COLLABORATIVE MASTERS PROGRAMME IN ECONOMICS FOR ANGLOPHONE AFRICA (CMAP)

JOINT FACILITY FOR ELECTIVES

ECONOMETRICS THEORY AND PRACTICE PART TWO:TOPICS IN MICRO-ECONOMETRICS

AUGUST/SEPTEMBER 2011

DR. MOSES SICHEI*

* KSMS Research Centre
Mobile:+254 723383505;Email: sichei@yahoo.co.uk or Sicheimm@centralbank.go.ke

Date: 22/8/2011

PRACTICAL 2: MULTINOMIAL LOGIT

1. OBJECTIVE:

Get a hands-on-experience in specifying, estimating and interpreting results from multinomial logit model. Specifically the practical focuses on;

(i) Specification of MNL
(ii) Estimation of these models
(iii) Interpretation of the models
(iv) Diagnostic statistics
(v) Presentation of the results

2. DATA:

- We use Uganda's household survey data dealing with enterprises' source of income.
- The data is called **Uganda enterprise data**
- It is a survey of 4093 firms (after cleaning the file)

- The main research problem is to ***examine the factors that determine the source of funds for setting up the business***.
- In the file, this corresponds to question h9q6 which was stated in the questionnaire as

What was the main source of money for setting up the business?
(You can check this by right clicking and viewing the variable properties)

- The possible alternatives (look at variable definitions) are:

1. Did not need money
2. Own savings
3. Commercial/development bank
4. Microfinance institutions
5. Local group
6. NGO
7. Other

You can check this in stata by using tabulate
tabulate h9q6

```
tabulate h9q6

 what was the main source |
  of money for setting up |
              the busines |      Freq.     Percent        Cum.
-------------------------+-----------------------------------
      dint need any money |        476       11.65       11.65
              own savings |      2,989       73.15       84.80
     commercial/devt bank |         32        0.78       85.58
microfinance institutions |         74        1.81       87.40
              local group |         59        1.44       88.84
                      ngo |          9        0.22       89.06
                    other |        447       10.94      100.00
-------------------------+-----------------------------------
                    Total |      4,086      100.00
```

- The following are the factors that explain the choice are;
  - Average expenditure on wages (h9q13)- x_{1i}
  - Number of people hired (h9q12)- x_{2i}
  - Average expenditure on raw materials (h9q14)- x_{3i}
  - Other operating expenses such as fuel, kerosene, electricity (h9q15)- x_{4i}
  - Monthly gross revenue (h9q11)- x_{5i}
  - Months of enterprise operation (h9q10)- x_{6i}
  - Age of the enterprise (h9q5 recorded as time)- x_{7i}

3.1 Specification of the MNL

3.1.1 Specify the multinomial density for one observation

Assume that the source of income is equal to y

$$f(y) = p_1^{y_1} \times p_2^{y_2} \times \ldots \times p_7^{y_7} = \prod_{j=1}^{7} p_j^{y_j}$$

3.1.2 Specify the likelihood function

$$L = \prod_{i=1}^{4093} \prod_{j=1}^{7} p_{ij}^{y_{ij}}$$

Where

$$p_{ij} = \frac{e^{\beta'_j x_{ij}}}{\sum_j^6 e^{\beta'_j x_i}}$$ for **multinomial logit model** (non-alternative varying regressors

- Average expenditure on wages (h9q13)- x_{1i}
- Number of people hired (h9q12)- x_{2i}
- Average expenditure on raw materials (h9q14)- x_{3i}
- Other operating expenses such as fuel, kerosene, electricity (h9q15)- x_{4i}
- Monthly gross revenue (h9q11)- x_{5i}
- Months of enterprise operation (h9q10)- x_{6i}
- Age of the enterprise (h9q5 recorded as time)- x_{7i}

$$x = (x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i})$$

$$\beta^j = (\beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \beta_5^1, \beta_6^j, \beta_7^j)$$

$$x\beta^{j} = \left(\beta_1^j x_{1i} + \beta_2^j x_{2i} + \beta_3^j x_{3i} + \beta_4^j x_{4i} + \beta_5^j x_{5i} + \beta_6^j x_{6i} + \beta_7^j x_{7i}\right)$$

The coefficients for alternative 1(did not need any money) is
$\beta^1 = \left(\beta_1^1, \beta_2^1, \beta_3^1, \beta_4^1, \beta_5^1, \beta_6^1, \beta_7^1\right)$

