Estimating heterogeneous choice models with Stata

Richard Williams
Department of Sociology, University of Notre Dame, Notre Dame, IN
Richard.A.Williams.5@ND.Edu
Last revised August 11, 2008 – Currently under review at *The Stata Journal*

Abstract. When a binary or ordinal regression model incorrectly assumes that error variances are the same for all cases, the standard errors are wrong and (unlike OLS regression) the parameter estimates are biased. Heterogeneous choice (also known as location-scale or heteroskedastic ordered) models explicitly specify the determinants of heteroskedasticity in an attempt to correct for it. Such models are also useful when the variance itself is of substantive interest. This paper illustrates how the user-written Stata routine `oglm` (Ordinal Generalized Linear Models) can be used to estimate heterogeneous choice and related models. It shows that two other models that have appeared in the literature (Allison's model for group comparisons and Hauser and Andrew's logistic response model with proportionality constraints) are either special cases of a heterogeneous choice model or else algebraically equivalent to it. The paper further argues that heterogeneous choice models may sometimes be an attractive alternative to other ordinal regression models, such as the generalized ordered logit model estimated by `gologit2`. Finally, the paper offers guidelines on how to interpret, test and modify heterogeneous choice models.

Keywords. oglm, heterogeneous choice model, location-scale model, gologit2, ordinal regression, heteroskedasticity, generalized ordered logit model

1 Introduction

When a binary or ordinal regression model incorrectly assumes that error variances are the same for all cases, the standard errors are wrong and (unlike OLS regression) the parameter estimates are biased (Yatchew & Griliches 1985). Heterogeneous choice (also known as location-scale or heteroskedastic ordered) models explicitly specify the determinants of heteroskedasticity in an attempt to correct for it.

In addition, most regression-type analyses focus on the conditional mean of a variable or on conditional probabilities, e.g. E(Y|X), Pr(Y=1|X). Sometimes, however, determinants of the conditional variance are also of interest. For example, Allison (1999) speculated that unmeasured variables affecting the chances of promotion may be more important for women scientists than for men, causing their career outcomes to be more variable and less predictable. Heterogeneous choice models make it possible to examine such issues.

Williams (forthcoming) provides an extensive critique of the strengths and weaknesses of heterogeneous choice models, including a more detailed substantive discussion of some of the examples presented here. This paper takes a more applied approach, and illustrates how the user-written Stata routine `oglm` (Ordinal Generalized Linear Models[1]) can be used to estimate heterogeneous choice and related models. The paper demonstrates how two other models that have appeared in the literature – Allison's (1999) model for comparing logit and probit coefficients across groups, and Hauser and Andrew's (2006) logistic response model with proportionality constraints (LRPC) – are special cases of the heterogeneous choice model or algebraically equivalent to it, and can also be estimated with `oglm`. The paper further argues that heterogeneous choice models may sometimes be an attractive alternative to other ordinal

[1] The name is slightly misleading in that `oglm` can also estimate the nonlinear models presented here.

regression models, such as the generalized ordered logit model estimated by `gologit2`. Finally, the paper offers guidelines on how to interpret the parameters of such models, ways to make interpretation easier, and procedures for testing hypotheses and making model modifications.

2 The Heterogeneous Choice/ Location-Scale Model

Suppose there is an observed variable, y, with ordered categories, e.g. strongly disagree, agree, neutral, agree, strongly agree. One of the rationales for the ordered logit and probit models is that y is actually a "collapsed" or "limited" version of a latent variable, y*. As respondents cross thresholds or cutpoints on y*, their observed values on y change, e.g.

y = 1 if $-\infty < y^* < \kappa_1$,
y = 2 if $\kappa_1 < y^* < \kappa_2$,
y = 3 if $\kappa_2 < y^* < \kappa_3$,
y = 4 if $\kappa_3 < y^* < \kappa_4$,
y = 5 if $\kappa_4 < y^* < +\infty$

The model for the underlying y* can be written as

$$y_i^* = \alpha_0 + \alpha_1 x_{i1} + ... + \alpha_K x_{iK} + \sigma\varepsilon_i$$

where the x's are the explanatory variables, the α's are a vector of coefficients that give the effect of each x on y*, ε_i is a residual term often assumed to have either a logistic or normal(0, 1) distribution, and σ is a parameter that allows the variance to be adjusted upward or downward.

Because y* is a latent variable, its metric has to be fixed in some way. Typically, this is done by scaling the coefficients so that the residual variance is 3.29 (as in logit) or 1 (as in probit)[2]. Further, because y* is unobserved, we do not actually estimate the αs. Rather, we estimate parameters called βs. As Allison (1999, citing Amemiya 1985:269) notes, the αs and the βs are related this way:

$$\beta_k = \alpha_k / \sigma \quad k = 1, \ldots, K$$

This now leads us to a potential problem with the ordered logit/probit model. When σ is the same for all cases – residuals are homoskedastic – the ratio between the βs and the αs is also the same for all cases. But, when σ differs across cases – there is heteroskedasticity – the ratio also differs. As Hoetker (2004, p. 17) notes, "in the presence of even fairly small differences in residual variation, naive comparisons of coefficients [across groups] can indicate differences where none exist, hide differences that do exist, and even show differences in the opposite direction of what actually exists."

[2] This can be easily illustrated using Long and Freese's `fitstat` command, which is part of the `spost9` package available from Long's website. No matter what logit or probit model is estimated (e.g. you can add variables, subtract variables, change the variables completely), `fitstat` always reports a residual variance of 3.29 for logit models and 1.0 for probit.

We will illustrate this first by a series of hypothetical examples. Remember, σ is an adjustment factor for the residual variance. Therefore, σ is fixed at 1 for one group, and the σ for the other group reflects how much greater or smaller that group's residual variance is.

