Mixed logit modelling in Stata
-
An overview

Arne Risa Hole
University of Sheffield

UK Stata Users Group meeting
September 2013

Background

- The conditional logit model (McFadden, 1974) is the ‘workhorse’ model for analysing discrete choice data
- While widely used this model has several well-known limitations:
  - Cannot account for preference heterogeneity among respondes (unless it’s related to observables)
  - IIA property: can lead to unrealistic predictions
- This has led researchers in various disciplines to consider more flexible alternatives
- The mixed logit model extends the standard conditional logit model by allowing one or more of the parameters in the model to be randomly distributed

Mixed logit models in Stata (pre Stata 13)

- Official Stata:
  - *xtmelogit*
- User written:
  - *gllamm*
  - *mixlogit*
  - *lclogit*
  - *gmnl*
  - *bayesmlogit*
  - *lslogit*
- I will give examples of the use of some of these commands in this talk

Outline

- Theoretical foundations - the random utility model
- Mixed logit with continuous distributions (*mixlogit*)
- Mixed logit with discrete distributions (*lclogit*)
- Generalised multinomial logit (*gmnl*)
- Bayesian mixed logit (*bayesmlogit*)

The random utility model

- The utility decision maker n obtains from choosing alternative j is given by

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

  where V_{nj} is a function of observable attributes of the alternatives, $\mathbf{x}_{nj}$, and of the decision maker, $\mathbf{z}_n$
- ε_{nj} is unknown and treated as random
- The probability that decision maker n chooses alternative i is

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj}) \forall j \neq i \\ &= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \forall j \neq i \\ &= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) \forall j \neq i \end{aligned}$$

- Different discrete choice models are obtained from different assumptions about the distribution of the random terms

The conditional logit model

- If we make the assumption that the random terms are IID type I extreme value distributed we obtain the conditional logit model:

$$P_{ni} = \frac{\exp(\sigma_n V_{ni})}{\sum_{j=1}^{J} \exp(\sigma_n V_{nj})}$$

- Typically the representative utility is specified to be a linear-in-parameters function

$$V_{ni} = \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{n}\gamma_i$$

- σ_n is a scale parameter which is typically normalised to 1. Assumes ε_{nj} is homoscedastic (more on that later).
- To estimate the conditional logit model in Stata we use the *asclogit* ('alternative-specific conditional logit') command

Limitations of the conditional logit

- Assumes that respondents have the same preferences (or that their preferences depend on observable characteristics)
- Equal proportional substitution between the alternatives:

$$\frac{\partial P_{ni}}{\partial x^*_{nj}} \frac{x^*_{nj}}{P_{ni}} = -x^*_{nj} P_{nj} \beta^*$$

- Note that this expression does not depend on i
- This is due to the assumption that the error terms are independent. Another consequence is is the IIA property:

$$\frac{P_{ni}}{P_{nk}} = \frac{\exp(V_{ni}) / \sum_{j=1}^{J} \exp(V_{nj})}{\exp(V_{nk}) / \sum_{j=1}^{J} \exp(V_{nj})} = \frac{\exp(V_{ni})}{\exp(V_{nk})}$$

Extension: the mixed logit model

- The mixed logit model overcomes these limitations by allowing the coefficients in the model to vary across decision makers
- The mixed logit choice probability is given by:

$$P_{ni} = \int \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{nj}\boldsymbol{\beta})} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta}$$

  where $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is the density function of $\boldsymbol{\beta}$
- Allowing the coefficients to vary implies that we allow for the fact that different decision makers may have different preferences
- It can also be seen that the IIA property no longer holds

Panel data

- If we observe an individual making several choices this can be taken into account in the analysis
- The probability of a particular sequence of choices is given by:

$$S_n = \int \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt}\boldsymbol{\beta})} \right]^{y_{njt}} f(\boldsymbol{\beta}|\boldsymbol{\theta})d\boldsymbol{\beta}$$

where $y_{njt} = 1$ if the individual chose alternative j in choice situation t and 0 otherwise

Maximum simulated likelihood

- The $\boldsymbol{\theta}$ parameters can be estimated by maximising the simulated log-likelihood function

$$SLL = \sum_{n=1}^{N} \ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_n^{[r]})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_n^{[r]})} \right]^{y_{njt}} \right\}$$

  where $\boldsymbol{\beta}_n^{[r]}$ is the r-th draw for individual n from the distribution of $\boldsymbol{\beta}$
- This approach can be implemented in Stata using the *mixlogit* command (Hole, 2007)