Its probability is

$$\Pr[y=1] = \frac{e^{\beta_1^1 x_{1i} + \beta_2^1 x_{2i} + \beta_3^1 x_{3i} + \beta_4^1 x_{4i} + \beta_5^1 x_{5i} + \beta_6^1 x_{6i} + \beta_7^1 x_{7i}}}{e^{\beta_1^1 x_{1i} + \beta_2^1 x_{2i} + \beta_3^1 x_{3i} + \beta_4^1 x_{4i} + \beta_5^1 x_{5i} + \beta_6^1 x_{6i} + \beta_7^1 x_{7i}} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$

Notice here that we have not substituted for the other alternatives due to space limitations

$$\Pr[y=1] = \frac{e^{x\beta^1}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=2] = \frac{e^{x\beta^2}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=3] = \frac{e^{x\beta^3}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=4] = \frac{e^{x\beta^4}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=5] = \frac{e^{x\beta^5}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=6] = \frac{e^{x\beta^6}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=7] = \frac{e^{x\beta^7}}{e^{x\beta^1} + e^{x\beta^2} + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$

- However this model is unidentified in the sense that there is more than one solution to $\beta^1, \beta^2, \beta^3, \beta^4, \beta^5, \beta^6$ and β^7 that lead to the same probability for $y=1, y=2, y=3, y=4, y=5, y=6$ and $y=7$

- To identify the model, we arbitrarily set one of $\beta^1, \beta^2, \beta^3, \beta^4, \beta^5, \beta^6$ and β^7 equal to zero-it does not matter which one
- Suppose we choose $\beta^2 = 0$ i.e.

$x\beta^2 = (0x_{1i} + 0x_{2i} + 0x_{3i} + 0x_{4i} + 0x_{5i} + 0x_{6i} + 0x_{7i}) = 0$

- The remaining coefficients will measure the change relative to the second group (own savings)
- Once $\beta^2 = 0$, it means that $e^{x\beta^2} = e^{x0} = e^0 = 1$ and the probability equations become
-

$$\Pr[y=1] = \frac{e^{x\beta^1}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=2] = \frac{1}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=3] = \frac{e^{x\beta^3}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=4] = \frac{e^{x\beta^4}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=5] = \frac{e^{x\beta^5}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=6] = \frac{e^{x\beta^6}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$
$$\Pr[y=7] = \frac{e^{x\beta^7}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$

- The relative probability of $y=1$ (didn't need any money) to the base outcome (own savings) is

$$\frac{\Pr[y=1]}{\Pr[y=2]} = \frac{e^{x\beta^1}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}} \div \frac{1}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}$$

$$\frac{\Pr[y=1]}{\Pr[y=2]} = \frac{e^{x\beta^1}}{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}} \times \frac{e^{x\beta^1} + 1 + e^{x\beta^3} + e^{x\beta^4} + e^{x\beta^5} + e^{x\beta^6} + e^{x\beta^7}}{1}$$

$$\frac{\Pr[y=1]}{\Pr[y=2]} = e^{x\beta^1}$$

This is the relative risk

- The ratio of the relative risk for one unit change in the regressors is e^{β^1}. This is what stata presents when you invoke rrr option

3.2 What are the assumptions in your model?

- Sigmoid property
- Equivalence property i.e. that if the terms for each alternative are improved equally there is no change
- IIA assumption: recall the "blue/red bus paradox"

What are the FOC for our model?

Multinomial Logit

$$\frac{\partial LL}{\partial \beta_k} = \sum_i [y_{ik} - p_{ik}] x_i$$

$$\frac{\partial p_{ij}}{\partial \beta_j} = p_{ij} x_i - p_{ij} p_{ij} x_i$$

For $j \neq k$

$$\frac{\partial p_{ij}}{\partial \beta_k} = -p_{ij} p_{ij} x_i$$

S.O.C

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_k'} = -\sum_{i=1}^{N} \sum_{j=1}^{J} p_{ij} \left(\delta_{ij} - p_{ij} \right) x_i x_i'$$

Where δ_{ij} is an indicator variable equal to 1 if $j = k$ and equal to 0 if $j \neq k$

Marginal effects

$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij} \left(\beta_j - \overline{\beta}_i \right)$$

