Redistribution in Microsimulation Models with Behavioral Responses

Max Löffler

ZEW and University of Cologne

Canazei — January 14, 2015

What is this session about?

- Redistribution: "If top tax rate was 53 instead of 42 %, how would that affect net incomes, income inequality or work incentives?"
- Microsimulation: Ex ante analysis to evaluate counterfactual situations or policy reforms that are to be implemented in the future
- Models: Analysis starting from economic theory and theoretic models
- Behavioral Responses: Allows to study adjustments to changed incentives and predict responses to counterfactual situations

Motivation

- Why could this be interesting for inequality and welfare analysis?
  - Understand and predict behavioral responses to redistributive policies
  - Estimation of structural utility models allows various welfare analyses
- Applied session
  - How do these models work?
  - How to estimate them?
  - How to predict behavior?
- Several references included (Creedy and Kalb, 2005, 2006, Aaberge and Colombino, 2014)

Outline

Basic workflow

1. Specify a theoretical model of individual labor supply behavior
   - E.g. $U = U(C, L)$, individuals maximize their utility over potential jobs
   - Jobs defined by wages w, working hours h and other characteristics ϵ
2. Find suitable data and set up the choice set
   - Identify chosen jobs (hours, wages and characteristics observed in the data)
   - Add counterfactual choices and simulate hypothetical hours and net incomes
3. Estimate the model using this choice set, obtain preference estimates
4. Hold preferences constant and simulate counterfactual scenarios
   - Study behavioral adjustments if tax/benefit system or wages were different
   - Calculate welfare measures (e.g. compensating variation, utility differentials)

Theoretical model

- Agents receive utility from consumption C and leisure L
  - Mainly two different classes of models have been used over the last decades
  - Classic models assume maximization over continuous range of working hours, i.e. $\max_{h=[0,60]} U(C[h], L[h])$ (Burtless and Hausman, 1978, Hausman, 1981)
  - More recent models usually assume discrete choice between different jobs, i.e. $\max_{j \in J} U(C_j, L_j)$ (Aaberge et al., 1995, van Soest, 1995, Hoynes, 1996)
  - Individual subscripts omitted, but no time subscripts: "steady-state" models

- Agents choose jobs and maximize utility. How does it look like?
  - Individual n's utility of choosing job j is defined as $U_{nj} = U(C_{nj}, L_j, \epsilon_{nj})$
  - Researcher needs to identify form of U, distribution of ϵ_{nj} and job offers
  - Observed choices inform us about maximum utility and optimal working hours

Identification

- Wait, agents' labor supply is thus determined by consumption and leisure, both functions of working hours? Isn't that tautological?
  - Right. That's why econometricians care so much about identification
  - Non-parametric identification impossible with cross-sectional data (Manski, 2014)
  - But even cross-sectional data typically gives us variation in wages, non-labor incomes, taxes and transfers (=consumption) for given working hours

- Use structural model and impose specific functional form, e.g.
  - $U_{nj} = \alpha_1 \ln C_{nj} + \alpha_2 (\ln C_{nj})^2 + \alpha_3 \ln L_j + \alpha_4 (\ln L_j)^2 + \alpha_5 \ln C_{nj} \ln L_j + \epsilon_{nj}$
  - C_{nj} accounts for earnings $w_{nj} L_j$, non-labor income I and tax/benefit system
  - Assumes homogeneous preferences and specific functional form

Job offers and hours categories

- We assume that individuals choose among jobs, but what is a job?
  - A job j is defined by hours h_j, wages w_{nj} and unobservables ϵ_{nj}
  - Most often, job modeling boils down to three simple assumptions:
    1. Unobservables follow Gumbel distribution: $\epsilon_{nj} \sim GEV(0,1,0)$
    2. Wages assumed exogenous to hours or preferences: $w_{nj} = w_n$
    3. Hours discretized to interval means: $h_j \in [0, 10, 20, 30, 40, 50, 60]$

    (van Soest, 1995, Hoynes, 1996, Keane and Moffitt, 1998, Blundell et al., 2000, Bargain et al., 2014)
  - This simple model generates seven choices for each individual

- Modeling approach is rather restrictive. Alternative approaches:
  - Define random choice-set instead of discretized hours intervals (Aaberge et al., 2009)
  - Allow correlation between wages and preferences (Blundell and Shephard, 2012)
  - Allow hours-dependent wages and flexible correlation (Löffler et al., 2014)

Data sources

- What kind of data is required for the estimation?
  - The model requires information at least on wages and working hours
  - Ideally, the data also contains household characteristics, information on children, demographics, social background, occupational information, ...
  - Most models based on cross-sectional data, but panel data highly useful
- Which data sets are used in practice?
  - Usually household surveys, only rarely administrative data because working hours are seldom reported in tax registers or social security records
  - EU: SILC, US: CPS, PSID, NLSY, Germany: SOEP, Britain: BHPS, ...

