Using Heterogeneous Choice Models
To Compare Logit and Probit Coefficients Across Groups
Part II: A Better Solution to the Problem

[Note: This is adapted from the June 2006 Working Paper (later revised August 2007) of the same name by Richard Williams. This handout includes the code needed to estimate these models in Stata. The user-written routine `oglm` should be installed. Go to http://www.nd.edu/~rwilliam/oglm/index.html for more information.]

Improving on Allison: Heterogeneous Choice Models

As noted before, Allisons' heteroskedastic logit model, with a single dichotomous variable in the variance equation, is a special case of the larger class of models that are variously known as location-scale models and heterogeneous choice models. Turning to this larger class of models offers several ways to improve on Allison's approach and hopefully overcome its most significant weaknesses.

- Allison's paper was written in 1999. The specialized routines that Allison wrote are no longer necessary because today, major software packages include routines for estimating heterogeneous choice models. For example, SPSS has PLUM (Norusis, 2005) while Stata has the free user-written routine `oglm` (Williams, 2006b).
    - SPSS PLUM uses the location-scale terminology for its models, while oglm lets the user choose whichever terminology they prefer. Unfortunately, and for unclear reasons, SPSS requires that the variables in the scale equation be a subset of the variables in the location equation; `oglm` has no such constraint.

- With `oglm` it is also possible to do stepwise selection of variables in either the choice or variance equations, easily estimate a sequence of nested models, and do survey data analysis of data sets with complicated sampling schemes.
    - Using Allison's data and stepwise selection with `oglm`, the only variable that entered into the variance equation was number of articles. Of course, it is not unusual for heteroskedasticity to be an issue with continuous variables that have a broad range. Hence, while Allison may be correct in fearing that heteroskedasticity is a problem, he may have focused his concern on the wrong variable.
    - Here is `oglm` code for stepwise selection with the variance equation. As the help for `oglm` notes, you almost certainly want to use the `lr` (likelihood ratio test) option with stepwise. The `flip` option causes the placement of the choice and variance options to be reversed. This can all be a little confusing, so you should go over the examples in the help file carefully.

```
. sw, pe(.05) lr: oglm  tenure female year yearsq select  articles prestige ,
eq2(female  year yearsq select  articles prestige)  flip

LR test               begin with empty model
p = 0.0000 <  0.0500  adding  articles

Heteroskedastic Ordered Logistic Regression       Number of obs   =       2797
                                                  LR chi2(7)      =     428.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -828.81224                       Pseudo R2       =     0.2052

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
tenure       |
      female |  -.4179259   .1742083    -2.40   0.016    -.7593679   -.0764838
        year |   2.108752   .2486632     8.48   0.000     1.621381    2.596123
      yearsq |  -.1542213   .0208578    -7.39   0.000    -.1951019   -.1133406
      select |   .1744644   .0598623     2.91   0.004     .0571364    .2917924
    articles |   .0628408    .015785     3.98   0.000     .0319026    .0937789
    prestige |  -.6118688   .1307262    -4.68   0.000    -.8680875   -.3556502
-------------+----------------------------------------------------------------
lnsigma      |
    articles |    .030149   .0091448     3.30   0.001     .0122256    .0480724
-------------+----------------------------------------------------------------
       /cut1 |   7.959555   .7637104    10.42   0.000      6.46271      9.4564
------------------------------------------------------------------------------
```

- There is no need to limit the variance equation to a single dichotomous grouping variable. Multiple grouping variables can be used, as can continuous variables that may be sources of heteroskedasticity. Indeed, the variables in the variance equation need not even be a subset of the variables in the choice equation, as is the case with Allison's procedure. This hopefully reduces or even eliminates problems caused by specification error in the variance equation.

- The variance may itself be of substantive interest. The variance equation lets you examine the determinants of variability. Alvarez and Brehm (1995), for example, argued that individuals whose core values are in conflict will have a harder time making a decision about abortion and will hence have greater variability/error variances in their responses. In the case of Allison's example, we might be interested in whether gender or other factors affect the variability in careers.

- Allison's procedure works with a dichotomous dependent variable. Heterogeneous choice models also allow for ordinal dependent variables. There are several advantages to using ordinal variables when possible.

    - As Keele and Park (2006) note, ordinal variables contain more information and models using them are much less prone to problems than are models with dichotomous dependent variables. Based on their Monte Carlo simulations, they concluded that, unlike the heteroskedastic probit model, when the model was correctly specified, "The heteroskedastic ordered probit model can be given a clean bill of health, as both the level of overconfidence and coverage rates are close to ideal."