Example: Households' choice of electricity supplier

- Subset of the data from Huber and Train (2000)
- Residential electricity customers presented with a series of experiments with four alternative electricity suppliers
- Price is either fixed or a variable rate that depends on the time of day or the season
- The following attributes are included in the experiment:
  - Price in cents per kWh if fixed price, 0 if TOD or seasonal rates
  - Contract length in years
  - Whether a local company (0-1 dummy)
  - Whether a well-known company (0-1 dummy)
  - TOD rates (0-1 dummy)
  - Seasonal rates (0-1 dummy)

First 16 records in dataset

```
. use http://fmwww.bc.edu/repec/bocode/t/traindata.dta, clear

. list in 1/12, sepby(gid)

     +---------------------------------------------------------------------+
     | y   price   contract   local   wknown   tod   seasonal   gid   pid |
     |---------------------------------------------------------------------|
  1. | 0       7          5       0        1     0          0     1     1 |
  2. | 0       9          1       1        0     0          0     1     1 |
  3. | 0       0          0       0        0     0          1     1     1 |
  4. | 1       0          5       0        1     1          0     1     1 |
     |---------------------------------------------------------------------|
  5. | 0       7          0       0        1     0          0     2     1 |
  6. | 0       9          5       0        1     0          0     2     1 |
  7. | 1       0          1       1        0     1          0     2     1 |
  8. | 0       0          5       0        0     0          1     2     1 |
     |---------------------------------------------------------------------|
  9. | 0       9          5       0        0     0          0     3     1 |
 10. | 0       7          1       0        1     0          0     3     1 |
 11. | 0       0          0       0        1     1          0     3     1 |
 12. | 1       0          0       1        0     0          1     3     1 |
     +---------------------------------------------------------------------+
```

Independent normally distributed coefficients

```
. global randvars "contract local wknown tod seasonal"

. mixlogit y price, rand($randvars) group(gid) id(pid) nrep(500)

Iteration 0:   log likelihood = -1321.8371  (not concave)
 (output omitted)
Iteration 5:   log likelihood = -1105.2832

Mixed logit model                                 Number of obs   =       4780
                                                  LR chi2(5)      =     502.21
Log likelihood = -1105.2832                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
       price |  -.9585486    .065665   -14.60   0.000     -1.08725   -.8298476
    contract |  -.2664931   .0463752    -5.75   0.000    -.3573869   -.1755994
       local |   2.138131   .2286567     9.35   0.000     1.689973     2.58629
      wknown |   1.551129    .176043     8.81   0.000     1.206091    1.896167
         tod |  -9.324015   .6113342   -15.25   0.000    -10.52221   -8.125822
    seasonal |  -9.354167   .6126139   -15.27   0.000    -10.55487   -8.153466
-------------+----------------------------------------------------------------
SD           |
    contract |   .3851452   .0411142     9.37   0.000     .3045629    .4657275
       local |   1.871411   .2237016     8.37   0.000     1.432964    2.309858
      wknown |   1.241902   .1698206     7.31   0.000     .9090594    1.574744
         tod |   2.470736   .3040799     8.13   0.000      1.87475    3.066721
    seasonal |   2.261269   .2508061     9.02   0.000     1.769698     2.75284
------------------------------------------------------------------------------
The sign of the estimated standard deviations is irrelevant: interpret them as
being positive

. *Save coefficients for later use
. matrix b = e(b)
```

Correlated normally distributed coefficients

```
. *Starting values
. matrix start = b[1,1..7],0,0,0,0,b[1,8],0,0,0,b[1,9],0,0,b[1,10],0,b[1,11]

. mixlogit y price, rand($randvars) group(gid) id(pid) nrep(500) ///
> corr from(start, copy)

Iteration 0:   log likelihood = -1105.2832  (not concave)
 (output omitted)
Iteration 8:   log likelihood = -1052.5628

Mixed logit model                                 Number of obs   =       4780
                                                  LR chi2(15)     =     607.65
Log likelihood = -1052.5628                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.9787741    .070082   -13.97   0.000    -1.116132   -.8414159
    contract |  -.2720893   .0461627    -5.89   0.000    -.3625666    -.181612
       local |   2.465699   .3208127     7.69   0.000     1.836918    3.094481
      wknown |   1.909235   .2441677     7.82   0.000     1.430675    2.387794
         tod |  -9.482937   .6453369   -14.69   0.000    -10.74777     -8.2181
    seasonal |  -9.550856   .6487147   -14.72   0.000    -10.82231   -8.279398
-------------+----------------------------------------------------------------
        /l11 |   .4042082   .0538859     7.50   0.000     .2985938    .5098225
        /l21 |   .8208457   .3444456     2.38   0.017     .1457448    1.495947
        /l31 |   .6398508   .2402662     2.66   0.008     .1689378    1.110764
        /l41 |  -.1434189   .3279705    -0.44   0.662    -.7862293    .4993915
        /l51 |   .4056873   .3557024     1.14   0.254    -.2914765    1.102851
        /l22 |   2.557771   .3423882     7.47   0.000     1.886702    3.228839
        /l32 |   1.603419   .2885535     5.56   0.000     1.037865    2.168974
        /l42 |   .5591408   .3596689     1.55   0.120    -.1457972    1.264079
        /l52 |   .3354436   .4288803     0.78   0.434    -.5051464    1.176034
        /l33 |   .6870104   .1541434     4.46   0.000     .3848949    .9891259
        /l43 |  -.5710674   .2779083    -2.05   0.040    -1.115758   -.0263772
        /l53 |  -.0338141   .3143597    -0.11   0.914    -.6499477    .5823196
        /l44 |   2.666976   .3572386     7.47   0.000     1.966801    3.367151
        /l54 |   1.994679   .3533701     5.64   0.000     1.302086    2.687271
        /l55 |   1.664762    .293901     5.66   0.000     1.088727    2.240797
------------------------------------------------------------------------------
```

