Structural Econometrics: Dynamic Discrete Choice

Jean-Marc Robin

Plan

1. Dynamic discrete choice models
2. Application: college and career choice

Dynamic discrete choice models

See for example the presentation by Wolpin (*AER*, 1996).

At each date t discrete, an individual has to choose one action among K possible actions.

Let

$$d_k(t) = \begin{vmatrix} 1 & \text{if } k \text{ is the chosen action,} \\ 0 & \text{otherwise.} \end{vmatrix}$$

Let $d(t) = (d_1(t), ..., d_K(t))$ or $d(t) = \sum_{k=1}^{K} k d_k(t)$ be the choice variable.

Let $S(t) \in \mathcal{S}$ be the state variable (i.e; the information at the beginning of period t when the action is chosen). Assume $\mathcal{S}$ discrete: $\mathcal{S} = \{s_1, ..., s_N\}$ (in any case the computer will require a discrete state space).

Action k yields payoff $R_k(S(t), t)$.

The state transition probability matrix is

$$p_{ij}(k,t) = \Pr\{S(t+1) = s_j | S(t) = s_i, d_k(t) = 1\}.$$

Strategies

A strategy is a sequence of functions

$$\begin{array}{rcl} D(\cdot,t): \mathcal{S} & \longrightarrow & \{0,1\}^K \\ s & \longmapsto & D(s,t) = (D_1(s,t), ..., D_K(s,t)) \end{array}$$

Individuals seek for the strategy D to maximise the expected discounted sum of future payoffs:

$$V(S(t),t) = \max_{D(\cdot,\cdot)} \mathbb{E}\left[\sum_{\tau=t}^{T} \beta^{\tau-t} \sum_{k=1}^{K} D_k\left(S(\tau),\tau\right) R_k(S(\tau),\tau) \middle| S(t) \right].$$

Bellman principle

Write, for $s \in \mathcal{S}$,

$$V(s,t) = \max\{V_1(s,t), ..., V_K(s,t)\}$$

where $V_k(s,t)$ is the present value if action k is chosen at t when $S(t) = s$:

$$V_k(s,t) = R_k(S(t),t) + \beta \mathbb{E}\left[V(S(t+1),t+1) | \, S(t) = s, d_k(t) = 1\right]$$

and

$$V_k(s,T) = R_k(s,T).$$

The optimal strategy is

$$D_k(s,t) = 1 \text{ iff } V_k(s,t) = \max\{V_1(s,t), ..., V_K(s,t)\}$$

and then

$$V(s,t) = \sum_{k=1}^{K} D_k(s,t) V_k(s,t).$$

Solution

Start from terminal period T and, for all $s \in \mathcal{S}$, determine the action which maximises payoff $R_k(s,T)$:

$$D_k(s,T) = 1 \text{ iff } R_k(s,T) = \max\{R_1(s,T), ..., R_K(s,T)\}$$

and

$$V(s,T) = \sum_{k=1}^{K} D_k(s,T) R_k(s,T).$$

Then determine $D(s,t)$ recursively: for all $s \in \mathcal{S}$,

$$D_k(s,t) = 1 \text{ iff } V_k(s,t) = \max\{V_1(s,t), ..., V_K(s,t)\}$$

where, for all $s_1, ..., s_N$,

$$\begin{aligned} V_k(s_i,t) &= R_k(s_i,t) + \beta \mathbb{E}\left[V(S(t+1),t+1) \middle| S(t) = s_i, d_k(t) = 1\right] \\ &= R_k(s,t) + \beta \sum_{j=1}^{N} p_{ij}(k,t) \underbrace{V(s_j,t+1)}_{= \sum_{k=1}^{K} D_k(s_j,t+1) V_k(s_j,t+1)} \end{aligned}$$

Curse of dimensionality: huge number of computations and large memory size required to compute $V_k(s,t)\ \forall k, s, t$.