Where $\overline{\beta}_i = \sum_j p_{ij} \beta_j$

In the command space type

mlogit funds h9q12 h9q14 h9q15 h9q11 h9q10 age, nolog

```
. mlogit funds h9q12 h9q14 h9q15 h9q11 h9q10 age, nolog

Multinomial logistic regression                   Number of obs   =       3336
                                                  LR chi2(30)     =     359.15
                                                  Prob > chi2     =     0.0000
Log likelihood = -1915.3927                       Pseudo R2       =     0.0857

------------------------------------------------------------------------------
       funds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
       h9q12 |   .0636653   .0323901     1.97   0.049     .0001819    .1271486
       h9q14 |  -8.92e-06   1.57e-06    -5.68   0.000     -.000012   -5.84e-06
       h9q15 |  -.0000278   6.22e-06    -4.47   0.000      -.00004   -.0000156
       h9q11 |  -1.31e-06   5.08e-07    -2.59   0.010    -2.31e-06   -3.20e-07
       h9q10 |  -.0071651    .014215    -0.50   0.614     -.035026    .0206959
         age |  -.0084486   .0056344    -1.50   0.134    -.0194919    .0025946
       _cons |  -.4174905   .4813677    -0.87   0.386    -1.360954    .5259729
-------------+----------------------------------------------------------------
3            |
       h9q12 |   .0631741   .0326544     1.93   0.053    -.0008273    .1271756
       h9q14 |  -7.73e-08   4.28e-08    -1.80   0.071    -1.61e-07    6.66e-09
       h9q15 |   7.67e-07   3.78e-07     2.03   0.042     2.71e-08    1.51e-06
       h9q11 |   6.30e-08   3.30e-08     1.91   0.056    -1.75e-09    1.28e-07
       h9q10 |   .0848217    .060003     1.41   0.157    -.0327821    .2024255
         age |   .0996839   .0451355     2.21   0.027     .0112198    .1881479
       _cons |  -13.51407   3.866806    -3.49   0.000    -21.09287   -5.935268
-------------+----------------------------------------------------------------
4            |
       h9q12 |   .0631793    .032445     1.95   0.052    -.0004118    .1267704
       h9q14 |   4.12e-08   1.26e-07     0.33   0.743    -2.05e-07    2.87e-07
       h9q15 |   7.78e-07   3.72e-07     2.09   0.037     4.87e-08    1.51e-06
       h9q11 |  -5.03e-08   1.09e-07    -0.46   0.645    -2.64e-07    1.64e-07
       h9q10 |    .085028   .0390105     2.18   0.029     .0085689    .1614872
         age |   .0603862   .0232125     2.60   0.009     .0148905     .105882
```

```
       _cons |  -9.378446   1.998056    -4.69   0.000    -13.29456   -5.462328
-------------+----------------------------------------------------------------
5            |
       h9q12 |   .0632242   .0324871     1.95   0.052    -.0004494    .1268978
       h9q14 |  -2.88e-07   3.76e-07    -0.77   0.443    -1.02e-06    4.48e-07
       h9q15 |  -9.63e-08   8.88e-07    -0.11   0.914    -1.84e-06    1.65e-06
       h9q11 |   6.76e-08   4.96e-08     1.36   0.173    -2.96e-08    1.65e-07
       h9q10 |  -.0638766   .0364832    -1.75   0.080    -.1353825    .0076292
         age |   .0483246   .0265675     1.82   0.069    -.0037468     .100396
       _cons |   -7.30361   2.260212    -3.23   0.001    -11.73354   -2.873675
-------------+----------------------------------------------------------------
6            |
       h9q12 |   .0636869   .0345752     1.84   0.065    -.0040792     .131453
       h9q14 |   4.86e-09   1.38e-06     0.00   0.997    -2.70e-06    2.71e-06
       h9q15 |  -.0000935   .0000885    -1.06   0.291     -.000267      .00008
       h9q11 |  -9.95e-08   1.40e-06    -0.07   0.943    -2.84e-06    2.64e-06
       h9q10 |   .0523304   .1050051     0.50   0.618    -.1534758    .2581367
         age |   .5404869    .309054     1.75   0.080    -.0652478    1.146222
       _cons |  -50.72461   26.38474    -1.92   0.055    -102.4378    .9885384
------------------------------------------------------------------------------
(funds==2 is the base outcome)
```

- The multinomial logit is equivalent to running a series of binomial logits:

3.3 Interpretation

- The output has six parts, labelled with the categories of the outcome funds. They correspond to 6 equations. For instance equation 1 is

$$\log\left[\frac{\Pr(y=1)}{\Pr(y=2)}\right] = \beta_0 + \beta_1^1 x_{1i} + \beta_2^1 x_{2i} + \beta_3^1 x_{3i} + \beta_4^1 x_{4i} + \beta_5^1 x_{5i} + \beta_6^1 x_{6i} + \beta_7^1 x_{7i}$$

$$\log\left[\frac{\Pr(y=3)}{\Pr(y=2)}\right] = \beta_0 + \beta_1^3 x_{1i} + \beta_2^3 x_{2i} + \beta_3^3 x_{3i} + \beta_4^3 x_{4i} + \beta_5^3 x_{5i} + \beta_6^3 x_{6i} + \beta_7^3 x_{7i}$$

.

