Conditional Choice Probability Estimation of Dynamic Discrte Choice Models with Unobserved Heterogeneity

Arcidiacono and Miller

ECMA 2011

April 8, 2019

Motivation

Combining two worlds

Estimation of dynamic discrete choice models

- Rust (1987), Aguirregabiria & Mira (2007), Bajari, Benkard & Levin (2007)
- Value function not observed but enters in likelihood
  - Requires simulation of value functions through fixed point algorithms
  - If CCP inversion is used econometrician needs to know all states

Models with unobserved heterogeneity: finite mixture models

- Keane & Wolpin (1997), Eckstein & Wolpin (1999), tons of Heckman papers
- FIML computationally intense
- No closed form solution of Emax function due to correlation
  - Requires numerical integration and simulation of value functions

This paper

Allows for unobserved heterogeneity in dynamic discrete choice models

- Value function = sum of flow payoffs and CCP for any choice sequence
- Extends the concept of "replacing the engine" (Altug & Miller 1998)
    - Finite dependence ~ future cancels out after a few periods
    - Difference in value functions can be written as finite sequence of CCP
- Builds on CCP inversion (Hotz & Miller 1993) and two-step estimators
- Builds on EM algorithm for finite mixtures (Arcidiacono & Jones 2003)

Fusion!

Arcidiacono & Jones (2003) + Hotz & Miller (1993) = Arcidiacono and Miller (2011)

Rust (1987) with unobserved heterogeneity

Revisiting Rust (1987) with time-invariant unobserved heterogeneity

Variables: accumulated mileage x_t, brand of the bus s, and idiosyncratic shocks.

If engine is replaced

$$d_{1t} = 1 \implies \text{mileage } x_{t+1} = 0$$

If engine not replaced

$$d_{2t} = 1 \implies \text{mileage } x_{t+1} = x_t + 1.$$

Discounted utility given by

$$\mathbb{E}\Big[\sum_{t=1}^{\infty} \beta^{t-1}\{d_{2t}(\theta_1 x_t + \theta_2 s + \epsilon_{2t}) + d_{1t}\epsilon_{1t}\}\Big]$$

Choice-specific value fucntion

$$v_j(x, s) = \begin{cases} \beta V(0, s) & \text{if } j = 1 \\ \theta_1 x + \theta_2 s + \beta V(x+1, s) & \text{if } j = 2 \end{cases}$$

If idiosyncratic shocks are Type I EV errors, the CCP of replacing the engine is

$$p_1(x, s) = \frac{1}{1 + \exp\big(v_2(x, s) - v_1(x, s)\big)}$$

Revisiting Rust (1987) (cont)

Since errors are type I EV, Emax function has closed form

$$V(x+1,s) = \log\Big(\exp\big(v_1(x+1,s)\big) + \exp\big(v_2(x+1,s)\big)\Big)$$

Then

$$\begin{aligned} v_2(x,s) &= \theta_1 x + \theta_2 s + \beta \log\Big(\exp\big(v_1(x+1,s)\big) + \exp\big(v_2(x+1,s)\big)\Big) \\ &= \theta_1 x + \theta_2 s + \beta \log\Big(\exp\big(v_1(x+1,s)\big)\big(1 + \exp\big(v_2(x+1,s) - v_1(x+1,s)\big)\big)\Big) \\ &= \theta_1 x + \theta_2 s + \beta v_1(x+1,s) - \beta \log p_1(x+1,s) \end{aligned}$$

Similarly

$$v_1(x,s) = \beta v_1(0,s) - \beta \log p_1(0,s)$$

Recall that $v_1(0,s) = v_1(x+1,s)$, hence

$$v_2(x,2) - v_1(x,s) = \theta_1 x + \theta_2 s + \beta \log p_1(0,s) - \beta \log p_1(x+1,s),$$

and so

$$p_1(x,s) = \frac{1}{1 + \exp\big(\theta_1 x + \theta_2 s + \beta \log p_1(0,s) - \beta \log p_1(x+1,s)\big)}$$

Estimation

Finite mixture model: S types, probability of type s is π_s

Assume we have N buses over T periods.