Coefficient covariance matrix

- The parameters in the bottom panel of the output are the elements of the lower-triangular matrix L, where the covariance matrix for the random coefficients is given by $\Sigma = LL'$
- The *mixlcov* command can be used postestimation to obtain the elements in the Σ matrix along with their standard errors:

```
. mixlcov
 (output omitted)
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v11 |   .1633843   .0435622     3.75   0.000     .0780039    .2487646
         v21 |   .3317926   .1402982     2.36   0.018     .0568131     .606772
         v31 |   .2586329   .1021618     2.53   0.011     .0583996    .4588663
         v41 |  -.0579711   .1317639    -0.44   0.660    -.3162236    .2002814
         v51 |   .1639821   .1503037     1.09   0.275    -.1306078     .458572
         v22 |   7.215978   1.779245     4.06   0.000     3.728722    10.70323
         v32 |   4.626398   1.258291     3.68   0.000     2.160193    7.092603
         v42 |   1.312429   .9886205     1.33   0.184    -.6252314     3.25009
         v52 |   1.190995   1.217745     0.98   0.328    -1.195743    3.577732
         v33 |   3.452346   .9864134     3.50   0.000     1.519012    5.385681
         v43 |   .4124413   .5923773     0.70   0.486    -.7485968     1.57348
         v53 |   .7742056    .712642     1.09   0.277    -.6225471    2.170958
         v44 |   7.772087   2.067076     3.76   0.000     3.720692    11.82348
         v54 |   5.468447   1.586649     3.45   0.001     2.358672    8.578222
         v55 |   7.028424   1.777315     3.95   0.000     3.544951     10.5119
------------------------------------------------------------------------------
```

Predicted probabilities and marginal effects

- We may want to investigate how the probability of choosing an alternative changes if the company is well-known:

```
. preserve

. set seed 12345

. gen rnd = runiform()

. bysort pid gid (rnd): gen alt = _n

. replace wknown = 0 if alt==1
(483 real changes made)

. mixlpred p0, nrep(500)

. replace wknown = 1 if alt==1
(1195 real changes made)

. mixlpred p1, nrep(500)

. gen p_diff = p1-p0

. sum p_diff if alt==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      p_diff |      1195    .1523634     .075117   .0141819   .3574491

. restore
```

- Note: this does not give us standard errors. Can use the bootstrap, but normally too time consuming

Willingness to pay estimates

- Since price is assumed to be a fixed parameter, we have the convenient result that

$$E(WTP^k) = -\frac{E(\beta^k)}{\beta^{price}}$$

- This can be calculated using the *wtp* command (SSC):

```
. wtp price $randvars

        contract       local      wknown         tod    seasonal
wtp   -.27798991   2.5191709   1.9506387  -9.6885863  -9.7579781
 ll   -.37131675   1.8956186    1.482387  -10.369959    -10.4398
 ul   -.18466306   3.1427233   2.4188904  -9.0072139  -9.0761563
```

- This shows that the average respondent is willing to pay 2.5 cents per kWh more if the company is local, for example

Individual-level coefficients

- The mixed logit model can be used to estimate individual-level coefficients
- The expected value of β conditional on a given response pattern $\mathbf{y}_n$ and a set of alternatives characterised by $\mathbf{x}_n$ is given by:

$$E[\boldsymbol{\beta}|\mathbf{y}_n, \mathbf{x}_n] = \frac{\int \boldsymbol{\beta} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt}\boldsymbol{\beta})} \right]^{y_{njt}} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta}}{\int \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt}\boldsymbol{\beta})} \right]^{y_{njt}} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta}}$$

