'MATH+ECON+CODE' MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2018
Day 4, January 18 2018: "Multinomial choice"
Block 12. Dynamic discrete choice

Learning objectives: block 12

- Finite-horizon Rust's model
- identification and estimation
- normalization issues

References for block 12

- Rust (1987), "Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica*.
- Chiong, Galichon and Shum (2016). Duality in discrete choice models. *Quantitative economics*.

Section 1

THEORY

- Recall the dynamic programming model seen in block 3. The setting is the same: there are n_x units (buses) in state x at the initial period ($t=1$); at each period, one must choose for each unit some alternative $y \in \mathcal{Y}$; the probability of transiting to state x' at period $t+t$ conditional on being in state x and choosing alternative y at time t is $P^t_{x'|xy}$.
- The difference with the setting seen in block 3 is that, following Rust, the utility associated with choosing y in state x at t is no longer deterministic, but includes an additional random term $\varepsilon_y \sim \mathbf{P}_{xt}$, so it is

$$u^t_{xy} + \varepsilon_y.$$

The stochastic structure is such that $x_t, (x_t, \varepsilon), (x_t, y_t)$ is a Markov chain – which rules out persistent shocks, i.e. there cannot be correlaton between ε_t and ε_{t+1} conditional on (x_{t+1}, y_{t+1}).

- As a result of the random utility term, Bellman's equation becomes

$$\begin{aligned} U_x^t &= \mathbb{E}_{\mathbf{P}_{xt}} \left[\max_{y \in \mathcal{Y}} \left\{ u_{xy}^t + \sum_{x'} U_{x'}^{t+1} P_{x'|xy} + \varepsilon_y \right\} \right] \\ &= G_{xt}(u_{x.}^t + \sum_{x'} U_{x'}^{t+1} P_{x'|x.}). \end{aligned}$$

- Set $W_{xt}\left(U\right) = G_{x(t-1)}(u_{x.}^{t-1} + \sum_{x'} U_{x'}^t P_{x'|x.})$ for $1 < t < T$, the equation becomes

$$U_x^{t-1} = W_{xt}\left(U^t\right).$$

- Note that

$$\frac{\partial W_{xt}\left(U\right)}{\partial U_{x'}} = \sum_y P_{x'|xy} \sigma_{x(t-1),y}$$

is the conditional probability of a transition to x' given being at x at time $t-1$, denoted $\mu_{x'|x}^{t-t}$.

- In the logit case, one has

$$U_x^t = \log \sum_{y \in \mathcal{Y}} \exp \left(u_{xy}^t + \sum_{x'} U_{x'}^{t+1} P_{x'|xy} \right)$$

- Setting $V_x^t = \exp\left(U_x^t\right)$ and $v_{xy}^t = \exp\left(v_{xy}^t\right)$, this becomes an algebraic expression

$$V_x^t = \sum_{y \in \mathcal{Y}} v_{xy}^t \prod_{x' \in \mathcal{X}} \left(V_x^t\right)^{P_{x'|xy}}.$$

- The dual problem can be expressed as:

$$\begin{aligned}
&\min_{U_x^t,} \sum_{x\in\mathcal{X}} n_x U_x^1 \qquad (1)\\
&s.t.\ U_x^{t-1} = W_{xt}(U^t)\ 1 < t \leq T\\
&U_x^T = G_{xT}(u_{x.}^T)
\end{aligned}$$

- In the logit case, with $V_x^t = \exp\left(U_x^t\right)$ and $v_{xy}^t = \exp\left(v_{xy}^t\right)$, one has

$$\begin{aligned}
&\min_{U_x^t,\ t\in\mathcal{T},\ x\in\mathcal{X}} \sum_{x\in\mathcal{X}} n_x \log V_x^1\\
&s.t.\ V_x^t = \sum_{y\in\mathcal{Y}} v_{xy}^t \prod_{x'\in\mathcal{X}} \left(V_x^t\right)^{P_{x'|xy}},\ t < T\\
&V_x^T = \sum_{y\in\mathcal{Y}} v_{xy}^t.
\end{aligned}$$

- Set n_x^t the Lagrange multipliers associated with the constraints. First order conditions in the dual problem yield

$$n_x = n_x^1$$
$$n_x^t = \sum_{x'} \frac{\partial W_{x(t-1)}}{\partial U_{x'}^{t-1}} n_{x'}^{t-1}, 1 < t \leq T$$

- The second line are Kolmogorov-forward equations (forward propagation of mass)

$$\sum_{x'} \mu_{x'|x}^{t-t} n_{x'}^{t-1} = n_x^t.$$

- Let $W_t\left(U^t;n^{t-1}\right)=\sum_x n_x^{t-1}W_{xt}(U_x^t)$, and let $W_t^*\left(n^t;n^{t-1}\right)$ be its Legendre transform with respect its first variable.
- **Theorem**: the value of the primal problem is

$$\max_{n^t}\left\{\sum_{x\in\mathcal{X}} n_x^T G_{xt}(u_{x.}^T) - \sum_{\substack{x\in\mathcal{X}\\ 1<t\leq T}} W_t^*\left(n^t;n^{t-1}\right)\right\}$$
$$s.t.\ n_x^1 = n_x$$