Recall

$$p_1(x,s) = \frac{1}{1 + \exp\left(\theta_1 x + \theta_2 s + \beta \log p_1(0,s) - \beta \log p_1(x+1,s)\right)}$$

If we knew $\hat{p}_1(x,s)$, likelihood contribution is

$$l(d_{nt}|x_{nt}, s_n, \hat{p}_1, \theta) = \frac{d_{1nt} + d_{2nt} e^{\theta_1 x_{nt} + \theta_2 s_n + \beta \log \hat{p}_1(0,s_n) - \beta \log \hat{p}_1(x_{nt}+1,s_n)}}{1 + e^{\theta_1 x_{nt} + \theta_2 s_n + \beta \log \hat{p}_1(0,s_n) - \beta \log \hat{p}_1(x_{nt}+1,s_n)}}$$

ML estimator given by

$$(\hat{\theta}, \hat{\pi}) = \sum_{i=1}^{N} \log\left(\sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s_n, \hat{p}_1, \theta)\right)$$

Recovering separability

Observation n likelihood

$$g(d_n, x_n; \hat{p}_1, \theta, \pi) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s, \hat{p}_1, \theta)$$

Maximize in (θ, p)

$$\frac{1}{N} \sum_n \log g(d_n, x_n; \hat{p}_1, \theta, \pi) \qquad \text{s.t. } \sum_s \pi_s = 1$$

$\implies$ non linear, non-separable.

It follows

$$\hat{\pi}_s = \frac{1}{N} \sum_n \hat{q}_{ns} \qquad \text{with} \qquad \hat{q}_{ns} = \mathbb{P}(s|d_n, x_n, \hat{p}_1, \hat{\theta}, \hat{\pi}) = \frac{\hat{\pi}_s \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s, \hat{p}_1, \hat{\theta})}{g(d_n, x_n; \hat{p}_1, \hat{\theta}, \hat{\pi})}$$

and

$$\hat{\theta} = \arg\max \sum_n \sum_s \sum_t \hat{q}_{ns} \log l(d_{nt}|x_{nt}, s, \hat{p}_1, \theta)$$

$\implies$ linear and separable!

EM algorithm

EM as an alternative to estimating FIML. At Step m

1. Expectation

    1 Given values $\theta^{(m)}, \pi^{(m)}$, update

    $$q_{ns}^{(m+1)} = \frac{\pi_s^{(m)} \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s, \hat{p}_1, \theta^{(m)})}{\sum_{s'=1}^{S} \pi_{s'}^{(m)} \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s', \hat{p}_1, \theta^{(m)})}$$

    2 Update $\pi^{(m+1)}$

    $$\pi_s^{(m+1)} = \frac{1}{N} \sum_{i=1}^{N} q_{ns}^{(m+1)}$$

2. Maximization

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} q_{ns}^{(m+1)} \log l(d_{nt}|x_{nt}, s, \hat{p}_1, \theta^{(m)})$$

Keep iterating until convergence.

But we have assumed $\hat{p}_1$ is known...

EM algorithm with unobserved heterogeneity

Same as before but we also need to update $p_1^{(m)}(x,s)$ in the Expectation step

Can be shown that

$$p_1(x,s) = \frac{\mathbb{E}[d_{1nt} q_{ns} | x_{nt} = x]}{\mathbb{E}[q_{ns} | x_{nt} = x]}$$

Hence, at step m, update $p_1^{(m+1)}(x,s)$

$$p_1^{(m+1)}(x,s) = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T} d_{1nt} q_{ns}^{(m+1)} \mathbb{1}(x_{nt} = x)}{\sum_{n=1}^{N} \sum_{t=1}^{T} q_{ns}^{(m+1)} \mathbb{1}(x_{nt} = x)}$$

or

$$p_1^{(m+1)}(x,s) = l(d_{nt} = 1 | x_{nt}, s, p_1^{(m)}, \theta^{(m)})$$

Maximization becomes

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} q_{ns}^{(m+1)} \log l(d_{nt} | x_{nt}, s, p_1^{(m+1)}, \theta)$$

Rust (1987) with unobserved heterogeneity: recap

- CCP inversion used in

$$p_1(x,s) = \frac{1}{1 + \exp\left(v_2(x,s) - v_1(x,s)\right)}$$

- Finite Dependence used in

$$v_1(0,s) = v_1(x+1,s)$$

- EM algorithm gives better computational performance:
  - Recover separability in Maximization step

