Multinomial Logit

Categorical and Limited Dependent Variables

Paul A. Jargowsky

3rd Motivation for Binary Models

- First we just viewed cumulative normal and cumulative logistic distributions as *convenient* functions
- Then we derived the models again using a *latent variable approach,* based on the distribution of the disturbance term.
- This week we derive the models based on individual choice by agents trying to *maximize utility*.

Utility Maximization Approach

$$Y_i = \begin{cases} 0 \\ 1 \end{cases}$$

In binary, Y=1 as often defined as the event of interest, but we can view both cases (0,1) as potential choices. Each choice has utility.

$$U_{1i} = \beta_{11} + \beta_{12}X_{2i} + \ldots + \beta_{1K}X_{Ki} + \varepsilon_{1i} = \mathbf{x}_i\boldsymbol{\beta}_1 + \varepsilon_{1i}$$

$$U_{0i} = \beta_{01} + \beta_{02}X_{2i} + \ldots + \beta_{0K}X_{ki} + \varepsilon_{0i} = \mathbf{x}_i\boldsymbol{\beta}_0 + \varepsilon_{0i}$$

You are going to choose the outcome (*Y=1*) if and only if the utility of that choice is greater than the utility of the other choice (*Y=0*). In other words,

Choose 1 if $U_{1i} > U_{0i}$

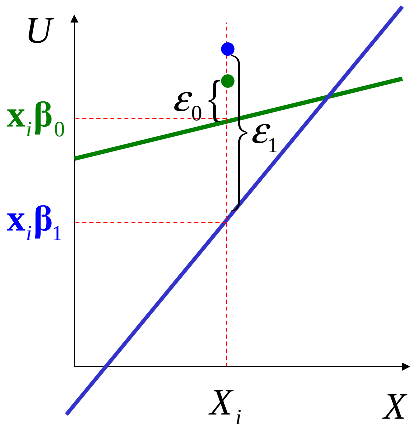
Probability of Choosing Y=1

Express the utilities of the choices in terms of the data and parameters we wish to estimate.

$$\begin{aligned}\Pr(Y_i = 1 \mid X_i) &= \Pr(U_{1i} > U_{0i} \mid X_i) \\ &= \Pr\left[(\mathbf{x}_i\boldsymbol{\beta}_1 + \varepsilon_{1i}) > (\mathbf{x}_i\boldsymbol{\beta}_0 + \varepsilon_{0i})\right] \\ &= \Pr\left[(\varepsilon_{1i} - \varepsilon_{0i}) > (\mathbf{x}_i\boldsymbol{\beta}_0 - \mathbf{x}_i\boldsymbol{\beta}_1)\right]\end{aligned}$$

To choose alternative one, the net difference in the disturbance terms must be at least large enough to offset the difference in the utilities.

Elements Determining Choice

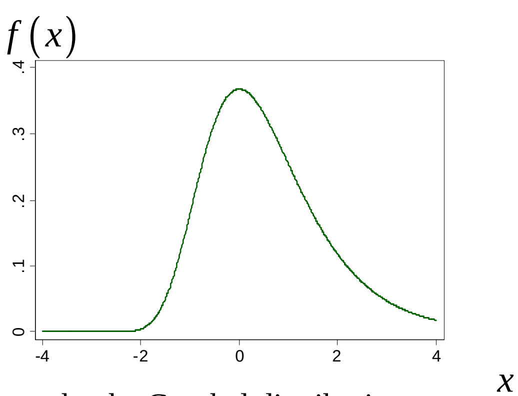


$$\Pr\left[\left(\varepsilon_{1i} - \varepsilon_{0i}\right) > \left(\mathbf{x}_i\boldsymbol{\beta}_0 - \mathbf{x}_i\boldsymbol{\beta}_1\right)\right]$$

In this example, the systematic part (expected value) of the utility favors choice 0, but the difference in the random elements is enough to offset that and leads to choice 1.

