Week 9

Discrete Choice Models

Rich Frank
University of New Orleans
October 18, 2012

- Welcome back!

- Two weeks ago we talked about multi-category variables that we believed were ordinal.

- We ended by concluding that constraining the effects of an X on all categories of Y was often a strong assumption.

- Today we look at several ways of using non-ordered data.

- These models can be motivated in a similar way to logit, probit, and ordered models.

- If we use these methods for discrete choice data and the data are actually ordered, this relaxes the parallel regression assumption unnecessarily.
- It also is inefficient (because we are neglecting some relevant information in these data), but it is less of an offense than if we impose order on unordered data.
- This would likely lead to biased estimates and is unlikely to make sense.

To recap…

- In previous weeks we were trying to estimate models to capture some latent variable y*, for which we can only see binary realizations of (0 and 1). So:

$$y^* = X\beta + \varepsilon$$

Where we see realizations of y*, y as resulting from:

$$y_i = 1 \text{ if } y^* > 0$$
$$y_i = 0 \text{ if } y^* \leq 0$$

- So we treat y as occurring with a certain probability given as:

$$P(y_i = 1) = P(y^*>0)$$
$$= P(\beta x_{ij} + \varepsilon_{ij} > 0)$$
$$= P(\varepsilon_{i,j} > -\beta x_{ij})$$
$$= F(\beta x_{ij})$$

Where F is the link function.

- Therefore, if we use the logistic then we have the binary logit model.

- Let’s take a step beyond this binary choice to examine data with more than two outcomes.

- Think of some group of unordered outcomes, J, for which each individual i has some utility.

- This utility is grounded in the economic rational choice literature where consumers buy goods that maximize their perceived utility amongst discrete choices. e.g.:
  - Flavors of ice cream
  - Automotive brands
  - Political candidates

- The utility that the individual *i* has for a choice *j* can be written as:

$$U_{ij} = \mu_{ij} + \varepsilon_{ij}$$

- Thus, the utility has both a systematic component (μ_{ij}) as well as a stochastic one (ε_{ij}).

- We can then parameterize the systematic component as being a function of some variables.

$$U_{ij} = \beta X_{ij} + \varepsilon_{ij}$$

- In English, this means that a variable X has an effect β on *i*'s utility for option *j*.
- Broadening our scope to looking at all outcomes, we can assume that an individual has a complete set of preferences over the *J* outcomes, and these preferences are transitive.

- This means that each individual makes choices to maximize his or her utility while comparing different choices, *j, k,* …*J* in a pairwise fashion.

- Thus if *i* went to Brocato's to get some gelato, *i* would have preferences for different flavors:

$$Chocolate > Vanilla$$

$$Vanilla > Strawberry$$

$$Chocolate > Strawberry$$

- Importantly, this also means that if the owners decide to mix up a batch of mango sorbet, our preference ordering above will stay the same. More on this later.

- Put more formally:

$$P(y_i = j \mid x_i) = P\left(U_{ij} > U_{ik}\right)$$
$$= P(\beta x_{ij} + \varepsilon_{ij} > \beta x_{ik} + \varepsilon_{ik})$$
$$= P(\beta x_{ij} - \beta x_{ik} > \varepsilon_{ij} - \varepsilon_{ik})$$

- This gives us a theoretical model, but for us to be able to estimate a model we have to make an assumption about how the errors are distributed—just like with ordinal models.

- Usually, statisticians assume the $\varepsilon_{ij}s$ are i.i.d. (independent and identically distributed) and Weibull (Type I extreme value) distributed:

$$F(\varepsilon_{ij}) \sim e^{[-\varepsilon_{ij} - e^{-\varepsilon_{ij}}]}$$

  - See Long (1997: 156)
  - Google it if you want to see a graphical representation of the Type 1 extreme value.

- The probability $y_i = j$ for an observed outcome m is:

$$P(y_i = m \mid x_i) = \frac{e^{\beta_m x_i}}{\sum_{j=1}^{J} e^{\beta_j x_i}}$$

- However, it is impossible to do this for each value of J—the model is not identified because there is no reference category.

- So we typically constrain $\beta_{ij=0} = 0$ and estimate $\beta_{ij=0,J}$.

- Therefore:

$$P(y_i = 1 \mid x_i) = \frac{1}{1 + \sum_{j=2}^{J} e^{\beta_j x_i}}$$

$$P(y_i = m \mid x_i) = \frac{e^{\beta_m x_i}}{1 + \sum_{j=1}^{J} e^{\beta_j x_i}}$$

for $m > 1$

- As you can see, if J = 1 the model is the binary logit.
- The binary logit is a special case of the multinomial logit.

- For the multinomial logit the likelihood function is given as follows (Long 1997: 157):

$$L(\beta_{2,} \; \ldots, \beta_J \mid \boldsymbol{y}, \mathbf{X}) = \prod_{m=1}^{J} \prod_{y_i=m} \frac{e^{\beta_m x_i}}{\sum_{j=1}^{J} e^{\beta_j \, x_i}}$$

Therefore for a variable with 3 outcomes we only estimate 2 $\hat{\beta}s$.

- Why do we only have to estimate $J - 1$ $\hat{\beta}$s?

- Because as Long (1997: 150) writes:

$$\ln\left[\frac{\Pr(A \mid x)}{\Pr(B \mid x)}\right] + \ln\left[\frac{\Pr(B \mid x)}{\Pr(C \mid x)}\right] = \ln\left[\frac{\Pr(A \mid x)}{\Pr(C \mid x)}\right]$$

Which means: $\hat{\beta}_{1,A|B} + \hat{\beta}_{1,B|C} = \hat{\beta}_{1,A|C}$

- Let's take a step back.