Estimation

Parameters: in the payoff functions $R_k(s,t)$ and transition probabilities $p_{ij}(k,t)$.

Inference: maximum likelihood or (simulated) method of moments.

Data: individual sequences $y^h = \left\{x^h(t_0^h), d^h(t_0^h), x^h(t_0^h+1), d^h(t_0^h+1), ..., x^h(t_1^h), d^h(t_1^h)\right\}$ for individuals $h = 1, ..., H$ and $t \in \left\{t_0^h, t_0^h+1, ..., t_1^h\right\}$, where $x^h(t) \in \{x_1, ..., x_I\}$ is the observed part of the state variables, i.e. $S^h(t) = \left(x^h(t), \varepsilon^h(t)\right)$, with the following...

...Assumptions on the process of shocks $\varepsilon^h(t)$:

- $\varepsilon^h(t) = \left(\varepsilon_1^h(t), ..., \varepsilon_K^h(t)\right)$ iid;
- $R_k(S^h(t), t) = \overline{R}_k(x^h(t), t) + \varepsilon_k^h(t)$;
- **conditional independence:**

$$\Pr\left\{x^h(t+1), \varepsilon^h(t+1) | x^h(t), \varepsilon^h(t), d_k(t) = 1\right\} = \Pr(\varepsilon^h(t+1)) \times \underbrace{\Pr\left\{x^h(t+1) = x_j | x^h(t) = x_i, d_k^h(t) = 1\right\}}_{\equiv \overline{p}_{ij}(k,t)}.$$

Likelihood

The conditional likelihood of y^h given $x^h(t_0^h)$ is

$$\begin{aligned}\ell(y^h|x^h(t_0^h)) = \Pr\left\{d^h(t_0^h)|x^h(t_0^h)\right\} \times \Pr\left\{x^h(t_0^h+1)|x^h(t_0^h), d^h(t_0^h)\right\} \\ \times \Pr\left\{d^h(t_0^h+1)|x^h(t_0^h+1)\right\} \times P\left\{x^h(t_0^h+2)|x^h(t_0^h+1), d^h(t_0^h+1)\right\} \\ \times \cdots \times \Pr\left\{d^h(t_1^h)|x^h(t_1^h)\right\}\end{aligned}$$

where

$$\Pr\left\{d_k^h(t) = 1|x^h(t)\right\} = \Pr\left\{\varepsilon^h(t) \text{ s.t. } D_k(x^h(t), \varepsilon^h(t), t) = 1|x^h(t)\right\}.$$

The conditional likelihood of the sample is

$$\prod_{h=1}^{H} \ell(y^h|x^h(t_0^h)).$$

Choice probabilities

$$\Pr\left\{d_k^h(t) = 1 | x^h(t) = x_i\right\} = \Pr\left\{\varepsilon_k^h(t) \geq \varepsilon_m^h(t) + \overline{V}_m(x_i,t) - \overline{V}_k(x_i,t), \forall m \neq k | x^h(t) = x_i\right\}.$$

where

$$\begin{aligned}
\overline{V}_k(x_i,t) &= \overline{R}_k(x_i,t) + \beta \sum_{j=1}^{N} \overline{p}_{ij}(k,t)\overline{V}(s_j,t+1), \\
\overline{p}_{ij}(k,t) &= \Pr\left\{x^h(t+1) = x_j | x^h(t) = x_i, d_k^h(t) = 1\right\}, \\
\overline{V}(x_j,t+1) &= \mathbb{E}\max_{k=1}^{K}\left\{\overline{V}_k(x_j,t+1) + \varepsilon_k^h(t+1)\right\}.
\end{aligned}$$

For instance, for (X_1, X_2) Gaussian,

$$\begin{aligned}
\mathbb{E}\max\{X_1, X_2\} &= X_2 + \mathbb{E}\max\{X_1 - X_2, 0\} \\
&= m_2 + (m_1 - m_2)\,\Phi\left(\frac{m_1 - m_2}{\sigma}\right) + \sigma\varphi\left(\frac{m_1 - m_2}{\sigma}\right)
\end{aligned}$$

where $\sigma = \operatorname{Std}(X_1 - X_2) = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$.