Type I Extreme Value Distribution

We assume that ε_1 and ε_0 are IID and have a "type I extreme value distribution".



aka the Gumbel distribution

$$f(x) = e^{-x-e^{-x}}$$

"The choice of the distribution is motivated by the simplicity, tractability, and usefulness of the resulting model." Long, p. 156.

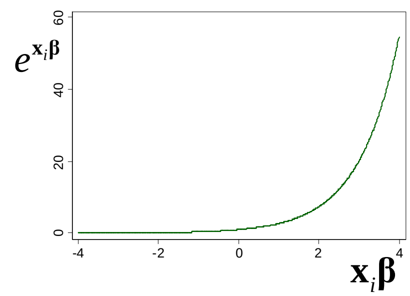
Probabilities

McFadden (1973) shows the resulting probability statements:

$$P_1 = \Pr\left(Y_1 = 1 \mid \mathbf{x}_i\right) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_1}}{e^{\mathbf{x}_i\boldsymbol{\beta}_0} + e^{\mathbf{x}_i\boldsymbol{\beta}_1}}$$

$$P_0 = \Pr\left(Y_1 = 0 \mid \mathbf{x}_i\right) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_0}}{e^{\mathbf{x}_i\boldsymbol{\beta}_0} + e^{\mathbf{x}_i\boldsymbol{\beta}_1}}$$

1) The numerator is always positive, therefore all Ps > 0.
2) Dividing by the total ensures all Ps < 1 and the total is one.



Identification Issue

$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \Pr\left[\left(\varepsilon_{1i} - \varepsilon_{0i}\right) > \left(\mathbf{x}_i\boldsymbol{\beta}_0 - \mathbf{x}_i\boldsymbol{\beta}_1\right)\right]$$
$$= \Pr\left[\left(\varepsilon_{1i} - \varepsilon_{0i}\right) > \mathbf{x}_i\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1\right)\right]$$

X is known. However, the expression on the right side of the inequality will be the same for an infinite number of combinations of β_0 and β_1. Only *differences* in coefficients are identified. Solution: fix one set of coefficients.

$$\boldsymbol{\beta}_0 = \mathbf{0} \quad \rightarrow \quad \beta_1 = 0, \beta_2 = 0, \ldots, \beta_K = 0$$

$$\rightarrow \quad \mathbf{x}_i\boldsymbol{\beta}_0 = 0 + 0X_{2i} + \ldots + 0X_{Ki} = 0$$

With $\boldsymbol{\beta}_0$=0, the Model Simplifies to the Standard Logit Model

Anything to the zero power is 1, therefore: $e^{\mathbf{x}_i\boldsymbol{\beta}_0} = e^0 = 1$

$$P_1 = \Pr(Y_1 = 1 \mid \mathbf{x}_i) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_1}}{e^{\mathbf{x}_i\boldsymbol{\beta}_0} + e^{\mathbf{x}_i\boldsymbol{\beta}_1}} = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_1}}{1 + e^{\mathbf{x}_i\boldsymbol{\beta}_1}} = \frac{1}{1 + e^{-\mathbf{x}_i\boldsymbol{\beta}}}$$

$$P_0 = \Pr(Y_1 = 0 \mid \mathbf{x}_i) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_0}}{e^{\mathbf{x}_i\boldsymbol{\beta}_0} + e^{\mathbf{x}_i\boldsymbol{\beta}_1}} = \frac{1}{1 + e^{\mathbf{x}_i\boldsymbol{\beta}_1}}$$

$$0 < P_0 < 1, \quad 0 < P_1 < 1, \quad P_0 + P_1 = \frac{1}{1 + e^{\mathbf{x}_i\boldsymbol{\beta}_1}} + \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_1}}{1 + e^{\mathbf{x}_i\boldsymbol{\beta}_1}} = 1$$