.

- With the betas being the raw regression coefficients from the output above
- For instance we can say that for a one unit change in the variable h9q12 (number of people hired), the log of the ratio of the two probabilities $\left[\frac{\Pr(y=1)}{\Pr(y=2)}\right]$ i.e. didn't need any money vs own savings will be increased by 0.064

Let’s rerun the model with **base category as 4**

mlogit funds h9q12 h9q14 h9q15 h9q11 h9q10 age, base(4)

```
. mlogit funds h9q12 h9q14 h9q15 h9q11 h9q10 age, base(4)

Iteration 0:   log likelihood = -2094.9693
Iteration 1:   log likelihood = -2076.3472
Iteration 2:   log likelihood =  -2064.924
Iteration 3:   log likelihood =  -2056.514
Iteration 4:   log likelihood = -2043.4081
Iteration 5:   log likelihood =  -2024.253
Iteration 6:   log likelihood = -2006.9297
Iteration 7:   log likelihood = -1999.0699
Iteration 8:   log likelihood = -1985.7619
Iteration 9:   log likelihood = -1979.0517
Iteration 10:  log likelihood = -1948.6815
Iteration 11:  log likelihood = -1932.0869
Iteration 12:  log likelihood = -1927.9138
Iteration 13:  log likelihood = -1916.7655
Iteration 14:  log likelihood = -1915.4309
Iteration 15:  log likelihood = -1915.3938
Iteration 16:  log likelihood = -1915.3927
Iteration 17:  log likelihood = -1915.3927

Multinomial logistic regression                   Number of obs   =       3336
                                                  LR chi2(30)     =     359.15
                                                  Prob > chi2     =     0.0000
Log likelihood = -1915.3927                       Pseudo R2       =     0.0857

------------------------------------------------------------------------------
       funds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
       h9q12 |   .0004859   .0024853     0.20   0.845    -.0043851     .005357
       h9q14 |  -8.97e-06   1.58e-06    -5.69   0.000    -.0000121   -5.88e-06
       h9q15 |  -.0000286   6.23e-06    -4.59   0.000    -.0000408   -.0000164
       h9q11 |  -1.26e-06   5.17e-07    -2.45   0.014    -2.28e-06   -2.51e-07
       h9q10 |  -.0921931   .0408577    -2.26   0.024    -.1722728   -.0121134
         age |  -.0688349   .0236984    -2.90   0.004    -.1152829   -.0223868
       _cons |   8.960956   2.038385     4.40   0.000     4.965795    12.95612
-------------+----------------------------------------------------------------
2            |
       h9q12 |  -.0631794    .032445    -1.95   0.052    -.1267704    .0004117
       h9q14 |  -4.12e-08   1.26e-07    -0.33   0.743    -2.87e-07    2.05e-07
       h9q15 |  -7.78e-07   3.72e-07    -2.09   0.037    -1.51e-06   -4.86e-08
       h9q11 |   5.03e-08   1.09e-07     0.46   0.645    -1.64e-07    2.64e-07
       h9q10 |   -.085028   .0390105    -2.18   0.029    -.1614871   -.0085689
         age |  -.0603862   .0232124    -2.60   0.009    -.1058818   -.0148907
       _cons |   9.378447   1.998047     4.69   0.000     5.462347    13.29455
-------------+----------------------------------------------------------------
3            |
       h9q12 |  -5.22e-06   .0051624    -0.00   0.999    -.0101234     .010113
       h9q14 |  -1.18e-07   1.32e-07    -0.90   0.371    -3.78e-07    1.41e-07
       h9q15 |  -1.11e-08   6.58e-08    -0.17   0.866    -1.40e-07    1.18e-07
       h9q11 |   1.13e-07   1.11e-07     1.02   0.308    -1.05e-07    3.31e-07
       h9q10 |  -.0002062   .0710356    -0.00   0.998    -.1394334     .139021
         age |   .0392977   .0504232     0.78   0.436    -.0595299    .1381253
       _cons |  -4.135626   4.322286    -0.96   0.339    -12.60715    4.335898
-------------+----------------------------------------------------------------
5            |
       h9q12 |   .0000449   .0038893     0.01   0.991     -.007578    .0076677
       h9q14 |  -3.29e-07   3.96e-07    -0.83   0.405    -1.10e-06    4.46e-07
       h9q15 |  -8.74e-07   9.41e-07    -0.93   0.353    -2.72e-06    9.70e-07
       h9q11 |   1.18e-07   1.19e-07     0.99   0.323    -1.16e-07    3.52e-07
       h9q10 |  -.1489046   .0528149    -2.82   0.005      -.25242   -.0453893
```

```
         age |  -.0120617   .0349543    -0.35   0.730    -.0805707    .0564474
       _cons |   2.074837   2.987156     0.69   0.487    -3.779881    7.929554
-------------+----------------------------------------------------------------
6            |
       h9q12 |   .0005075   .0124897     0.04   0.968    -.0239719     .024987
       h9q14 |  -3.63e-08   1.39e-06    -0.03   0.979    -2.75e-06    2.68e-06
       h9q15 |  -.0000943   .0000885    -1.06   0.287    -.0002677    .0000792
       h9q11 |  -4.92e-08   1.40e-06    -0.04   0.972    -2.80e-06    2.70e-06
       h9q10 |  -.0326976   .1117159    -0.29   0.770    -.2516567    .1862616
         age |   .4801006   .3098751     1.55   0.121    -.1272434    1.087445
       _cons |  -41.34616   26.45603    -1.56   0.118    -93.19904    10.50671
------------------------------------------------------------------------------
(funds==4 is the base outcome)
```