Case 1: Underlying alphas are equal, residual variances differ

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1, \sigma_0 = 1$
αs & σ for group 1	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1, \sigma_1 = 2$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	.5X1 + .5X2 + .5X3	$\beta_1 = \beta_2 = \beta_3 = .5$

In Case 1, the underlying αs are equal. But, because the residual variances differ, the βs will only be half as large for group 1 as for group 0. Naive comparisons of coefficients can indicate differences where none exist.

Case 2: Underlying alphas differ, residual variances differ

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1, \sigma_0 = 1$
αs & σ for group 1	2X1 + 2X2 + 2X3	$\alpha_1 = \alpha_2 = \alpha_3 = 2, \sigma_1 = 2$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$

In Case 2, the αs are twice as large in group 1 as in group 0. But, because the residual variances also differ, the βs are the same. Differences in residual variances obscure the differences in the underlying effects. Naive comparisons of coefficients can hide differences that do exist.

Case 3: Underlying alphas differ, residual variances differ even more

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1, \sigma_0 = 1$
αs & σ for group 1	2X1 + 2X2 + 2X3	$\alpha_1 = \alpha_2 = \alpha_3 = 2, \sigma_1 = 3$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	2/3X1 + 2/3X2 + 2/3X3	$\beta_1 = \beta_2 = \beta_3 = 2/3$

In Case 3, the αs are again twice as large in group 1 as in group 0. But, because of the large differences in residual variances, the βs are smaller for group 0 than group 1. Differences in residual variances make it look like the Xs have smaller effects on group 1 when really the effects are larger. Naive comparisons of coefficients can even show differences in the opposite direction of what actually exists.

To think of the problem another way, the βs that are estimated are basically standardized coefficients, and hence when doing cross-group comparisons we encounter problems that are very similar to those that occur when comparing standardized coefficients for different groups in OLS regression. Since coefficients are always scaled so that the residual variance is the same no matter what variables are in the model, the scaling of coefficients will differ across groups if the residual variances are different, making cross-group comparisons of effects invalid.

The heterogeneous choice model provides us with a means for dealing with these problems. With this model, σ can differ across cases, hence correcting for heteroskedasticity. The heterogeneous choice model accomplishes this by estimating two equations: one for the determinants of the outcome, or choice, and another for the determinants of the residual variance. For an ordered variable Y with M categories coded 1 to M[3], the heterogeneous choice model (using logit link) can be written as

$$P(Y_i > j) = \text{invlogit}\left(\frac{x_i\beta - \kappa_j}{\exp(z_i\gamma)}\right) = \text{invlogit}\left(\frac{x_i\beta - \kappa_j}{\sigma_i}\right), \text{ j} = 1, 2, ..., \text{M-1}, \qquad (1)$$

where

$$\text{invlogit}(x) = \text{inverse logit function of x} = \exp(x)/[1+\exp(x)],$$
$$\exp(z_i\gamma) = \exp(\ln(\sigma_i)) = \sigma_i \;,$$
$$\kappa_0 = -\infty \text{ and } \kappa_M = \infty$$

In the above formula,

- The numerator is referred to as the choice equation, while the denominator is the variance equation. These are also referred to as the location and scale equations. Neither equation includes a constant.
- the logit link function is used, but others are possible, such as probit, complementary log-log, log-log and cauchit.
- x is a vector of values for the ith observation. The x's are the explanatory variables and are said to be the determinants of the choice, or outcome.
- z is a vector of values for the ith observation. The z's can define groups with different error variances in the underlying latent variable, e.g. the z's might include dummy variables for gender or race. But, the z's can also include continuous variables that are related to the error variances, e.g. as income increases, the error variances may increase. The z's and x's need not include any of the same variables, although they can. Note that, when the z's all equal 0, $\sigma_i = 1$.
- β and γ are vectors of coefficients. They show how the x's affect the choice and the z's affect the variance (or more specifically, the log of σ; estimating the log of σ guarantees that σ itself will always have a positive value).
- κs are the cutpoints. As is the case with `logit` and `ologit`, when the dependent variable is a 0-1 dichotomy, the model can be rewritten to add a constant (β_0) rather than subtract a cutpoint. The end result is the same because the cutpoint and constant are opposite in sign.

When $\sigma_i = 1$ for all cases and links logit or probit are used, the heteregenous choice model becomes the same as the ordered logit or probit models estimated by `ologit` and `oprobit`. When the dependent variable is a dichotomy and the link is probit, the heterogeneous choice model becomes the same as the heteroskedastic probit model estimated by `hetprob` (except that `hetprob` uses an intercept rather than a cutpoint.)

[3] The actual coding does not matter so long as the categories are ordered, e.g. Y could be coded -2 to 2, or Y could be a dichotomy coded 0-1.

3 Empirical Examples

A series of empirical examples will help to illustrate the utility of heterogeneous choice models and the capabilities of the `oglm` program. These examples require that Richard Williams' `oglm` and `gologit2` routines and Ben Jann's (2005) `esttab` program (all available from SSC) be installed. The first three examples demonstrate the equivalencies between the heterogeneous choice model and two other models that have appeared in the literature: Allison's (1999) model for group comparisons and Hauser and Andrew's (2006) logistic response model with proportionality constraints (LRPC). The fourth example compares and contrasts heterogeneous choice models and generalized ordered logit models as a means for dealing with violations of assumptions in the ordered logit model. The final two examples deal with practical issues in estimating and interpreting heterogeneous choice models. They illustrate (a) how to interpret coefficients, and how to make the interpretation more straightforward, (b) why likelihood ratio tests, when possible, are generally preferable to Wald tests for hypothesis testing, (c) the use of stepwise regression with the variance equation, and (d) the use of heterogeneous choice models as a diagnostic device even when the researcher does not want to use a heterogeneous choice model for the final analysis.