Extend to 3 choices

$$Y_i = \begin{cases} 1 & U_{1i} = \mathbf{x}_i\boldsymbol{\beta}_1 + \varepsilon_{1i}, \quad \boldsymbol{\beta}_1 = 0 \\ 2 & U_{2i} = \mathbf{x}_i\boldsymbol{\beta}_2 + \varepsilon_{2i} \\ 3 & U_{3i} = \mathbf{x}_i\boldsymbol{\beta}_3 + \varepsilon_{3i} \end{cases}$$

Choose Y=1 IFF $(U_{1i} > U_{2i}) \cap (U_{1i} > U_{3i})$

Choose Y=2 IFF $(U_{2i} > U_{1i}) \cap (U_{2i} > U_{3i})$

Choose Y=3 IFF $(U_{3i} > U_{1i}) \cap (U_{3i} > U_{2i})$

Assume ε_{ji} distributed IID Type I extreme value

MNL Probabilities with 3 Choices

$$P_1 = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_1}}{e^{\mathbf{x}_i \boldsymbol{\beta}_1} + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}} = \frac{1}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}}$$

$$P_2 = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_2}}{e^{\mathbf{x}_i \boldsymbol{\beta}_1} + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}} = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}}$$

$$P_3 = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_3}}{e^{\mathbf{x}_i \boldsymbol{\beta}_1} + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}} = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}_2} + e^{\mathbf{x}_i \boldsymbol{\beta}_3}}$$

$$0 < P_1 < 1$$
$$0 < P_2 < 1$$
$$0 < P_3 < 1$$

$$P_1 + P_2 + P_3 = 1$$

Utility Functions

The Utility functions for the different choices are:

car: $U_{1i} = \mathbf{x}_i \boldsymbol{\beta}_1 = \beta_{11} + \beta_{12} income_i + \varepsilon_{1i}$

bus: $U_{2i} = \mathbf{x}_i \boldsymbol{\beta}_2 = \beta_{21} + \beta_{22} income_i + \varepsilon_{2i}$

walk: $U_{3i} = \mathbf{x}_i \boldsymbol{\beta}_3 = \beta_{31} + \beta_{32} income_i + \varepsilon_{3i}$

A person chooses the option with highest utility. If we assume that the disturbance terms are IID (independently and identically distributed) with an Type I Extreme Value distribution:

$$\hat{P}_{ji} = \frac{e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}_j}}{\sum_{j=1}^{3} e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}_j}} \qquad L = \prod_{j=1}^{3} \left[\prod_{Y_i = j} \left(\hat{P}_{ji} \right) \right]$$

Maximize $\ln L$ with respect to $\hat{\boldsymbol{\beta}}_j$, with $\hat{\boldsymbol{\beta}}_1 = \mathbf{0}$

Example: Transit Choice

The choices are car, bus, and walk. They are coded 1, 2, and 3 respectively, but the coding does matter.

$$choice_i = \begin{cases} 1 \text{ if the person drives a car} \\ 2 \text{ if the person takes the bus} \\ 3 \text{ if the person walks} \end{cases}$$

We want to examine how personal characteristics (X) affect the choice. We will use *income* as the characteristic, but there could be many more than one, including dummies, interactions terms, etc.

transit.dta

```
. list, noobs clean
person     income   pwalk       pbus       pcar   car   walk   bus   choice
     1   11.26115       0   1.400565   7.050649     0      1     0     Walk
     2    19.8307       0   1.415419   6.960087     1      0     0      Car
     3   34.89569       0   1.726712   8.057094     0      0     1      Bus
     4    24.6605       0    2.88912   8.799998     0      0     1      Bus
     5   22.78829       0   1.198922   8.043923     0      0     1      Bus
     6    13.6575       0   2.927421   6.364377     0      0     1      Bus
     7   33.06366       0   1.867234   5.479687     1      0     0      Car
     :
     :

. sum
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      person |      1000       500.5    288.8194          1       1000
      income |      1000    20.16592    8.854359   5.004433    34.9992
       pwalk |      1000           0           0          0          0
        pbus |      1000    1.960678    .5682552   1.000583   2.992725
        pcar |      1000     6.99019    .9699899   4.263136   10.05263
-------------+--------------------------------------------------------
         car |      1000        .361    .4805309          0          1
        walk |      1000        .292    .4549098          0          1
         bus |      1000        .347    .4762539          0          1
      choice |      1000       1.931    .8055358          1          3
```

