Introduction

Robert A. Miller

Structural Econometrics

September 2017

Overview of the Course

Course website, topics, themes and assessment

- The course material can be found at: comlabgames.com/899/
- It is organized around three topics:
  1. Dynamic discrete choice
  2. Auctions and optimal contracting
  3. Market structure
- Four methodological themes permeate this course:
  1. Summarizing data using economic structure
  2. Analyzing empirical content and identification
  3. Estimating and testing structural models
  4. Conducting counterfactuals
- Your grades will come from:
  1. Three assignments (10 percent each)
  2. Presentation to class of published work (20 percent)
  3. Midterm in-class test on lecture material (20 percent)
  4. Final examination on lecture and reading material (30 percent)

- Each period $t \in \{1, 2, \ldots, T\}$ for $T \leq \infty$, an individual chooses among J mutually exclusive actions.
- Let d_{jt} equal one if action $j \in \{1, \ldots, J\}$ is taken at time t and zero otherwise:

$$d_{jt} \in \{0, 1\}$$

$$\sum_{j=1}^{J} d_{jt} = 1$$

- At an abstract level assuming that choices are mutually exclusive is innocuous, because two combinations of choices sharing some features but not others can be interpreted as two different choices.
- For example in a female labor supply and fertility model, suppose:

$$j \in \{(\text{work, no birth}), (\text{work, birth}), (\text{no work, no birth}), (\text{no work, birth})\}$$

Dynamic Discrete Choice

Information and states

- Suppose that actions taken at time t can potentially depend on the state $z_t \in Z$.
- For Z finite denote by $f_{jt}(z_{t+1}|z_t)$, the probability of z_{t+1} occurring in period $t+1$ when action j is taken at time t.
- For example in the example above, suppose $z_t = (w_t, k_t)$ where:
  - $k_t \in \{0, 1, \ldots\}$ are the number of births before t
  - $w_t \equiv d_{1,t-1} + d_{2,t-1}$, so $w_t = 1$ if the female worked in period $t-1$, and $w_t = 0$ otherwise.
- Note that Z must be defined compatible to the transition matrix: for example setting $z_t = (w_t, k_t)$ where $k_t \in \{0, 1, \ldots\}$ are the number of births before $t-1$, is incompatible with assumption about transitions and choices.
- With up to 5 offspring, 3 levels of experience, the number of states including age (say 50 years) is 750. Add in 4 levels of education (less than high school, high school, some college and college graduate) and 3 racial categories, increases this number to 7800.

- When Z is finite there is a $Z \times Z$ transition matrix for each (j, t).
- In the example above, the matrices are of dimension 180 but very sparse; only 180 elements are nonzero and all the nonzero elements are one.
- Given a deterministic sequence of actions sequentially taken over S periods, we form the S period transition matrix by producting the one period transitions.
- If Z is a Euclidean space $f_{jt}(z_{t+1}|z_t)$ is the probability (density function) of z_{t+1} occurring in period $t+1$ when j is picked at time t.
- With almost identical notation we could model $z_t \in Z_t$ and in this way generalize from states of the world to histories, or information known at t, or t-measurable events.
- For example in a health application we might define $z_t \equiv \{h_s\}_{s=1}^{t-1}$ as a medical record with $h_s \in \{\text{healthy at } s, \text{ sick at } s\}$.

- The individual's current period payoff from choosing j at time t is determined by z_t, which is revealed to the individual at the beginning of the period t.
- The current period payoff at time t from taking action j is $u_{jt}(z_t)$.
- Given choices $(d_{1t}, \ldots, d_{Jt})$ in each period $t \in \{1, 2, \ldots, T\}$ the individual's expected utility is:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1}d_{jt}u_{jt}(z_t)\right\}$$

  where $\beta \in (0, 1)$ is the subjective discount factor, and at each period t the expectation is taken over $z_{t+1}, \ldots, z_T$.
- Formally β is redundant if u is subscripted by t; we typically include a geometric discount factor so that infinite sums of utility are bounded, and the optimization problem is well posed.

- Write the optimal decision at period t as a decision rule denoted by $d_t^o(z)$ formed from its elements $d_{jt}^o(z_t)$.
- Let $V_t(z_t)$ denote the value function in period t, conditional on behaving according to the optimal decision rule:

$$V_t(z_t) \equiv E\left[\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t} d_{j\tau}^o(z_\tau)\, u_{j\tau}(z_\tau)\right]$$

- In terms of period $t+1$:

$$\beta V_{t+1}(z_{t+1}) \equiv \beta E\left\{\sum_{\tau=t+1}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1} d_{j\tau}^o(z_\tau)\, u_{j\tau}(z_\tau)\right\}$$

Dynamic Discrete Choice Models

Recursive Representation

- Appealing to Bellman's (1958) principle we obtain:

$$
\begin{aligned}
V_t(z_t) &= \sum_{j=1}^{J} d_{jt}^{o} u_{jt}(z_t) \\
&\quad + \sum_{j=1}^{J} d_{jt}^{o} \sum_{z=1}^{Z} E\left\{ \sum_{\tau=t+1}^{T} \sum_{j=1}^{J} \beta^{\tau-t} d_{j\tau}^{o}\left(z_\tau\right) u_{j\tau}(z_\tau) \,|z \right\} f_{jt}\left(z|z_t\right) \\
&= \sum_{j=1}^{J} d_{jt}^{o} \left[u_{jt}(z_t) + \beta \sum_{z=1}^{Z} V_{t+1}(z) f_{jt}\left(z|z_t\right) \right]
\end{aligned}
$$

when Z is finite with a similar expression holding (using an integral) when Z is Euclidean.

