $\pi(s)$, κ.

Example: Entry & Exit

- State variables, X, is $S \times \{\text{previously } in, \text{ previously } out\}$
  - $F_1(\{s', 1\}|\{s, 1\}) = f(s'|s)$
  - $F_1(\{s', 1\}|\{s, 0\}) = f(s'|s)$
  - $F_1(\{s', 0\}|\{s, 1\}) = 0$
  - $F_1(\{s', 0\}|\{s, 0\}) = 0.$
- Similarly for F_2.
- Note that $f(s'|s)$ is readily identified.

Example: Entry & Exit

- Value function is,

$$
\begin{aligned}
V(\{s,1\}) &= \mathbf{E}_{\varepsilon}\left[\max\left\{\begin{array}{c} \pi_s + \varepsilon_1 + \beta\int V(s',1)f(s'|s)ds', \\ \beta\int V(s',0)f(s'|s)ds' + \varepsilon_2 \end{array}\right\}\right] \\
V(\{s,0\}) &= \mathbf{E}_{\varepsilon}\left[\max\left\{\begin{array}{c} \pi_s - \kappa + \varepsilon_1 + \beta\int V(s',1)f(s'|s)ds', \\ \beta\int V(s',0)f(s'|s)ds' + \varepsilon_2\} \end{array}\right\}\right]
\end{aligned}
$$

- Let

$$
\begin{aligned}
v_1(s) &= \int V(s',1)f(s'|s)ds' \\
v_2(s) &= \int V(s',0)f(s'|s)ds',
\end{aligned}
$$

then

$$
\begin{aligned}
V(\{s,1\}) &= \mathbf{E}_{\varepsilon}\left[\max\{\pi_s + \varepsilon_1 + \beta v_1(s),\ \beta v_2(s) + \varepsilon_2\right] \\
V(\{s,0\}) &= \mathbf{E}_{\varepsilon}\left[\max\{\pi_s - \kappa + \varepsilon_1 + \beta v_1(s),\ \beta v_2(s) + \varepsilon_2\}\right]
\end{aligned}
$$

Example: Entry & Exit

- We know that

$$\begin{aligned}\Pr(j=1|\{s,0\}) &= \Pr(\pi_s-\kappa+\varepsilon_1+\beta v_1(s)>\beta v_2(s)+\varepsilon_2) \\ &= \frac{\exp(\pi_s-\kappa+\beta v_1(s))}{\exp(\pi_s-\kappa+\beta v_1(s))+\exp(\beta v_2(s))}\end{aligned}$$

$$\begin{aligned}\Pr(j=2|\{s,0\}) &= \Pr(\beta v_2(s)+\varepsilon_2>\pi_s-\kappa+\varepsilon_1+\beta v_1(s)) \\ &= \frac{\exp(\beta v_2(s))}{\exp(\pi_s-\kappa+\beta v_1(s))+\exp(\beta v_2(s))}\end{aligned}$$

- Hence,

$$\begin{aligned}&\log\frac{\Pr(j=1|\{s,0\})}{\Pr(j=2|\{s,0\})} \\ =\ & \pi_s-\kappa+\beta v_1(s)-\beta v_2(s)\end{aligned}$$

Example: Entry & Exit

- Note that

$$
\begin{aligned}
v_2(s) &= \int V(s',0) f(s'|s) ds' \\
&= \int \mathbf{E}_\varepsilon \left[\max \left\{ \begin{array}{c} \pi_{s'} - \kappa + \beta v_1(s') + \varepsilon_1, \\ \beta v_2(s') + \varepsilon_2 \end{array} \right\} \right] f(s'|s) ds'.
\end{aligned}
$$

Using the previous expression, this can be rewritten as

$$
\begin{aligned}
v_2(s) &= \int \mathbf{E}_\varepsilon \left[\max \left\{ \begin{array}{c} \beta v_2(s') + \varepsilon_1 + \log \frac{\Pr(j=1|\{s',0\})}{\Pr(j=2|\{s',0\})}, \\ \beta v_2(s') + \varepsilon_2 \end{array} \right\} \right] f(s'|s) ds' \\
&= \int \beta v_2(s') + \mathbf{E}_\varepsilon \left[\max\{\varepsilon_1 + \log \frac{\Pr(j=1|\{s',0\})}{\Pr(j=2|\{s',0\})}, \varepsilon_2\} \right] f(s'|s) ds' \\
&= \cdots = \int \beta v_2(s') + \gamma + \log \left(1 + \frac{\Pr(j=1|\{s',0\})}{\Pr(j=2|\{s',0\})} \right) f(s'|s) ds'.
\end{aligned}
$$

where γ is a known constant (Euler's constant).