Stata: `mlogit`

```
Iteration 0:   log likelihood = -1094.5425
Iteration 1:   log likelihood = -859.91246
Iteration 2:   log likelihood = -832.43375
Iteration 3:   log likelihood = -830.68069
Iteration 4:   log likelihood = -830.66645
Iteration 5:   log likelihood = -830.66645

Multinomial logistic regression                   Number of obs   =       1000
                                                  LR chi2(2)      =     527.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -830.66645                       Pseudo R2       =     0.2411
------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Bus          |
      income |  -.1864629   .0139343   -13.38   0.000    -.2137737   -.1591521
       _cons |   4.295364   .3437584    12.50   0.000      3.62161    4.969119
-------------+----------------------------------------------------------------
Walk         |
      income |  -.2794017   .0167763   -16.65   0.000    -.3122827   -.2465207
       _cons |   5.582236    .368069    15.17   0.000     4.860834    6.303638
------------------------------------------------------------------------------
(choice==Car is the base outcome)
```

```
. use transit
. mlogit choice income
```

Initial Log Likelihood for `mlogit`

```
. tab choice
     choice |      Freq.     Percent        Cum.
------------+-----------------------------------
        Car |        361       36.10       36.10
        Bus |        347       34.70       70.80
       Walk |        292       29.20      100.00
------------+-----------------------------------
      Total |      1,000      100.00
```

$$\ln L = (361)\ln(0.361) + (347)\ln(0.347) + (292)\ln(0.292)$$
$$= -1094.5425$$

$$\chi^2 = -2\Delta \ln L = -2(-1094.54 + 830.67) = 527.74$$

$$\tilde{R}^2 = 1 - \frac{\ln L_F}{\ln L_0} = 1 - \frac{-830.67}{-1094.54} = 0.241$$

Results for Income Equal to $20,000

$$\hat{U}_{1i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_1 = 0 + 0(income_i) = 0$$
$$\hat{U}_{2i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_2 = 4.30 - 0.19(income_i)$$
$$\hat{U}_{3i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_3 = 5.58 - 0.28(income_i)$$

If income = 20 ($20,000), we get the following utilities:

$$\hat{U}_{1i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_1 = 0 + 0(20) = 0$$
$$\hat{U}_{2i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_2 = 4.30 - 0.19(20) = 0.50$$
$$\hat{U}_{3i} = \mathbf{x}_i\hat{\boldsymbol{\beta}}_3 = 5.58 - 0.28(20) = -0.02$$

Predicted Probabilities

From the utilities, we can calculate predicted probabilities as follows:

$$\hat{P}_1 = \frac{e^0}{e^0 + e^{0.50} + e^{-0.02}} = \frac{1}{1 + 1.65 + 0.98} = \frac{1}{3.63} = 0.28$$
$$\hat{P}_2 = \frac{1.65}{3.63} = 0.45$$
$$\hat{P}_3 = \frac{0.98}{3.63} = 0.27$$

Check: $0.28 + 0.45 + 0.27 = 1$

Increase Income to $21,000

We repeat these calculations for income = 21 ($21,000), and summarize the predicted probabilities in the table below.

$Choice$	$income = 20$	$income = 21$	$Difference$
$Car\ (j = 1)$	0.28	0.32	+0.04
$Bus\ (j = 2)$	0.45	0.44	–0.01
$Walk\ (j = 3)$	0.27	0.24	–0.03

Probabilities in Stata

```
. margins, at(income=20)

1._predict   : Pr(choice==Car), predict(pr outcome(1))
2._predict   : Pr(choice==Bus), predict(pr outcome(2))
3._predict   : Pr(choice==Walk), predict(pr outcome(3))
at           : income          =          20

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _predict |
          1  |   .2662681   .0196601    13.54   0.000      .227735    .3048011
          2  |   .4690033   .0208092    22.54   0.000     .4282179    .5097886
          3  |   .2647286   .0187753    14.10   0.000     .2279297    .3015276
------------------------------------------------------------------------------
```