Tax/benefit calculator

- What we need to do with the data
  - Clean up sample and keep only working-age population (or subgroups)
  - Create choice set according to job model (for now: simple approach)
  - Calculate individual wage rates (and impute wage for non-workers)
  - Duplicate individual observations by number of hypothetical choices
- How to create hypothetical job choices?
  - Define weekly/monthly working hours for every hypothetical job choice
  - Based on hypothetical hours and wages, calculate gross earnings for all jobs
  - Apply tax/benefit calculator to these hypothetical households with their hypothetical working hours and gross earnings to get disposable income
  - Popular calculators: EUROMOD for Europe, TAXSIM for the US, ...
- Observed working hours inform us about the job actually chosen

Browsing through the data

- The resulting data set should look somehow like this one:

#	id	hours	wage	gross	dpi	leisure	choice	...
1	1	0	7.5	0	400	80	0	...
2	1	10	7.5	300	500	70	0	...
3	1	20	7.5	600	600	60	1	...
4	1	30	7.5	900	700	50	0	...
5	1	40	7.5	1,200	800	40	0	...
6	1	50	7.5	1,500	900	30	0	...
7	1	60	7.5	1,800	1,000	20	0	...
8	2	0	9.7	0	400	80	0	...
9	2	10	9.7	388	550	70	0	...
10	2	20	9.7	766	700	60	0	...
11	2	30	9.7	1,164	850	50	0	...
12	2	40	9.7	1,552	1,000	40	1	...
13	2	50	9.7	1,940	1,150	30	0	...
14	2	60	9.7	2,328	1,300	20	0	...
...	...	...	...	...	...	...	...	...

Estimation approach

- Bringing the theory to the data
  - We know the choice and we know the consumption and leisure for all jobs
  - We want to outback the preference coefficients α of our utility function

- Recall the utility function we assumed for our model
  - $U_{nj} = \underbrace{\alpha_1 \ln C_{nj} + \alpha_2 (\ln C_{nj})^2 + \alpha_3 \ln L_j + \alpha_4 (\ln L_j)^2 + \alpha_5 \ln C_{nj} \ln L_j}_{V_{nj}} + \epsilon_{nj}$
  - Moreover we assumed that ϵ_{nj} is extreme value distributed
  - Yields well known multinomial or conditional logit problem (McFadden, 1974)
  - $P\left(U_{nj} > U_{nk} \forall k \neq j\right) = \frac{\exp V_{nj}}{\sum_{h \in J} \exp V_{nh}} \quad \forall \quad n = 1, \ldots, N, \quad j \in J$
  - We can estimate this model using maximum likelihood methods

Estimation using Stata

- Define the likelihood function of this simple model
  - Log-likelihood given by: $\ln L = \sum_{n=1}^{N} \sum_{j \in J} y_{nj} \left(V_{nj} - \ln \sum_{h \in J} \exp V_{nh} \right)$
  - Where $y_{nj} = 1$ if individual n chose alternative j and $y_{nj} = 0$ otherwise
- Turning to Stata
  - The model defined above can be estimated using Stata's `clogit` command
  - Call by: `clogit choice ln_dpi ln_leisure ... [fw=...], group(id)`
- Rather restrictive when it comes to more complicated models
  - User-written command `lslogit` for labor supply estimation (Löffler, 2013)
  - Supports many heavily used utility specifications and extensions (Löffler et al., 2014)
  - Call by: `lslogit choice [fw=...], group(id) con(dpi) lei(leisure)`

Example output

```
. lslogit choice, group(id) c(dpi) l(leisure) boxcox cx(age*_m) lx1(age*_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 id)
------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Cx     age_m |   63.48062   10.78075     5.89   0.000     42.35074     84.6105
      age2_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    age_m |   1.314325   1.334258     0.99   0.325    -1.300773    3.929424
      age2_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

.
```

Postestimation

- Maximum likelihood gives us estimates for preference coefficients $\hat{\alpha}$
  - This allows us to predict utility levels $\hat{V}_{nj}$ from choosing the different jobs
    - Call by: `predict util, xb`
  - Using these we can predict choice probabilities or utility maximizing choice
    - Call by: `predict prob, pc1`
  - Enables us to check estimation fit by comparing observed and predicted choices

#	id	hours	choice	util	prob	...
1	1	0	0	1.0	0.100	...
2	1	10	0	1.5	0.165	...
3	1	20	1	2.0	0.271	...
4	1	30	0	1.5	0.165	...
5	1	40	0	1.0	0.100	...
6	1	50	0	1.0	0.100	...
7	1	60	0	1.0	0.100	...
...	...	...	...	...	...	...

Counterfactual analysis

- Holding preferences $\hat{\alpha}$ constant, we can simulate policy reforms
  - E.g. by calculating new disposable incomes for all job categories
    - `gen dpi2 = ...`
    - `replace dpi = dpi2`
    - `predict prob2, pc1`

#	id	hours	dpi	dpi2	choice	prob	prob2	...
1	1	0	400	300	0	0.100	0.07	...
2	1	10	500	400	0	0.165	0.10	...
3	1	20	600	500	1	0.271	0.14	...
4	1	30	700	700	0	0.165	0.14	...
5	1	40	800	900	0	0.100	0.20	...
6	1	50	900	1,100	0	0.100	0.18	...
7	1	60	1,000	1,300	0	0.100	0.18	...
...	...	...	...	...	...	...	...	...