- However, even for a heteroskedastic ordered probit model, they stressed the importance of the model being correctly specified; a mis-specified model, e.g. a variance equation with omitted variables might be worse than a model that made no correction at all for heteroskedasticity.

o Also, unlike with Allison's procedure, with ordinal variables you do NOT need to make the questionable assumption that at least one coefficient is the same across groups. You do, however, need to make the assumption that the cutpoints are the same for both groups. This is a less questionable assumption, in that it implies that both groups interpret the question the same way.

  - Nonetheless, researchers should realize that the assumption may be wrong in some cases; for example, Lindeboom & Doorslaer (2004) note (p. 1084) that sometimes "sub-groups of a population use systematically different threshold levels when assessing their health, despite having the same level of 'true' health. These differences may be influenced by, among other things, age, sex, education, language and personal experience of illness. It means that different groups appear to 'speak different languages' and to use different reference points when they are responding to the same question." Of course, any procedure can have problems if different groups interpret and answer questions differently.

Example. We'll once again use the Working Mothers data. Respondents are asked to evaluate the following statement: "A working mother can establish just as warm and secure a relationship with her child as a mother who does not work." Responses were coded as 1 = Strongly Disagree (1SD), 2 = Disagree (2D), 3 = Agree (3A), and 4 = Strongly Agree (4SA). Explanatory variables are yr89 (survey year; 0 = 1977, 1 = 1989), male (0 = female, 1 = male), white (0 = nonwhite, 1 = white), age (measured in years), ed (years of education), and prst (occupational prestige scale). In Table 5, we present a series of models for these data; but first is the `oglm` code for model 1.

```
. * oglm code for Model 1, table 5
. use http://www.nd.edu/~rwilliam/xsoc73994/long2006/ordwarm2.dta
(77 & 89 General Social Survey)

. oglm  warm yr89 male white age ed prst, store(ologit)

Ordered Logistic Regression                       Number of obs   =       2293
                                                  LR chi2(6)      =     301.72
                                                  Prob > chi2     =     0.0000
Log likelihood = -2844.9123                       Pseudo R2       =     0.0504

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |   .5239025   .0798989     6.56   0.000     .3673036    .6805014
        male |  -.7332998   .0784827    -9.34   0.000     -.887123   -.5794765
       white |  -.3911595   .1183808    -3.30   0.001    -.6231817   -.1591374
         age |  -.0216655   .0024683    -8.78   0.000    -.0265032   -.0168278
          ed |   .0671727    .015975     4.20   0.000     .0358624    .0984831
        prst |   .0060727   .0032929     1.84   0.065    -.0003813    .0125267
-------------+----------------------------------------------------------------
       /cut1 |  -2.465362   .2389129   -10.32   0.000    -2.933623   -1.997102
       /cut2 |  -.6309044   .2333157    -2.70   0.007    -1.088195    -.173614
       /cut3 |   1.261854   .2340181     5.39   0.000     .8031864    1.720521
------------------------------------------------------------------------------
```

Table 5: Ordered Logit & Heterogeneous Choice Models for the Working Mothers' Data

Equation/ Variable	(1) Ordered Logit	(2) Heterogeneous Choice	(3) Heterogeneous Choice + Gender Interactions
Choice			
yr89	0.524***	0.453***	0.413***
	(0.080)	(0.069)	(0.10)
male	-0.733***	-0.635***	-0.418
	(0.078)	(0.070)	(0.40)
white	-0.391***	-0.309**	-0.496***
	(0.12)	(0.10)	(0.15)
age	-0.0217***	-0.0186***	-0.0184***
	(0.0025)	(0.0022)	(0.0032)
ed	0.0672***	0.0536***	0.0831***
	(0.016)	(0.014)	(0.024)
prst	0.00607	0.00529	0.00530
	(0.0033)	(0.0028)	(0.0044)
Male*yr89			0.0689
			(0.13)
Male*white			0.371
			(0.20)
Male*age			0.000110
			(0.0042)
Male*ed			-0.0437
			(0.029)
Male*prst			-0.00105
			(0.0057)
Variance			
yr89		-0.149**	-0.147**
		(0.046)	(0.046)
male		-0.191***	-0.194***
		(0.045)	(0.045)
Thresholds			
Cutpoint 1	-2.465***	-2.151***	-1.959***
	(0.24)	(0.21)	(0.33)
Cutpoint 2	-0.631**	-0.570**	-0.382
	(0.23)	(0.20)	(0.32)
Cutpoint 3	1.262***	1.067***	1.259***
	(0.23)	(0.20)	(0.32)
Observations	2293	2293	2293
Pseudo R^2	0.0504	0.0552	0.0564
Model chi-square	301.7	331.0	338.1
Model d.f.	6	8	13