- Intuitively this can be thought of as the conditional mean of the coefficient distribution for the sub-group of individuals who face the same alternatives and make the same choices

- Revelt and Train (2000) suggest approximating $E[\boldsymbol{\beta}|\mathbf{y}_n, \mathbf{x}_n]$ using simulation:

$$\widehat{\boldsymbol{\beta}_n} = \frac{\frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\beta}_n^{[r]} \prod_{t=1}^{T}\prod_{j=1}^{J}\left[\frac{\exp(\mathbf{x}'_{njt}\boldsymbol{\beta}_n^{[r]})}{\sum_{j=1}^{J}\exp(\mathbf{x}'_{njt}\boldsymbol{\beta}_n^{[r]})}\right]^{y_{njt}}}{\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T}\prod_{j=1}^{J}\left[\frac{\exp(\mathbf{x}'_{njt}\boldsymbol{\beta}_n^{[r]})}{\sum_{j=1}^{J}\exp(\mathbf{x}'_{njt}\boldsymbol{\beta}_n^{[r]})}\right]^{y_{njt}}}$$

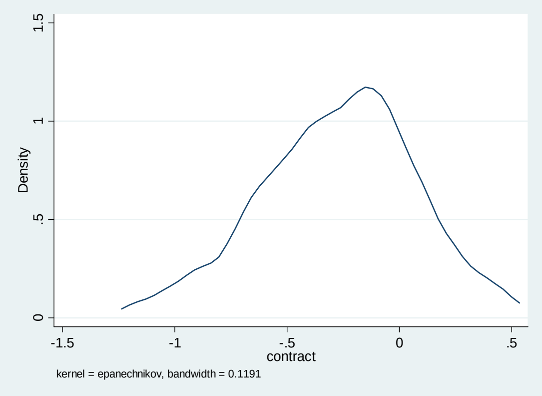
  where $\boldsymbol{\beta}_n^{[r]}$ is the r-th draw for individual n from the estimated distribution of $\boldsymbol{\beta}$
- This approach can be implemented with the *mixlbeta* command after estimating a model using *mixlogit*

Example: the distribution of the individual-level coefficient for contract length

```
. mixlbeta contract, nrep(500) saving(contract_data) replace

. use contract_data, clear

. kdensity contract, title("")
```



Discrete parameter distributions

- So far we have assumed that the distribution of the coefficients in the model is continuous
- Alternatively the coefficients may be discrete, which leads to the **latent class** model
- Each respondent is assumed to belong to a class q, where preferences vary across, but not within, classes
- In this case the probability of a particular sequence of choices is given by:

$$S_n = \sum_{q=1}^{Q} H_{nq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)} \right]^{y_{njt}}$$

- The probability of belonging to class q, H_{nq}, is typically specified as

$$H_{nq} = \frac{\exp(\mathbf{z}_n' \gamma_q)}{\sum_{q=1}^{Q} \exp(\mathbf{z}_n' \gamma_q)}$$

  where $\gamma_Q = 0$

- The log-likelihood for this model is

$$S_n = \sum_{n=1}^{N} \ln \left\{ \sum_{q=1}^{Q} H_{nq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}_{njt}' \boldsymbol{\beta}_q)}{\sum_{j=1}^{J} \exp(\mathbf{x}_{njt}' \boldsymbol{\beta}_q)} \right]^{y_{njt}} \right\}$$

- This expression can be maximised directly using standard methods, or indirectly using the EM algorithm
- In Stata the former approach is implemented by *gllamm* (Rabe-Hesketh and Skrondal, 2012) and the latter by the *lclogit* command (Pacifico and Yoo, 2012)

3-class model

```
. lclogit y price contract local wknown tod seasonal, id(pid) group(gid) ///
> nclasses(3)

Iteration 0:  log likelihood = -1336.3051
 (output omitted)
Iteration 34:  log likelihood = -1118.2358

Latent class model with 3 latent classes

-----------------------------------------
    Variable |  Class1   Class2   Class3
-------------+---------------------------
       price |  -0.805   -0.212   -1.136
    contract |  -0.508    0.023   -0.228
       local |   0.497    3.086    1.684
      wknown |   0.355    2.307    1.650
         tod |  -6.004   -1.870  -12.487
    seasonal |  -6.662   -1.951  -11.721
-------------+---------------------------
 Class Share |   0.274    0.362    0.364
-----------------------------------------

Note: Model estimated via EM algorithm
```

- Note: no standard errors. We can use these estimates as starting values for *gllamm*