Model extensions

- These are the basics. Model can be extended in several dimensions
  - Alternative representation of job offers (Aaberge et al., 2009, Dagsvik et al., 2014)
  - Different assumptions on the wage distribution and exogeneity (Löffler et al., 2014)
  - Functional form of the utility function (van Soest et al., 2002, Löffler et al., 2014)
  - Part-time restrictions and fixed costs (van Soest, 1995, Euwals and van Soest, 1999)
  - Random preferences and unobserved heterogeneity (van Soest, 1995, Pacifico, 2013)
  - Correlated wages and preferences (Breunig et al., 2008, Blundell and Shephard, 2012)
  - Welfare stigma from benefit participation (Hoynes, 1996, Keane and Moffitt, 1998)
  - Job offers and different economic sectors (Dagsvik and Strøm, 2006)
- Huge literature, but all follow the same outline and workflow
  - Good intro: "Labour supply and microsimulation" (Creedy and Kalb, 2005, 2006)
  - Handbook chapter: "Labour Supply Models" (Aaberge and Colombino, 2014)
  - Comprehensive sensitivity check with respect to assumptions (Löffler et al., 2014)

Comments or questions? — loeffler@zew.de

References I

Aaberge, R. and Colombino, U. (2014). Labour Supply Models, *in* C. O'Donoghue (ed.), *Handbook of Microsimulation*, Vol. 293, Emerald Group Publishing Limited, pp. 167–221.

Aaberge, R., Colombino, U. and Wennemo, T. (2009). Evaluating Alternative Representations of the Choice Sets in Models of Labor Supply, *Journal of Economic Surveys* **23**(3): 586–612.

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.

Bargain, O., Orsini, K. and Peichl, A. (2014). Comparing Labor Supply Elasticities in Europe and the US: New Results, *Journal of Human Resources* **49**(3): 723–838.

Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (2000). The Labour Market Impact of the Working Families' Tax Credit, *Fiscal Studies* **21**(1): 75–104.

Blundell, R. and Shephard, A. (2012). Employment, Hours of Work and the Optimal Taxation of Low-Income Families, *Review of Economic Studies* **79**(2): 481–510.

Breunig, R., Cobb-Clark, D. A. and Gong, X. (2008). Improving the Modelling of Couples' Labour Supply, *Economic Record* **84**(267): 466–485.

Burtless, G. and Hausman, J. A. (1978). The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment, *The Journal of Political Economy* **86**(6): 1103–1130.

Creedy, J. and Kalb, G. (2005). Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation, *Journal of Economic Surveys* **19**(5): 697–734.

References II

Creedy, J. and Kalb, G. (2006). *Labour Supply and Microsimulation*, Edward Elgar Publishing Limited.

Dagsvik, J. K., Jia, Z., Kornstadt, T. and Thoresen, T. O. (2014). Theoretical and Practical Arguments For Modeling Labor Supply as a Choice Among Latent Jobs, *Journal of Economic Surveys* **28**(1): 134–151.

Dagsvik, J. K. and Strøm, S. (2006). Sectoral Labor Supply, Choice Restrictions and Functional Form, *Journal of Applied Econometrics* **21**(6): 803–826.

Euwals, R. and van Soest, A. (1999). Desired and actual labour supply of unmarried men and women in the Netherlands, *Labour Economics* **6**(1): 95–118.

Hausman, J. A. (1981). Labor Supply, *in* H. J. Aaron and J. A. Pechman (eds), *How taxes affect economic behavior*, Brookings Institution, pp. 27–72.

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 Moffitt, R. (1998). A Structural Model of Multiple Welfare Program Participation and Labor Supply, *International Economic Review* **39**(3): 553–589.

Löffler, M. (2013). Fitting Complex Mixed Logit Models: An Application to Labor Supply. Mimeo, presented at the Stata Conference in New Orleans, July 2013.

Löffler, M., Peichl, A. and Siegloch, S. (2014). Structural Labor Supply Models and Wage Exogeneity, *ZEW Discussion Paper* **14-040**.

References III

Manski, C. F. (2014). Identification of income-leisure preferences and evaluation of income tax policy, *Quantitative Economics* **5**(1): 145–174.

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

Pacifico, D. (2013). On the role of unobserved preference heterogeneity in discrete choice models of labour supply, *Empirical Economics* **45**(2): 929–963.

van Soest, A. (1995). Structural Models of Family Labor Supply – A Discrete Choice Approach, *Journal of Human Resources* **30**(1): 63–88.

van Soest, A., Das, M. and Gong, X. (2002). A structural labour supply model with flexible preferences, *Journal of Econometrics* **107**(1-2): 345–374.