Model 1 is an ordered logit model, with no controls for heteroskedasticity. (`ologit` produces the same results as `oglm` in this case.) As we have previously discussed, the results from Model 1 are relatively straightforward, intuitive and easy to interpret. People tended to be more supportive of working mothers in 1989 than in 1977. Males, whites and older people tended to be less supportive of working mothers, while better educated people and people with higher occupational prestige were more supportive.

But, while the results may be straightforward, intuitive, and easy to interpret, are they correct? As we have shown before, yr89 and male violate the parallel lines assumptions of the ordered logit model. Earlier, we demonstrated how a generalized ordered logit model provided a superior fit. Model 2 presents another alternative: a heterogeneous choice model, where yr89 and male (the two variables that violated the parallel lines assumption) are included in the variance equation.

- NOTE: `oglm`'s stepwise selection procedure also identifies yr89 and male as statistically significant variables for inclusion in the variance equation. This implies that variability in attitudes toward working mothers differed by year and by gender, both of which are substantively plausible.

```
. * oglm code for Model 2, table 5
. oglm  warm yr89 male white age ed prst, store(oglm) het( yr89 male)

Heteroskedastic Ordered Logistic Regression       Number of obs   =       2293
                                                  LR chi2(8)      =     331.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -2830.2563                       Pseudo R2       =     0.0552

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
warm         |
        yr89 |   .4531574   .0686839     6.60   0.000     .3185394    .5877755
        male |  -.6345402   .0697638    -9.10   0.000    -.7712748   -.4978057
       white |  -.3087676    .102739    -3.01   0.003    -.5101323   -.1074029
         age |  -.0186098   .0021728    -8.56   0.000    -.0228684   -.0143512
          ed |   .0535685   .0135944     3.94   0.000     .0269239     .080213
        prst |   .0052866     .00278     1.90   0.057    -.0001622    .0107353
-------------+----------------------------------------------------------------
lnsigma      |
        yr89 |  -.1486188   .0458169    -3.24   0.001    -.2384183   -.0588192
        male |  -.1909211    .044807    -4.26   0.000    -.2787412   -.1031011
-------------+----------------------------------------------------------------
       /cut1 |  -2.151122   .2114069   -10.18   0.000    -2.565472   -1.736772
       /cut2 |  -.5696264   .1992724    -2.86   0.004    -.9601932   -.1790596
       /cut3 |   1.066508   .2022099     5.27   0.000     .6701839    1.462832
------------------------------------------------------------------------------
```

```
. * oglm code for Model 2, table 5, using mindless sw regression for the variance equation
. sw, pe(.05) lr: oglm  warm yr89 male white age ed prst, store(oglm) eq2( yr89 male
white age ed prst) flip

LR test                begin with empty model
p = 0.0000 <  0.0500  adding  male
p = 0.0012 <  0.0500  adding  yr89

Heteroskedastic Ordered Logistic Regression       Number of obs   =       2293
                                                  LR chi2(8)      =     331.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -2830.2563                       Pseudo R2       =     0.0552

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
warm         |
        yr89 |   .4531574   .0686839     6.60   0.000     .3185394    .5877755
        male |  -.6345402   .0697638    -9.10   0.000    -.7712748   -.4978057
       white |  -.3087676    .102739    -3.01   0.003    -.5101323   -.1074029
         age |  -.0186098   .0021728    -8.56   0.000    -.0228684   -.0143512
          ed |   .0535685   .0135944     3.94   0.000     .0269239     .080213
        prst |   .0052866     .00278     1.90   0.057    -.0001622    .0107353
-------------+----------------------------------------------------------------
lnsigma      |
        male |  -.1909211    .044807    -4.26   0.000    -.2787412   -.1031011
        yr89 |  -.1486188   .0458169    -3.24   0.001    -.2384183   -.0588192
-------------+----------------------------------------------------------------
       /cut1 |  -2.151122   .2114069   -10.18   0.000    -2.565472   -1.736772
       /cut2 |  -.5696264   .1992724    -2.86   0.004    -.9601932   -.1790596
       /cut3 |   1.066508   .2022099     5.27   0.000     .6701839    1.462832
------------------------------------------------------------------------------
```

The negative coefficients for these variables in the variance equation tell us that the variance in attitudes towards working mothers declined across time, and that men were less variable in their attitudes than were women. The addition of the two heteroskedasticity parameters improves the model fit significantly (29.3 chi-square with only 2 d.f.).