Two stage estimation

One can proceed in two stages to save computer time, although t the cost of some efficiency loss.

1. Maiximise partial likelihood of state changes:

$$\prod_{h=1}^{H} \Pr\left\{x^h(t_0^h+1)|x^h(t_0^h), d^h(t_0^h)\right\} \Pr\left\{x^h(t_0^h+2)|x^h(t_0^h+1), d^h(t_0^h+1)\right\}$$
$$\cdots \Pr\left\{x^h(t_1^h)|x^h(t_1^h-1), d^h(t_1^h-1)\right\},$$

   with respect to parameters of $\Pr\{x(t+1)|x(t), d(t), t\}$.

2. Maximise the likelihood of the sequence of decisions:

$$\prod_{h=1}^{H} \Pr\left\{d^h(t_0^h)|x^h(t_0^h)\right\} \times \cdots \times \Pr\left\{d^h(t_1^h)|x^h(t_1^h)\right\}$$

   using the estimated $\Pr\{x(t+1)|x(t), d(t), t\}$ to compute the present value functions necessary to calculate choice probabilities.

Unobserved heterogeneity

The two-stage estimation procedure does not work if there exists unobserved heterogeneity.

Assume that $S^h(t) = \left(x^h(t), \varepsilon^h(t), \mu^h\right)$ where $\mu^h \in \{1, ..., M\}$ indicates a particular way of grouping individuals. All individuals with the same μ^h have a specific value of the parameters governing payoff functions and state probabilities.

Let $\Pr\left\{\mu^h = m\right\} = \pi_m$, $m \in \{1, ..., M\}$.

The likelihood becomes

$$\prod_{h=1}^{H} \ell(y^h|x^h(t_0^h)) = \prod_{h=1}^{H} \left(\sum_{m=1}^{M} \pi_m \ell(y^h|x^h(t_0^h), m) \right)$$

where

$$\begin{aligned}\ell(y^h|x^h(t_0^h), \mu^h) = \Pr\left\{d^h(t_0^h)|x^h(t_0^h), \mu^h\right\} \times \Pr\left\{x^h(t_0^h+1)|x^h(t_0^h), \mu^h, d^h(t_0^h)\right\} \\ \times \Pr\left\{d^h(t_0^h+1)|x^h(t_0^h+1), \mu^h\right\} \times \Pr\left\{x^h(t_0^h+2)|x^h(t_0^h+1), \mu^h, d^h(t_0^h+1)\right\} \\ \times \cdots \times \Pr\left\{d^h(t_1^h)|x^h(t_1^h), \mu^h\right\}.\end{aligned}$$

EM algorithm

Let $\mathbf{y} = (y_1, \cdots, y_H)$ be a vector of observations. Let $\mathbf{z} = (z_1, \cdots, z_H)$ be unobserved covariates. The likelihood of $(\mathbf{y}, \mathbf{z})$ is $f(\mathbf{y}, \mathbf{z}; \theta)$.

Since $\mathbf{z}$ is not observed one estimates θ by maximixing the integrated likelihood:

$$f(\mathbf{y}; \theta) = \int f(\mathbf{y}, \mathbf{z}; \theta)\mu(d\mathbf{z}).$$

This integral may be difficult to compute and the numerical approximation may yield unstable Newton-type optimisation algorithms (numerical errors accumulate instead of averaging). The EM algorithm is often preferable.

The EM algorithm iterates the following steps until numerical convergence (generally slowly)

$$\theta^{(p)} = \arg\max_{\theta} Q(\theta|\theta^{(p-1)}),$$

where

$$\begin{aligned} Q(\theta|\theta^{(p-1)}) &= \mathbb{E}\left[\ln f(\mathbf{y}, \mathbf{z}; \theta)|\mathbf{y}; \theta^{(p-1)}\right] \\ &= \int p\left\{\mathbf{z}|\mathbf{y}; \theta^{(p-1)}\right\} \ln f(\mathbf{y}, \mathbf{z}; \theta)\mu(d\mathbf{z}). \end{aligned}$$

Each iteration increases the likelihood and converges toward a local maximum of the likelihood.