- What the multinomial logit model (MNL) does is try and model the utility for different options given different characteristics of the individual (let's say age, education, whether his/her parents' really like ice cream).

- What it does not do is look at characteristics of the different flavors of ice cream…

  - Maybe chocolate costs more than vanilla.
  - Maybe the strawberry looks like it has been sitting there all summer….
  - Etc.

- If you are more interesting in how the characteristics of the ice cream affects the probability of an alternative being chosen, then you need to look to another type of model: the Conditional Logit (CL).
  - “conditional” in this case meaning conditional on the characteristics of the alternatives.

- These two models—the multinomial logit (MNL) and the conditional logit (CL) have identical error structures.

- Often, the names MNL and the CL are used interchangeably.

- Therefore before I start into interpretation, it is easier to describe the similarities and differences of the MNL and the CL.

- The crucial difference is how the two treat our expectations about how the *X*'s influence the choices in *Y*.

- Before we begin to discuss their differences, it is crucial that we are all on the same page in our nomenclature.

- For data with nominal DVs we need to differentiate between *cases* and *alternatives*.
  - *Cases* are individual observations
  - *Alternatives* are the different outcome choices.

- So using our ice cream example:
  - Cases are the individuals (Bill, Shirley, Paul) going into Brocato's.
  - Alternatives are the ice cream flavors.

- Trying to tie this to political phenomena…
- The cases in Alvarez and Nagler (1998) are ______________?
- The alternatives are then ________________?

Comparing MNL and CL

Independent Variable	**With respect to cases**	**With respect to Y**	**Model**	**# of $\widehat{\beta}$ s**
Characteristics of the individual	Vary across cases	Constant across choices (*Y*= *j*)	MNL	J-1
Characteristics of the outcome j	Constant across cases	Vary across outcomes *J*= 1 to m	CL	1
Individual and case characteristics	Vary across cases	Vary across outcomes *J* =1 to m	Modified CL	$\beta_x, \beta_0, \beta_{x,0}$

Case specific data example

Voter	Choice	Party ID	Education
1	0	Rep.	14
2	0	Dem.	12
3	2	Rep.	6

Alternative-specific data example

Voter	Choice	Spending	Campaign Stops
1	0	145	10
2	0	145	10
3	2	130	12

Both Case and Alternatives

Voter	Alternative	Choice	D	Party ID	Education	Spending	Campaign Stops
1	0	1	0	R	14	11	0
1	1	1	1	R	14	45	2
1	2	1	0	R	14	54	2
2	0	0	1	D	12	11	1
2	1	0	0	D	12	45	6
2	2	0	0	D	12	54	4
3	0	2	0	D	6	11	2
3	1	2	0	D	6	45	3
3	2	2	1	D	6	54	5

- Therefore the conditional logit model includes information about the choices and not about the individuals.

- This means that the predicted probability looks a bit different.

- Remember the MNL probability:

$$P(y_i = m \mid x_i) = \frac{e^{\beta_m x_i}}{\sum_{j=1}^{J} e^{\beta_j x_i}}$$

Where the X's were characteristics of the unit.

- The probabilities of the outcome in a CL model are given by information in a vector of parameters (**Z**) about the choices and we estimate a vector of coefficients ($\boldsymbol{\gamma}$) that is most likely to have produced the observed y.

$$P(y_i = m \mid z_i) = \frac{e^{\gamma_{mi} z_i}}{\sum_{j=1}^{J} e^{\gamma_{mi} z_i}}$$

- Can we estimate a model that includes both case and alternative specific data?

- Yes!

- What you have to do then, is estimate both βs and γs.

- This requires a modification of the conditional logit predicted probability model.

$$P(y_i = m \mid x_i, z_i) = \frac{e^{\gamma_{mi} z_i + \beta_m x_i}}{\sum_{j=1}^{J} e^{\gamma_{ji} z_i + \beta_j x_i}}$$

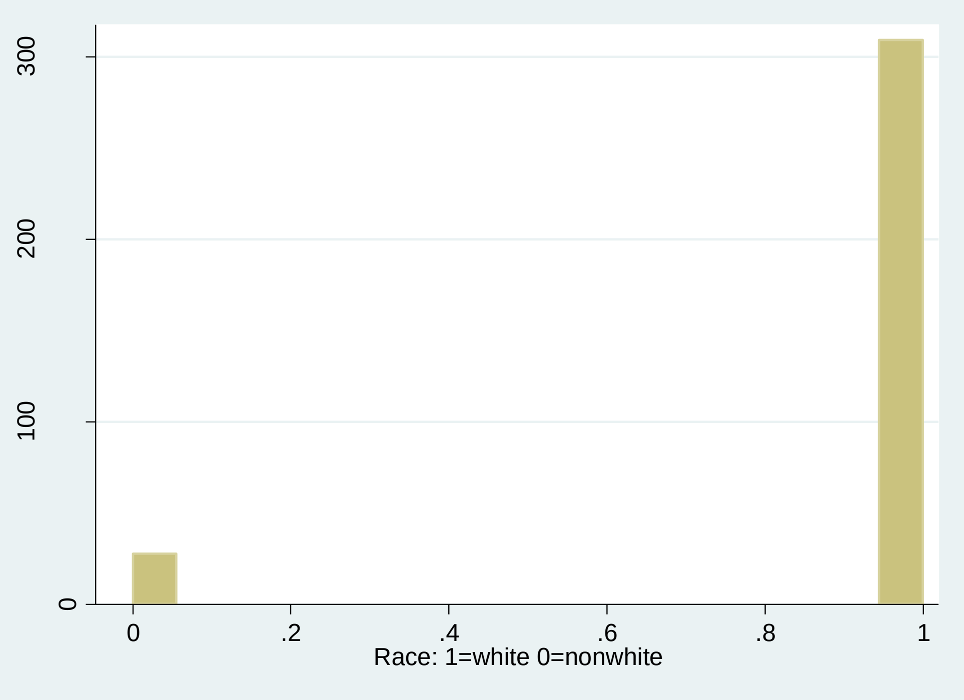
Where $\beta_1 = 0$.

