Advanced Applied Econometrics: Structural econometric techniques in labor economics

Maximilian Blesch & Maxi Schaller

July 5, 2025

Part III: Labor Economics - Lectures

Overview - Dynamic discrete choice models:

- Lecture 1: Keane and Wolpin (1997)
  The Career Decisions of Young Men. Journal of Political Economy 105 (3), 473-522.
- Lecture 2: Blundell et al. (2016)
  Female Labour Supply, Human Capital and Welfare Reform. Econometrica 84(5), 1705-1753.

Part III: Labor Economics - Exercises

Exercises:

- Exercise 1 (today): Static discrete choice of labor supply
  - Static choice of working hours from finite set of alternatives
  - Simulation of simple dataset; recovering parameters using maximum likelihood estimation
- Exercise 2: Dynamic discrete choice of labor supply I
  - Simple finite-horizon Robinson economy with deterministic evolution of states
  - Implementation of state-space, indexer and model solution with discrete choice dynamic programming techniques (backwards induction)
- Exercise 3 (**graded**): Dynamic discrete choice of labor supply II
  - More complex state-space; introduction of uncertainty/stochastic elements
  - Expectation maximization, multidimensional numerical integration, estimation using simulated method of moments

Organization

Graded Problem Set 4 - Labor:

- Due date: **August 15**
- Additional tutorial: coordinate date next week!

Exam:

- Date: **July 18, 9am**; Elinor Ostrom Hall; duration 2hrs
- All material covered in the course is relevant for the exam.
- 1/2 Part I (FW), 1/2 Part II: IO (HU) & labor (PH)
- On Part II: There will be no explicit questions on Python programming, however, you need to be able discuss the conceptual implementation of the models and methods discussed in the problem sets.
- Materials allowed:
  - Hand-written or printed notes: 10 pages, double-sided, A4. **No** lecture slides/papers etc.
  - **No** electronic devices
  - Non-programmable calculator
- **Master students** receive the same exam questions. Their final grade will only based on the exam. Passing the graded problem sets is sufficient.

Exercise 1: Static discrete choice labor supply

- Stochastic discrete choice labour supply models are very popular in empirical research and have frequently been used as a tool to perform (tax) policy analysis using counterfactual simulations.
  $\Rightarrow$ see next week's lecture!
- Today: basic static discrete choice model of working hours decisions, abstracting from any demographic heterogeneity
  $\Rightarrow$ Recall consumption-leisure models: utility maximization problem with budget constraint

Setup - Choice alternatives

- Why discrete choice instead of continuous choice models?
  - Tax-transfer systems: Non-linearities and non-convexities in budget sets
  - Empirical reality/institutional constraints: Many jobs offer fixed hour contracts; hours cluster at specific values (20, 40 hours)
  - Computational advantages: Easier integration, faster estimation (closed-form solutions given assumptions on unobservables)

- **Here:** Individuals choose working hours per week from finite set of alternatives $h \in \{0, 10, 20, 30, 40\}$.
- Considerations for specification of choice set?
  - Review empirical distribution of working hours in the target sample
  - Capture observed clustering in data (e.g., 40-hour work week; differences by gender)
  - Parsimony for computational tractability (not relevant in this simple exercise)

Setup - Utility

Random utility model:

- In a static setting: individuals i choose working hours per week from a finite set of alternatives $j = 1, ..., 5$ associated with (avg.) working hours $h \in \{0, 10, 20, 30, 40\}$.
- Preferences over these discrete alternatives may be described by a parametric utility function:

$$U_{ij} = \gamma \left[\frac{c_i^\theta}{\theta} - \alpha \ h_{ij} \right] + \varepsilon_{ij} \quad \forall i, j$$

Notes:

- CRRA preferences over consumption with linear disutility of working hours
- θ coefficient of relative risk aversion; α captures disutility of working hours
- ε_{ij} are the unobserved components of utility (e.g. preference shocks)
  $\Rightarrow$ need to make assumption about distribution of unobervables
- γ weight on systematic utility from consumption and leisure rel. to preference shocks

Discrete choice - Random utility models

Random utility models (RUM) are a class of models that describe the choice of an individual i from a finite set of alternatives $j = 1, ..., J$ as follows:

$$U_{ij} = x'_{ij}\beta + \varepsilon_{ij} \quad \forall i = 1, ..., N \quad j = 1, ..., J$$

$$\begin{aligned} \Pr(y_i = j) &= \Pr(U_{ij} = \max\{U_{i1}, ..., U_{iJ}\}) \\ &= \Pr(x'_{ij}\beta + \varepsilon_{ij} > \max_{k \neq j}\{x'_{ik}\beta + \varepsilon_{ik}\}) \end{aligned}$$

Multinomial Probit (MNP): assume unobservables are joint normal $\varepsilon_{ij} \sim N(0, \Sigma)$, where Σ is a $J \times J$ covariance matrix.

$$\Pr(y_i = j) = \Pr(\varepsilon_{i1} - \varepsilon_{ij} < x'_{ij}\beta - x'_{i1}\beta, \forall k \neq j)$$

... given by cumulative probability of a (J-1) variate normal distribution. Estimation of this model (obtaining choice probabilities) is computationally expensive, as it requires numerical integration over the joint distribution of the unobservables.