- Start from the dual

$$\begin{aligned} &\min_{U_x^t} \sum_{x\in\mathcal{X}} n_x U_x^1 \\ s.t.\ &U_x^{t-1} = W_{xt}(U^t)\ 1 < t \leq T \\ &U_x^T = G_{xT}(u_{x.}^T) \end{aligned} \tag{2}$$

- Write the saddlepoint formulation

$$\min_{U^t} \max_{n^t} \left\{ \begin{array}{c} \sum_{x\in\mathcal{X}} n_x U_x^1 - \sum_{x,1\leq t\leq T} n_x^t U_x^t \\ +\sum_{x,1<t\leq T} n_x^{t-1} W_{xt}(U^t) \\ +\sum_x n_x^T G_{xT}(u_{x.}^T) \end{array} \right\}$$

- Saddlepoint rewrites

$$\max_{n^t} \min_{U^t} \left\{ \begin{array}{c} \sum_x n_x^T G_{xt}(u_{x.}^T) + \sum_x \left(n_x - n_x^1\right) U_x^1 \\ + \sum_{1<t\leq T} n_x^{t-1} W_{xt}(U_x^t) - n_x^t U_x^t \end{array} \right\}$$

- Recall that $W_t\left(U^t; n^{t-1}\right) = \sum_x n_x^{t-1} W_{xt}(U_x^t)$, and $W_t^*\left(n^t; n^{t-1}\right)$ be its Legendre transform with respect its first variable, one has

$$\begin{array}{l} \max_{n_x^t, t\geq 1} \left\{ \sum_{t<T} n_x^T G_{xt}(u_{x.}^T) - \sum_{1<t\leq T} W_t^*\left(n^t; n^{t-1}\right) \right\} \\ s.t.\ n_x^1 = n_x \end{array}$$

QED.

- Recall

$$\max_{n_x^t, t\geq 1} \left\{ \sum_{t<T} n_x^T G_{xt}(u_{x.}^T) - \sum_{1<t\leq T} W_t^* \left(n^t ; n^{t-1} \right) \right\}$$
$$s.t.\ n_x^1 = n_x$$

- For $1 \leq t < T$, one has

$$\frac{\partial W_t^*}{\partial n_x^t} \left(n^t ; n^{t-1} \right) + \frac{\partial W_{t+1}^*}{\partial n_x^t} \left(n^{t+1} ; n^t \right) = 0,$$

and note that

$$\frac{\partial W_t^*}{\partial n_x^t} = U_x^t \text{ and } \frac{\partial W_{t+1}^*}{\partial n_x^t} \left(n^{t+1} ; n^t \right) = -W_{xt+1}(U^{t+1})$$

hence the first order condition recovers the Bellman equation.

- One has

$$W_t\left(U;n^{t-1}\right)=\sum n_x^{t-1}\log\sum_{y\in\mathcal{Y}}\exp\left(u_{xy}^{t-1}+\sum_{x'}U_{x'}^{t}P_{x'|xy}\right)$$

and thus

$$\begin{aligned}&W_t\left(n^t;n^{t-1}\right)\\&=\max_U\left\{\sum_x n_x^tU_x-\sum_x n_x^{t-1}\log\sum_{y\in\mathcal{Y}}\exp\left(u_{xy}^{t-1}+\sum_{x'}U_{x'}^{t}P_{x'|xy}\right)\right\}\end{aligned}$$

- Sadly, no closed-form formula.

- Rust studies the infinite-horizon version of the problem, in which case u^t_{xy} does not depend on t, and, if $\beta > 0$ is a discount factor, then the intertemporal utility is given by the set of equations

$$U_x = W_x(\beta U),$$

where $W_x(\beta U) = \mathbb{E}\left[\max_y \left\{ u_{xy} + \sum_{x'} \beta U_{x'} P_{x'|xy} + \varepsilon_y \right\}\right]$.

- It's possible to show that $(U_x) \rightarrow (W_x(\beta U))$ is a contraction mapping, so the above equation has a unique solution.

- Assume agents choose actions $y \in \mathcal{Y}$ from a finite space $\mathcal{Y} = \{0, 1, \ldots, m\}$. At each period, agents get utility flow from choosing y given by
$$u_{xy} + \varepsilon_y$$
where ε_y denotes the utility shock associated to action y, which differs across agents and is assumed to be i.i.d. Gumbel distributed.
- Assume the model is stationary, so dependence on time is omitted.
- In Rust's setting, y is the decision to perform maintenance on the bus ($y = 1$ if maintenance, $y = 0$ if not); x is the mileage of the bus since last maintenance (reset to zero if maintenance and random otherwise); u_{x1} is the average profit if no maintenance and mileage x, and u_{0x} if maintenance.