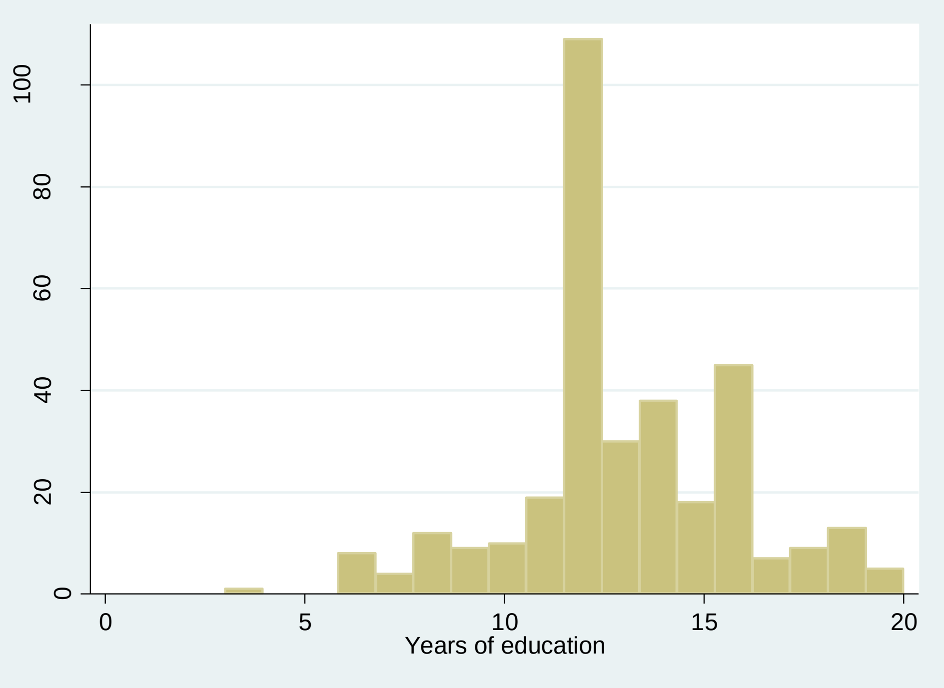
- As Long (1997) and Greene (2008) mention, estimating and interpreting these types of models requires a bit of effort.

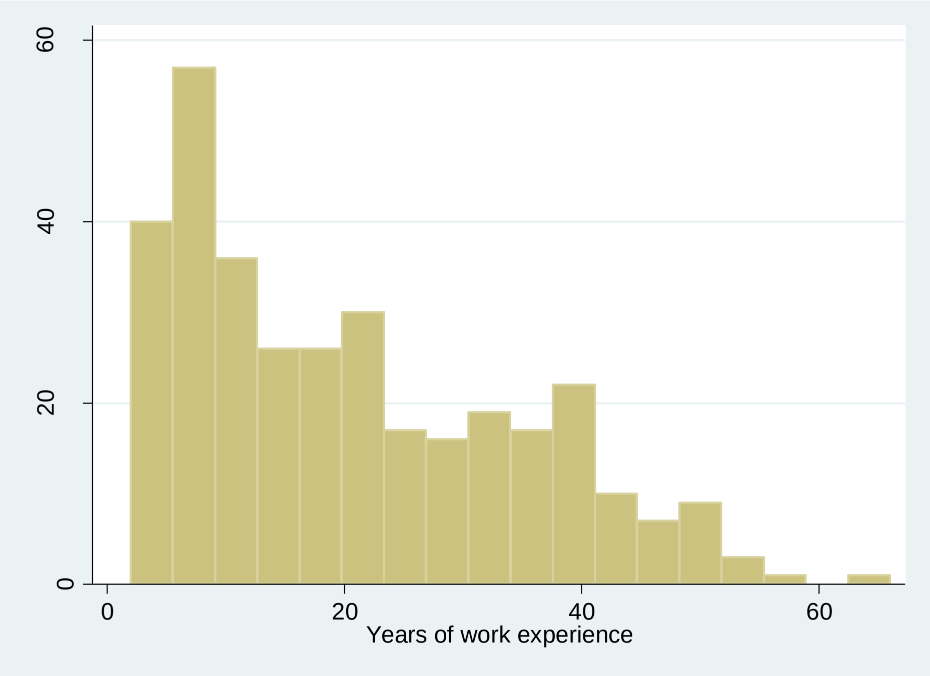
- There are a lot of moving parts!

- Let’s start with a simple MNL: Long’s(1997) example of occupational attainment…

```
. sum  occ white ed exper

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         occ |       337    3.397626    1.367913          1          5
       white |       337    .9169139    .2764227          0          1
          ed |       337    13.09496    2.946427          3         20
       exper |       337    20.50148    13.95936          2         66


. tab occ

 Occupation |      Freq.     Percent        Cum.
------------+-----------------------------------
     Menial |         31        9.20        9.20
    BlueCol |         69       20.47       29.67
      Craft |         84       24.93       54.60
   WhiteCol |         41       12.17       66.77
       Prof |        112       33.23      100.00
------------+-----------------------------------
      Total |        337      100.00
```


