Graduate Labor Economics

Discrete Choice Dynamic Programming

ScPo, Spring 2018

Based on Keane and Wolpin (2009)

March 7, 2018

Table of contents

Intro

- Dynamic Discrete Choice models have become a hallmark of empirical ecomomics.
- Applications in Labor, IO, healh, development, political economy, ...
- Aguirregabiria and Mira (2010) is an excellent survey
- Today we use Keane and Wolpin (2009) to develop a simple dynamic labor supply model.

Common Setup

- We focus on a binary choice $d_{it} \in 0, 1$. D_{it} is history of past choices.
- The latent variable v_{it} is the difference between payoffs.
- X_{it} are observed $\boxed{\text{state variables}}$, ϵ is unobserved (by us!)
- Whether v_{it} depends on entire D_{it} or just d_{it}, and how X evolves, determines whether static or dynamic model.

$$d_{it} = \begin{cases} 1 & \text{if } v_{it}(d_{it}, X_{it}, \epsilon_{it}) > 0 \\ 0 & \text{if } v_{it}(d_{it}, X_{it}, \epsilon_{it}) \leq 0 \end{cases}$$

Labor Supply of Married Women

- Consider a static model: no intertemporal dependencies.
- Utility for married woman i in period t from working (option 1) vs not working (option 0) with n_i small children is

$$U_{it}^1 = y_{it} + w_{it} - \pi n_{it} \tag{1}$$
$$U_{it}^0 = y_{it} + x_{it}\beta + \epsilon_{it} \tag{2}$$

- where y_{it} is the husband's income. Let's write the difference in those utilities as $U_{it}^1 - U_{it}^0$

Latent Value

$$v_{it}(x_{it}, w_{it}, n_{it}, \epsilon_{it}) = w_{it} - \pi n_{it} - x_{it}\beta - \epsilon_{it}$$

- define the work indicator $d_{it} = \mathbf{1}[U_{it}^{1} > U_{it}^{0}]$
- Observed state space: $\Omega_{it} = (x_{it}, w_{it}, n_{it})$
- Unobserved by us: ϵ

Discrete Choice

- This is a threshold-crossing problem.
- Woman i will work in t if $U_{it}^1 > U_{it}^0$
- I.e. if $v_{it}(x_{it}, w_{it}, n_{it}, \epsilon_{it}) > 0$. At $v_{it}(x_{it}, w_{it}, n_{it}, \epsilon_{it}) = 0$ she is indifferent
- Call the ϵ that solves this *the critical epsilon* $\epsilon^*(\Omega_{it})$.

$$i \text{ chooses to } \begin{cases} \text{work in t} & \text{if } \epsilon_{it} < \epsilon^*(\Omega_{it}) \Rightarrow U_{it}^1 > U_{it}^0 \\ \text{not work in t} & \text{if } \epsilon_{it} > \epsilon^*(\Omega_{it}) \Rightarrow U_{it}^1 < U_{it}^0 \end{cases}$$

Setup

- Assume ϵ is independent of Ω
- then,

$$\Pr[d_{it} = 1|\Omega it] = \int_{-\infty}^{\epsilon_{it}^*} dF_\epsilon(\epsilon|\Omega it) = \int_{-\infty}^{\epsilon_{it}^*} dF_\epsilon(\epsilon)$$

- We have $\Pr[d_{it} = 0|\Omega it] = 1 - \Pr[d_{it} = 1|\Omega it]$
- Likelihood for a random sample of N females observed for T periods is

$$L(\beta, \pi, F_\epsilon; x_{it}, w_{it}) = \Pi_{i=1}^N \Pi_{t=1}^T \Pr[d_{it} = 1|\Omega it]^{d_{it}} \Pr[d_{it} = 0|\Omega it]^{1-d_{it}}$$

Notice that there are no dynamics in the model up to now!