Discrete choice - Random utility models

Multinomial Logit/Conditional Logit (MNL/CL): assume unobservables are independent and identically distributed (IID) Type I extreme value (Gumbel) distributed. (independent across individuals & alternatives!)

$$\mathrm{CDF}_{\text{EV-I}}(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}} \quad \forall i, j$$

Economic foundation by McFadden (1973): differences between errors in RUM follow the logit distribution. Maximization of stochastic utility function implies closed form solution for choice probabilities (proof see Train, 2009 (p.85)):

$$\Pr(y_i = j) = \frac{e^{x_{ij}'\beta}}{\sum_{k=1}^{J} e^{x_{ik}'\beta}} \quad \forall i, j$$

Setup

Utility function:

$$U_{ij} = \gamma \left[\frac{c_i^\theta}{\theta} - \alpha \; h_{ij} \right] + \varepsilon_{ij} \quad \forall i, j$$

... where unobserved components of utility ε_{ij} are assumed to follow a Type-I extreme value distribution.

Budget constraint and wage equation:

$$\begin{aligned} c_i &= y_i + \omega_i \cdot h_{ij} - T(\omega_i \cdot h_{ij}) \\ \log(\omega_i) &= \mu_\omega + \varepsilon_i^\omega \end{aligned}$$

- non-labour income y_i is uniformly distributed over $[10, 100]$
- unobserved component of wages: $\varepsilon_i^\omega \sim N(0, \sigma_\omega^2)$, calibrate $\mu_\omega = 1$ and $\sigma_\omega = 0.55$
- Tax system T(): earnings below 80eur per week are not taxed; any earnings greater than 80eur are taxed at the constant marginal tax rate $t = 0.3$. Non-labour income is not taxed.

Exercise 1: Additional notes

- Rescaling of utility/value function on state space (grid)

Rescaling utility functions: Static setting

- Recall our static discrete choice example: the closed-form solution for choice probabilities requires taking exponentials of the utility function

$$\Pr(y_i = j) = \frac{e^{x_{ij}'\beta}}{\sum_{k=1}^{J} e^{x_{ik}'\beta}} \quad \forall i, j$$

- In principle, values of the utility function may be arbitrarily large/small depending on the scale of the variables and the parameters. This can lead to quite easily numerical issue (over-/underflows) when computing the exponentials (or taking logs) (e.g. e^{100}).[1]
- Note also that the choice probabilities depend on the relative differences in utility across alternatives, not on the absolute values.
  $\Rightarrow$ We can rescale the utility function during evaluation by subtracting a constant.
- Usually overflows are more of a concern, hence we can rescale the utility function by subtracting the maximum value of the utility function across observations in this simple static setting.

[1] General note: in practice it is often a good idea to rescale variables in the data to have similar scales. E.g. if you have yearly earnings/consumption divide by 10000; if you have hours worked per week divide by 10. This disciplines the scale of parameters during the estimation procedure and avoids numerical issues.

Rescaling value functions: Dynamic setting

⇒ Rescaling in dynamic setting (as discussed before in the lecture):

- In a dynamic settings, the dynamic programming problem is characterized by a value function representation. Given the assumptions on unobservables (EV-I), we can obtain a closed-form expression (log-sum-exp) for the expectation of the maximum.
- Compare the expression derived on the slide 40 of lecture `3_lecture_dynamics.pdf`. In general, we can arrive for the dynamic programming conditional logit problem at an analytical expression for the integrated Bellman equation of form:[2]

$$\bar{V}(x_{it}) = \log\left(\sum_{i_{it}\in D(x_{it})} \exp\left\{u(x_{it}, i_{it}, \theta) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) p(x_{i,t+1}|x_{it}, i_{it})\right\}\right)$$

[2] Assuming a discrete state space of observables x_{it} and a finite set of alternatives i_{it}. Details on the derivation see Aguirregabiria and Mira (2010) "Dynamic discrete choice structural estimation: A survey".

Log-sum-exp

- We can exploit the log-sum-exp structure to implement a simple rescaling scheme (which can be executed at each iteration step of the backwards induction).
- To illustrate this, consider some function $q(a, b)$ where a, b are input arguments and the second equivalence holds for any constant scalar z:

$$q(a, b) = \exp(a) + \exp(b) = \left[\exp(a - z) + \exp(b - z)\right] \cdot \exp(z)$$

Taking logs:

$$\log\left[\exp(a) + \exp(b)\right] = \log\left[\exp(a - z) + \exp(b - z)\right] + z$$

- if a takes on very large positive values, numerical overflows may occur from exponentiation.
- if a (and b) takes on large negative values, the application of logarithms can cause underflows as $\lim_{\exp(a)\to 0} \log\left[\exp(a)\right] = -\infty$.

Rescaling value functions: Dynamic setting

$\Rightarrow$ Thus we can use the log-sum-exp structure and use simple rescaling scheme:

$$\bar{V}(x_{it}) = \log\left(\sum_{i_{it}\in D(x_{it})} \exp\left\{u(x_{it}, i_{it}, \theta) + \beta \sum_{x_{i,t+\mathbf{1}}} \bar{V}(x_{i,t+1}) p(x_{i,t+1}|x_{it}, i_{it}) - z_t\right\}\right) + z_t$$

- where z_t is derived as the minimum/maximum value of the value function across all grid points of the state space at period t.