3.1 Example 1: Allison's Model of Group Comparisons

Allison (1999) analyzes a data set of 301 male and 177 female biochemists[4]. The units of analysis are person-years rather than persons. Each person has one record for each year they were an assistant professor, for up to ten years; once a person achieves tenure no further records are added. This results in 1,741 person-years for men and 1,056 person-years for women. The dependent variable in his analysis, `tenure`, is promotion to associate professor, coded 1 if the person was promoted in that year, 0 otherwise. For the independent variables, `year` is the number of years since the beginning of the assistant professorship, `yearsq` is years squared, `select` is a measure of the selectivity of the colleges where scientists received their bachelor's degrees, `articles` is the cumulative number of articles published by the end of each person-year, and `prestige` is a measure of prestige of the department in which scientists were employed. The primary substantive interest of the analysis is whether the determinants of tenure differ for men (group 0) and women (group 1). Williams (forthcoming) provides an extended discussion of the strengths and weaknesses of Allison's proposed strategy and raises concerns that researchers should be aware of when considering Allison's suggestions. For the purposes of this paper, we will show how Allison's analysis can be replicated using specialized code presented in his appendix, and then show how the same analysis can be easily done using `oglm`. We begin by showing how to reproduce the estimates presented in his Tables 1 and 2.

[4] The data were originally collected by J. Scott Long (Long, Allison and McGinnis 1993) and are available on his website.

```
. capture program drop palogit

. program define palogit
  1.          version 6
  2.          args lnf theta delta
  3.          quietly replace `lnf' = ///
>                  $ML_y1*`theta'*(1+`delta') - ///
>                  ln(1+exp(`theta'*(1+`delta')))
  4. end

. use "http://www.indiana.edu/~jslsoc/stata/spex_data/tenure01.dta", clear
(Gender differences in receipt of tenure (Scott Long 06Jul2006))

. * Allison restricted the sample to the first 10 years as an Assistant Prof
. keep if year <= 10
(148 observations deleted)

. * men only
. quietly logit tenure female year yearsq select articles prestige if female==0

. quietly estimates store male

. * females only
. quietly logit tenure female year yearsq select articles prestige if female==1

. quietly estimates store female

. * Pooled model with delta added
. quietly ml model lf palogit (tenure = year yearsq select articles prestige female )
(delta: female,nocons), maximize

. quietly est store delta1

. * pooled model with delta and interaction term
. quietly ml model lf palogit (tenure = year yearsq select articles prestige female
f_articles) (delta: female,nocons), maximize

. quietly est store delta2

. esttab male female delta1 delta2, stats(N ll) mtitle
```

	(1) male	(2) female	(3) delta1	(4) delta2
main				
year	1.909***	1.408***	1.910***	1.838***
	(8.92)	(5.47)	(9.56)	(9.06)
yearsq	-0.143***	-0.0956***	-0.140***	-0.134***
	(-7.70)	(-4.36)	(-8.24)	(-7.89)
select	0.216***	0.0551	0.182***	0.170**
	(3.51)	(0.77)	(3.45)	(3.29)
articles	**0.0737*** **	**0.0340** **	0.0635***	0.0720***
	(6.37)	(2.69)	(6.22)	(6.31)
prestige	-0.431***	-0.371*	-0.446***	-0.420***
	(-3.96)	(-2.38)	(-4.60)	(-4.37)
female			-0.939*	-0.378
			(-2.53)	(-0.84)
f_articles				-0.0305
				(-1.63)
_cons	-7.680***	-5.842***	-7.491***	-7.365***
	(-11.27)	(-6.75)	(-11.36)	(-11.25)
delta				
female			-0.261*	-0.163
			(-2.41)	(-1.19)
N	1741	1056	2797	2797
ll	-526.5	-306.2	-836.3	-835.1

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Allison starts by estimating separate logistic regression models for men and women. Of key interest is the effect of articles: the effect is twice as great for men (.0737) as it is for women (.0340) and separate tests reveal that this difference is statistically significant. Allison (p. 188) says “If accurate, this difference suggests that men get a greater payoff from their published work than do females, a conclusion that many would find troubling.”

Allison notes, however, that differences in effects could be artifacts of differences in residual variability. There are reasons for believing that women have more heterogeneous career patterns than men, especially during the period covered by his data. “Hence, unmeasured variables affecting the chances of promotion may be more important for women than for men. That difference could explain why the coefficients… are larger for men than for women.” (Allison p. 190). Using our earlier terminology, Allison is arguing that this may fall under Case I, Underlying Alphas are equal but the residual variances differ.

To examine this possibility, Allison uses the `palogit`[5] program presented in the appendix of his paper to estimate a single model for men and women that includes a new parameter he calls δ. δ adjusts for the differences in residual variability between men and women. Allison's model can be written as

$$P(Y_i = 1) = \text{invlogit}\ \left((x_i\beta + \beta_0) * (1 + \delta G_i)\right) = \text{invlogit}\ \left(\frac{(x_i\beta + \beta_0)}{1/(1+\delta G_i)}\right) = \text{invlogit}\ \left(\frac{(x_i\beta + \beta_0)}{\sigma_i}\right) \quad (2)$$

where X is a vector of explanatory variables, G_i is a grouping variable (in this case female) coded either 1 or 0, and $\delta > -1$. The traditional logistic regression model is a special case of the above, where $\delta = 0$. Under Allison's approach, the σ for group 0 equals 1 and the σ for group 1 equals $1/(1 + \delta)$.

The value of -.261 for δ presented above in the model labeled delta1 means that the residual standard deviation for men is 26% smaller than it is for women (or equivalently, the standard deviation for women is 1/(1 -.261) = 1.353 times larger than it is for men). In his final model, labeled delta2 above, Allison includes both δ and an interaction term (f_articles = female * articles). Having controlled for differences in residual variability, the interaction term is not significant. Allison concludes (p. 196) that "The apparent difference in the coefficients for article counts… does not necessarily reflect a real difference in causal effects. It can be readily explained by differences in the degree of residual variation between men and women."