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} q_{ns}^{(m+1)} \log l(d_{nt}|x_{nt}, s, p_1^{(m+1)}, \theta)$$

  - Objective function becomes globally concave in θ

Outline

Motivation

Rust (1987) with unobserved heterogeneity

Monte Carlo Simulations

Using Hotz and Miller (1993) inversion in dynamic models

Conclusions

Monte Carlo Simulations: DGP

- Assume $s \in \{1, 2\}$ with probability 0.5 each.
- $u_2(x_{1t}, s) = \theta_0 + \theta_1 \min\{x_{1t}, 25\} + \theta_2 s$
- Need to specify how mileage evolves

$$f(x_{1t+1}|x_{1t}, j)$$

- Decision-maker lives 30 periods and makes decisions over 1000 buses.
- Derive value function by backward recursion.

Monte Carlo Simulations: Results

TABLE I

MONTE CARLO FOR THE OPTIMAL STOPPING PROBLEM[a]

		s Observed		Ignoring *s*	*s* Unobserved		Time Effects	
							s Observed	*s* Unobserved
	DGP	FIML	CCP	CCP	FIML	CCP	CCP	CCP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
θ_0 (intercept)	2	2.0100	1.9911	2.4330	2.0186	2.0280		
		(0.0405)	(0.0399)	(0.0363)	(0.1185)	(0.1374)		
θ_1 (mileage)	−0.15	−0.1488	−0.1441	−0.1339	−0.1504	−0.1484	−0.1440	−0.1514
		(0.0074)	(0.0098)	(0.0102)	(0.0091)	(0.0111)	(0.0121)	(0.0136)
θ_2 (unobs. state)	1	0.9945	0.9726		1.0073	0.9953	0.9683	1.0067
		(0.0611)	(0.0668)		(0.0919)	(0.0985)	(0.0636)	(0.1417)
β (discount factor)	0.9	0.9102	0.9099	0.9115	0.9004	0.8979	0.9172	0.8870
		(0.0411)	(0.0554)	(0.0591)	(0.0473)	(0.0585)	(0.0639)	(0.0752)
Time (minutes)		130.29	0.078	0.033	275.01	6.59	0.079	11.31
		(19.73)	(0.0041)	(0.0020)	(15.23)	(2.52)	(0.0047)	(5.71)

[a]Mean and standard deviations for 50 simulations. For columns 1–6, the observed data consist of 1000 buses for 20 periods. For columns 7 and 8, the intercept (θ_0) is allowed to vary over time and the data consist of 2000 buses for 10 periods. See the text and the Supplemental Material for additional details.

Using Hotz and Miller (1993) inversion in dynamic models

Using the inversion result in dynamic models

Assume state variable is z_t and choice-specific value function

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

Lemma

Define $p \equiv (p_1, ..., p_J)$, where $\sum_j p_j = 1$ and $p_j > 0$ for all j. Then, there exists a real-valued function $\psi_k(p)$ for every $k \in \{1, ..., J\}$ such that

$$\psi_k(p_t(z_t)) \equiv V_t(z_t) - v_{kt}(z_t)$$

Substituting in $v_{jt}(z_t)$

$$v_{jt}(z_t) = u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} \Big(v_{kt+1}(z_{t+1}) + \psi_k(p_{t+1}(z_{t+1})) \Big) f(z_{t+1}|j, z_t),$$

which is true for any choice k.

Keep doing this for a sequence of choices $(j_t, j_{t+1}, ..., j_{t+T})$

$$v_{jt}(z_t) = u_{jt}(z_t) + \sum_{\tau=t+1}^{T} \sum_{z_\tau=1}^{Z} \beta^{\tau-1} \Big(u_{j_\tau}(z_\tau) + \psi_{j_\tau}(p_\tau(z_\tau)) \Big) f(z_\tau|j_t, z_t)$$

Finite dependence

Define ρ–period dependence as

$$f(z_{t+\rho}|z_t, j) = f(z_{t+\rho}|z_t, j')$$

In that case, expand $v_{jt}(z_t)$ and $v_{j't}(z_t)$ until period ρ.