Probabilities in Stata

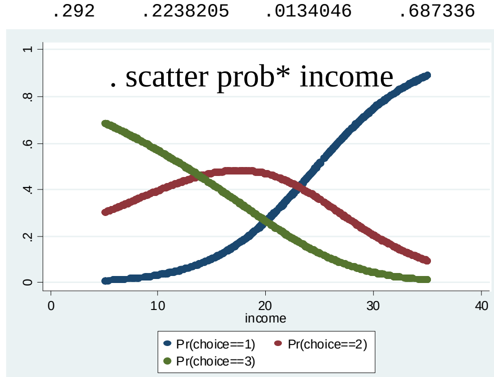
```
. margins, predict(outcome(1)) at(income==(0 10 20 30 40 50 60 70))
Adjusted predictions                              Number of obs    =       1000
Model VCE    : OIM
Expression   : Pr(choice==Car), predict(outcome(1))
1._at        : income          =           0
Etc.
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    .002941   .0010358     2.84   0.005     .0009109     .004971
          2  |   .0349415   .0070406     4.96   0.000     .0211422    .0487408
          3  |   .2662681   .0196601    13.54   0.000      .227735    .3048011
          4  |   .7497624   .0224937    33.33   0.000     .7056755    .7938493
          5  |   .9560111   .0100583    95.05   0.000     .9362972     .975725
          6  |   .9932647   .0025015   397.07   0.000     .9883619    .9981675
          7  |   .9989717   .0005255  1901.00   0.000     .9979417    1.000002
          8  |   .9998418   .0001028  9728.19   0.000     .9996404    1.000043
------------------------------------------------------------------------------
```

Mode Choice Probabilities by Income

```
. predict probcar probbus probwalk
(option pr assumed; predicted probabilities)

. sum prob*
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     probcar |      1000        .361    .3066615   .0104734   .8908678
     probbus |      1000        .347    .1188307   .0957276   .4834927
    probwalk |      1000        .292    .2238205   .0134046    .687336
```

Note that P_{bus} first increases as income rises, then falls.
In other words, the marginal effect changes signs!



Marginal Effects

- Marginal effect varies
- Previous slide shows they can change direction

$$\frac{\partial P_{1i}}{\partial X_k} = \frac{\partial\left(\dfrac{e^{\mathbf{x}_i\boldsymbol{\beta}_1}}{e^{\mathbf{x}_i\boldsymbol{\beta}_1}+e^{\mathbf{x}_i\boldsymbol{\beta}_2}+e^{\mathbf{x}_i\boldsymbol{\beta}_3}}\right)}{\partial X_k}$$

Apply the quotient rule from calculus.

$$= \beta_{1k}P_1 - P_1\left(\beta_{1k}P_1+\beta_{2k}P_2+\beta_{3k}P_3\right)$$

$$= \beta_{1k}P_1\left(1-P_1\right) - P_1\left(\beta_{2k}P_2+\beta_{3k}P_3\right)$$

The first term indicates that the marginal effect varies with P1, and the second term indicates an interaction of the effects.

The adding up constraint

The sum of all Ps is 1, both before and after an increase in Xk. Therefore, any increases or decreases in P1must be just offset by decreases or increases in other Ps.

$$\left.\begin{aligned}
\frac{\partial P_{1i}}{\partial X_k} &= \beta_{1k}P_1 - P_1\left(\beta_{1k}P_1+\beta_{2k}P_2+\beta_{3k}P_3\right)\\
\frac{\partial P_{2i}}{\partial X_k} &= \beta_{2k}P_2 - P_2\left(\beta_{1k}P_1+\beta_{2k}P_2+\beta_{3k}P_3\right)\\
\frac{\partial P_{3i}}{\partial X_k} &= \beta_{3k}P_3 - P_3\left(\beta_{1k}P_1+\beta_{2k}P_2+\beta_{3k}P_3\right)
\end{aligned}\right\}$$

Adds to zero

(Confirm that it adds to zero.)