3-class model using the *lclogitml* wrapper for *gllamm*

```
. lclogitml, iterate(10)
-gllamm- is initialising. This process may take a few minutes.

Iteration 0:   log likelihood = -1118.2357  (not concave)
 (output omitted)
Iteration 3:   log likelihood = -1118.2348

Latent class model with 3 latent classes
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice1      |
       price |  -.8034478   .1178941    -6.81   0.000    -1.034516   -.5723797
    contract |  -.5066114   .0592648    -8.55   0.000    -.6227682   -.3904546
       local |   .4995165   .2047987     2.44   0.015     .0981185    .9009146
      wknown |   .3566455   .1823778     1.96   0.051    -.0008085    .7140995
         tod |   -5.99186   .9318469    -6.43   0.000    -7.818246   -4.165473
    seasonal |  -6.647019   .9928728    -6.69   0.000    -8.593014   -4.701024
-------------+----------------------------------------------------------------
choice2      |
       price |  -.2115514   .0847294    -2.50   0.013     -.377618   -.0454848
    contract |   .0234125   .0301625     0.78   0.438    -.0357049    .0825298
       local |   3.086841   .2504324    12.33   0.000     2.596003     3.57768
      wknown |   2.308963   .2336502     9.88   0.000     1.851017    2.766909
         tod |  -1.878439   .7354142    -2.55   0.011    -3.319824   -.4370538
    seasonal |  -1.964792   .7642616    -2.57   0.010    -3.462717   -.4668668
-------------+----------------------------------------------------------------
choice3      |
       price |  -1.139812   .1411815    -8.07   0.000    -1.416522   -.8631009
    contract |  -.2304551   .0654821    -3.52   0.000    -.3587976   -.1021126
       local |   1.675052   .3361127     4.98   0.000     1.016283     2.33382
      wknown |    1.64531   .2618367     6.28   0.000     1.132119      2.1585
         tod |   -12.5301   1.386537    -9.04   0.000    -15.24767   -9.812542
    seasonal |  -11.74968   1.138413   -10.32   0.000    -13.98093   -9.518429
-------------+----------------------------------------------------------------
share1       |
       _cons |  -.2791875   .2910487    -0.96   0.337    -.8496326    .2912575
-------------+----------------------------------------------------------------
share2       |
       _cons |   .0002157   .2867051     0.00   0.999    -.5617159    .5621473
------------------------------------------------------------------------------
```

Choosing the number of latent classes

- We have seen that the number of latent classes needs to be specified prior to estimating the latent class model
- In practice the following procedure is often used to determine the optimal number of classes:
  - Estimate models with different numbers of classes, say 2-10
  - Choose the preferred model using the AIC, CAIC and/or BIC information criteria

```
. forvalues c = 2/10 {
  2.         quietly lclogit y price contract local wknown tod seasonal, ///
>         id(pid) group(gid) nclass(`c')
  3.         matrix b = e(b)
  4.         matrix ic = nullmat(ic) \ `e(nclasses)', `e(ll)', ///
>         `=colsof(b)', `e(aic)', `e(caic)', `e(bic)'
  5. }
(output omitted)

. matrix colnames ic = "Classes" "LLF" "Nparam" "AIC" "CAIC" "BIC"

. matlist ic, name(columns)

  Classes        LLF     Nparam        AIC       CAIC        BIC
-----------------------------------------------------------------
        2  -1211.352         13   2448.704   2495.571   2482.571
        3  -1118.236         20   2276.471   2348.575   2328.575
        4  -1085.303         27   2224.607   2321.946   2294.946
        5  -1040.488         34   2148.976   2271.552   2237.552
        6   -1028.56         41   2139.121   2286.933   2245.933
        7  -1006.369         48   2108.738   2281.786   2233.786
        8  -990.2386         55   2090.477   2288.761   2233.761
        9  -983.6419         62   2091.284   2314.804   2252.804
       10  -978.0925         69   2094.185   2342.942   2273.942
```

Posterior class membership probabilities

- As we have seen the class membership probability is given by

$$H_{nq} = \frac{\exp(\mathbf{z}'_n \gamma_q)}{\sum_{q=1}^{Q} \exp(\mathbf{z}'_n \gamma_q)}$$

- This is the **prior** class membership probability
- The **posterior** class membership probability is given by

$$G_{nq} = \frac{H_{nq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)} \right]^{y_{njt}}}{\sum_{q=1}^{Q} H_{nq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{njt} \boldsymbol{\beta}_q)} \right]^{y_{njt}}}$$

Posterior class membership probabilities

- The prior and posterior class membership probabilities can be calculated using the postestimation command *lclogitpr* with the *up* and *cp* options

```
. lclogitpr H, up

. sum H*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          H1 |      4780     .273921           0    .273921    .273921
          H2 |      4780    .3620063           0   .3620063   .3620063
          H3 |      4780    .3640727           0   .3640727   .3640727

. lclogitpr G, cp

. sum G*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          G1 |      4780    .2717595    .4252277   9.61e-13          1
          G2 |      4780    .3634778    .4545172   3.94e-12          1
          G3 |      4780    .3647628    .4625874   4.22e-14   .9999977
```

