Fitting Complex Mixed Logit Models with Particular Focus on Labor Supply Estimation

Max Löffler

IZA and University of Cologne

Potsdam — June 7, 2013

Motivation

- Logit models estimate discrete choice situations (e.g., yes/no)
- Discrete choice models are frequently used in many fields of economic research (and other disciplines as well)
  - Consumer demand literature, transport economics, ...
  - Labor supply estimation (Aaberge et al., 1995, van Soest, 1995, Hoynes, 1996)

Motivation

- Logit models estimate discrete choice situations (e.g., yes/no)
- Discrete choice models are frequently used in many fields of economic research (and other disciplines as well)
  - Consumer demand literature, transport economics, ...
  - Labor supply estimation (Aaberge et al., 1995, van Soest, 1995, Hoynes, 1996)
- Simple logit models are very easy to fit, but estimation of more complex logit models can become quite cumbersome
- In labor supply context: Models often chosen because of computational convenience instead of theoretical considerations
  - New command `lslogit` to estimate complex mixed logit models in a rather simple way, focus on labor supply estimation

Agenda

Conditional logit

- Individual n faces J_n alternatives and chooses alternative i (observed)
- Utility of individual n when choosing i: $U_{ni} = v(\boldsymbol{x_{ni}}|\boldsymbol{\beta}) + \epsilon_{ni}$
- Random error terms ϵ_{ni} are independent and identically distributed according to extreme value type I distribution: $\epsilon_{ni} \sim GEV(0, 1, 0)$

Conditional logit

- Individual n faces J_n alternatives and chooses alternative i (observed)
- Utility of individual n when choosing i: $U_{ni} = v(\boldsymbol{x_{ni}}|\boldsymbol{\beta}) + \epsilon_{ni}$
- Random error terms ϵ_{ni} are independent and identically distributed according to extreme value type I distribution: $\epsilon_{ni} \sim GEV(0, 1, 0)$
- Likelihood function can be solved analytically (McFadden, 1974)

$$L = \prod_{n=1}^{N} P(U_{ni} > U_{nj}, \forall j \neq i) = \prod_{n=1}^{N} \frac{\exp\left(v(\boldsymbol{x_{ni}}|\boldsymbol{\beta})\right)}{\sum_{j=1}^{J_n} \exp\left(v(\boldsymbol{x_{nj}}|\boldsymbol{\beta})\right)} \quad (1)$$

- Known as (McFadden's) conditional logit, multinomial logit, ...

IIA assumption

- Conditional logit setup easy to estimate but exhibits Independence from Irrelevant Alternatives (IIA) assumption (Luce, 1959)
  - Implies: ratio of choice probabilities, i.e., preference between two alternatives is independent of the presence of a third alternative
  - Working fulltime compared to unemployment:

$$\frac{P(h_n = 40)}{P(h_n = 0)} = \frac{\exp\left(v(\boldsymbol{x}_{n,h=40}|\boldsymbol{\beta})\right)}{\exp\left(v(\boldsymbol{x}_{n,h=0}|\boldsymbol{\beta})\right)} \tag{2}$$

  - Independent from existence and characteristics of other alternatives
- Sometimes justified but unrealistic and restrictive in many cases

Mixed logit I

- Mixed logit models allow to introduce unobserved heterogeneity in preferences and/or alternatives
  - Assume distribution for β (e.g., normal, log-normal or uniform, ...)
  - Integrate over conditional probabilities
- Likelihood given by

$$L = \prod_{n=1}^{N} P(U_{ni} > U_{nj}, \forall j \neq i) = \prod_{n=1}^{N} \int_{-\infty}^{+\infty} \frac{\exp\left(v(x_{ni}|\beta)\right)}{\sum_{j=1}^{J_n} \exp\left(v(x_{nj}|\beta)\right)} f(\beta)\, \mathrm{d}\beta \quad (3)$$

- Overcomes IIA assumption

Mixed logit II

- Likelihood function cannot be solved analytically anymore
- Approximate by maximum simulated likelihood methods (Train, 2009)
  - Draw R times from distribution of $\boldsymbol{\beta}$, average choice probabilities

$$\ln(SL) = \sum_{n=1}^{N} \ln \left(\frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left(v(\boldsymbol{x}_{ni} | \boldsymbol{\beta}^r)\right)}{\sum_{j=1}^{J_n} \exp\left(v(\boldsymbol{x}_{nj} | \boldsymbol{\beta}^r)\right)} \right) \tag{4}$$

- Any discrete choice model can be approximated by mixed logit (McFadden and Train, 2000)

Logit models in Stata

- Flexible framework to estimate maximum likelihood models in Stata (`ml`), extensive documentation (Gould et al., 2010, Haan and Uhlendorff, 2006)
- Conditional logit can be estimated via built-in command `clogit`
- User-written command `mixlogit` to fit mixed logit models (Hole, 2007)