300
200
100
0
0
.2
.4
.6
.8
1
Race: 1=white 0=nonwhite


100
80
60
40
20
0
0
5
10
15
20
Years of education


60
40
20
0
0
20
40
60
Years of work experience

```
. mlogit occ white ed exper

Iteration 0:   log likelihood = -509.84406
Iteration 1:   log likelihood = -432.18549
Iteration 2:   log likelihood = -426.88668
Iteration 3:   log likelihood = -426.80057
Iteration 4:   log likelihood = -426.80048
Iteration 5:   log likelihood = -426.80048

Multinomial logistic regression                   Number of obs   =        337
                                                  LR chi2(12)     =     166.09
                                                  Prob > chi2     =     0.0000
Log likelihood = -426.80048                       Pseudo R2       =     0.1629

------------------------------------------------------------------------------
         occ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Menial       |
       white |  -1.774306   .7550543    -2.35   0.019    -3.254186   -.2944273
          ed |  -.7788519   .1146293    -6.79   0.000    -1.003521   -.5541826
       exper |  -.0356509    .018037    -1.98   0.048    -.0710028    -.000299
       _cons |   11.51833   1.849356     6.23   0.000     7.893659      15.143
-------------+----------------------------------------------------------------
BlueCol      |
       white |  -.5378027   .7996033    -0.67   0.501    -2.104996    1.029391
          ed |  -.8782767   .1005446    -8.74   0.000     -1.07534   -.6812128
       exper |  -.0309296   .0144086    -2.15   0.032      -.05917   -.0026893
       _cons |   12.25956   1.668144     7.35   0.000     8.990061    15.52907
-------------+----------------------------------------------------------------
Craft        |
       white |  -1.301963    .647416    -2.01   0.044    -2.570875   -.0330509
          ed |  -.6850365   .0892996    -7.67   0.000    -.8600605   -.5100126
       exper |  -.0079671   .0127055    -0.63   0.531    -.0328693    .0169351
       _cons |   10.42698   1.517943     6.87   0.000     7.451864    13.40209
-------------+----------------------------------------------------------------
WhiteCol     |
       white |  -.2029212   .8693072    -0.23   0.815    -1.906732     1.50089
          ed |  -.4256943   .0922192    -4.62   0.000    -.6064407   -.2449479
       exper |   -.001055   .0143582    -0.07   0.941    -.0291967    .0270866
       _cons |   5.279722   1.684006     3.14   0.002     1.979132    8.580313
-------------+----------------------------------------------------------------
Prof         |  (base outcome)
------------------------------------------------------------------------------
```

- We can interpret the direction of the coefficients directly.
  - As education increases, it decreases the probability of being in any category but professional.
- Or we can see if a particular X has a significant effect on different pairs of categories…

```
. listcoef, pvalue(.05)

mlogit (N=337): Factor Change in the Odds of occ when P>|z| < 0.05

Variable: white (sd=.27642268)

Odds comparing    |
Alternative 1     |
to Alternative 2  |      b         z     P>|z|     e^b   e^bStdX
------------------+---------------------------------------------
Menial   -Prof    |  -1.77431   -2.350   0.019   0.1696   0.6123
Craft    -Prof    |  -1.30196   -2.011   0.044   0.2720   0.6978
Prof     -Menial  |   1.77431    2.350   0.019   5.8962   1.6331
Prof     -Craft   |   1.30196    2.011   0.044   3.6765   1.4332
----------------------------------------------------------------

Variable: ed (sd=2.9464271)

Odds comparing    |
Alternative 1     |
to Alternative 2  |      b         z     P>|z|     e^b   e^bStdX
------------------+---------------------------------------------
Menial   -WhiteCol |  -0.35316   -3.011   0.003   0.7025   0.3533
Menial   -Prof     |  -0.77885   -6.795   0.000   0.4589   0.1008
BlueCol  -Craft    |  -0.19324   -2.494   0.013   0.8243   0.5659
BlueCol  -WhiteCol |  -0.45258   -4.425   0.000   0.6360   0.2636
BlueCol  -Prof     |  -0.87828   -8.735   0.000   0.4155   0.0752
Craft    -BlueCol  |   0.19324    2.494   0.013   1.2132   1.7671
Craft    -WhiteCol |  -0.25934   -2.773   0.006   0.7716   0.4657
Craft    -Prof     |  -0.68504   -7.671   0.000   0.5041   0.1329
WhiteCol-Menial    |   0.35316    3.011   0.003   1.4236   2.8308
WhiteCol-BlueCol   |   0.45258    4.425   0.000   1.5724   3.7943
WhiteCol-Craft     |   0.25934    2.773   0.006   1.2961   2.1471
WhiteCol-Prof      |  -0.42569   -4.616   0.000   0.6533   0.2853
Prof     -Menial   |   0.77885    6.795   0.000   2.1790   9.9228
Prof     -BlueCol  |   0.87828    8.735   0.000   2.4067  13.3002
Prof     -Craft    |   0.68504    7.671   0.000   1.9838   7.5264
Prof     -WhiteCol |   0.42569    4.616   0.000   1.5307   3.5053
----------------------------------------------------------------

Variable: exper (sd=13.959364)

Odds comparing    |
Alternative 1     |
to Alternative 2  |      b         z     P>|z|     e^b   e^bStdX
------------------+---------------------------------------------
Menial   -Prof    |  -0.03565   -1.977   0.048   0.9650   0.6079
BlueCol  -Prof    |  -0.03093   -2.147   0.032   0.9695   0.6494
Prof     -Menial  |   0.03565    1.977   0.048   1.0363   1.6449
Prof     -BlueCol |   0.03093    2.147   0.032   1.0314   1.5400
----------------------------------------------------------------
```

- You can do similar hypothesis testing using LR tests or Wald tests to see the significance of individual variables or groups of variables.

```
. mlogtest, lr

**** Likelihood-ratio tests for independent variables (N=337)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+-------------------------
       white |      8.095    4    0.088
          ed |    156.937    4    0.000
       exper |      8.561    4    0.073
---------------------------------------
```