$$\Pr[y=1]=\frac{e^{x\beta^1}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=2]=\frac{e^{x\beta^2}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=3]=\frac{e^{x\beta^3}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=4]=\frac{1}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=5]=\frac{e^{x\beta^5}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=6]=\frac{e^{x\beta^6}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

$$\Pr[y=7]=\frac{e^{x\beta^7}}{e^{x\beta^1}+e^{x\beta^2}+e^{x\beta^3}+1+e^{x\beta^5}+e^{x\beta^6}+e^{x\beta^7}}$$

- Do you see the effect of the IIA assumption? (outcome 2 and outcome 4 results are exactly the same)
- We could do the same for other pairs

How do we present such results?

Multinomial logit estimates

Variable	**4 v.1**	**4 v.2**	**4 V.3**	**4 v.5**	**4 v.6**	**2 V.1**
h9q12	.0004859					
h9q14	-8.97e-06***					
h9q15	-.0000286**					

h9q11	-1.26e-06**					
h9q10	- .0921931**					
age	-.0688349**					

Alternatives

1. Did not need money
2. Own savings
3. Commercial/development bank
4. Microfinance institutions
5. Local group
6. NGO
7. Other

- The sign of a coefficient estimate reflects the direction of change in the risk ratio (the ratio between P(Y=ngo) / P(Y=own savings) in response to a ceteris paribus change in the value to which the coefficient is attached.
  - It does not reflect the direction of change in the individual probabilities P(Y=k).

Relative odds ratio

- Odds ratios: The odds ratios are simply the ratio of the exponentiated coefficients.
- This is computed in stata using the relative risk ratio (rrr) as follows

Type **mlogit,rrr**

```
mlogit,rrr

Multinomial logistic regression                   Number of obs   =       3336
                                                  LR chi2(30)     =     359.15
                                                  Prob > chi2     =     0.0000
Log likelihood = -1915.3927                       Pseudo R2       =     0.0857

------------------------------------------------------------------------------
       funds |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
       h9q12 |   1.000486   .0024865     0.20   0.845     .9956245    1.005371
       h9q14 |    .999991   1.58e-06    -5.69   0.000     .9999879    .9999941
       h9q15 |   .9999714   6.23e-06    -4.59   0.000     .9999592    .9999836
       h9q11 |   .9999987   5.17e-07    -2.45   0.014     .9999977    .9999997
       h9q10 |   .9119291   .0372594    -2.26   0.024     .8417495    .9879597
         age |   .9334808    .022122    -2.90   0.004      .891114    .9778619
```

```
-------------+----------------------------------------------------------------
2            |
       h9q12 |   .9387751   .0304586    -1.95   0.052     .8809359    1.000412
       h9q14 |          1   1.26e-07    -0.33   0.743     .9999997           1
       h9q15 |   .9999992   3.72e-07    -2.09   0.037     .9999985           1
       h9q11 |          1   1.09e-07     0.46   0.645     .9999998           1
       h9q10 |   .9184866   .0358306    -2.18   0.029     .8508775    .9914677
         age |   .9414009   .0218522    -2.60   0.009      .899531    .9852196
-------------+----------------------------------------------------------------
3            |
       h9q12 |   .9999948   .0051624    -0.00   0.999     .9899277    1.010164
       h9q14 |   .9999999   1.32e-07    -0.90   0.371     .9999996           1
       h9q15 |          1   6.58e-08    -0.17   0.866     .9999999           1
       h9q11 |          1   1.11e-07     1.02   0.308     .9999999           1
       h9q10 |   .9997938    .071021    -0.00   0.998     .8698509    1.149148
         age |    1.04008   .0524441     0.78   0.436     .9422073    1.148119
-------------+----------------------------------------------------------------
5            |
       h9q12 |   1.000045   .0038895     0.01   0.991     .9924506    1.007697
       h9q14 |   .9999997   3.96e-07    -0.83   0.405     .9999989           1
       h9q15 |   .9999991   9.41e-07    -0.93   0.353     .9999973    1.000001
       h9q11 |          1   1.19e-07     0.99   0.323     .9999999           1
       h9q10 |   .8616513    .045508    -2.82   0.005     .7769184    .9556254
         age |   .9880108   .0345352    -0.35   0.730     .9225896    1.058071
-------------+----------------------------------------------------------------
6            |
       h9q12 |   1.000508   .0124961     0.04   0.968     .9763132    1.025302
       h9q14 |          1   1.39e-06    -0.03   0.979     .9999972    1.000003
       h9q15 |   .9999058   .0000885    -1.06   0.287     .9997323    1.000079
       h9q11 |          1   1.40e-06    -0.04   0.972     .9999972    1.000003
       h9q10 |   .9678312   .1081221    -0.29   0.770     .7775116    1.204737
         age |   1.616237   .5008316     1.55   0.121     .8805193    2.966684
------------------------------------------------------------------------------
(funds==4 is the base outcome)

.
```