- In a one-period context, agents would simply maximize $u_{xy} + \varepsilon_y$ over y as in static discrete choice framework just seen.
- However, the action chosen at time t, y_t, has an impact on the state x_t at the next period. Agents take this into account and optimize instead

$$u_{xy} + \varepsilon_y + \beta \mathbb{E}\left[\bar{U}_{x'}\left(\varepsilon'\right)|x, y\right]$$

over $y \in \mathcal{Y}$, where $\beta \mathbb{E}\left[\bar{U}_{x'}\left(\varepsilon'\right)|x, y\right]$ is the expected discounted payoff from later periods conditional on being at state x and choosing action y at the current period. $\bar{U}$ is defined recursively.
- We use primes ($'$) to denote next-period values.

- We make the following conditional independence assumption on the evolution of (x, ε):

$$\begin{aligned} Pr(x', \varepsilon' | y, x, \varepsilon) =& Pr(\varepsilon' | x', y, x, \varepsilon) \cdot Pr(x' | y, x, \varepsilon) \\ =& Pr(\varepsilon' | x') \cdot Pr(x' | y, x). \end{aligned}$$

  which means that the distribution of ε' only depends on x', and the distribution of x' only depends on y and x at the previous period.

- The discount rate is β. At each period, agents do

$$y \in \arg\max_{y \in \mathcal{Y}} \left\{ u_{xy} + \varepsilon_y + \beta \mathbb{E} \left[\bar{U}_{x'} \left(\varepsilon' \right) | x, y \right] \right\}, \tag{3}$$

  where the value function $\bar{U}$ is defined in a recursive manner as

$$\bar{U}_x \left(\varepsilon \right) = \max_{y \in \mathcal{Y}} \left\{ u_{xy} + \varepsilon_y + \beta \mathbb{E} \left[\bar{U}_{x'} \left(\varepsilon' \right) | x, y \right] \right\}.$$

- Introduce $U(x)$, the ex-ante (or integrated) value function, defined as:

$$U_x = \mathbb{E}\left[\bar{U}_x\left(\varepsilon\right)|x\right].$$

- The choice-specific value functions therefore consists of two terms: the per-period utility flow and the discounted continuation payoff:

$$w_{xy} \equiv u_{xy} + \beta\mathbb{E}\left[U_{x'}|x,y\right].$$

and

$$U_x = \log \sum_{y \in \mathcal{Y}} e^{w_{xy}} \tag{4}$$

- The problem here is that we cannot impose the normalization $w_{0y} = 0$ as we did before. Indeed, $w_{0y} \equiv u_{0x} + \beta \mathbb{E}\left[U_{x'} | x, 0\right)]$; but the second term of the rhs is endogenous. Instead, we can e.g. normalize

$$u_{0x} = 0$$

which is an exogenous quantity.

- Then, the log-odds ratio formula implies that

$$\log \frac{\pi_{y|x}}{\pi_{y|0}} = w_{xy} - w_{0y}$$

thus

$$w_{xy} = \log \frac{\pi_{y|x}}{\pi_{y|0}} + w_{0y} \tag{5}$$

where the $\log(...)$ term is known, and the w_{0y} is to be identified.

- Plugging in the expression of w_{xy} into (4) yields

$$U_x = \log \sum_{y \in \mathcal{Y}} \frac{\pi_{y|x}}{\pi_{y|0}} + w_{0y}$$

thus

$$U_x = w_{0y} - \log \pi_{y|0} \tag{6}$$

- One has by definition of w

$$w_{0y} = \beta \mathbb{E}\left[U_{x'} | x, 0\right], \tag{7}$$

thus, combining (6) and (7),

$$U_x = \beta \mathbb{E}\left[U_{x'} | x, 0\right] - \log \pi_{y|0}$$

hence

$$\mathbb{E}\left[\beta U_{x'} - U_x | x, 0\right] = \log \pi_{y|0}. \tag{8}$$

- Let W be the column vector whose general term is $\left(\log \pi_{y|0}\right)_{x\in\mathcal{X}}$, let U be the column vector whose general term is $\left(U_x\right)_{x\in\mathcal{X}}$, and let Π be the $|X| \times |X|$ matrix whose general term is

$$\Pi_{iy} = Pr\left(x_{t+1} = y | x_t = i, y = 0\right).$$

Equation (8), rewritten in matrix notation, is

$$W = \left(\beta\Pi - I\right) U$$

- It is possible to show that for $\beta < 1$, matrix $I - \beta\Pi^0$ is invertible. Thus Equation (8) becomes

$$U = \left(\beta\Pi - I\right)^{-1} W. \tag{9}$$

- Therefore U_x is identified from data.

- It follows that all the remaining quantities are also identified by

$$\begin{aligned} w_{0y} &= U_x + \log \pi_{y|0} \\ w_{xy} &= w_{0y} + \log \frac{\pi_{y|x}}{\pi_{y|0}} \\ u_{xy} &= w_{xy} - \beta \mathbb{E}\left[U_{x'} | x, y \right] \end{aligned}$$

Section 2

CODING

- Today's application is based on Rust's (1994) bus engine dataset.