Wald Test

```
. test white

 ( 1)  [Menial]white = 0
 ( 2)  [BlueCol]white = 0
 ( 3)  [Craft]white = 0
 ( 4)  [WhiteCol]white = 0
 ( 5)  [Prof]o.white = 0
       Constraint 5 dropped

           chi2(  4) =    8.15
         Prob > chi2 =    0.0863

. test ed

 ( 1)  [Menial]ed = 0
 ( 2)  [BlueCol]ed = 0
 ( 3)  [Craft]ed = 0
 ( 4)  [WhiteCol]ed = 0
 ( 5)  [Prof]o.ed = 0
       Constraint 5 dropped

           chi2(  4) =   84.97
         Prob > chi2 =    0.0000

. test exper

 ( 1)  [Menial]exper = 0
 ( 2)  [BlueCol]exper = 0
 ( 3)  [Craft]exper = 0
 ( 4)  [WhiteCol]exper = 0
 ( 5)  [Prof]o.exper = 0
       Constraint 5 dropped

           chi2(  4) =    7.99
         Prob > chi2 =    0.0918
```

- Or more simply:

```
. mlogtest, wald

**** Wald tests for independent variables (N=337)

 Ho: All coefficients associated with given
variable(s) are 0.

             |       chi2    df   P>chi2
-------------+-------------------------
       white |      8.149     4    0.086
          ed |     84.968     4    0.000
       exper |      7.995     4    0.092
---------------------------------------
```

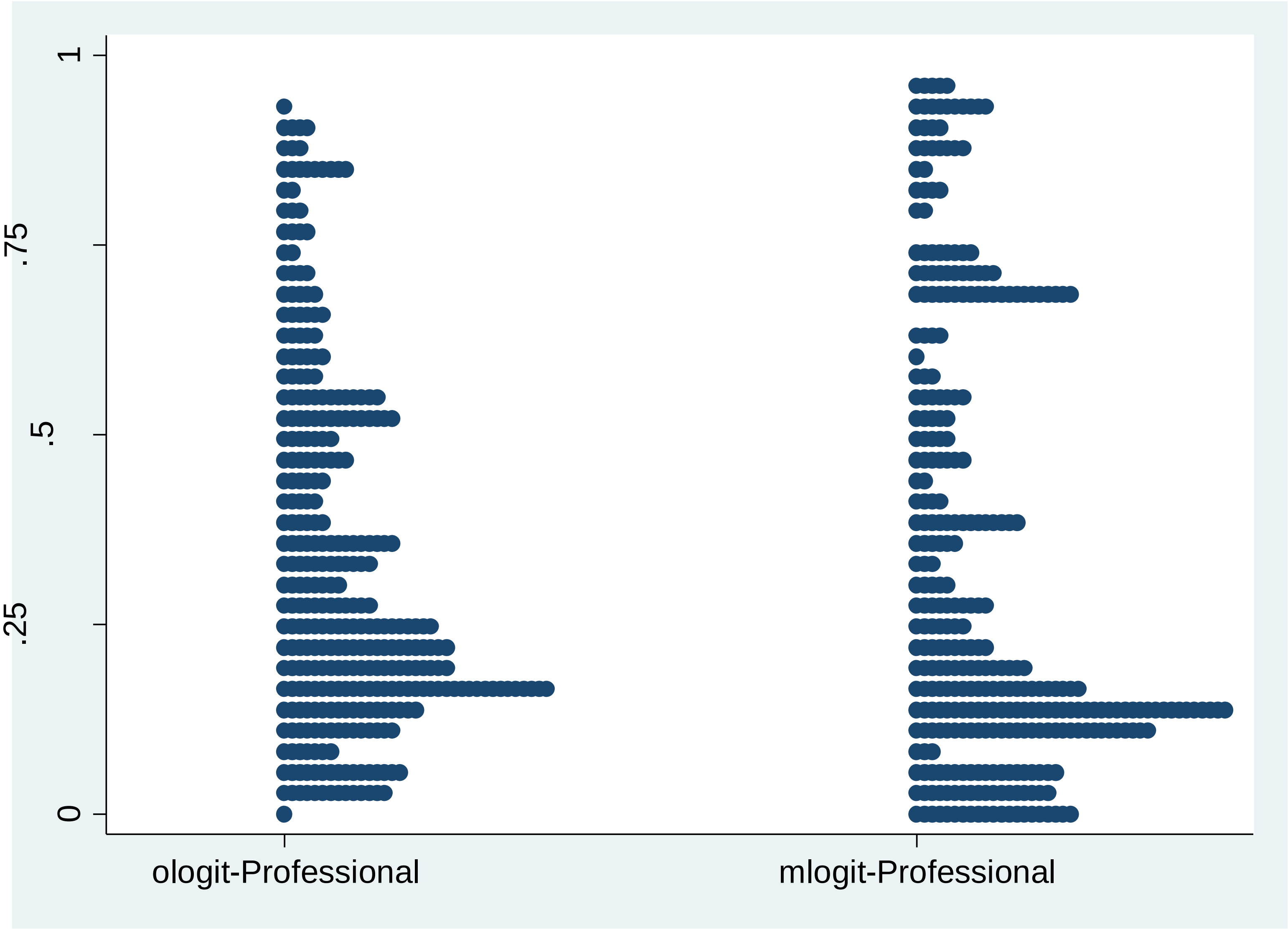
- We can also test if two categories can be combined….
  - Menial and blue collar sound pretty similar.
  - As does white collar and professional.

```
. mlogtest, combine

**** Wald tests for combining alternatives (N=337)

 Ho: All coefficients except intercepts associated with a given pair
     of alternatives are 0 (i.e., alternatives can be combined).

Alternatives tested|      chi2    df   P>chi2
-------------------+------------------------
  Menial- BlueCol  |     3.994     3    0.262
  Menial-   Craft  |     3.203     3    0.361
  Menial-WhiteCol  |    11.951     3    0.008
  Menial-    Prof  |    48.190     3    0.000
 BlueCol-   Craft  |     8.441     3    0.038
 BlueCol-WhiteCol  |    20.055     3    0.000
 BlueCol-    Prof  |    76.393     3    0.000
   Craft-WhiteCol  |     8.892     3    0.031
   Craft-    Prof  |    60.583     3    0.000
WhiteCol-    Prof  |    22.203     3    0.000
--------------------------------------------
```