- Notice that all the coefficients with negative signs have less than 1 odds ratio
- If the odds ratio is higher than 1 it favours the numerator outcome, if it is less than 1, it favours the base outcome
- These results show the relative risk ratio for one unit change in the corresponding variable (measured as the risk of the outcome relative to the base outcome)

$$\frac{\Pr[y=j]}{\Pr[y=2]} = e^{x\beta^2}$$

4. Computation of the Marginal effects (partial effects)

- We need the marginal effects to interpret the results of MNL effectively

- The marginal effects show how the probabilities of each outcome change with respect to changes in regressors
- To calculate the marginal effects we run the **mfx** command separately for each outcome

mfx,predict(outcome(1))

```
.  mfx,predict(outcome(1))

Marginal effects after mlogit
      y  = Pr(funds==1) (predict, outcome(1))
         =  .00228193
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   h9q12 |   .0001379       .0001    1.35   0.177  -.000062  .000338   2.00779
   h9q14 |  -2.03e-08      .00000   -2.26   0.024  -3.8e-08 -2.7e-09    255965
   h9q15 |  -6.34e-08      .00000   -2.58   0.010  -1.1e-07 -1.5e-08   74137.4
   h9q11 |  -2.99e-09      .00000   -1.51   0.132  -6.9e-09  9.0e-10    506088
   h9q10 |  -.0000201      .00003   -0.58   0.561  -.000088  .000048   9.25869
     age |  -.0000263      .00002   -1.37   0.172  -.000064  .000011   78.4206
------------------------------------------------------------------------------

.
mfx,predict(outcome(2))

Marginal effects after mlogit
      y  = Pr(funds==2) (predict, outcome(2))
         =  .94850262
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   h9q12 |  -.0030878      .00171   -1.81   0.071  -.006437  .000261   2.00779
   h9q14 |   2.36e-08      .00000    2.18   0.029   2.4e-09  4.5e-08    255965
   h9q15 |   3.80e-08      .00000    1.26   0.208  -2.1e-08  9.7e-08   74137.4
   h9q11 |   2.39e-09      .00000    0.72   0.473  -4.1e-09  8.9e-09    506088
   h9q10 |  -.0015592      .00115   -1.35   0.176  -.003818  .000699   9.25869
     age |  -.0029235      .00073   -4.01   0.000  -.004351 -.001496   78.4206
------------------------------------------------------------------------------
```

We can do for the rest of the categories

Change in predicted probabilities/Marginal effects

Variable	**Did not need money**	**Own savings**	**Commercial /development bank**	**Microfinance**	**Local group**	**NGO**
h9q12	.0001379	-.0030878*	.0005089*	.001425*	.001016*	2.35e-08
h9q14	-2.03e-08**	2.36e-08 **	-4.45e-10	1.57e-09	-4.46e-09	1.16e-14
h9q15	-6.34e-08**	3.80e-08	6.86e-09*	1.95e-08**	-9.52e-10	-3.64e-11
h9q11	-2.99e-09	2.39e-09	5.57e-10**	-1.14e-09	1.19e-09	-3.78e-14
h9q10	-.0000201	-.0015592	.0007064	.0019829**	-.00111*	1.97e-08

age	-.0000263	-.0029235 ***	.0008204**	.0013627	.0007665*	2.09e-07
Mean value	.00228193	.94850262	.00849295	.02378039	.01694173	3.894e-07