Example: Entry & Exit

- If S is finite (say 2×2), then $v_2(s)$ is a S-dim vector, and,

$$v_2(s) = \begin{pmatrix} v_{2g} \\ v_{2b} \end{pmatrix}$$

$$= \begin{pmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{pmatrix} \begin{pmatrix} \beta v_{2g} + \gamma + \log\left(1 + \frac{\Pr(j=1|\{g,0\})}{\Pr(j=2|\{g,0\})}\right) \\ \beta v_{2b} + \gamma + \log\left(1 + \frac{\Pr(j=1|\{b,0\})}{\Pr(j=2|\{b,0\})}\right) \end{pmatrix}.$$

$$\iff \begin{pmatrix} v_{2g} \\ v_{2b} \end{pmatrix} =$$

$$\left[I_2 - \beta \begin{pmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{pmatrix} \right]^{-1} \begin{pmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{pmatrix} \begin{pmatrix} \gamma + \log\left(1 + \frac{\Pr(j=1|\{g,0\})}{\Pr(j=2|\{g,0\})}\right) \\ \gamma + \log\left(1 + \frac{\Pr(j=1|\{b,0\})}{\Pr(j=2|\{b,0\})}\right) \end{pmatrix}.$$

- Hence $v_2(s)$ is identified.

Example: Entry & Exit

- If S is finite, $v_2(s)$ is a S-dim vector, and explicitly solving for the T function/its FP is easy.
- In the previous example, T is a linear function of (v_{2g}, v_{2b}).
- Solving for the fixed point simply means inverting a matrix.
- If S is not finite, then there is a unique fixed point of the above functional equation.

Example: Entry & Exit

- Now, consider identification of $v_1(s)$:

$$\begin{aligned} v_1(s) &= \int V(s',1)f(s'|s)ds' \\ &= \int \mathbf{E}_\varepsilon \left[\max \left\{ \begin{array}{c} \pi_{s'} + \varepsilon_1 + \beta v_1(s'), \\ \beta v_2(s') + \varepsilon_2 \end{array} \right\} \right] f(s'|s)ds' \end{aligned}$$

- Recall that

$$\begin{aligned} & \log \frac{\Pr(j=1|\{s,1\})}{\Pr(j=2|\{s,1\})} \\ = \; & \pi_s + \beta v_1(s) - \beta v_2(s), \end{aligned}$$

- Using this to substitue out $v_2(s)$ from the previous expression,

$$\begin{aligned} v_1(s) &= \int \beta v_2(s') + E_\varepsilon \left[\max\{\log \frac{\Pr(j=1|\{s',1\})}{\Pr(j=2|\{s',1\})} + \varepsilon_1, \; \varepsilon_2 \right] f(s'|s)ds' \\ &= \int \beta v_2(s') + \gamma + \log \left(1 + \frac{\Pr(j=1|\{s',1\})}{\Pr(j=2|\{s',1\})} \right) f(s'|s)ds'. \end{aligned}$$

- Hence $v_1(s)$ is identified.

Example: Entry & Exit

- Finally, to identify π_s and κ, use the fact that

$$\log \frac{\Pr(j = 1|\{s, 1\})}{\Pr(j = 2|\{s, 1\})} = \pi_s + \beta v_1(s) - \beta v_2(s)$$

and

$$\log \frac{\Pr(j = 1|\{s, 2\})}{\Pr(j = 2|\{s, 2\})} = \pi_s - \kappa + \beta v_1(s) - \beta v_2(s).$$

- Note that $v_1(s)$, $v_2(s)$ and β are known, so π_s and κ are identified.

Example: Entry & Exit

- CCP Estimation - how would you estimate π_s and κ when $|S| < \infty$.
  - 2 and 2'?

Is there a more general theory?