We can also test individual pairs

```
. test [Menial=Craft]

 ( 1)  [Menial]white - [Craft]white = 0
 ( 2)  [Menial]ed - [Craft]ed = 0
 ( 3)  [Menial]exper - [Craft]exper = 0

           chi2(  3) =    3.20
         Prob > chi2 =    0.3614


. test [Menial=Prof]

 ( 1)  [Menial]white - [Prof]o.white = 0
 ( 2)  [Menial]ed - [Prof]o.ed = 0
 ( 3)  [Menial]exper - [Prof]o.exper = 0

           chi2(  3) =   48.19
         Prob > chi2 =    0.0000
```

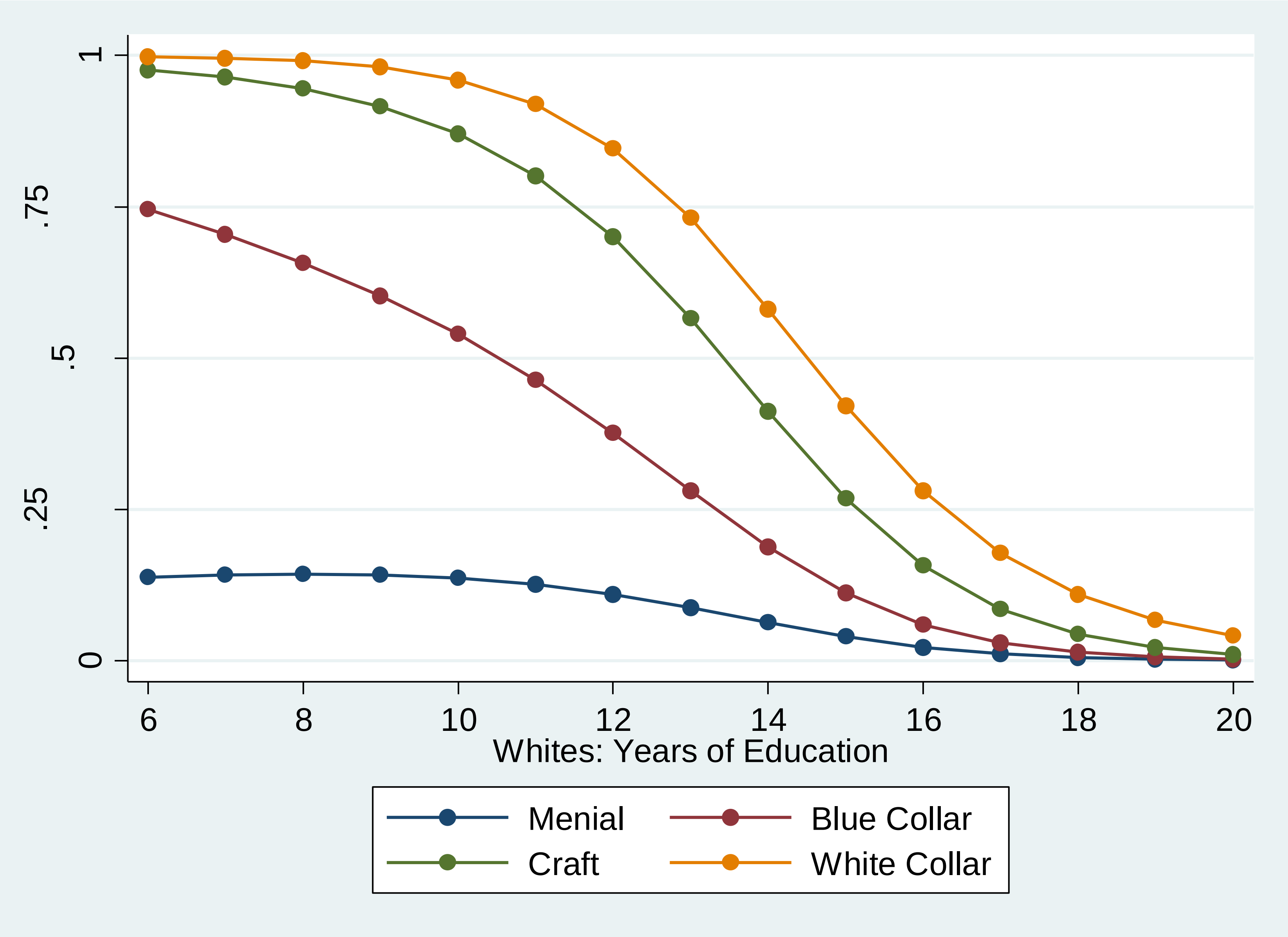
The predictions do differ from ologit



```
** comparing ologit and mlogit
ologit occ white ed exper, nolog
predict Menialo Blueo Crafto Whitecolo Profo
label var Profo "ologit-Professional"
mlogit occ white ed exper, baseoutcome(5) nolog
predict Menialm Bluem Craftm Whitecolm Profm
label var Profm "mlogit-Professional"
dotplot Profo Profm, ylabel(0(.25)1)
```