Attributes

- Average expenditure on wages (h9q13)- x_{1i}
- Number of people hired (h9q12)- x_{2i}
- Average expenditure on raw materials (h9q14)- x_{3i}
- Other operating expenses such as fuel, kerosene, electricity (h9q15)- x_{4i}
- Monthly gross revenue (h9q11)- x_{5i}
- Months of enterprise operation (h9q10)- x_{6i}
- Age of the enterprise (h9q5 recorded as time)- x_{7i}

Interpretation of marginal effects

- We interpret the results the same way we did for the logit and probit model

4.2 Predicting

Could predict

- Probabilities such as **predict p1 if e(sample),outcome(1)**

- Index values e.g. **predict idx2, outcome(2) xb**.

- Notice that own savings was our base case-the outcome for which all the coefficients were set to 0-so the index is always 0

4.3 Test of Hypothesis about coefficients

4.3.1 Testing whether variables have zero effects across all equations

- If we list variables after the test command, we are testing that the corresponding coefficients are all zeros across all equations

test age h9q11

```
. test age h9q11

 ( 1)  [1]age = 0
 ( 2)  [3]age = 0
 ( 3)  [4]age = 0
 ( 4)  [5]age = 0
 ( 5)  [6]age = 0
 ( 6)  [1]h9q11 = 0
 ( 7)  [3]h9q11 = 0
 ( 8)  [4]h9q11 = 0
 ( 9)  [5]h9q11 = 0
 (10)  [6]h9q11 = 0

           chi2( 10) =   29.37
         Prob > chi2 =    0.0011
```

- We reject the null that the age and monthly gross revenue do not affect the probability of choosing the different alternatives

4.3.2 Testing whether all the coefficients (except the constant) in a single equation are zero

- Simply use test and type the outcome in a square bracket
- Example
- The results are

. test [6]

```
 ( 1)  [6]h9q12 = 0
 ( 2)  [6]h9q14 = 0
```

(3) [6]h9q15 = 0
(4) [6]h9q11 = 0
(5) [6]h9q10 = 0
(6) [6]age = 0

chi2(6) =   7.61
Prob > chi2 =   0.2677

- We cannot reject the null implying that the factors do not affect the probability of getting funds from ngo

4.3.3 Testing whether a specific variable in a single equation are zero

Type
test [outcome]: var1 var2….varn

Example
test [6]:age h9q10

The results are;
test [6]:age h9q10

(1) [6]age = 0
(2) [6]h9q10 = 0

chi2(2) =   3.17
Prob > chi2 =   0.2048

- We reject the null hypothesis

4.3.4 Testing whether coefficients are equal across equations

For instance we can test whether all the coefficients except the constant are equal for the did not need money and commercial bank outcomes as follows

test [1=3]

```
 ( 1)  [1]h9q12 - [3]h9q12 = 0
 ( 2)  [1]h9q14 - [3]h9q14 = 0
 ( 3)  [1]h9q15 - [3]h9q15 = 0
 ( 4)  [1]h9q11 - [3]h9q11 = 0
 ( 5)  [1]h9q10 - [3]h9q10 = 0
 ( 6)  [1]age - [3]age = 0

           chi2(  6) =  103.40
         Prob > chi2 =    0.0000
```

Can we reject the null hypothesis?

4.3.5 Testing whether some variables are equal across equations

To test that only the age and h9q10 are equal in the local group and ngo outcomes. We type the following

Test [5=6]: age h9q10

The results are as follows

```
test [other=ngo]: age h9q10

. test [5=6]: age h9q10

 ( 1)  [5]age - [6]age = 0
 ( 2)  [5]h9q10 - [6]h9q10 = 0

           chi2(  2) =    2.67
         Prob > chi2 =    0.2628
```

5. **Testing for the IIA Assumption**

- The strategy is to estimate the model with the choice you want to check (unconstrained model) and without the choice (the constrained model)

- If the IIA assumption is true, the constrained and the unconstrained estimated coefficients on the remaining categories should not be statistically different
- The Hausman test statistic is
  $(b_c - b_u)'[Cov(b_c) - Cov(b_u)]^{-1}(b_c - b_u)$
- Where b_c and b_u are the constrained and unconstrained coefficients and $Cov(b_c)$ and $Cov(b_u)$ are the estimated covariances
- The statistic has an approximate chi-square distribution with the number of degrees of freedom equal to the number of coefficients in the constrained model
- Let's do this for a simple mlogit model of the form
- We use Hausman test in stata to do this

