Dynamic discrete choice models: extensions on assignment

Olivier De Groote

TSE

MRes: EEE, 2021-2022

Outline

Introduction I

- CCP estimation has been used in several papers in education, especially after a crucial contribution by Arcidiacono and Miller (2011)
- Arcidiacono and Miller (2011) allows us to add Heckman and Singer (1984) types to a CCP estimator
- Why so important in education/labor?
- Today I'll add unobserved types to the model you had in the assignment

Model

- 2 period model
- In $t = 1$, can go to high school ($j = 1$) and get utility

$$u_{HS}(x_i) + \varepsilon_{i11}.$$

- x_i are time-invariant state variables, ε_{ijt} is an extreme value type 1 distributed taste shock
- After $t = 1$, i can obtain a high school degree ($g_i = 1$) if

$$f(x_i) + \eta_i > 0$$

with $f(x_i)$ an estimated index and η_i a logit shock, realized in $t = 2$.

Model

- The probability to obtain a degree is then

$$\Pr(g_i|x_i) = \frac{\exp(f(x_i))}{1+\exp(f(x_i))}.$$

- Students that drop out in $t = 1$ never return to school and receive the lifetime utility of dropping out, normalized to 0
- In $t = 2$, students that obtained a high school degree have the option to stay in school ($j = 1$) by going to college and get utility

$$u_c(x_i) + \varepsilon_{i12}.$$

Model

- The probability to go to school in $t = 2$ for students with a high school degree is given by:

$$\Pr(d_{i2} = 1 | x_i, g_i = 1) = \frac{\exp(u_c(x_i))}{1 + \exp(u_c(x_i))} .$$

- The conditional value function of choosing high school in $t = 1$

$$v_{hs}(x_i) = u_{hs}(x_i) + \beta \left(\gamma + \Pr(g_i | x_i) \ln \left(1 + \exp u_c(x_i) \right) \right)$$

- and

$$\Pr(d_{i1} = 1 | x_i) = \frac{\exp(v(x_i))}{1 + \exp(v(x_i))} .$$

Model

- Note that we could also write the conditional value function in CCP form

$$v_{hs}(x_i) = u_{hs}(x_i) + \beta \left(\gamma + \Pr(g_i|x_i) \left(-\ln \left(\Pr(d_{i2} = 0|x_i, g_i = 1) \right) \right) \right)$$

- This is especially useful if the model of college goes on for multiple periods as there is no need to solve it before estimation, we can use predictions (without requiring structural model) of $\Pr(d_{i2} = 0|x_i, g_i = 1)$ and proceed to estimation of high school utility
- We could proceed similarly in college
- Because of the 2 period nature of the model, we actually assumed:

$$v_c(x_i) = u_c(x_i)$$

Model

- Say we change the model and allow students to keep on making choices from period 3 on
- Let's say that in period 3, students can choose to continue college, or drop out, with the same functional form assumptions as before
- We do not say what happens after period 3 (but remember that drop out is terminal)

Model

- We could then write the conditional value function of college in period 2 as follows:

$$v_{c2}(x_i) = u_c(x_i) + \beta\left(\gamma + \left(-\ln\left(\Pr(d_{i3} = 0|x_i, d_{i2} = c)\right)\right)\right)$$

- $=>$ if we observe data in period 3, we can obtain some (reduced form) prediction for the CCP and therefore estimate $u_c(x_i)$ which should now be interpreted as the flow utility of college during one year only because $\beta\left(\gamma + \left(-\ln\left(\Pr(d_{i3} = 0|x_i, d_{i2} = c)\right)\right)\right)$ is capturing the expected utility after that

Model I

- What if we want to allow for an unobserved type ν_i such that $u_{hs}(x_i, \nu_i)$,$u_c(x_i, \nu_i)$ and $f(x_i, \nu_i)$?
- See Arcidiacono and Miller (2011)
- You can think of the likelihood contribution of one observation when adding unobserved types in Arcidiacono (2005) was of the form

$$L_i(\alpha, \gamma, \pi)$$
$$= \sum_{r=1}^{R} \pi_r \prod_t L_{itr}^{choice}(\alpha, \gamma) \times L_{itr}^{transition}(\gamma)$$

- Here: $L_{itr}^{transition}$ is given by the probability of a obtaining a degree, while $L_{itr}^{choice}(\alpha, \gamma)$ comes from the choice model