- However, what we are probably interested from a theoretical perspective is how the probability of one category changes relative to another over some range of an independent variable.

```
. prgen ed, x(white=1) from(6) to(20) gen(wht) ncases(15)

mlogit: Predicted values as ed varies from 6 to 20.

        white          ed      exper
x=          1   13.094955  20.501484

. desc wht*

              storage  display     value
variable name   type   format      label      variable label
---------------------------------------------------------------------------
------------------
whtx            float  %9.0g                  Years of education
whtp1           float  %9.0g                  pr(Menial)=Pr(1)
whtp2           float  %9.0g                  pr(BlueCol)=Pr(2)
whtp3           float  %9.0g                  pr(Craft)=Pr(3)
whtp4           float  %9.0g                  pr(WhiteCol)=Pr(4)
whtp5           float  %9.0g                  pr(Prof)=Pr(5)
whts1           float  %9.0g                  pr(y<=1)
whts2           float  %9.0g                  pr(y<=2)
whts3           float  %9.0g                  pr(y<=3)
whts4           float  %9.0g                  pr(y<=4)
whts5           float  %9.0g                  pr(y<=5)
```


1
.75
.5
.25
0
6
8
10
12
14
16
18
20
Whites: Years of Education
Menial
Blue Collar
Craft
White Collar

```
label var        whts1   "Menial"
         label var whts2 "Blue Collar"
         label var whts3 "Craft"
         label var whts4 "White Collar"
graph twoway connected whts1 whts2 whts3 whts4 whtx, ///
         ytitle("Summed Probability")       ///
         xtitle("Whites: Years of Education") ///
         xlabel(6(2)20) ylabel(0(.25)1) ///
```

- There are a number of other means of interpretation described in Long(1997: Ch. 6) and Long and Freese (2006: Ch.6-7).

- Let's move to CL models.

- Many econometrics text books use the example from Greene and Hensher (1995) of transport options.

- These data can be structured in several ways:

```
. use "http://www.indiana.edu/~jslsoc/stata/spex_data/travel2.dta", clear
. list id mode choice train bus time invc in 1/6, nolabel sepby(id)

     +--------------------------------------------------+
     | id   mode   choice   train   bus   time   invc |
     |--------------------------------------------------|
  1. |  1      1        0       1     0    406     31 |
  2. |  1      2        0       0     1    452     25 |
  3. |  1      3        1       0     0    180     10 |
     |--------------------------------------------------|
  4. |  2      1        0       1     0    398     31 |
  5. |  2      2        0       0     1    452     25 |
  6. |  2      3        1       0     0    255     11 |
     +--------------------------------------------------+

. use "http://www.stata-press.com/data/lf2/travel2case.dta", clear
(Greene & Hensher 1997 data in one-row-per-case format)

. list id time1 time2 time3 invc1 invc2 invc3 choice in 1/2, nolabel

     +----------------------------------------------------------------+
     | id   time1   time2   time3   invc1   invc2   invc3   choice |
     |----------------------------------------------------------------|
  1. |  1     406     452     180      31      25      10        3 |
  2. |  2     398     452     255      31      25      11        3 |
     +----------------------------------------------------------------+
```

```
. tab mode

     Mode of |
transportat |
        ion |      Freq.     Percent        Cum.
------------+-----------------------------------
      Train |        152       33.33       33.33
        Bus |        152       33.33       66.67
        Car |        152       33.33      100.00
------------+-----------------------------------
      Total |        456      100.00
```

```
. clogit choice train bus time invc, group(id)

Iteration 0:   log likelihood = -142.24059
Iteration 1:   log likelihood = -84.116723
Iteration 2:   log likelihood = -80.965361
Iteration 3:   log likelihood = -80.961135
Iteration 4:   log likelihood = -80.961135

Conditional (fixed-effects) logistic regression   Number of obs   =        456
                                                  LR chi2(4)      =     172.06
                                                  Prob > chi2     =     0.0000
Log likelihood = -80.961135                       Pseudo R2       =     0.5152

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |   2.671238   .4531611     5.89   0.000     1.783058    3.559417
         bus |   1.472335   .4007152     3.67   0.000     .6869474    2.257722
        time |  -.0191453   .0024509    -7.81   0.000    -.0239489   -.0143417
        invc |  -.0481658   .0119516    -4.03   0.000    -.0715905   -.0247411
------------------------------------------------------------------------------
```

Interpreting odds ratios from clogit

```
. listcoef, help

clogit (N=456): Factor Change in Odds

  Odds of: 1 vs 0

--------------------------------------------------
      choice |        b         z     P>|z|    e^b
-------------+------------------------------------
       train |   2.67124    5.895   0.000  14.4579
         bus |   1.47233    3.674   0.000   4.3594
        time |  -0.01915   -7.812   0.000   0.9810
        invc |  -0.04817   -4.030   0.000   0.9530
--------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
   SDofX = standard deviation of X
```

- Increasing travel *time* of an alternative by 1 minute decreases the odds of using that option by a factor of .98 (2%) holding other alternative values constant.

- If cost and time were equal, travelers would be 4.36 times more likely to travel by train than car.