The unconstrained model is

mlogit funds gender

```
mlogit funds gender

Iteration 0:   log likelihood = -2293.3044
Iteration 1:   log likelihood = -2290.1421
Iteration 2:   log likelihood = -2289.9838
Iteration 3:   log likelihood = -2289.9833

Multinomial logistic regression                   Number of obs   =       3639
                                                  LR chi2(5)      =       6.64
                                                  Prob > chi2     =     0.2486
Log likelihood = -2289.9833                       Pseudo R2       =     0.0014

------------------------------------------------------------------------------
       funds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
      gender |   .0111597   .0987067     0.11   0.910    -.1823018    .2046212
       _cons |  -1.842911   .0702081   -26.25   0.000    -1.980516   -1.705306
-------------+----------------------------------------------------------------
3            |
      gender |  -.9523214   .3948708    -2.41   0.016    -1.726254   -.1783889
       _cons |  -4.167002   .2101237   -19.83   0.000    -4.578837   -3.755167
-------------+----------------------------------------------------------------
4            |
      gender |  -.0140518   .2353558    -0.06   0.952    -.4753407    .4472372
       _cons |  -3.691579   .1664358   -22.18   0.000    -4.017787    -3.36537
-------------+----------------------------------------------------------------
5            |
      gender |  -.0479533   .2629723    -0.18   0.855    -.5633695    .4674628
```

```
       _cons |  -3.901299   .1844104   -21.16   0.000    -4.262737   -3.539861
-------------+----------------------------------------------------------------
6            |
      gender |   .2090918   .6718172     0.31   0.756    -1.107646    1.525829
       _cons |  -5.916202   .5006734   -11.82   0.000    -6.897504     -4.9349
------------------------------------------------------------------------------
(funds==2 is the base outcome)
```

Lets store the coefficients and the covariances using the command

est store all

Let's estimate the constrained model excluding category 1

mlogit funds gender if funds !=1

```
mlogit funds gender if funds !=1

Iteration 0:   log likelihood = -881.68375
Iteration 1:   log likelihood = -878.54174
Iteration 2:   log likelihood = -878.38548
Iteration 3:   log likelihood = -878.38503

Multinomial logistic regression                   Number of obs   =       3163
                                                  LR chi2(4)      =       6.60
                                                  Prob > chi2     =     0.1588
Log likelihood = -878.38503                       Pseudo R2       =     0.0037

------------------------------------------------------------------------------
       funds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
3            |
      gender |  -.9523214   .3948709    -2.41   0.016    -1.726254   -.1783887
       _cons |  -4.167002   .2101237   -19.83   0.000    -4.578837   -3.755167
-------------+----------------------------------------------------------------
4            |
      gender |  -.0140518   .2353558    -0.06   0.952    -.4753407    .4472372
       _cons |  -3.691579   .1664358   -22.18   0.000    -4.017787    -3.36537
-------------+----------------------------------------------------------------
5            |
      gender |  -.0479533   .2629723    -0.18   0.855    -.5633695    .4674628
       _cons |  -3.901299   .1844104   -21.16   0.000    -4.262737   -3.539861
-------------+----------------------------------------------------------------
6            |
      gender |   .2090918   .6718172     0.31   0.756    -1.107646    1.525829
       _cons |  -5.916202   .5006734   -11.82   0.000    -6.897504     -4.9349
------------------------------------------------------------------------------
(funds==2 is the base outcome)
```

Let's store the coefficients and the covariances
est store partial

Finally, let's do the Hausman test
hausman partial all, alleqs constant

```
hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
3            |
      gender |   -.9523214    -.9523214       -3.67e-11         .000329
       _cons |   -4.167002    -4.167002       -5.63e-13        .0000838
-------------+----------------------------------------------------------------
4            |
      gender |   -.0140518    -.0140518        1.91e-15        .0000292
       _cons |   -3.691579    -3.691579       -5.77e-15        .0000114
-------------+----------------------------------------------------------------
5            |
      gender |   -.0479533    -.0479533       -4.71e-14        .0000329
       _cons |   -3.901299    -3.901299        5.06e-14        .0000126
-------------+----------------------------------------------------------------
6            |
      gender |    .2090918     .2090918        7.24e-15        .0000803
       _cons |   -5.916202    -5.916202       -5.33e-15        .0000345
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.00
                Prob>chi2 =      1.0000
```

Interpretation

- Look at the way the Hausman test is computed
- It has 8 degrees of freedom
- We cannot reject the null that the constrained and unconstrained coefficients are the same
- This implies that IIA assumption is true
- This calls for other methods of estimation such as nested logit or multinomial probit