Contingent on the thresholds being the same for both men and women, we can further test whether any of the coefficients for the choice equation differ by gender. Model 3 adds interaction terms for gender to Model 2. We see that none of the interaction terms for gender are significant. Further, the chi-square contrast between the two models is 7.04 with 5 d.f., which is also insignificant.

```
. * Model 3, table 5
. gen male89 = male*yr89
. gen malewhit = male*white
. gen maleage = male*age
. gen maleed = male*ed
. gen maleprst = male*prst

. oglm  warm yr89 male white age ed prst male89 malewhit maleage maleed maleprst, het(
yr89 male) store(oglm2)

Heteroskedastic Ordered Logistic Regression       Number of obs   =       2293
                                                  LR chi2(13)     =     338.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -2826.7163                       Pseudo R2       =     0.0564

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
warm         |
        yr89 |   .4129525   .1006924     4.10   0.000     .2155989     .610306
        male |  -.4178799   .4036235    -1.04   0.301    -1.208967    .3732077
       white |  -.4960901   .1452637    -3.42   0.001    -.7808017   -.2113785
         age |  -.0183564   .0031543    -5.82   0.000    -.0245388    -.012174
          ed |   .0830669   .0240416     3.46   0.001     .0359462    .1301876
        prst |   .0053009   .0044468     1.19   0.233    -.0034146    .0140164
      male89 |   .0688798   .1344997     0.51   0.609    -.1947349    .3324944
    malewhit |   .3706861   .2032988     1.82   0.068    -.0277723    .7691445
     maleage |   .0001101   .0041793     0.03   0.979    -.0080812    .0083014
      maleed |  -.0437119   .0288247    -1.52   0.129    -.1002072    .0127833
    maleprst |  -.0010469   .0057186    -0.18   0.855    -.0122551    .0101614
-------------+----------------------------------------------------------------
lnsigma      |
        yr89 |  -.1474389   .0459572    -3.21   0.001    -.2375133   -.0573644
        male |  -.1943191   .0447506    -4.34   0.000    -.2820287   -.1066096
-------------+----------------------------------------------------------------
       /cut1 |  -1.959263   .3255178    -6.02   0.000    -2.597267    -1.32126
       /cut2 |  -.3823006   .3189892    -1.20   0.231    -1.007508    .2429067
       /cut3 |   1.259111   .3224691     3.90   0.000     .6270835    1.891139
------------------------------------------------------------------------------
```

In this case, the interaction effects involving group membership are not significant. *Nonetheless*, the heterogeneous choice model yields important insights into the effects of gender and year that would be overlooked in a mis-specified ordered logit model. An examination of marginal effects helps to clarify what the substantive differences are between the two models.

- With marginal effects, all variables except one are set equal to their means, and we see how changes in the remaining variables affect the probability of each possible outcome occurring.
    - For a dichotomous explanatory variable, we measure the effect as the variable changes from 0 to 1. For continuous variables, the instantaneous rate of change is measured. (See Long and Freese 2006 for a more detailed discussion of marginal effects in categorical models.)

Table 6 presents the marginal effects for the ordered logit and heterogeneous choice models. The table illustrates important differences and similarities for the two models.