Slightly More Realistic Wage

Potential Experience

- let's add *potential experience*
- potential experience: (age - educ - 6)

$$w_{it} = \gamma z_{it} + \eta_{it}$$

Now:

$$\begin{aligned} v_{it}(x_{it}, z_{it}, n_{it}, \epsilon_{it}, \eta_{it}) &= \gamma z_{it} - \pi n_{it} - x_{it}\beta + \eta_{it} - \epsilon_{it} && (3) \\ &= \xi^*_{it}(\Omega_{it}) + \xi_{it} && (4) \end{aligned}$$

where $\xi = \eta - \epsilon$ is your new composite iid error, and $\xi^*_{it}(\Omega_{it}) = \gamma z_{it} - \pi n_{it} - x_{it}\beta$.

Likelihood conditional on wage

- The likelihood now is the likelihood of observing work, $d_{it} = 1$, and a certain wage w_{it}.

$$L(\beta, \pi, F_\epsilon; x_{it}, w_{it}) = \Pi_{i=1}^{N} \Pi_{t=1}^{T} \Pr[d_{it} = 1, w_{it} | \Omega it]^{d_{it}} \times \Pr[d_{it} = 0 | \Omega it]^{1-d_{it}} \quad (5)$$

and we have

$$\Pr[d_{it} = 1, w_{it} | \Omega it] = \Pr[d_{it} = 1, \eta_{it} = w_{it} - \gamma z_{it}]$$

Identification

- We have two latent processes:

$$w_{it} = \begin{cases} \gamma z_{it} + \eta_{it} & \text{if work} \\ 0 & \text{else} \end{cases}$$

$$d_{it} = \begin{cases} 1 & \text{if } \eta_{it} - \epsilon_{it} = \xi_{it} > -\xi^*_{it}(\Omega_{it}) \\ 0 & \text{else.} \end{cases}$$

- If we Assume that (ϵ, η) are joint normally distributed with

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Delta = \begin{pmatrix} \sigma^2_\epsilon & \cdot \\ \sigma_{\epsilon\eta} & \sigma^2_\eta \end{pmatrix}$$

then we get a standard Heckman selection model.

Identification

Identified Parameters

1. β_π and π are not separately identified: set $\pi = 0$.
2. Want to identify $\beta, \gamma, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_{\epsilon\eta}$.
3. Let's remind ourselves of the workings of the Heckman model first.

Suppose our latent process of interest is

$$y_{1i}^* = x_{1i}'\beta_1 + u_{1i}$$

with $E(u_1|x_1) = 0$. We observe y_1 as

$$y_{1i} = \begin{cases} y_{1i}^* & \text{if } y_{2i}^* = x_{2i}'\beta_2 + u_{2i} > 0 \\ 0 & \text{else.} \end{cases}$$

Heckman Model

- We can't run OLS on the selected sample where $y^*_{2i} > 0$:

$$\begin{aligned} E(u_{1i}|y^*_{2i} > 0) &= E(u_{1i}|x'_{2i}\beta_2 + u_{2i} > 0) \\ &= E(u_{1i}|u_{2i} > -x'_{2i}\beta_2) \\ &\neq 0 \end{aligned} \tag{6}$$

  if u_1, u_2 correlated.

- But with our joint normaltiy assumption, we can write

$$E[u_1|u_2] = 0 + \frac{\sigma_{12}}{\sigma_2^2} u_2$$

- hence, any u_{1i} can be decomposed into a conditional mean and an error:

$$u_{1i} = \frac{\sigma_{12}}{\sigma_2^2} u_{2i} + \epsilon_{1i}, \text{with } \epsilon_{1i} \sim \mathcal{N}(0, \sigma_\epsilon^2), \text{independent of } u \tag{7}$$

Heckman Selection

- Substitute (7) into (6):

$$
\begin{aligned}
E(u_{1i}|y_{2i}^* > 0) &= E(\frac{\sigma_{12}}{\sigma_2^2}u_{2i} + \epsilon_{1i}|u_{2i} > -x_{2i}'\beta_2) \\
&= \frac{\sigma_{12}}{\sigma_2^2}E(u_{2i}|u_{2i} > -x_{2i}'\beta_2) + E(\epsilon_{1i}|u_{2i} > -x_{2i}'\beta_2) \\
&= \frac{\sigma_{12}}{\sigma_2^2}E(u_{2i}|u_{2i} > -x_{2i}'\beta_2)
\end{aligned}
$$