Calculating Marginal Effects at Income = 20

$$\frac{\partial \hat{P}_{ji}}{\partial X_k} = \underbrace{\hat{\beta}_{jk}\hat{P}_j - \hat{P}_j}_{\text{Depends on } j}\underbrace{\left(\hat{\beta}_{1k}\hat{P}_1 + \hat{\beta}_{2k}\hat{P}_2 + \hat{\beta}_{3k}\hat{P}_3\right)}_{\text{Same for all three.}}$$

$$= \hat{\beta}_{jk}\hat{P}_j - \hat{P}_j\left(0 + (-0.19)(0.45) + (-0.28)(.27)\right)$$

$$= \hat{\beta}_{jk}\hat{P}_j - \hat{P}_j(-0.16)$$

$\hat{P}_1 = 0.28$

$\hat{P}_2 = 0.45$

$\hat{P}_3 = 0.27$

$\hat{\beta}_1 = 0$

$\hat{\beta}_2 = -0.19$

$\hat{\beta}_3 = -0.28$

$$\frac{\partial \hat{P}_{1i}}{\partial X_k} = 0 - 0.28(-0.16) = +0.045$$

$$\frac{\partial \hat{P}_{2i}}{\partial X_k} = (-0.19)(0.45) - 0.45(-0.16) = -0.014$$

$$\frac{\partial \hat{P}_{3i}}{\partial X_k} = (-0.28)(0.27) - 0.27(-0.16) = -0.032$$

They add to zero, except for a mall rounding error

Marginal Effects: `margins`

```
. margins, predict(outcome(1)) dydx(income)
Average marginal effects                          Number of obs   =       1000
Model VCE    : OIM
Expression   : Pr(choice==Car), predict(outcome(1))
dy/dx w.r.t. : income
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0292565   .0006236    46.92   0.000     .0280344    .0304787
------------------------------------------------------------------------------

. margins, predict(outcome(1)) at(income==(0 20 40)) dydx(income)
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
income       |
         _at |
          1  |   .0007603    .000222     3.42   0.001     .0003251    .0011955
          2  |   .0429803   .0023756    18.09   0.000     .0383243    .0476363
          3  |   .0081607   .0012472     6.54   0.000     .0057162    .0106052
------------------------------------------------------------------------------
```

Marginal Effects: `mchange`

```
. mchange, at(income =20)

mlogit: Changes in Pr(y) | Number of obs = 1000

Expression: Pr(choice), predict(outcome())

             |       Car        Bus       Walk
-------------+---------------------------------
income       |
          +1 |     0.045     -0.014     -0.031
     p-value |     0.000      0.000      0.000
         +SD |     0.437     -0.231     -0.206
     p-value |     0.000      0.000      0.000
    Marginal |     0.043     -0.012     -0.031
     p-value |     0.000      0.000      0.000

Average predictions

             |       Car        Bus       Walk
-------------+---------------------------------
  Pr(y|base) |     0.266      0.469      0.265

Base values of regressors

             |    income
-------------+-----------
          at |        20
```

Average Marginal Effects

```
. mchange

mlogit: Changes in Pr(y) | Number of obs = 1000

Expression: Pr(choice), predict(outcome())

             |       Car        Bus       Walk
-------------+---------------------------------
income       |
          +1 |     0.012      0.016     -0.028
     p-value |     0.000      0.000      0.000
         +SD |     0.225      0.048     -0.273
     p-value |     0.000      0.035      0.000
    Marginal |     0.029     -0.007     -0.022
     p-value |     0.000      0.000      0.000

Average predictions

             |       Car        Bus       Walk
-------------+---------------------------------
  Pr(y|base) |     0.361      0.347      0.292
```