As noted, Allison used specialized code to estimate his model. However, as Williams (forthcoming) points out, although he did not label it as such, Allison actually estimated a heteroskedastic logit model, which in turn is a special case of a heterogeneous choice model: the link is logit, the dependent variable is a 0-1 dichotomy and the variance equation is limited to a single 0-1 dichotomous grouping variable that also appears in the choice equation. Under these conditions, the heterogeneous choice model simplifies to

$$P(Y_i = 1) = \text{invlogit}\left(\frac{x_i\beta - \kappa}{\exp(G_i\gamma)}\right) = \text{invlogit}\left(\frac{x_i\beta - \kappa}{\exp(\ln(\sigma_i))}\right) = \text{invlogit}\left(\frac{x_i\beta - \kappa}{\sigma_i}\right) \quad (3)$$

Note the similarities between the formulas for the heterogeneous choice model (equation 3) and for Allison's (equation 2). In Allison's approach, a constant (β_0) is added in the numerator while in the heterogeneous choice model a cutpoint (κ) is subtracted. This is a trivial difference because one number is the negative of the other. In both models the numerator is divided by σ_i. The main difference is how the two methods arrive at their estimate of σ_i. Neither method estimates σ_i directly, but σ_i is easily computed from the numbers they do estimate. The heterogeneous choice model estimates the log of σ_i, which guarantees that σ_i will be a positive number. Under Allison's approach, δ is estimated, where δ is the difference between the values

[5] Allison called the program `glogit` but we have renamed it since a program by that name already exists in Stata. We have also corrected a typographical error that appeared in the program. The complete sequence of models and tests suggested by Allison can be estimated using Glenn Hoetker's `complogit` program, available from SSC.

of σ in the two groups. Not surprisingly, then, `oglm` can easily reproduce the estimates from Allison's model.

```
. * oglm replication of Allison's delta models
. quietly oglm tenure year yearsq select articles prestige female, het(female)
store(oglm1)

. quietly oglm tenure year yearsq select articles prestige female f_articles,
het(female) store(oglm2)

. esttab delta1 oglm1 delta2 oglm2, stats(N ll) mtitle
```

```
----------------------------------------------------------------------------
                      (1)             (2)             (3)             (4)
                   delta1           oglm1          delta2           oglm2
----------------------------------------------------------------------------
main
year                1.910***        1.910***        1.838***        1.838***
                   (9.56)          (9.56)          (9.06)          (9.06)

yearsq             -0.140***       -0.140***       -0.134***       -0.134***
                  (-8.24)         (-8.24)         (-7.89)         (-7.89)

select              0.182***        0.182***        0.170**         0.170**
                   (3.45)          (3.45)          (3.29)          (3.29)

articles           0.0635***       0.0635***       0.0720***       0.0720***
                   (6.22)          (6.22)          (6.31)          (6.31)

prestige           -0.446***       -0.446***       -0.420***       -0.420***
                  (-4.60)         (-4.60)         (-4.37)         (-4.37)

female             -0.939*         -0.939*         -0.378          -0.378
                  (-2.53)         (-2.53)         (-0.84)         (-0.84)

f_articles                                        -0.0305         -0.0305
                                                  (-1.63)         (-1.63)

_cons              -7.491***                       -7.365***
                 (-11.36)                        (-11.25)
----------------------------------------------------------------------------
delta
female             -0.261*                         -0.163
                  (-2.41)                         (-1.19)
----------------------------------------------------------------------------
lnsigma
female                              0.302*                          0.177
                                   (2.07)                          (1.09)
----------------------------------------------------------------------------
cut1
_cons                               7.491***                        7.365***
                                  (11.36)                         (11.25)
----------------------------------------------------------------------------
N                    2797            2797            2797            2797
ll                 -836.3          -836.3          -835.1          -835.1
----------------------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

The models labeled oglm1 and oglm2 correspond to Allison's models delta1 and delta2. The log likelihoods for the corresponding models are identical, as are the coefficients for the variables in the choice equation. Similar to the difference between `logit` and `ologit` with a binary dependent variable, `oglm` reports cutpoints rather than constants, and the cutpoints equal the negative of the constants. The main, less obvious difference in the results is that Allison's model reports δ while `oglm` reports γ, which in this case is $\ln(\sigma_{Group1})$. As the following statements show, these results are algebraically equivalent: $\delta = (1 - \exp(\gamma))/\exp(\gamma) = (1 - \sigma_{Group1})/ \sigma_{Group1}$.

```
. * Convert oglm's lnsigma parameters into Allison's delta
. forval i = 1/2 {
  2.         display
  3.         quietly est restore oglm`i'
  4.         display as yellow "Model oglm`i': sigma = exp([lnsigma]female) = exp("
%5.3f `=[#2]female' ") = " %5.3f exp([#2]female)
  5.         display as yellow "Model delta`i': Allison's delta = (1 - sigma)/ sigma =
  6.         nlcom (1 - exp([lnsigma]female)) / exp([lnsigma]female)
  7. }

Model oglm1: sigma = exp([lnsigma]female) = exp(0.302) = 1.353
Model delta1: Allison's delta = (1 - sigma)/ sigma =

       _nl_1:  (1 - exp([lnsigma]female)) / exp([lnsigma]female)

------------------------------------------------------------------------------
      tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.2608323   .1080501    -2.41   0.016    -.4726065   -.0490581
------------------------------------------------------------------------------

Model oglm2: sigma = exp([lnsigma]female) = exp(0.177) = 1.194
Model delta2: Allison's delta = (1 - sigma)/ sigma =

       _nl_1:  (1 - exp([lnsigma]female)) / exp([lnsigma]female)