Choice probabilities

- The probability of choosing an alternative in a choice situation can be calculated using *lclogitpr* with the *pr* option

```
. lclogitpr P, pr

. sum P*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           P |      4780         .25    .1838437   .0116864   .8217464
          P1 |      4780         .25    .2142197   .0064628   .8243505
          P2 |      4780         .25    .2187802    .016932   .8660929
          P3 |      4780         .25    .3238693   .0006114   .9779206
```

- Here P1, P2 and P3 are the choice probabilities conditional on belonging to class 1, 2 and 3
- P is the unconditional choice probability, or the average of P1-P3 weighted by the prior class membership probabilities

Individual-level parameters

- The latent class model can be used to estimate individual-level coefficients
- The expected value of $\boldsymbol{\beta}$ conditional on a given response pattern $\mathbf{y}_n$ and a set of alternatives characterised by $\mathbf{x}_n$ is given by:

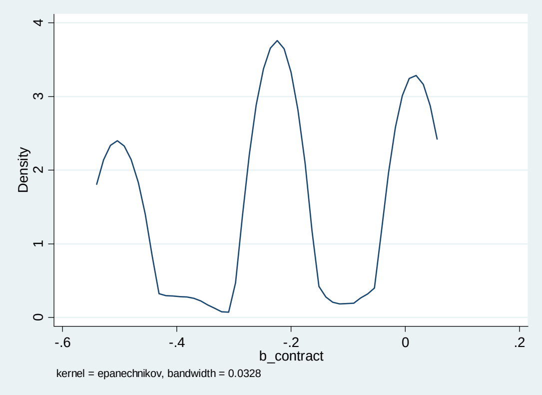
$$E[\boldsymbol{\beta}|\mathbf{y}_n, \mathbf{x}_n] = \sum_{q=1}^{Q} \boldsymbol{\beta}_q G_{nq}$$

- We can obtain an estimate of $\boldsymbol{\beta}_n$ by plugging in our estimates of $\boldsymbol{\beta}_q$ and G_{nq} into this formula

Example: the distribution of the individual-level coefficient for contract length

```
. gen b_contract = [choice1]contract*G1 + [choice2]contract*G2 + ///
>                    [choice3]contract*G3

. kdensity b_contract, title("")
```



Scale heterogeneity

- In the preceding analysis we have assumed that preference heterogeneity is the main driver behind individuals making different choices
- Recently some researchers (e.g. Louviere et al. 1999) have suggested that much of the preference heterogeneity may be better described as "scale" heterogeneity
- That is, with attribute coefficients fixed, the scale of the idiosyncratic error term is greater for some consumers than it is for others
- Since the scale of the error term is inversely related to the error variance, this argument implies that choice behavior is more random for some consumers than it is for others

The Generalised Multinomial Logit (G-MNL) model

- The G-MNL model (Fiebig et al., 2010) extends the mixed logit by specifying

$$\boldsymbol{\beta}_n = \sigma_n \boldsymbol{\beta} + \{\gamma + \sigma_n(1 - \gamma)\}\boldsymbol{\eta}_n$$

  where $\boldsymbol{\beta}$ is a constant vector, γ is a scalar parameter, $\boldsymbol{\eta}_n$ is distributed $\mathrm{MVN}(0, \Sigma)$

- $\sigma_n = \exp(\overline{\sigma} + \boldsymbol{\theta}\mathbf{z}_n + \tau v_n)$, where $\mathbf{z}_n$ is a vector of characteristics of individual n and v_n is distributed $N(0, 1)$.

- $\overline{\sigma}$ is a normalizing constant which is set so that the mean of σ_n is equal to 1 when $\theta = 0$

Special cases of G-MNL

- G-MNL-I: $\boldsymbol{\beta}_n = \sigma_n \boldsymbol{\beta} + \boldsymbol{\eta}_n$ (when $\gamma = 1$)
- G-MNL-II: $\boldsymbol{\beta}_n = \sigma_n(\boldsymbol{\beta} + \boldsymbol{\eta}_n)$ (when $\gamma = 0$)
- S-MNL: $\boldsymbol{\beta}_n = \sigma_n \boldsymbol{\beta}$ (when $var(\boldsymbol{\eta}_n) = 0$)
- Mixed logit: $\boldsymbol{\beta}_n = \boldsymbol{\beta} + \boldsymbol{\eta}_n$ (when $\sigma_n = 1$)
- Standard conditional logit: $\boldsymbol{\beta}_n = \boldsymbol{\beta}$ (when $\sigma_n = 1$ and $var(\boldsymbol{\eta}_n) = 0$)