EM algorithm: discrete mixtures

Assume $z_i \in \{1, ..., M\}$ and $\pi_m = \Pr\{z_i = m\}$.

Then $\theta = (\beta, \pi)$ where β indexes $f(y_i|z_i;\beta)$ and $\pi = (\pi_1, ..., \pi_M)$.

We have

$$f(\mathbf{y}, \mathbf{z}; \theta) = \prod_{i=1}^{H} f(y_i, z_i; \theta) = \prod_{i=1}^{H} \left[\sum_{m=1}^{M} \pi_m f(y_i|z_i = m; \beta) \right].$$

Step E (expectation): Use Bayes rule to compute posterior probabilities:

$$p\left\{z_i = m|y_i; \theta^{(p-1)}\right\} = \frac{\pi_m^{(p-1)} f(y_i|z_i = m; \beta^{(p-1)})}{\sum_{m=1}^{M} \pi_m^{(p-1)} f(y_i|z_i = m; \beta^{(p-1)})}$$

and

$$\begin{aligned} Q(\theta|\theta^{(p-1)}) &= \int p\left\{\mathbf{z}|\mathbf{y}; \theta^{(p-1)}\right\} \ln f(\mathbf{y}, \mathbf{z}; \theta) \mu(d\mathbf{z}) \\ &= \sum_{i=1}^{H} \sum_{m=1}^{M} p\left\{z_i = m|y_i; \theta^{(p-1)}\right\} \ln \left[\pi_m f(y_i|z_i = m; \beta)\right]. \end{aligned}$$

Step M (maximisation): Update β by constrained ML:

$$\beta^{(p)} = \arg\max_{\beta} \sum_{i=1}^{H} \sum_{m=1}^{M} p\left\{z_i = m|y_i;\theta^{(p-1)}\right\} \ln f(y_i|z_i = m;\beta),$$

(i.e. duplicate individual observations K times and affect a weight equal to posterior probability $p\left\{z_i = m|y_i;\theta^{(p-1)}\right\}$) and update π as

$$\pi_m^{(p)} = \frac{1}{H}\sum_{i=1}^{H} p\left\{z_i = m|y_i;\theta^{(p-1)}\right\}.$$

Application: education and career choice

See for example the presentation by Keane et Wolpin (*JPE*, 1997).

Model of education and career choices.

- **Data:** 11-year panel (National Longitudinal Survey of Youths): cohort of youths aged 16 in 1979 and followed until 1990.
- **Objective**: evaluate policy effects such as education subsidies.
- **Population studied** is a cohort of individuals starting at the age of 16 and retiring at 65.
- **Choices**: blue collar worker ($k = 1$), white collar worker ($k = 2$), military ($k = 3$), education ($k = 4$) or inactivity ($k = 5$).

Model

- Payoffs associated to choices $k = 1, 2, 3$ are the corresponding wages, the log of which are

$$\ln R_k(t) = e_k(16) + e_{k1} EDUC(t) + e_{k2} EXP_k(t) - e_{k3} \left[EXP_k(t)\right]^2 + \varepsilon_k(t)$$

where $e_k(16)$ is the intercept (initial condition), $EDUC(t)$ is the number of years of education, $EXP_k(t)$ is occupation-k specific experience (= nb of years spent working as k; with $EXP_k(16) = 0$).