- To compute the optimum for T finite, we first solve a static problem in the last period to obtain $d_T^o(z)$.
- Applying backwards induction $i \in \{1, \ldots, J\}$ is chosen to maximize:

$$u_{it}(z_t) + E\left\{\sum_{\tau=t+1}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^{o}(z_\tau)\,u_{j\tau}(z_\tau)\,|z_t, d_{it} = 1\right\}$$

- In the stationary infinite horizon case we assume $u_{jt}(z) \equiv u_j(z)$ and that $u_j(z) < \infty$ for all (j, z).
- Consequently expected utility each period is bounded and the contraction mapping theorem applies, proving $d_t^o(z) \to d^o(z)$ for large T.

- Let $v_{jt}(z_t)$ denote the flow payoff of action j plus the expected future utility of behaving optimally from period $t+1$ on:

$$v_{jt}(z_t) \equiv u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1}|z_t)$$

- By definition:

$$d_{jt}^{o}(z_t) \equiv I\{v_{jt}(z_t) \geq v_{kt}(z_t) \forall\ k\}$$

- Suppose we observe the states z_{nt} and decisions d_{nt} of individuals $n \in \{1, \ldots, N\}$ over time periods $t \in \{1, \ldots, T\}$.
- If two people with the same z_t made different decisions, say j and k, then $v_{jt}(z_t) = v_{kt}(z_t)$.
- Such equalities imply that large data sets typically impose many restrictions on $u_{jt}(z_t)$, $f_{jt}(z_{t+1}|z_t)$ and β.
- Can they all be satisfied in a finite data set without rejecting a model that has empirical content?

- A more modest objective is to predict the probability distribution of choices margined over factors that individuals observe, but data analysts do not.
- In this respect we seek to predict the behavior of a population, not each individual, essentially obliterating that difference between macroeconomics and microeconomics.
- We now assume the states can be partitioned into those which are observed, x_t, and those that are not, ϵ_t.
- Thus $z_t \equiv (x_t, \epsilon_t)$.
- Suppose the data consist of N independent and identically distributed draws from the string of random variables $(X_1, D_1, \ldots, X_T, D_T)$.
- The n^{th} observation is given by $\left\{x_1^{(n)}, d_1^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)}\right\}$ for $n \in \{1, \ldots, N\}$.

- Denote the probability (density) of the pair $(x_{t+1}, \epsilon_{t+1})$, conditional on $\left(x_t^{(n)}, \epsilon_t\right)$ and the optimal action taken by n at t, as:

$$H_{nt}\left(x_{t+1}, \epsilon_{t+1} \middle| x_t^{(n)}, \epsilon_t\right) \equiv$$

$$\sum_{j=1}^{J} I\left\{d_{jt}^{(n)} = 1\right\} d_{jt}^{o}\left(x_t^{(n)}, \epsilon_t\right) f_{jt}\left(x_{t+1}, \epsilon_{t+1} \middle| x_t^{(n)}, \epsilon_t\right)$$

- Conditional on $x_1^{(n)}$ the joint probability of $\left\{d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)}\right\}$ is:

$$\Pr\left\{d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \middle| x_1^{(n)}\right\} =$$

$$\int_{\epsilon_T} \cdots \int_{\epsilon_1} \left[\begin{array}{c} \sum_{j=1}^{J} I\left\{d_{jT}^{(n)} = 1\right\} d_{jT}^{o}\left(x_T^{(n)}, \epsilon_T\right) \times \\ \prod_{t=1}^{T-1} H_{nt}\left(x_{t+1}^{(n)}, \epsilon_{t+1} \middle| x_t^{(n)}, \epsilon_t\right) g\left(\epsilon_1 \middle| x_1^{(n)}\right) \end{array}\right] d\epsilon_1 \ldots d\epsilon_T$$

- Let $\theta \in \Theta$ uniquely index a specification of $u_{jt}(z_t)$, $f_{jt}(z_{t+1}|z_t)$ and β under consideration.
- Conditional on $x_1^{(n)}$ suppose $\left\{d_1^{(n)}, x_2^{(n)}, \ldots, , d_T^{(n)}\right\}_{n=1}^{N}$ was generated by $\theta_0 \in \Theta$.
- Define $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_T)$. The maximum likelihood estimator, θ_{ML}, selects $\theta \in \Theta$ to maximize the joint probability of the observed occurrences conditional on the initial conditions:

$$\theta_{ML} \equiv \underset{\theta \in \Theta}{\arg\max} \left\{ N^{-1} \sum_{n=1}^{N} \log \left(\Pr \left\{ d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \left| x_1^{(n)}; \theta \right.\right\} \right) \right\}$$

- If there is a unique maximum in $\theta \in \Theta$ to:

$$\int_{x_1^{(n)}} \log \left(\Pr \left\{ d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \left| x_1^{(n)}; \theta \right.\right\} \right) dF\left(x_1^{(n)}\right)$$

then the model is identified, and under standard conditions θ_{ML} is $\sqrt{N}$ consistent, asymptotically normal, and efficient.