Taking differences

$$\begin{aligned} v_{jt}(z_t) - v_{j't}(z_t) = & u_{jt}(z_t) - u_{j't}(z_t) \\ & + \sum_{\tau=t+1}^{t+\rho} \sum_{z_\tau=1}^{Z} \beta^{\tau-1}\left(u_{j_\tau}(z_\tau) + \psi_{j_\tau}(p_\tau(z_\tau))\right)\Big(f(z_\tau|j, z_t) - f(z_\tau|j', z_t)\Big) \end{aligned}$$

$v_{jt}(z_t) - v_{j't}(z_t)$ almost in terms of observables. What about $\psi_{j_\tau}(p_\tau(z_\tau))$?

When ψ has a closed-form solution: Generalized EV Distributions

When ϵ_t is drawn from a GEV distribution

$$\epsilon_t \sim exp\Big[-H\Big(\exp(-\epsilon_{1t}), ..., \exp(-\epsilon_{Jt})\Big)\Big]$$

choice probabilities are given by

$$p_{jt}(z_t) = \frac{e^{v_{jt}(z_t)} H_j\Big(e^{v_{1t}(z_t)}, ..., \exp v_{Jt}(z_t)\Big)}{H\Big(e^{v_{1t}(z_t)}, ..., e^{v_{Jt}(z_t)}\Big)}$$

and it can be shown that $\psi_j(p)$ has a closed-form expression.

- In the case of Type I EV errors $H(Y) = \sum_j Y_j$ and $\psi_j(p) = -\log p_j$
- In the case of a nested logit $H(Y) = \sum_k \Big(\sum_{j\in k} Y_j^{\frac{1}{\lambda_k}}\Big)^{\lambda_k}$ and

$$\psi_j(p) = \gamma - \lambda_{k(j)} \ln(p_j) - (1 - \lambda_{k(j)}) \ln(\sum_{j' \in k} p_{j'})$$

Why is all of this useful?

Recall

$$v_{jt}(z_t) - v_{j't}(z_t) = u_{jt}(z_t) - u_{j't}(z_t) + \sum_{\tau=t+1}^{t+\rho} \sum_{z_\tau=1}^{Z} \beta^{\tau-1} \big(u_{j_\tau}(z_\tau) + \psi_{j_\tau}(p_\tau(z_\tau)) \big) \Big(f(z_\tau | j, z_t) - f(z_\tau | j', z_t) \Big)$$

and with GEV errors also closed-form solution for $\psi(p)$.

Two approaches

- Approach 1: estimate via ML
- Approach 2: estimate via regression equations

Approach 1: estimate via ML

1. $v_{jt}(z_t) - v_{j't}(z_t)$ has a closed-form solution.
2. Substitute inside likelihood
3. Estimate using FIML or EM.

Approach 2: estimate via regression equations

Recall

$$\psi_k(p_t(z_t)) = V_t(z_t) - v_{kt}(z_t) \implies v_{jt}(z_t) - v_{j't}(z_t) = \psi_{j'}(p_t(z_t)) - \psi_j(p_t(z_t))$$

1. With simple logit $\psi_j(p) = -\log p_j$

   $$\ln p_{jt}(z_t) - \ln p_{j't}(z_t) = v_{jt}(z_t) - v_{j't}(z_t) =$$

   $$u_{jt}(z_t) - u_{j't}(z_t) + \sum_{\tau=t+1}^{t+\rho} \sum_{z_\tau=1}^{Z} \beta^{\tau-1}\left(u_{j_\tau}(z_\tau) + \ln p_{j\tau}(z_\tau)\right)\left(f(z_\tau|j, z_t) - f(z_\tau|j', z_t)\right)$$

2. Similarly, with nested logit subsitute

   $$\psi_j(p) = \gamma - \lambda_{k(j)} \ln(p_j) - (1 - \lambda_{k(j)}) \ln(\sum_{j' \in k} p_{j'})$$

3. Needs extra structure or assumptions to deal with unobserved heterogeneity.

Conclusions

Conclusions

This paper shows us

- How to use HM inversion to write likelihood in terms of observables and parameters with unobserved heterogeneity
- Hot to estimate via ML using the likelihood above via EM
- How EM estimation outperforms in time FIML estimation
- How to use HM inversion more generally in dynamic models