------------------------------------------------------------------------------
      tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.1625714   .1362569    -1.19   0.233    -.4296301    .1044872
------------------------------------------------------------------------------
```

The oglm1 model says that the standard deviation of the residuals is 1.35 times larger for women than men, while the delta1 model makes the equivalent statement that the standard deviation for men is 26% smaller than it is for women. In the oglm2 model, the standard deviation is 1.194 times larger for women, which is the same as saying that the standard deviation for men is 16.25% smaller.

While either Allison's code or `oglm` can be used for this problem, there are several advantages to using `oglm`. `oglm` allows for both ordinal and binary dependent variables, and the variance equation is not limited to a single binary dependent variable. `oglm` has several other powerful features which we describe later, such as the ability to obtain predicted probabilities. Finally, the use of `oglm` makes it clear that the model estimated falls within the broader class of heterogeneous choice/location scale models that have already been well-documented in the literature.

3.2 Example 2: Hauser and Andrew's LRPC and LRPPC models

Mare (1980) applied a logistic response model to school continuation. Contrary to prior supposition, Mare's estimates suggested the effects of some socioeconomic background variables declined across six successive transitions including completion of elementary school through entry into graduate school. Hauser & Andrew (2006) replicate & extend Mare's analysis using the same data he did, the 1973 Occupational Changes in a Generation (OCG) survey data. Rather than analyzing each educational transition separately as Mare did, Hauser & Andrew estimate a single model across all educational transitions. They take the original data set of 21,682 white men and restructure it into 88,768 person-transition records. For example, somebody who completed the first three educational transitions would have four records. On the first three records, the dependent variable, `outcome`, would be coded 1 because the person made the transition, while on the record for the uncompleted 4th transition the dependent variable would be coded 0. The person would have no records for the 5th and 6th transitions because you cannot make those transitions if you haven't made the 4th. To each record they also added variables `trans1-trans6`, each of which is coded 1 if the record is from the transition in question, 0 otherwise (e.g. `trans3` is coded 1 for each person-transition record where the individual has completed the 2nd transition and is now eligible to complete the 3rd; otherwise `trans3` is coded 0).

Hauser and Andrew argue that the relative effects of some (but not necessarily all) background variables are the same at each transition, and that multiplicative scalars express proportional change in the effect of those variables across successive transitions. Specifically, Hauser & Andrew estimate two new types of models. We primarily focus on the first of these, the *logistic response model with proportionality constraints* (LRPC).

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_k \beta_k X_{ijk}\,,\ \ j = 1, 2, \ldots, 6 \qquad (4)$$

The λ_j introduce proportional increases or decreases in the βk across transitions; thus the LRPC model implies proportional changes in main effects across transitions. Instead of having to estimate a different set of betas for each transition, a single set of betas is estimated, along with one λ_j proportionality factor for each of the J = 6 transitions (λ_1 is constrained to equal 1). The proportionality constraints would hold if, say, the coefficients for the 2nd transition were all 2/3 as large as the corresponding coefficients for the first transition, the coefficients for the 3rd transition were all half as large as for the first transition, etc. Put another way, if the model holds, the items can be viewed as forming a composite scale. If it is valid, the model is both parsimonious and substantively interesting.

Hauser & Andrew note, however, that "one cannot distinguish empirically between the hypothesis of uniform proportionality of effects across transitions and the hypothesis that group differences between parameters of binary regressions are artifacts of heterogeneity between groups in residual variation." (p. 8). Similarly, Mare (2006, p.32) points out that "the constants of proportionality, λj , are estimable, but their values incorporate both differences across equations in the effects of the regressors and also differences in the variances of the underlying dependent variables."

Indeed, even though the rationales behind the models are totally different, the heterogeneous choice model estimated by `oglm` produces a fit identical to the LRPC model estimated by Hauser and Andrew. The models are algebraically equivalent. Informally, we note that, in the heterogeneous choice model (equation 1), the Xβ's are divided by σs, while in the LRPC (equation 4) the Xβ's are multipled by λs. This suggests that the λs of the LRPC model are simply the reciprocals of the σs in the heterogeneous choice model. To illustrate this empirically, we reproduce Hauser & Andrew's LRPC using a modified version of the Stata program presented in their appendix. The main difference is that their program estimated (and their paper reported) $\lambda_j - 1$, whereas our program estimates λ_j directly[6]. We then show the corresponding code using `oglm`.

```
. capture program drop lrpc01

. program define lrpc01
  1.         tempvar theta
  2.         version 8
  3.         args lnf betas intercepts lambdas
  4.         gen double `theta' = `intercepts' + (`lambdas' * `betas')
  5.         quietly replace `lnf' = ln(exp(`theta')/(1+exp(`theta'))) if $ML_y1==1
  6.         quietly replace `lnf' = ln(1/(1+exp(`theta'))) if $ML_y1==0
  7. end

. * The next line constrains lamda 1 to be 1
. constraint 1 [lambdas]trans1 = 1

. ml model lf lrpc01 ///
>         (betas: dunc sibsttl9 ln_inc_trunc edhifaom edhimoom broken farm16 south,
nocons) ///
>         (intercepts: outcome = trans1 trans2 trans3 trans4 trans5 trans6, nocons)
///
>         (lambdas: trans1 trans2 trans3 trans4 trans5 trans6, nocons)  ///
>          , maximize constraint(1) nolog

. ml display
```

[6] Also, our code uses double- rather than single-precision, which makes the routine run faster and give more accurate results. We have also made changes to make the program and output easier to read.