Model II

- Arcidiacono & Jones (2003) show that this can be estimated using an EM algorithm

- Start from arbitrary parameter values (α^0,γ^0,π_r^0)

- Step 1: Calculate the probability to belong to each type, conditional on all data for i ($\mathbf{X_i}$) and parameters

$$Pr(r|\mathbf{X_i}, \alpha^0, \gamma^0, \pi^0) = \frac{\pi_r \prod_t L_{itr}^{choice} \times L_{itr}^{transition}}{\sum_{r'=1}^{R} \pi_{r'} \prod_t L_{itr'}^{choice} \times L_{itr'}^{transition}}$$

Model III

- Step 2: Find new parameters using the expected log-likelihood function, holding the conditional probabilities fixed

$$\sum_i \sum_{r=1}^{R} Pr(r|\mathbf{X_i}, \theta^0, \pi^0) \left(\sum_t \ln L_{itr}^{choice}(\alpha^1, \gamma^1) + \sum_t \ln L_{itr}^{transition}(\gamma^1) \right)$$

- Repeat until convergence
- With a CCP estimator: L_{itr}^{choice} depends on a type-specific CCPs, so we also use $Pr(r|\mathbf{X_i}, \theta^0, \pi^0)$ as weights to obtain a type-specific CCP before step 1, and update it each round
- Arcidiacono & Miller (2011) also propose an alternative estimator where L_{itr}^{choice} in step 2 is not coming from the structural model, but is replaced by CCPs

Some applications in education

- Arcidiacono P., Aucejo, E., Maurel, A. and Ransom, T. "College Attrition and the Dynamics of Information Revelation." R&R JPE (2016).
  - Allows for learning about ability by additional vector A to be revealed through grades and wages
  - Uses a different form of finite dependence to not have to assume that drop out is a terminal action
- Closer to example above
  - Declercq, K., and Verboven, F. "Enrollment and Degree Completion in Higher Education without Admission Standards." Economics of Education Review 66 (October 2018): 223–44
- Like above but endogenizing state transitions through effort
  - De Groote, O. "A Dynamic Model of Effort Choice in High School." TSE Working Paper - 1002, 2020.

Outline

A black box approach

- Remember the likelihood function of a static model using

$$Pr(d_{ijt} = 1|x_{it}) = \frac{exp(u_{ijt})}{\sum_{j'} exp(u_{ij't})}$$

- with $u_{ijt} = u_j(x_{it})$ to estimate
- Dynamic models complicated things because now we have

$$Pr(d_{ijt} = 1|x_{it}) = \frac{exp(v_{ijt})}{\sum_{j'} exp(v_{ij't})}$$

- with

$$v_{ijt} = u_j(x_{it}) + \beta \int \bar{V}_{t+1}(x_{it+1}) f(x_{it+1}|x_{it}, d_{it}) dx_{it+1}$$

- or with unobserved types

$$v_{ijt} = u_j(x_{it}, \nu_i) + \beta \int \bar{V}_{t+1}(x_{it+1}, \nu_i) f(x_{it+1}|x_{it}, \nu_i, d_{it}) dx_{it+1}$$

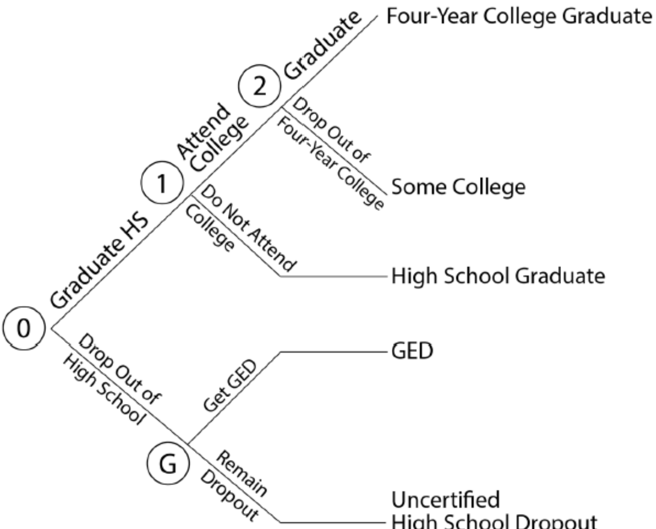
Introduction

- But what if we directly estimate $v_{ijt} = v_j(x_{it}, \nu_i)$?
- What can we still do?
- What is no longer possible?
- When is this better/worse/easier?