- Education's instantaneous payoff (or cost if negative):

$$R_4(t) = e_4(16) - c_1 \underbrace{\mathbf{1}\left[EDUC(t) \geq 12\right]}_{\text{HS graduate}} - c_2 \underbrace{\mathbf{1}\left[EDUC(t) \geq 16\right]}_{\text{college graduate}} + \varepsilon_4(t).$$

- Leisure utility:

$$R_5(t) = e_5(16) + \varepsilon_5(t).$$

- State variable: $S(t) = (e(16), EDUC(t), EXP(t), \varepsilon(t))$ with

$$\begin{cases} e(16) = \left(e_1(16), ..., e_5(16)\right), \\ EXP(t) = \left(EXP_1(t), EXP_2(t), EXP_3(t)\right), \\ \varepsilon(t) = \left(\varepsilon_1(t), ..., \varepsilon_5(t)\right). \end{cases}$$

Model (continued)

- Heterogeneity μ
  - four groups $m = 1, 2, 3, 4$.
  - $e(16) = (e_1(16), ..., e_5(16))$ group-specific.
  - as $EDUC(16) = 9$ or 10, assume different proportions of each type given $EDUC(16)$:

$$\Pr\{\mu = m | EDUC(16)\} \equiv \pi_{m,EDUC(16)}.$$

- State probabilities:
  - $\varepsilon(t) = (\varepsilon_1(t), ..., \varepsilon_5(t))$ iid and $\sim \mathcal{N}(0, \Omega)$, with $\text{Cov}(\varepsilon_k(t), \varepsilon_\ell(t)) = 0$ for ℓ or $k \geq 4$ (i.e. only $\varepsilon_1(t), \varepsilon_2(t), \varepsilon_3(t)$ corresponding to employment spells are correlated).
  - Education: $EDUC(t+1) = EDUC(t) + d_4(t)$.
  - Expérience: $EXP_k(t+1) = EXP_k(t) + d_k(t)$.
- Value functions:

$$V_k(S(t), t) = R_k(t) + \beta \mathbb{E}\left[V(S(t+1), t+1) | d_k(t) = 1\right]$$

where $\varepsilon(t+1)$ is the only risk factor (not predetermined) in $V(S(t+1), t+1)$ given $d(t)$.

Value functions

$$V_k(S(t), t) = R_k(t) + \beta \mathbb{E}\left[V(S(t+1), t+1) | d_k(t) = 1\right]$$

where $\varepsilon(t+1)$ is the only risk factor (not predetermined) in $V(S(t+1), t+1)$ given $d(t)$, a

- Pour $k = 1, 2, 3,$

$$\begin{cases} EXP_\ell(t+1) = EXP_\ell(t) + \mathbf{1}(\ell = k), \ \ell = 1, 2, 3, \\ EDUC(t+1) = EDUC(t). \end{cases}$$

- Pour $k = 4,$

$$\begin{cases} EXP_\ell(t+1) = EXP_\ell(t), \ \ell = 1, 2, 3, \\ EDUC(t+1) = EDUC(t) + 1. \end{cases}$$

- Pour $k = 5,$

$$\begin{cases} EXP_\ell(t+1) = EXP_\ell(t), \ \ell = 1, 2, 3, \\ EDUC(t+1) = EDUC(t). \end{cases}$$

Likelihood

- **Individual observations:** $y^h(t) = \left(d^h(t), w^h(t)\right)$, $t = 16, ..., 26$, where $d^h(t) = \left(d_1^h(t), ..., d_5^h(t)\right)$ is occupation choice and $w^h(t) = \sum_{k=1}^{3} d_k^h(t) R_k^h(t)$ is current wage (missing if not working).

- Sample likelihood:

$$L = \prod_{h=1}^{H} \left[\sum_{h=1}^{H} \pi_{m,EDUC^h(16)} \ell^h(y^h(16), ..., y^h(26)|e^h(16), EDUC^h(16)) \right].$$

- Likelihood for individual h:

$$\ell^h(y^h(16), y^h(17), ..., y^h(26)|e^h(16), EDUC^h(16)) = \prod_{t=16}^{26} \ell^h \left(y^h(t)|e^h(16), EDUC^h(t), EXP^h(t)\right).$$

Likelihood (continued)

Likelihood for individual h at time t: $\ell^h\left(y^h(t)|e^h(16), EDUC^h(t), EXP^h(t)\right)$ is computed as follows (we omit conditioning to simplify notations).