S-MNL example

```
. gmnl y price contract local wknown tod seasonal, group(gid) id(pid) ///
> nrep(500)

Iteration 0:   log likelihood = -1348.2029  (not concave)
 (output omitted)
Iteration 5:   log likelihood = -1318.5702

Generalized multinomial logit model                Number of obs    =       4780
                                                   Wald chi2(6)     =      68.64
Log likelihood = -1318.5702                        Prob > chi2      =     0.0000

                                    (Std. Err. adjusted for clustering on pid)
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.7580152   .0986035    -7.69   0.000    -.9512745    -.564756
    contract |  -.1586749   .0285772    -5.55   0.000    -.2146852   -.1026645
       local |   1.585564   .2219014     7.15   0.000     1.150645    2.020483
      wknown |   1.226283    .174271     7.04   0.000     .8847181    1.567848
         tod |  -7.423988   .9242881    -8.03   0.000    -9.235559   -5.612417
    seasonal |  -7.484115   .9345008    -8.01   0.000    -9.315703   -5.652527
-------------+----------------------------------------------------------------
        /tau |   1.013072    .127856     7.92   0.000     .7624783    1.263665
------------------------------------------------------------------------------
The sign of the estimated standard deviations is irrelevant: interpret them as
being positive
```

G-MNL-II example

```
. gmnl y price, rand($randvars) group(gid) id(pid) nrep(500) gamma(0)

Iteration 0:   log likelihood = -1316.6538  (not concave)
 (output omitted)
Iteration 11:  log likelihood = -1097.9075

Generalized multinomial logit model               Number of obs    =       4780
                                                  Wald chi2(6)     =     126.96
Log likelihood = -1097.9075                       Prob > chi2      =     0.0000

                                   (Std. Err. adjusted for clustering on pid)
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
       price |  -1.114902   .1050251   -10.62   0.000    -1.320747   -.9090566
    contract |  -.3132268   .0693406    -4.52   0.000    -.4491318   -.1773218
       local |   2.465006    .321147     7.68   0.000     1.835569    3.094442
      wknown |   1.867701   .2279987     8.19   0.000     1.420832    2.314571
         tod |  -10.73435   .9956536   -10.78   0.000    -12.68579   -8.782904
    seasonal |  -10.72091    .983361   -10.90   0.000    -12.64826   -8.793559
-------------+----------------------------------------------------------------
SD           |
    contract |   .4326061    .064926     6.66   0.000     .3053536    .5598587
       local |   2.111132   .3134762     6.73   0.000      1.49673    2.725534
      wknown |   1.360522   .1999992     6.80   0.000     .9685305    1.752513
         tod |     2.8647   .4243209     6.75   0.000     2.033046    3.696353
    seasonal |   2.601677   .3417888     7.61   0.000     1.931783     3.27157
-------------+----------------------------------------------------------------
        /tau |  -.4978479   .0990645    -5.03   0.000    -.6920107   -.3036851
------------------------------------------------------------------------------
The sign of the estimated standard deviations is irrelevant: interpret them as
being positive
```

- *gmnl* has a similar suite of post-estimation commands to *mixlogit*
- See Gu et al. (2013) for more info

An alternative to MSL: Bayesian estimation

- We have seen that the parameters in the mixed logit model can be estimated using maximum simulated likelihood
- Alternatively we can use Bayesian procedures to obtain the estimates
- The results can be interpreted in the same way as if they were maximum likelihood estimates
- The Bayesian approach can therefore be viewed as an alternative algorithm to obtain the estimates

- By Bayes rule, we have that the posterior parameter distribution is given by

$$K(\theta|Y) = \frac{L(Y|\theta)k(\theta)}{L(Y)}$$

  where $L(Y|\theta)$ is the likelihood function (the probability of observing the data given the parameters), $k(\theta)$ is the prior parameter distribution and $L(Y) = \int L(Y|\theta)k(\theta)d\theta$
- The mean of the posterior distribution can be shown to have the same asymptotic properties as the maximum likelihood estimator
- The Bayesian approach involves taking many draws from the posterior distribution and averaging these draws

- Train (2009) describes an algorithm for taking draws from the posterior distribution of the coefficients in a mixed logit with normally distributed coefficients
- This algorithm is implemented by the *bayesmlogit* command (Baker, 2013)
- The mean and variance of β as well as the individual-level parameters, β_n, are treated as parameters to be estimated
- The values from the algorithm converges to draws from the posterior distribution
- The iterations prior to convergence are called the ‘burn-in’
- Even after convergence the draws are correlated, so only a portion of the draws are kept