Introduction

- Paper: Heckman, Humphries and Veramendi (JPE 2018): Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking
- Mix between structural DDC and reduced-form TE literature
- Interested in returns to education in a broad sense
  - High school and college
  - Monetary and non-monetary outcomes
  - Heterogeneity in TE
  - Policy-relevant simulations
- Importantly: interested in impact of education on outcomes, not the reverse!

Model: overview



Model: components

- Potential outcome k in final state s

$$Y_s^k = \left\{ \begin{array}{cc} \widetilde{Y}_s^k & \text{if ctu} \\ 1\left(\widetilde{Y}_s^k\right) \geq 0 & \text{if binary} \end{array} \right\}$$

with

$$\widetilde{Y}_s^k = \tau_s^k(X) + U_s^k$$

Selection equation to go from node j to the next one

$$I_j = \phi_j(Z) - \eta_j$$

- Interpretation $\phi_j(Z)$?

Model: unobservables

$$\eta_j = -(\theta' \lambda_j - v_j)$$

$$U_s^k = \theta' \alpha_s^k + \omega_s^k$$

- θ : unobserved endowments
- Assume conditional independence: equations are related but only through Z, X, θ
- We have used this before, in 2-type model θ is dummy to be type 2
- Here: richer and with interpretation by allowing for two "factors"
  - Cognitive skills
  - Non-cognitive skills
- How? adding additional equations of noisy measurements as functions of Z, X, θ and measurement error

Discussion

- "Counterfactuals" look for treatment effects: what if we make someone go to a certain option, how does this change the outcome?
- No need to solve backwards, only forward
- Is this structural?
- Which assumptions were different/avoided? (see in particular p38 and compare to what you did in assignment)
- What can we still do?
- What is no longer possible?
- Tip in STATA: check out "GSEM" and "heterofactor"
- Other examples using this (more focused on application instead of method)
  - Same but on other outcomes: Heckman, Humphries, Veramendi (JHC 2018)
  - Changes in education and experience premium: Ashworth et al (JOLE 2020)

Outline

Introduction

- Think of data, similar to your assignment
  - Educational choices (going to school)
  - Performance (high school degree)
  - Student characteristics
- Can identify utility and state transitions
- But what if study effort changes in counterfactuals?
  - Observe it and include it: Todd and Wolpin (2018), Fu and Mehta (2018)
  - Change the primitives: De Groote (2020)
- Today
  - Focus on simplified model to explain methodological contribution of De Groote (2020)

Model

- 2 period model (similar to your assignment)
- In $t = 1$, can go to high school ($j = 1$) and get utility

$$u(x_i) + \varepsilon_{i11}.$$

- x_i are time-invariant state variables, ε_{ijt} is an extreme value type 1 distributed taste shock
- After $t = 1$, i can obtain a high school degree ($g_i = 1$) if

$$f(x_i) + \eta_i > 0$$

with $f(x_i)$ an estimated index and η_i a logit shock, realized in $t = 2$.

Model

- The probability to obtain a degree is then

$$\Pr(g_i|x_i) = \frac{\exp(f(x_i))}{1+\exp(f(x_i))}.$$

- Students that drop out in $t = 1$ never return to school and receive the lifetime utility of dropping out, normalized to 0

- In $t = 2$, students that obtained a high school degree have the option to stay in school ($j = 1$) by going to college and get utility

$$\Psi(x_i) + \varepsilon_{i12}.$$

Model

- The probability to go to school in $t = 2$ for students with a high school degree is given by:

$$\Pr(d_{i2} = 1 | x_i, g_i = 1) = \frac{\exp(\Psi(x_i))}{1 + \exp(\Psi(x_i))} .$$

- The conditional value function of choosing high school in $t = 1$

$$v(x_i) = u(x_i) + \beta \left(\gamma + \Pr(g_i | x_i) \ln\left(1 + \exp \Psi(x_i)\right)\right)$$