```
                                                  Number of obs   =      88768
                                                  Wald chi2(8)    =    2191.33
Log likelihood = -33529.654                       Prob > chi2     =     0.0000

 ( 1)  [lambdas]trans1 = 1
------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
betas        |
        dunc |   .2751199   .0130478    21.09   0.000     .2495468    .3006931
    sibsttl9 |  -.1744805   .0072242   -24.15   0.000    -.1886396   -.1603214
ln_inc_trunc |    .538349   .0216585    24.86   0.000     .4958992    .5807989
    edhifaom |   .0942193   .0067319    14.00   0.000     .0810249    .1074136
    edhimoom |   .1470294   .0068439    21.48   0.000     .1336156    .1604432
      broken |  -.2778074   .0524071    -5.30   0.000    -.3805234   -.1750914
      farm16 |  -.1634613   .0427207    -3.83   0.000    -.2471924   -.0797303
       south |  -.1850324    .037429    -4.94   0.000    -.2583918    -.111673
-------------+----------------------------------------------------------------
intercepts   |
      trans1 |   .5622383   .0691995     8.12   0.000     .4266097    .6978669
      trans2 |   .7709552   .0639569    12.05   0.000      .645602    .8963084
      trans3 |  -.1755367   .0539172    -3.26   0.001    -.2812126   -.0698609
      trans4 |  -1.880549   .0568775   -33.06   0.000    -1.992027   -1.769071
      trans5 |  -.9066807   .0658453   -13.77   0.000    -1.035735   -.7776262
      trans6 |  -.4451361   .0955024    -4.66   0.000    -.6323175   -.2579548
-------------+----------------------------------------------------------------
lambdas      |
      trans1 |          1          .        .       .            .           .
      trans2 |   .7479284   .0260956    28.66   0.000      .696782    .7990748
      trans3 |   .5880246   .0190159    30.92   0.000      .550754    .6252951
      trans4 |   .5442034   .0174114    31.26   0.000     .5100777    .5783291
      trans5 |   .2055068   .0146818    14.00   0.000      .176731    .2342825
      trans6 |   .0923081   .0193412     4.77   0.000        .0544    .1302162
------------------------------------------------------------------------------
```

In the corresponding `oglm` code, all of the variables in Hauser and Andrew's betas and intercepts equation are included in `oglm`'s choice equation (except for `trans1`, since its inclusion would result in perfect multicollinearity). The variables in their lambdas equation are included in `oglm`'s heteroskedasticity equation.

```
. oglm outcome dunc sibsttl9 ln_inc_trunc edhifaom edhimoom broken farm16 south trans2
trans3 trans4 trans5 trans6 , het(trans2 trans3 trans4 trans5 trans6) store(olrpc)

Heteroskedastic Ordered Logistic Regression       Number of obs   =      88768
                                                  LR chi2(18)     =   26602.23
                                                  Prob > chi2     =     0.0000
Log likelihood = -33529.654                       Pseudo R2       =     0.2840

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
outcome      |
        dunc |   .2751199   .0130478    21.09   0.000     .2495466    .3006931
    sibsttl9 |  -.1744805   .0072242   -24.15   0.000    -.1886396   -.1603213
ln_inc_trunc |   .5383488   .0216585    24.86   0.000     .4958989    .5807987
    edhifaom |   .0942192   .0067319    14.00   0.000     .0810249    .1074136
    edhimoom |   .1470293   .0068439    21.48   0.000     .1336155    .1604431
      broken |  -.2778073   .0524071    -5.30   0.000    -.3805232   -.1750913
      farm16 |  -.1634613   .0427207    -3.83   0.000    -.2471923   -.0797303
       south |  -.1850324   .0374289    -4.94   0.000    -.2583918    -.111673
      trans2 |    .468548    .102289     4.58   0.000     .2680652    .6690308
      trans3 |  -.8607577   .0742938   -11.59   0.000    -1.006371   -.7151445
      trans4 |  -4.017835   .0674156   -59.60   0.000    -4.149967   -3.885702
      trans5 |  -4.974159   .1330155   -37.40   0.000    -5.234865   -4.713454
      trans6 |  -5.384518    .345992   -15.56   0.000     -6.06265   -4.706387
-------------+----------------------------------------------------------------
lnsigma      |
      trans2 |   .2904472   .0348906     8.32   0.000     .2220628    .3588316
      trans3 |   .5309857   .0323389    16.42   0.000     .4676026    .5943688
      trans4 |   .6084307   .0319945    19.02   0.000     .5457226    .6711389
      trans5 |   1.582275   .0714418    22.15   0.000     1.442251    1.722298
      trans6 |    2.38262   .2095284    11.37   0.000     1.971952    2.793288
-------------+----------------------------------------------------------------
       /cut1 |  -.5622391   .0691998    -8.12   0.000    -.6978682   -.4266101
------------------------------------------------------------------------------
```

Several equivalencies between the two models are immediately apparent. Both models produce identical log likelihoods of -33529.654. The coefficients in Hauser and Andrew's betas equation have exact counterparts in `oglm`'s choice equation. Although less obvious, the coefficients in their intercepts and lambdas equations can also be easily computed using `oglm`. Following we present the necessary code and output for the first few calculations.

```
. * Convert oglm's choice equations trans1-trans6 into H & A's intercepts
. * Do trans1 separately - it is just the negative of oglm's cutpoint
. forval i = 1/6 {
  2.         display
  3.         display as yellow "trans`i':"
  4.         if "`i'"=="1" {
  5.                 display "oglm's cutpoint coefficient = [#3]_cons = " %6.3f
[#3]_cons
  6.                 display "Hauser and Andrew's intercept ="
  7.                 nlcom -[#3]_cons
  8.         }
  9.         else {
 10.                 display "oglm's choice coefficient = [#1]trans`i' = " %6.3f
[#1]trans`i'
 11.                 display "Hauser and Andrew's intercept ="
 12.                 nlcom ([#1]trans`i' - [#3]_cons) * 1/exp([lnsigma]trans`i')
 13.         }
 14. }
```

```
trans1:
oglm's cutpoint coefficient = [#3]_cons = -0.562
Hauser and Andrew's intercept =

       _nl_1:  -[#3]_cons

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .5622391   .0691998     8.12   0.000     .4266101    .6978682
------------------------------------------------------------------------------

trans2:
oglm's choice coefficient = [#1]trans2 =  0.469
Hauser and Andrew's intercept =

       _nl_1:  ([#1]trans2 - [#3]_cons) * 1/exp([lnsigma]trans2)

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .7709556   .0639569    12.05   0.000     .6456024    .8963088
------------------------------------------------------------------------------

trans3:
oglm's choice coefficient = [#1]trans3 = -0.861
Hauser and Andrew's intercept =

       _nl_1:  ([#1]trans3 - [#3]_cons) * 1/exp([lnsigma]trans3)

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.1755364   .0539172    -3.26   0.001    -.2812122   -.0698606
------------------------------------------------------------------------------

[Rest of output deleted]
```

The above results are identical to those reported by the LRPC. The complexity of the calculations reflects the fact that `oglm` uses a cutpoint, the LRPC does not multiply the intercepts by lambda while the heterogeneous choice model does divide them by sigma, etc. The computation of the lambda coefficients is more straightforward since, as suggested before, they are simply the reciprocals of the corresponding sigma parameters in `oglm`[7].