Logit models in Stata

- Flexible framework to estimate maximum likelihood models in Stata (`ml`), extensive documentation (Gould et al., 2010, Haan and Uhlendorff, 2006)
- Conditional logit can be estimated via built-in command `clogit`
- User-written command `mixlogit` to fit mixed logit models (Hole, 2007)
- Awesome tool, but two drawbacks
  1. So called `derivative1` evaluator, gradient $\partial \ln L / \partial \boldsymbol{\beta}$ derived analytically, Hessian matrix $\partial^2 \ln L / \partial \boldsymbol{\beta}^2$ numerically approximated
  2. Assumes linear utility specification: $v(\boldsymbol{x}_{nj}|\boldsymbol{\beta}) = \boldsymbol{x}_{nj}\boldsymbol{\beta}'$

Gradient and Hessian Matrix of equation (4)

$$\frac{\partial \ln(SL)}{\partial \beta_k} = \sum_{n=1}^{N} \underbrace{\frac{1}{\sum_{r=1}^{R} p_{ni}^r}}_{=1/L_{sum,n}} \underbrace{\sum_{r=1}^{R} p_{ni}^r \sum_{j=1}^{J_n} \left(d_{nj} - p_{nj}^r\right) \frac{\partial v_{nj}^r}{\partial \beta_k}}_{=G_{sum,n}} = \sum_{n=1}^{N} \frac{G_{sum,n}}{L_{sum,n}} \quad (5)$$

$$\frac{\partial^2 \ln(SL)}{\partial \beta_k \partial \beta_m'} = \sum_{n=1}^{N} \left(\frac{1}{L_{sum,n}} \frac{\partial G_{sum,n}}{\partial \beta_m'} - \frac{\partial L_{sum,n}}{\partial \beta_m'} \frac{G_{sum,n}}{L_{sum,n}^2} \right) \quad (6)$$

$$\frac{\partial G_{sum,n}}{\partial \beta_m'} = \sum_{r=1}^{R} \left(p_{ni}^r \left\{ \frac{\partial v_{ni}^r}{\partial \beta_m'} - \sum_{j=1}^{J_n} p_{nj}^r \frac{\partial v_{nj}^r}{\partial \beta_m'} \right\} \sum_{j=1}^{J_n} (d_{nj} - p_{nj}^r) \frac{\partial v_{nj}^r}{\partial \beta_k} \right.$$
$$- p_{ni}^r \sum_{j=1}^{J_n} p_{nj}^r \left\{ \frac{\partial v_{nj}^r}{\partial \beta_m'} - \sum_{s=1}^{J_n} p_{ns}^r \frac{\partial v_{ns}^r}{\partial \beta_m'} \right\} \frac{\partial v_{nj}^r}{\partial \beta_k}$$
$$\left. + p_{ni}^r \sum_{j=1}^{J_n} (d_{nj} - p_{nj}^r) \frac{\partial^2 v_{nj}^r}{\partial \beta_k \partial \beta_m'} \right) \quad (7)$$

$$\frac{\partial L_{sum,n}}{\partial \beta_m'} = \sum_{r=1}^{R} p_{ni}^r \left(\frac{\partial v_{ni}^r}{\partial \beta_m'} - \sum_{j=1}^{J_n} p_{nj}^r \frac{\partial v_{nj}^r}{\partial \beta_m'} \right) \quad (8)$$

Yet another logit?

- Linear utility specification not necessarily a problem, x_{nj} may include interaction terms, squares and terms of higher order
- But more complex models cause problems, e.g., labor supply context
  - Error components in variables like measurement errors in wages
  - Simultaneous estimation of labor supply and wages
  - Box-Cox models where power parameters have to be estimated

Estimating labor supply

- Most generally, discrete choice labor supply models can be written as

$$L = \prod_{n=1}^{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\exp\left(v(\boldsymbol{x_{ni}} | \hat{w}_{ni}, \boldsymbol{\beta})\right)}{\sum_{j=1}^{J_n} \exp\left(v(\boldsymbol{x_{nj}} | \hat{w}_{nj}, \boldsymbol{\beta})\right)} f(\hat{\boldsymbol{w}}_{\boldsymbol{n}}, \boldsymbol{\beta})\, \mathrm{d}\hat{\boldsymbol{w}}_{\boldsymbol{n}} \mathrm{d}\boldsymbol{\beta}$$

$$\cdot \left(\frac{1}{\sigma} \phi \left\{ \frac{\ln w_{ni} - \boldsymbol{x_{ni}} \boldsymbol{\beta}'_{\boldsymbol{w}}}{\sigma} \right\} \right)^{1(h_{ni}>0)} \tag{9}$$

