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

Max Löffler

IZA and University of Cologne

New Orleans — July 18, 2013

Motivation

- Logit models estimate discrete choice situations (e.g., yes/no)
- Discrete choice models are used in many fields of (economic) research
  - 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 used in many fields of (economic) research
  - 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: Used models often chosen because of computational convenience instead of theoretical considerations
  - New command `lslogit` to estimate complex mixed logit models in a common framework, focus on labor supply estimation

Agenda

Conditional or multinomial logit

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

Conditional or multinomial logit

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

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

- Conditional logit models are very convenient and easy to estimate

IIA assumption

- Conditional logit setup exhibits assumption on independence from irrelevant alternatives (IIA) (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 non-participation/unemployment:

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

  - 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 error components in preferences, variables, alternative or individual specific taste shifters
- Heavily used in practice: unobserved heterogeneity in preferences
  - Assume distribution for $\boldsymbol{\beta}$ (e.g., normal, log-normal or uniform, ...)
  - Integrate over conditional choice probabilities
- Likelihood given by

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

- Mixed logit setup overcomes IIA assumption

Mixed logit II

- Likelihood function (3) cannot be solved analytically anymore
- Approximate by maximum simulated likelihood methods (Train, 2009)
  - Draw randomly R times from distribution of $\boldsymbol{\beta}$
  - Average choice probabilities $p^r_{ni_n}|\boldsymbol{\beta}^r$ over set of draws

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

Mixed logit II

- Likelihood function (3) cannot be solved analytically anymore
- Approximate by maximum simulated likelihood methods (Train, 2009)
  - Draw randomly R times from distribution of β
  - Average choice probabilities $p^r_{ni_n}|\beta^r$ over set of draws

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

- Various more general mixed logit setups (Keane and Wasi, 2012)
- 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), more general models can be estimated via `gmnl` (Gu et al., 2013)

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), more general models can be estimated via `gmnl` (Gu et al., 2013)
- Awesome tools, but two drawbacks
  1. `mixlogit`: so-called `d1` evaluator, gradient $\partial \ln L / \partial \beta$ derived analytically, Hessian matrix $\partial^2 \ln L / \partial \beta^2$ numerically approximated

     `gmnl`: both derivates have to be approximated numerically
  2. Both assume linear utility specification: $v(x_{nj}|\beta) = x_{nj}\beta'$

Gradient and Hessian Matrix of equation (4)

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

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

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

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

Yet another logit?

- Choice of ML evaluator is a matter of time (and precision)
- Linear utility specification not necessarily a problem, x_{nj} may include interaction terms, squares and terms of higher order

Yet another logit?

- Choice of ML evaluator is a matter of time (and precision)
- 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

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

$$\times \left(\frac{1}{\sigma} \phi \left\{ \frac{\ln w_{ni_n} - \boldsymbol{x_{ni_n}} \boldsymbol{\beta}'_w}{\sigma} \right\} \right)^{1(h_{ni_n} > 0)} \quad (9)$$

where, e.g.,

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

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

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

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)

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)
- Technically: `derivative2`-evaluator, written in Mata
  - Rather quick, but more flexible and thus also slower than `clogit`
  - By now: beta version available on request

How to use it

Use of clogit

```
. clogit depvar varlist, group(varname)
. clogit depvar consumption#(varlist) leisure#(varlist)
                alternatives#(varlist), group(varname)
```

How to use it

Use of clogit

```
. clogit depvar varlist, group(varname)
. clogit depvar consumption#(varlist) leisure#(varlist)
                alternatives#(varlist), group(varname)
```

Use of lslogit

```
. lslogit depvar [if] [in] [weight], group(varname)
                  consumption(varname) leisure(varlist)
                  [{translog|quadratic|boxcox} ...
                   cx(varlist) lx1(varlist) ind(varlist)
                   randvars(numlist) corr draws(integer)
                   hwage(varlist) wagepred(varlist)
                   taxreg(name) tria1(varlist) ...]
```

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.

Gu, Y., Hole, A. R. and Knox, S. (2013). Fitting the generalized multinomial logit model in stata, *Stata Journal* **13**(2): 382–397.

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.

Keane, M. P. and Wasi, N. (2012). Comparing alternative models of heterogeneity in consumer choice behavior, *Journal of Applied Econometrics* . DOI: 10.1002/jae.2304.

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.

References II

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.