Relative Risk in MNL

The *odds* of choosing the bus *relative to* driving a car, aka "the relative risk" of bus vs. car, is:

$$\Omega_{bus|car} = \frac{P_{bus}}{P_{car}} = \frac{\left(\dfrac{e^{\mathbf{x}_i\boldsymbol{\beta}_{bus}}}{e^{\mathbf{x}_i\boldsymbol{\beta}_{car}} + e^{\mathbf{x}_i\boldsymbol{\beta}_{bus}} + e^{\mathbf{x}_i\boldsymbol{\beta}_{walk}}}\right)}{\left(\dfrac{e^{\mathbf{x}_i\boldsymbol{\beta}_{car}}}{e^{\mathbf{x}_i\boldsymbol{\beta}_{car}} + e^{\mathbf{x}_i\boldsymbol{\beta}_{bus}} + e^{\mathbf{x}_i\boldsymbol{\beta}_{walk}}}\right)} = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_{bus}}}{e^{\mathbf{x}_i\boldsymbol{\beta}_{car}}}$$

$$= e^{\left(\mathbf{x}_i\boldsymbol{\beta}_{bus} - \mathbf{x}_i\boldsymbol{\beta}_{car}\right)} = e^{\mathbf{x}_i\left(\boldsymbol{\beta}_{bus} - \boldsymbol{\beta}_{car}\right)}$$

($\boldsymbol{\beta}_{\mathbf{bus}} - \boldsymbol{\beta}_{\mathbf{car}}$) is known as the *contrast* of bus with car. Since we have constrained the coefficients in the car equation to zero, then the estimated coefficients (beta hat for bus) reported above are already the contrasts with car. Note the coefficients for walk do not appear in the equation for the relative risk of bus vs. car.

Relative Risk Ratio

$$\hat{\Omega}_{bus|car} = \frac{\hat{P}_{bus}}{\hat{P}_{car}} = e^{\mathbf{x}_i\hat{\boldsymbol{\beta}}_{bus|car}} = e^{\hat{\beta}_{11}+\hat{\beta}_{12}income_i}$$

If income, increases by 1 unit ($1,000 in this case), the new odds will be:

$$\hat{\Omega}'_{bus|car} = e^{\hat{\beta}_{11}+\hat{\beta}_{12}\left(income_i+1\right)} = e^{\hat{\beta}_{11}+\hat{\beta}_{12}income_i+\hat{\beta}_{12}}$$

The *relative risk ratio* (RRR) for a one unit change in income for bus vs. car is the *ratio of the relative risks before and after the one unit increase*:

$$RRR = \frac{\hat{\Omega}'_{bus|car}}{\hat{\Omega}_{bus|car}} = \frac{e^{\hat{\beta}_{11}+\hat{\beta}_{12}income_i+\hat{\beta}_{12}}}{e^{\hat{\beta}_{11}+\hat{\beta}_{12}income_i}} = e^{\hat{\beta}_{12}} = e^{-0.19} = 0.83$$

RRR: the Long Way

$$RRR = \frac{\left(\dfrac{P'_{bus}}{P'_{car}}\right)}{\left(\dfrac{P_{bus}}{P_{car}}\right)} = \frac{\left(\dfrac{0.44}{0.32}\right)}{\left(\dfrac{0.45}{0.28}\right)} = \frac{1.38}{1.61} = 0.86$$

Using the probabilities predicted earlier the answer is the same, within rounding error. The first calculation is both easier and more accurate. So the odds ratio of *m* vs. *n* for the variable *X* depends on the *m* vs. *n* contrast of the *X* variable, and does not depend on anything else – not the values of *X*, the other *X*s, or even the utility functions of the other choices.

Some points about RRR

- Long calls this the "odds ratio of *m* vs. *n*"
- That's OK, but "relative risk ratio" is more accepted and less subject to confusion.
- Some people define the odds of *m* as $P_m/(1-P_m)$, in other words the relative risk of *m* vs. all other choices (not *m*).
- See **discussion on statalist**.
- I suggest you stick with RRR.