Correlated normally distributed coefficients

```
. bayesmlogit y price, rand($randvars) group(gid) id(pid) draws(20000) ///
> burn(10000) thin(10) saving(beta1) replace


Bayesian Mixed Logit Model                           Observations    =      4780
                                                     Groups          =       100
Acceptance rates:                                    Choices         =      1195
 Fixed coefs              = 0.087                    Total draws     =     20000
 Random coefs(ave,min,max)= 0.230, 0.195, 0.270      Burn-in draws   =     10000
                                                     *One of every 10 draws kept
------------------------------------------------------------------------------------------
                   y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------------
Fixed                |
               price |  -.2010195    .0121183   -16.59   0.000    -.2247997   -.1772393
---------------------+--------------------------------------------------------------------
Random               |
            contract |  -.2891212    .0596938    -4.84   0.000    -.4062607   -.1719817
               local |    2.64075    .3331781     7.93   0.000     1.986942    3.294558
              wknown |   2.048036    .2474377     8.28   0.000     1.562479    2.533592
                 tod |  -9.774435    .5827792   -16.77   0.000    -10.91805   -8.630824
            seasonal |  -9.807829    .5833574   -16.81   0.000    -10.95257   -8.663084
---------------------+--------------------------------------------------------------------
Cov_Random           |
        var_contract |   .2605256    .0494688     5.27   0.000      .163451    .3576002
   cov_contractlocal |   .3446803    .1849132     1.86   0.063    -.0181821    .7075426
  cov_contractwknown |   .2895433    .1418135     2.04   0.041     .0112571    .5678296
     cov_contracttod |  -.1531769    .1834363    -0.84   0.404    -.5131412    .2067874
cov_contractseasonal |  -.0339583    .1600677    -0.21   0.832    -.3480654    .2801489
           var_local |   7.454412    1.762436     4.23   0.000     3.995915    10.91291
     cov_localwknown |   4.680826    1.177204     3.98   0.000     2.370753      6.9909
        cov_localtod |  -.3037282    1.097876    -0.28   0.782    -2.458133    1.850676
   cov_localseasonal |   .1552508    .9601524     0.16   0.872    -1.728894    2.039395
          var_wknown |   3.827949    .9450303     4.05   0.000     1.973479    5.682419
      cov_wknowntod  |   -.604678    .8000253    -0.76   0.450    -2.174599     .965243
  cov_wknownseasonal |  -.1699213    .6828207    -0.25   0.804    -1.509847    1.170004
             var_tod |   6.888885    1.701412     4.05   0.000     3.550138    10.22763
     cov_todseasonal |   4.245384    1.145296     3.71   0.000     1.997926    6.492842
        var_seasonal |   5.743988    1.362853     4.21   0.000     3.069607    8.418369
------------------------------------------------------------------------------------------
(output omitted)
```

References and further reading I

- Baker MJ. 2013. Adaptive Markov chain Monte Carlo sampling and estimation in Mata. Hunter College Department of Economics Working Paper 440. Available at http://ideas.repec.org/p/htr/hcecon/440.html.
- Fiebig DG, Keane MP, Louviere J, Wasi N. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Marketing Science 29: 393–421.
- Greene WH, Hensher DA. 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. Transportation 37: 413–428.
- Gu Y, Hole AR, Knox S. 2013. Fitting the generalized multinomial logit model in Stata. Stata Journal 13: 382-397.
- Hole AR. 2007. Fitting mixed logit models by using maximum simulated likelihood. Stata Journal 7: 388-401.
- Huber J, Train K. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. Marketing Letters 12: 259-269.
- Louviere JJ, Meyer RJ, Bunch DS, Carson RT, Dellaert B, Hanemann WM, Hensher D, Irwin J. 1999. Combining sources of preference data for modeling complex decision processes. Marketing Letters 10: 205–217.

References and further reading II

- McFadden DL. 1974. Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics, Zarembka P (ed.). Academic Press: New York.
- Pacifico D, Yoo H. 2012. lclogit: A Stata module for estimating latent class conditional logit models via the Expectation-Maximization algorithm. Stata Journal, forthcoming. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2174146.
- Rabe-Hesketh, S and Skrondal, A. 2012. Multilevel and Longitudinal Modeling Using Stata, Third Edition. Stata Press: College Station.
- Revelt D, Train K. 2000. Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. Working Paper, Department of Economics, University of California, Berkeley.
- Train KE, Weeks M. 2005. Discrete choice models in preference space and willingness-to-pay space. In Application of simulation methods in environmental and resource economics, Scarpa R, Aberini A (eds.). Springer: Dordrecht.
- Train KE. 2009. Discrete Choice Methods with Simulation, Second Edition. Cambridge University Press: Cambridge.