Heckman Selection

- Then use a well-known result about truncated normals:

$$
\begin{aligned}
\frac{\sigma_{12}}{\sigma_2^2} E(u_{2i} | u_{2i} > -x'_{2i}\beta_2) &= \frac{\sigma_{12}}{\sigma_2} \frac{\phi\left(\frac{-x'_{2i}\beta_2}{\sigma_2}\right)}{1 - \Phi\left(\frac{-x'_{2i}\beta_2}{\sigma_2}\right)} \\
&= \frac{\sigma_{12}}{\sigma_2} \frac{\phi\left(\frac{x'_{2i}\beta_2}{\sigma_2}\right)}{\Phi\left(\frac{x'_{2i}\beta_2}{\sigma_2}\right)} \\
&= \frac{\sigma_{12}}{\sigma_2} \lambda\left(\frac{x'_{2i}\beta_2}{\sigma_2}\right)
\end{aligned}
$$

- $\lambda(\cdot)$ is the **inverse Mills ratio**.
- So we identify $\frac{\beta_2}{\sigma_2}$. Heckman model usually imposes $\sigma_2 = 1$.

Identification Again

- Similarly here:
- Work choices identify $\frac{\beta}{\sigma_\xi}, \frac{\gamma}{\sigma_\xi}$

$$
\begin{aligned}
\Pr[d_{it} = 0] &= \Pr[\xi_{it} < -(z_{it}\gamma - x_{it}\beta)] \\
&= \Phi\left(-\frac{z_{it}\gamma - x_{it}\beta}{\sigma_\xi}\right)
\end{aligned}
$$

- The wage equation identifies γ and σ_η^2
- Identifying σ_ξ requires an exclusion restriction.

Exclusion Restriction

- In the likelihood function (5) there are 3 types of variables.
  1. things only in z (i.e. wage-related)
  2. things only in x (i.e. leisure-related)
  3. things in both.
- with (wage params) γ in hand, we need at least 1 element in z that is **not** in x to identify σ_{ξ} and $\sigma_{\eta\epsilon}$
  - here, either education or experience should not affect leisure.

Why Assume Structure?

Part 1

- One could estimate $\Pr[d_{it} = 0]$ non-parametrically, right?
- True. **But**, separating budget set from preferences allows us to do counterfactual analysis.
- Suppose we want to know effect of implementing childcare subsidy. Couple gets $\tau > 0$ dollars if woman works.
- new budget of couple is then:

$$C_{it} = w_{it}d_{it} + y_{it} + \tau d_{it}n_{it}$$

Why Assume Structure?

Part 2

- Then resulting probability of work is

$$\begin{aligned}\Pr[d_{it} = 1 | z_{it}, x_{it}, n_{it}] &= \Pr[\xi_{it} > -(z_{it}\gamma - x_{it}\beta + \tau n_{it})] \\ &= \Phi\left(\frac{z_{it}\gamma - x_{it}\beta + \tau n_{it}}{\sigma_\xi}\right)\end{aligned}$$

- Without an estimate of σ_ξ we cannot say $\boxed{\text{anything}}$ about the expected effect of the reform!

Introduction

A Static Model

Identification I

Heckman Model

Identification II

Dynamic Model

Dynamic Version

- Up until now, there was no connection between **today's** choices and **tomorrow's** value.
- Even if we imposed forward-looking behaviour, this was a static optimization problem.
- Now suppose the woman's wage increases with experience h:

$$w_{it} = z_{it}\gamma_1 + h_{it}\gamma_2 + \eta_{it} \tag{8}$$

where $h_{it} = \sum_{j=1}^{j=t-1} d_{ij}$

- Clearly working today has implications for the value of tomorrow (through higher wage).