- and

$$\Pr(d_{i1} = 1 | x_i) = \frac{\exp(v(x_i))}{1 + \exp(v(x_i))} .$$

Model

- Now do a counterfactual: increase in $\Psi(x_i)$, what changes and what doesn't?
- Assume now a different utility function of going to high school:
$$u(x_i, y_i) = -C^0(x_i) - c(x_i)y_i$$
- with y_i "effective" study effort by defining the probability to obtain a degree as
$$\phi(y_i) = \frac{y_i}{1 + y_i}$$
- and conditional value functions
$$v(x_i, y_i) = u(x_i, y_i) + \beta\left(\gamma + \phi(y_i)\ln\left(1 + \exp\Psi(x_i)\right)\right)$$

Model

- Now choose y_i optimally:

$$\frac{\partial v_i(x_i, y_i)}{\partial y_i} = \frac{\partial u(x_i, y_i)}{\partial y_i} + \beta \left(\frac{\partial \phi(y_i)}{\partial y_i} \ln\left(1 + \exp \Psi(x_i)\right) \right) = 0$$
$$\text{if } y_i = y_i^*$$

with $\frac{\partial u(x_i, y_i)}{\partial y_i} = -c(x_i)$ and $\frac{\partial \phi(y_i)}{\partial y_i} = (1 + y_i)^{-2}$. Then

$$y_i^* = \sqrt{\frac{\beta \ln\left(1 + \exp \Psi(x_i)\right)}{c(x_i)}} - 1.$$

- What if we know $C^0(x_i)$, $c(x_i)$, $\Psi(x_i)$ and run the counterfactual?

Identification

- Standard model: state transitions nonparametrically identified and flow utility identified after specifying (Magnac and Thesmar, 2002)
  - Utility of $j = 0$ normalized to 0
  - Distribution of ε_{ijt}
  - Discount factor β
- Here: same data, same number of primitives but now these are fixed costs $C^0(x_i)$ and marginal costs $c(x_i)$ instead of utility functions $u(x_i)$ and state transitions $f(x_i)$

Identification

- Make use of the FOC at y_i to identify $c(x_i)$:

$$c(x_i) = \beta \left((1 + y_i)^{-2} \ln \left(1 + \exp \Psi(x_i)\right)\right) \text{ if } y_i = y_i^* \qquad (1)$$

- $\Psi(x_i)$ can be found using data in period 2
- Optimal choice of $y_i = y^*(x_i)$ in the data can be found using $\phi(y_i) = \frac{y_i}{1+y_i}$ because $\phi(y^*(x_i)) = \Pr(g_i|x_i)$
- Once you have $c(x_i)$ and $y^*(x_i)$, $u(x_i, y_i^*)$ only depends on $C^0(x_i)$
- Estimate $C^0(x_i)$ using the logit choice probabilities in period 1 with

$$v(x_i) = -C^0(x_i) - c(x_i)y_i^* + \beta \left(\gamma + \phi(y_i^*) \ln \left(1 + \exp \Psi(x_i)\right)\right)$$

References I

Peter Arcidiacono.
Affirmative action in higher education: How do admission and financial aid rules affect future earnings?
Econometrica, 73(5):1477–1524, 2005.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom.
College Attrition and the Dynamics of Information Revelation.
Working paper, 2016.

Peter Arcidiacono and Robert A. Miller.
Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity.
Econometrica, 79(6):1823–1867, 2011.

References II

Olivier De Groote.
A Dynamic Model of Effort Choice in High School.
TSE working paper - 1002, 2020.

Koen Declercq and Frank Verboven.
Enrollment and degree completion in higher education without admission standards.
Economics of Education Review, 66:223–244, October 2018.

Chao Fu and Nirav Mehta.
Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation.
Journal of Labor Economics, 36(4), October 2018.

References III

J. Heckman and B. Singer.
A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data.
Econometrica, 52(2):271, March 1984.

James J Heckman, John Eric Humphries, and Gregory Veramendi.
The Nonmarket Benefits of Education and Ability.
Journal of Human Capital, 12(2):282–304, 2018.

James J Heckman, John Eric Humphries, and Gregory Veramendi.
Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking.
Journal of Political Economy, 126:51, 2018.

References IV

Thierry Magnac and David Thesmar.
Identifying dynamic discrete decision processes.
Econometrica, 70(2):801–816, March 2002.

Petra E Todd and Kenneth I Wolpin.
Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom.
Journal of Political Economy, (forthcoming), 2018.