```
. * Convert oglm's lnsigma parameters into H & A's lambdas
. * Remember that lambda 1 is fixed at 1
. forval i = 2/6 {
  2.         display
  3.         display as yellow "trans`i':"
  4.         display "oglm's sigma = exp([lnsigma]trans`i') = exp(" %6.3f
`=[lnsigma]trans`i '' ")= " %9.3f exp([lnsigma]trans`i')
  5.         display "Hauser and Andrew's lambda = 1/sigma ="
  6.         nlcom 1/exp([lnsigma]trans`i')
  7. }
```

[7] In their paper, Hauser and Andrew reported $\lambda j_i - 1$ rather than λ_j. To get the same estimates, change the `nlcom` command to `nlcom 1/exp([lnsigma]trans`i') - 1`.

```
trans2:
oglm's sigma = exp([lnsigma]trans2) = exp( 0.290)=     1.337
Hauser and Andrew's lambda = 1/sigma =

       _nl_1:  1/exp([lnsigma]trans2)

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |    .747929   .0260957    28.66   0.000     .6967824    .7990757
------------------------------------------------------------------------------

trans3:
oglm's sigma = exp([lnsigma]trans3) = exp( 0.531)=     1.701
Hauser and Andrew's lambda = 1/sigma =

       _nl_1:  1/exp([lnsigma]trans3)

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .5880251   .0190161    30.92   0.000     .5507542    .6252959
------------------------------------------------------------------------------

[Rest of output deleted]
```

Hauser and Andrew also propose a less restrictive model, which they call the *logistic response model with partial proportionality constraints* (LRPPC):

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_{k=1}^{k'} \beta_k X_{ijk} + \sum_{k'+1}^{K} \beta_{jk} X_{ijk} \qquad (5)$$

This model maintains the proportionality constraints for some variables, while allowing the effects of other variables to freely differ across transitions. For example, Hauser & Andrew say the LRPPC could apply to Mare's analysis where effects of socioeconomic variables appear to decline across transitions while those of farm origin, one-parent family, and Southern birth vary in other ways.

The LRPPC model can also be easily estimated using `oglm`. As Hauser and Andrew show in their appendix, this model is estimated by adding interaction terms involving transitions and the variables whose effects are allowed to freely vary across transitions[8]. In `oglm`, this is accomplished by adding the interaction terms to the choice equation. The code is shown below.

```
*** H & A Model 6: An intercept for each transition, proportional effects of
* socioeconomic variables, interactions of broken, farm, and south with transition.
* This is the second hetero choice model (equivalent to H & A's LRPPC).
quietly oglm outcome trans2 trans3 trans4 trans5 trans6 broken farm16 south
trans2Xbroken trans2Xfarm16 trans2Xsouth trans3Xbroken trans3Xfarm16 trans3Xsouth
trans4Xbroken trans4Xfarm16 trans4Xsouth trans5Xbroken trans5Xfarm16 trans5Xsouth
trans6Xbroken trans6Xfarm16 trans6Xsouth dunc sibsttl9 ln_inc_trunc edhifaom edhimoom,
het(trans2 trans3 trans4 trans5 trans6) store(m6)
```

[8] In the `lrpc01` program presented here, these variables would be added to the equation labeled "intercepts". Since the equation is no longer limited to just intercepts, a more appropriate name might be something like "freebetas".

Having noted these equivalences, it is important to realize that the substantive implications and rationales of the models are very different. The LRPC and LRPPC say that effects differ across transitions by scale factors. The algebraically-equivalent heterogeneous choice model says that effects do not differ across transitions; they only appear to differ when you estimate separate models because the variances of residuals change across transitions. Empirically, there is no way to distinguish between the two. In any event, there can be little arguing that, at least in these data, the effects of SES relative to other influences decline across transitions. The only question is whether this is because the absolute effects of SES decline, or because the influences of other (omitted) variables go up.

3.3 Example 3: The Biochemist Data and the LRPC

We have considered Allison's model for group comparisons, the heterogeneous choice model, and Hauser and Andrew's LRPC and LRPPC. We return one more time to the biochemist data to further highlight the similarities and differences between these models.

Allison (1999, p. 200) says of his method that "One possible flaw is that the tests cannot detect departures from the null hypothesis (of no difference between groups) if all the true coefficients differ by a constant multiple across groups. That is because such uniform differences are attributed to unequal disturbance variances rather than to real differences in the coefficients." Conversely, as noted before, Hauser and Andrew (2006, p. 8) warn that, even if the LRPC fits the data, "one cannot distinguish empirically between the hypothesis of uniform proportionality of effects across transitions and the hypothesis that group differences between parameters of binary regressions are artifacts of heterogeneity between groups in residual variation."

To illustrate exactly what they mean, we will estimate both of their models using the biochemist data. We begin by once again presenting Allison's analysis of the data, using his own program.