Dynamic Model Setup

- We define the remaining lifetime utility at age t as

$$V_t(\Omega_{it}) = \max_{d_{it}} \mathbb{E} \left\{ \sum_{j=t}^{j=T} \delta^{j-t} \left(U_{it}^1 d_{it} + U_{it}^0 (1 - d_{it}) \right) | \Omega_{it} \right\} \tag{9}$$

- with Ω as before plus h_{it}, and $h_{it} = h_{it-1} + d_{it-1}$
- We can write the value function also as

$$V_t(\Omega_{it}) = \max(V_t^0, V_t^1)$$

with

$$V_t(\Omega_{it})^j = U_{it}^j + \delta \mathbb{E} \left[V_{t+1}(\Omega_{it+1}) | \Omega_{it}, d_{it} = j \right], j = 0, 1$$

Dynamic Model

Latent Variable Formulation

- Similarly to before, the latent variable is

$$
\begin{aligned}
v_{it}(x_{it}, z_{it}, h_{it}, n_{it}, \epsilon_{it}, \eta_{it}) &= \gamma_1 z_{it} + \gamma_2 h_{it} - x_{it}\beta + \eta_{it} - \epsilon_{it} \\
&+ \delta \mathbb{E}\left[V_{t+1}(\Omega_{it+1}) | \Omega_{it}, d_{it} = 1\right] \\
&- \delta \mathbb{E}\left[V_{t+1}(\Omega_{it+1}) | \Omega_{it}, d_{it} = 0\right] \\
&= \xi^*_{it}(\Omega_{it}) + \xi_{it} \qquad (10)
\end{aligned}
$$

- The only difference between (10) and (3) is the difference in future values.
- So, estimation would proceed as in the static case...
- ... after having computed $\mathbb{E}\left[\max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it}))\right]$ at all Ω_{it}.

Estimation of the Dynamic Model

- Suppose we have a panel including h_{it} of work spells (t_{1i}, t_{Li})
- Our likelihood function becomes

$$L(\theta, x_{it}) = \Pi_{i=1}^{N} \Pi_{j=t_{1i}}^{t_{Li}} \Pr(d_{ij} = 1, w_{ij}|\Omega_{ij})^{d_{ij}} \Pr(d_{ij} = 0|\Omega_{ij})^{1-d_{ij}} \tag{11}$$

- and as before,

$$\Pr(d_{ij} = 1, w_{ij}|\Omega_{ij}) = \Pr(\xi_{ij} \geq -\xi_{ij}^{*}(\Omega_{ij}), \eta_{ij} = w_{ij} - z_{ij}\gamma_1 - \gamma_2 h_{ij})$$
$$\Pr(d_{ij} = 0|\Omega_{ij}) = 1 - \Pr(\xi_{ij} \geq -\xi_{ij}^{*}(\Omega_{ij}))$$

Identification of Dynamic Model

- The difference in future values in (10) is a non-linear function G:

$$\begin{aligned} v_{it}(x_{it}, z_{it}, h_{it}, n_{it}, \epsilon_{it}, \eta_{it}) = &\gamma_1 z_{it} + \gamma_2 h_{it} - x_{it}\beta + \eta_{it} - \epsilon_{it} \\ &+ \delta G(z_{it}, h_{it}, x_{it}) \end{aligned}$$

- Functional/distributional form assumptions on G alone may be enough to identify δ.
- We require the same exclusion restriction as before for identification not based on functional form.
- Experience h should not affect leisure.

References

Michael P. Keane and Kenneth I. Wolpin. Empirical applications of discrete choice dynamic programming models. *Review of Economic Dynamics*, 12(1):1 – 22, 2009. ISSN 1094-2025. doi: https://doi.org/10.1016/j.red.2008.07.001. URL `http://www.sciencedirect.com/science/article/pii/S1094202508000318`.

Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38 – 67, 2010. ISSN 0304-4076. doi: DOI:10.1016/j.jeconom.2009.09.007. URL `http://www.sciencedirect.com/science/article/B6VC0-4X7R85S-3/2/b78d013f2b1f3dbce3f2325f0c41d4e4`. Structural Models of Optimization Behavior in Labor, Aging, and Health.