Comments on ME in MNL

- Effect depends on P_j, which depends on all X's (not constant)
- Attenuation of the effect due to adding up constraint (sum of all ΔP=0)
- Sign is not a reliable indicator of the direction of the effect
- Suggestion: stick with odds/RRR or discrete changes for interesting/important cases

IIA Implications

- Odds of Bus to Car don't depend on Walk
- Known as the "Independence of Irrelevant Alternatives" problem.
- For example, if first bus was blue, and a red bus was added, the relative odds of car to blue bus should be affected.
- Carefully group alternatives into groups that seem independent
- Other models (e.g. multinomial probit, nested logit) can be used to overcome non-independence of alternatives. Talk about next week.

Alternative Base Category: Bus

. mlogit choice income, base(2)

```
Multinomial logistic regression                   Number of obs   =       1000
                                                  LR chi2(2)      =     527.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -830.66645                       Pseudo R2       =     0.2411

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Car          |
      income |   .1864629   .0139343    13.38   0.000     .1591521    .2137737
       _cons |  -4.295364   .3437584   -12.50   0.000    -4.969119    -3.62161
-------------+----------------------------------------------------------------
Walk         |
      income |  -.0929388   .0120633    -7.70   0.000    -.1165824   -.0692952
       _cons |   1.286871    .202701     6.35   0.000     .8895844    1.684158
------------------------------------------------------------------------------
(choice==Bus is the base outcome)
```

How do these results differ?

Calculations for Base Case = Bus

For income = 20:

$\hat{U}_{car} =$

$\hat{U}_{bus} =$

$\hat{U}_{walk} =$

$\hat{P}_{\text{car}} =$

$\hat{P}_{bus} =$

$\hat{P}_{walk} =$

$OR_{bus|car} =$

Alternative Base Category: Walk

. mlogit choice income, base(3)

```
Multinomial logistic regression                   Number of obs   =       1000
                                                  LR chi2(2)      =     527.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -830.66645                       Pseudo R2       =     0.2411

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Car          |
      income |          ?   .0167763    16.65   0.000     .2465207    .3122827
       _cons |  -5.582236    .368069   -15.17   0.000    -6.303638   -4.860834
-------------+----------------------------------------------------------------
Bus          |
      income |          ?   .0120633     7.70   0.000     .0692952    .1165824
       _cons |  -1.286871    .202701    -6.35   0.000    -1.684158   -.8895844
------------------------------------------------------------------------------
(choice==Walk is the base outcome)
```

What are the missing coefficients?

Generalizing

$$U_{ji} = \mathbf{x}_i \boldsymbol{\beta}_j + \varepsilon_{ji} \qquad j = \{1, 2, \ldots, J\} \qquad Y_i = choice$$

$$\Pr(Y_i = m \mid X_i) = \Pr(U_{mi} > U_{ni}) \quad \forall\, m, n \in j, m \neq n$$

$$\Pr(U_{mi} > U_{ni}) = \Pr\left[(\mathbf{x}_i \boldsymbol{\beta}_m + \varepsilon_{mi}) > (\mathbf{x}_i \boldsymbol{\beta}_n + \varepsilon_{ni})\right]$$

$$= \Pr\left[(\varepsilon_{mi} - \varepsilon_{ni}) > \mathbf{x}_i (\boldsymbol{\beta}_n - \boldsymbol{\beta}_m)\right] \qquad \boldsymbol{\beta}_1 = \mathbf{0}$$

If the ε_j have a “type I extreme value distribution,” then:

$$P_{ji} = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{j=1}^{J} \mathbf{x}_i \boldsymbol{\beta}_j} \qquad 0 < P_j < 1 \qquad \sum_{j=1}^{J} P_j = 1 \qquad \mathcal{L} = \prod_{i=1}^{N} \left[\frac{e^{\mathbf{x}_i \boldsymbol{\beta}_{Y_i}}}{\sum_{j=1}^{J} e^{\mathbf{x}_i \boldsymbol{\beta}_j}} \right]$$