Mixed Model

```
gen busXhinc = bus*hinc
gen trainXhinc =train*hinc
gen busXpsize=bus*psize
gen trainXpsize=train*psize

. clogit choice busXhinc  busXpsize bus  trainXhinc trainXpsize train ///
>          time invc, group(id) nolog

Conditional (fixed-effects) logistic regression   Number of obs   =        456
                                                  LR chi2(8)      =     178.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -77.504846                       Pseudo R2       =     0.5359

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    busXhinc |  -.0080174   .0200322    -0.40   0.689    -.0472798     .031245
   busXpsize |  -.5141037   .4007015    -1.28   0.199    -1.299464    .2712569
         bus |   2.486465   .8803649     2.82   0.005     .7609815    4.211949
  trainXhinc |  -.0342841   .0158471    -2.16   0.031    -.0653438   -.0032243
 trainXpsize |  -.0038421   .3098075    -0.01   0.990    -.6110537    .6033695
       train |   3.499641   .7579665     4.62   0.000     2.014054    4.985228
        time |  -.0185035   .0025035    -7.39   0.000    -.0234103   -.0135966
        invc |  -.0402791   .0134851    -2.99   0.003    -.0667095   -.0138488
------------------------------------------------------------------------------
```

Independence of Irrelevant Alternatives (IIA)

- Both the MNL and the CL make an important assumption that is actually pretty restrictive.
- The IIA assumption derives from the “ratcho” literature mentioned above.
- In this sense these models are rather direct links between theory (rational utility) and estimation (rational utility models).

- The IIA assumption is that the probability of an outcome is unaffected by the addition or subtraction of other irrelevant alternatives.

  - If you add (or take away) strawberry ice cream, and I prefer chocolate to vanilla, I should still prefer chocolate to vanilla.

- Importantly, this is an assumption about individual behavior rather than an econometric assumption (e.g. holding the variance to 1).

- Of course we are also making econometric assumptions by assuming that the errors are i.i.d. and the homogeneity between individuals and alternatives.

- If you do add an alternative and the preference ordering changes then IIA is violated.

- To be specific we are concerned about the *ratio* between the probabilities of alternatives (Long 1997: 182)

- For the MNL, the odds of *m* versus *n*

$$\frac{P(y=m|x)}{P(y=n|x)} = e^{(x[\beta_m - \beta_n])}$$

- For the CL, the odds of m versus *n*

$$\frac{P(y=m|\, z)}{P(y=n|\, z)} = e^{([z_m - z_n]\gamma)}$$

- The classic example is again transport options.
- Let's say that there are two options: take a car or a red bus.
- Let's also say that a person is indifferent between these two options. P(car)=1/2 and P(red bus)=1/2
- The implied odds are ½ / ½ =1.
- What happens if a new bus line (blue bus) opens that is identical to the red bus in every way but color?

- IIA assumes that the probabilities are now: P(car)=1/3; P(red bus)=1/3; and P(blue bus)=1/3.
- This is necessary to keep the same ratio (1) between car and red bus.
- Therefore, if a bunch of new bus companies start operating then the probability of using a car keeps decreasing.
- This is a strong assumption, because it is doubtful that people are going to keep giving up their cars for the bus, especially when the bus (regardless of color) is not any more attractive, cheaper, or faster.

Testing IIA

- There are several ways of testing IIA in Stata.
  - Hausman
  - Small-Hsiao
    - See Long and Freese (2006: 243-246)

```
. mlogtest, hausman base
**** Hausman tests of IIA assumption (N=337)
 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
 Omitted |      chi2   df   P>chi2   evidence
---------+------------------------------------
  Menial |     7.324   12    0.835   for Ho
 BlueCol |     0.320   12    1.000   for Ho
   Craft |   -14.436   12     ---    ---
WhiteCol |    -5.541   11     ---    ---
    Prof |    -0.119   12     ---    ---
----------------------------------------------
 Note: If chi2<0, the estimated model does not
 meet asymptotic assumptions of the test.

. mlogtest, smhsiao

**** Small-Hsiao tests of IIA assumption (N=337)
 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+-------------------------------------------------------
  Menial |   -173.287   -166.950  12.675   12    0.393   for Ho
 BlueCol |   -154.895   -150.543   8.705   12    0.728   for Ho
   Craft |   -133.658   -130.611   6.095   12    0.911   for Ho
WhiteCol |   -152.900   -148.357   9.086   12    0.696   for Ho
-----------------------------------------------------------------
```

The Small-Hsiao test is fragile.

- As Long and Freese (2005:244-246) suggest.

```
. set seed 911
. mlogtest, hausman base
**** Hausman tests of IIA assumption (N=337)
 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
 Omitted |      chi2   df   P>chi2   evidence
---------+------------------------------------
  Menial |     7.324   12    0.835   for Ho
 BlueCol |     0.320   12    1.000   for Ho
   Craft |   -14.436   12     ---    ---
WhiteCol |    -5.541   11     ---    ---
    Prof |    -0.119   12     ---    ---
----------------------------------------------
 Note: If chi2<0, the estimated model does not
 meet asymptotic assumptions of the test.

. mlogtest, smhsiao

**** Small-Hsiao tests of IIA assumption (N=337)
 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+-------------------------------------------------------
  Menial |   -246.322   -165.532 161.579   12    0.000   against Ho
 BlueCol |   -157.439   -129.881  55.117   12    0.000   against Ho
   Craft |   -204.042   -123.616 160.851   12    0.000   against Ho
WhiteCol |   -204.080   -147.249 113.662   12    0.000   against Ho
-----------------------------------------------------------------
```

- Are there ways of relaxing the IIA?
- Yes!
- The Alternative-specific Multinomial Probit (ASMP)
  - Stata: asmprobit
  - "Alternative specific" means that we need information about the different alternatives
  - E.g. how much it costs to ride the bus.
  - See Long and Freese (2006: Ch 7) for details and Lacy and Burden (1999) for an example.
- This model allows the errors to be correlated.

- There are a number of other models of discrete choice.

- Stereotype model (see Long and Freese 2006)

- Nested logit
  - Grouping alternatives to different branches and twigs
  - E.g. land and air transport

- Rank-ordered logit
  - If you have data in which cases actually explicitly order preferences

- A quick demonstration about creating tables:
  - Outreg2
  - Esttab

- Now I would like to spend some time working through the two substantive articles for today.
  - Alvarez & Nagler (1998)
  - Lacy & Burden (1999)

- Questions?