Table 6: Marginal Effects for the ordered logit and heterogeneous choice models

COEFFICIENT	Ordered Logit	Heterogeneous Choice
Strongly Disagree		
yr89	-0.0499***	-0.0786***
	(0.0076)	(0.011)
male	0.0746***	0.0355***
	(0.0086)	(0.012)
white	0.0345***	0.0319***
	(0.0094)	(0.0095)
age	0.00214***	0.00213***
	(0.00025)	(0.00025)
ed	-0.00664***	-0.00613***
	(0.0016)	(0.0016)
prst	-0.000600	-0.000605
	(0.00033)	(0.00032)
Disagree		
yr89	-0.0775***	-0.0618***
	(0.012)	(0.014)
male	0.105***	0.137***
	(0.012)	(0.014)
white	0.0594***	0.0543***
	(0.018)	(0.018)
age	0.00319***	0.00318***
	(0.00039)	(0.00039)
ed	-0.00990***	-0.00916***
	(0.0024)	(0.0023)
prst	-0.000895	-0.000904
	(0.00049)	(0.00048)
Agree		
yr89	0.0539***	0.0995***
	(0.0083)	(0.016)
male	-0.0814***	-0.0344***
	(0.0097)	(0.015)
white	-0.0356***	-0.0333***
	(0.0086)	(0.0090)
age	-0.00241***	-0.00240***
	(0.00031)	(0.00031)
ed	0.00746***	0.00691***
	(0.0018)	(0.0018)
prst	0.000675	0.000682
	(0.00037)	(0.00036)
Strongly Agree		
yr89	0.0735***	0.0409***
	(0.012)	(0.015)
male	-0.0979***	-0.138***
	(0.011)	(0.014)
white	-0.0583***	-0.0529***
	(0.019)	(0.019)
age	-0.00293***	-0.00291***
	(0.00034)	(0.00034)
ed	0.00908***	0.00839***
	(0.0022)	(0.0021)
prst	0.000821	0.000828
	(0.00045)	(0.00044)

Let us begin by noting the similarities.

- The marginal effects for white, age, ed and prst are very similar in both models and for all outcomes.

- These are the four variables that were not included in the variance equation of the heterogeneous choice model. It is not surprising that both models therefore largely agree on the effects of these four variables.

The story is very different for the variables yr89 and male.

- Both models agree that there was a shift toward more positive attitudes between 1977 and 1989, but they describe that shift differently.

- The ordered logit model provides the smallest estimate of the decline in strong disagreement (4.99% as opposed to 7.86%) and the largest estimate of the increase in strong agreement (7.35% versus a little over 4%).

- That is, the heterogeneous choice model says that the main reason attitudes became more favorable across time was because people shifted from extremely negative positions to more moderate positions; there was only a fairly small increase in people strongly agreeing that women should work. The ordered logit model, on the other hand, understates how much people moved from an extremely negative position and overstates how much they became extremely positive.

The models also provide different pictures of the effect of gender on attitudes.

- The ordered logit model provides a much larger estimate of how much men strongly disagree with a mother working (7.46% versus 3.55%).

- However, it also provides the smallest estimates of how much less likely men are to strongly agree that a woman should work.

- Again, the ordered logit model is creating a misleading image of why men were less supportive of working mothers; it isn't so much that men were extremely negative in their attitudes, it is more a matter of them being less likely than women to be extremely supportive.

VII Conclusions

Allison has provided a valuable service by alerting researchers to an important problem that has gone unnoticed by many. However, thanks, in part, to additional research that Allison's paper helped inspire, we know that his original proposed solution can sometimes have serious problems, and, counter to his advice, should NOT be applied on a routine basis.

- Under certain conditions, Allison's procedure can produce biased and inefficient estimates, and may be worse than doing nothing at all. Today, superior alternatives can be easily estimated using readily available commercial software.

- Luckily, Allison's procedures can be easily modified take advantage of the broader class of heterogeneous choice models, providing a powerful, and often more appropriate, way for addressing the problems that Allison presents.

At the same time, researchers need to realize that even with these methods, mis-specified models can be problematic.

- As Keele and Park (2006) show, ordinal models can also produce misleading results when the variance equation is mis-specified. The greater flexibility of heterogeneous choice models (which, unlike Allison's procedure, allow multiple variables in the variance equation) make omitted variable bias less likely, but it is still up to researchers to think through their models carefully.

- The inclusion of extraneous variables in the variance equation could still potentially distort estimates of group differences. Again, this seems less likely with a well thought-out model involving multiple variables, but it could still happen.

- Researchers should therefore estimate models both with and without controls for heteroskedasticity, and consider whether model mis-specification could be the cause of any seemingly-major differences in conclusions.

As part of this process, researchers may wish to vary the sequence in which they estimate nested models.

- For example, Allison first added the delta parameter to his model, noted that it was significant, and then added the interaction term for gender and articles, which he concluded was insignificant. (The delta term also became insignificant once the interaction term was entered, but Allison did not note this.)

- Had he first added the interaction term for articles, he would have found that it was significant, and that the delta term was insignificant when it was next added to the model.

- In other words, the sequence of models should not automatically give preference to the hypothesis of different residual variances over the hypothesis of differing coefficients. If the sequence of models does affect the conclusions reached, researchers should at least acknowledge this in their discussion if not rethink their models altogether.

In short, comparisons of logit and probit coefficients across groups do pose challenges to researchers. However, well thought out models, modern statistical software, and the methods described here can help to make those challenges manageable.