Different as general studied above as the wage information tells us about shocks $\varepsilon_k^h(t)$.

- Case $d^h(t) = k \in \{1,2,3\}$: one thus knows that $w^h(t) = R_k^h(t)$ and $V_k(S(t),t) \geq V_\ell(S(t),t)$, $\ell \neq k$:

$$\ell^h\left(y^h(t)\right) = \Pr\left\{V_k(S^h(t),t) \geq V_\ell(S^h(t),t),\ \forall \ell \neq k | \underbrace{R_k^h(t) = w^h(t)}_{\text{determines } \varepsilon_k^h(t)}\right\} \times \underbrace{\text{pdf}\left\{R_k^h(t) = w^h(t)\right\}}_{\text{i.e. density of } R_k^h(t) \text{ at observation } w^h(t)}.$$

- Other cases: one only knows that $V_k(S^h(t),t) \geq V_\ell(S^h(t),t)$, $\ell \neq k$:

$$\ell^h\left(y^h(t)\right) = \Pr\left\{V_k(S^h(t),t) \geq V_\ell(S^h(t),t),\ \forall \ell \neq k\right\}.$$

Likelihood (continued)

Given $e^h(16), EDUC^h(t), EXP^h(t)$,

$$\mathrm{pdf}\left\{R_k^h(t) = w^h(t)\right\} = \frac{1}{w^h(t)}\frac{1}{\sigma_k}\varphi\left(\frac{\overbrace{\ln w^h(t) - e_k(16) - e_{k1}EDUC(t) - e_{k2}EXP_k(t) + e_{k3}\left[EXP_k(t)\right]^2}^{\varepsilon_k^h(t)}}{\sigma_k}\right)$$

where $\sigma_k^2 = \mathrm{Var}(\varepsilon_k(t))$, and

$$\Pr\left\{V_k(S^h(t),t) \geq V_\ell(S^h(t),t),\ \forall \ell \neq k \mid \varepsilon_k^h(t)\right\} = \Pr\left\{\varepsilon_\ell^h(t) \leq g_\ell(t),\ \forall \ell \neq k \mid \varepsilon_k(t)\right\}$$

where

$$\begin{aligned} g_\ell(t) &= \ln\left(V_k(S^h(t),t) - \beta\mathbb{E}\left[V(S^h(t+1),t+1)|d_\ell(t)=1\right]\right) \\ &\quad - e_\ell(16) - e_{\ell 1}EDUC(t) - e_{\ell 2}EXP_k(t) + e_{\ell 3}\left[EXP_k(t)\right]^2,\ \ \ell = 1,2,3\,, \\ g_4(t) &= V_k(S^h(t),t) - e_4(16) - c_1\mathbf{1}\left[EDUC(t) \geq 12\right] - c_2\mathbf{1}\left[EDUC(t) \geq 16\right], \\ g_5(t) &= V_k(S^h(t),t) - e_5(16). \end{aligned}$$

One has to compute the cdf of a vector of 4 normal r.v.'s. (Computation simplified by the fact that $\varepsilon_4^h(t)$ and $\varepsilon_5^h(t)$ are assumed independent and independent of $\varepsilon_1^h(t), \varepsilon_2^h(t)$ and $\varepsilon_3^h(t)$.

Lastly, $\Pr\left\{V_k(S^h(t),t) \geq V_\ell(S^h(t),t),\ \forall \ell \neq k\right\}$ can be computed by numerical integration of $\Pr\left\{V_k(S^h(t),t) \geq\right.$
w.r.t. $\varepsilon_k^h(t)$.

Results

See article.

The fit is excellent.

They find a very limited effect of college tuition subsidies (exogenous change in c_2).