where, e.g.,

$$v(\boldsymbol{x_{nj}} | \hat{w}_{nj}, \boldsymbol{\beta}) = \boldsymbol{x_{C,nj}} \boldsymbol{\beta'_1} C_{nj}^{(\beta_2)} + \beta_3 C_{nj}^{(\beta_2)} L_{nj}^{(\beta_5)} + \boldsymbol{x_{L,nj}} \boldsymbol{\beta'_4} L_{nj}^{(\beta_5)} + \boldsymbol{x_{I,nj}} \boldsymbol{\beta'_6} \tag{10}$$

$$C_{nj}^{(\beta_2)} = \begin{cases} \frac{(C_{nj}/\bar{C})^{\beta_2} - 1}{\beta_2} & \text{if } \beta_2 \neq 0 \\ \ln(C_{nj}/\bar{C}) & \text{if } \beta_2 = 0 \end{cases} \qquad L_{nj}^{(\beta_5)} = \begin{cases} \frac{(L_{nj}/\bar{L})^{\beta_5} - 1}{\beta_5} & \text{if } \beta_5 \neq 0 \\ \ln(L_{nj}/\bar{L}) & \text{if } \beta_5 = 0 \end{cases} \tag{11}$$

$$C_{nj} = \hat{w}_{nj} h_j + I_n + T_{nj} - \tau(\hat{w}_{nj} h_j, I_n) \tag{12}$$

New estimation routine

- Estimates complex mixed logit models where households have preferences with regard to two or three goods
- Built-in utility functions (quadratic, log-quadratic/translog, Box-Cox)
- Allows to specify observed and unobserved heterogeneity in preferences and across alternatives

New estimation routine

- Estimates complex mixed logit models where households have preferences with regard to two or three goods
- Built-in utility functions (quadratic, log-quadratic/translog, Box-Cox)
- Allows to specify observed and unobserved heterogeneity in preferences and across alternatives
- Main focus on labor supply estimation, therefore additional options
  - Simultaneous estimation of preferences and wage equation
  - Integrates wage prediction error out during estimation process
  - Calculates marginal utility of consumption (and allows constraints)
  - . . . and naming conventions

How to use it

- Technically: `derivative2`-evaluator, written in Mata
- Includes prediction command and calculator of covariance matrix

Use on Stata command line

```
. lslogit depvar [if] [in] [weight], group(varname)
          consumption(varname) leisure(varlist)
          cx(varlist) lx1(varlist) ind(varlist) [options]
. predict newvar[list] [if] [in] [, options]
. lslogit, cov [options]
```

- By now: Beta version available on request

Output

```
. lslogit choice, group(id) c(dpi) l(freiz) boxcox cx(lage*_m) lx1(lage*_m)

Mixed Logit Labor Supply Model                   Number of obs   =       5761
                                                 LR chi2(2)      =     171.99
                                                 Prob > chi2     =     0.0000
Log likelihood = -1368.2779                      Pseudo R2       =     0.1456

                                (Std. Err. adjusted for clustering on hhnrakt)
------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Cx     lage_m |   63.48062   10.78075     5.89   0.000     42.35074    84.6105
     lage2_m |  -8.688809   1.483041    -5.86   0.000    -11.59552  -5.782102
       _cons |  -112.7283   19.40653    -5.81   0.000    -150.7644  -74.69218
-------------+----------------------------------------------------------------
CxL1   _cons |   .0862592   .0578086     1.49   0.136    -.0270435   .1995619
-------------+----------------------------------------------------------------
L1x   lage_m |   1.314325   1.334258     0.99   0.325    -1.300773   3.929424
     lage2_m |  -.1787296   .1835297    -0.97   0.330    -.5384413   .1809821
       _cons |   -2.32873    2.38518    -0.98   0.329    -7.003596   2.346136
-------------+----------------------------------------------------------------
        /l_C |    .593499   .0875811     6.78   0.000     .4218433   .7651548
       /l_L1 |  -2.624566   .5228987    -5.02   0.000    -3.649429  -1.599704
-------------+----------------------------------------------------------------
     [dudes] |   .0341955
------------------------------------------------------------------------------
Model: - Box-Cox utility function

.
```

Thank you for your attention!

Comments or questions? — loeffler@iza.org

Appendix

References I

Aaberge, R., Dagsvik, J. K. and Strøm, S. (1995). Labor supply responses and welfare effects of tax reforms, *The Scandinavian Journal of Economics* **97**(4): 635–659.

Gould, W., Pitblado, J. and Poi, B. (2010). *Maxiumum Likelihood Estimation with Stata*, fourth edn, Stata Press.

Haan, P. and Uhlendorff, A. (2006). Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood, *Stata Journal* **6**(2): 229–245.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood, *Stata Journal* **7**(3): 388–401.

Hoynes, H. W. (1996). Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP, *Econometrica* **64**(2): 295–332.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.

McFadden, D. and Train, K. E. (2000). Mixed mnl models for discrete response, *Journal of Applied Econometrics* **15**(5): 447–470.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*, second edn, Cambridge University Press.

van Soest, A. (1995). Structural models of family labor supply – a discrete choice approach, *Journal of Human Resources* **30**(1): 63–88.