```
. * Allison's original model, using his program & biochemist data
. ml model lf palogit (tenure = year yearsq select articles prestige female ) ///
>         (delta: female,nocons), maximize nolog
. ml display

                                                  Number of obs   =       2797
                                                  Wald chi2(6)    =     181.39
Log likelihood = -836.28235                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
        year |   1.909544   .1996935     9.56   0.000     1.518152    2.300936
      yearsq |  -.1396868   .0169425    -8.24   0.000    -.1728935   -.1064801
      select |   .1819201   .0526572     3.45   0.001     .0787139    .2851263
    articles |   .0635345    .010219     6.22   0.000     .0435055    .0835635
    prestige |  -.4462073    .096904    -4.60   0.000    -.6361356   -.2562791
      female |  -.9391901   .3705242    -2.53   0.011    -1.665404    -.212976
       _cons |  -7.490505   .6596628   -11.36   0.000     -8.78342    -6.19759
-------------+----------------------------------------------------------------
delta        |
      female |  -.2608321   .1080501    -2.41   0.016    -.4726064   -.0490579
------------------------------------------------------------------------------
```

We will now estimate Hauser and Andrew's LRPC model using these data. However, for reasons that will be apparent in a moment, this time we use Hauser and Andrew's original code, which estimated $\lambda_j - 1$ rather than λ_j.[9]

```
. capture program drop lrpc02
. * Hauser & Andrew's original LRPC program
. * Code has been made more efficient and readable,
. * but results are the same.
. program define lrpc02
  1.         tempvar theta
  2.         version 8
  3.         args lnf  intercepts lambdaminus1 betas
  4.         gen double `theta' = `intercepts' + `betas' + (`lambdaminus1' * `betas')
  5.         quietly replace `lnf' = ln(exp(`theta')/(1+exp(`theta'))) if $ML_y1==1
  6.         quietly replace `lnf' = ln(1/(1+exp(`theta'))) if $ML_y1==0
  7. end

. * Hauser & Andrews original LRPC parameterization used with Allison's data
. ml model lf lrpc02 ///
>         (intercepts: tenure = male female, nocons) ///
>         (lambdaminus1: female, nocons) ///
>         (betas: year yearsq select articles prestige, nocons), max nolog

. ml display

                                                  Number of obs   =       2797
                                                  Wald chi2(2)    =     180.60
Log likelihood = -836.28235                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
intercepts   |
        male |  -7.490506    .659663   -11.36   0.000    -8.783421    -6.19759
      female |   -6.23096   .6205867   -10.04   0.000    -7.447287   -5.014632
-------------+----------------------------------------------------------------
lambdaminus1 |
      female |  -.2608322   .1080501    -2.41   0.016    -.4726066   -.0490579
-------------+----------------------------------------------------------------
betas        |
        year |   1.909544   .1996936     9.56   0.000     1.518152    2.300936
      yearsq |  -.1396868   .0169425    -8.24   0.000    -.1728935   -.1064801
      select |   .1819201   .0526572     3.45   0.001     .0787139    .2851264
    articles |   .0635345    .010219     6.22   0.000     .0435055    .0835635
    prestige |  -.4462073    .096904    -4.60   0.000    -.6361356    -.256279
------------------------------------------------------------------------------
```

The similarities are obvious: other than the intercepts, which the two programs parameterize differently, the coefficient estimates are identical. Most critically, Allison's δ, which his program estimated and which he reported in his paper, is exactly identical to Hauser and Andrew's $\lambda - 1$, which their program estimated and which they reported in their paper. Hauser and Andrew's software is, in fact, a generalization of Allison's software for when there are two

[9] We have slightly modified their program to make it more efficient and readable, but the estimates are identical.

or more groups. It is not simply a matter of the two sets of results being algebraically equivalent to each other: they are (other than the intercepts) the exact same results.[10]

In short, Allison, and Hauser and Andrew, wrote almost the exact same program. But, the theoretical concerns that motivated their models and programs lead to radically different interpretations of the results. According to Allison's theory (and the theory behind the heterogeneous choice model) apparent differences in effects between men and women are an artifact of differences in residual variability. Once these differences are taken into account, there is no significant difference in the effect of articles across groups, implying there is no gender inequality in the tenure process.

Someone looking at these exact same numbers from the viewpoint of the LRPC, however, would conclude that the effect of articles (and every other variable for that matter) is 26 percent smaller for women than it is men. Those who believed that the LRPC was the theoretically correct model would likely conclude that there is substantial gender inequality in the tenure promotion process.

For any given problem, strong substantive arguments might be made for one perspective or the other. Researchers using any of these models should realize, however, that there is often if not always a radically different interpretation that, empirically, fits the data just as well.

3.4 Example 4: Heterogeneous choice versus generalized ordered logit models

Williams (2006) notes that the proportional odds/ parallel regressions/ parallel lines assumption of the ordered logit model is often violated[11]. He shows that generalized ordered logit models are one way of dealing with the problem. Williams (forthcoming) suggests that heterogeneous choice models may also be an attractive alternative. We now explicitly compare the two approaches.

Long and Freese (2006) present data from the 1977/1989 General Social Survey. 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),

[10] It is easy to show why the results are identical. As noted earlier, the LRPC's λs are the reciprocals of the heterogeneous choice model's σs. Thus, $\lambda_j - 1 = 1/\sigma_j - 1 = (1 - \sigma_j)/ \sigma_j$. Further, as was also noted earlier, Allison's $\delta = (1 - \sigma_1)/ \sigma_1$. Hence, the $\lambda_j - 1$that Hauser and Andrew estimated and reported is a generalization of Allison's δ to the more-than-two-groups case.

[11] As Williams (2006) notes, the parallel lines assumption goes by many different names. In Stata, Wolfe and Gould's (1998) `omodel` command calls it the *proportional odds* assumption, a terminology that is only appropriate when the logit link is used. Long and Freese's `brant` command refers to the *parallel regressions* assumption. Both SPSS's PLUM command (Norusis 2005) and SAS's PROC LOGISTIC (SAS Institute 2004) provide tests of what they call the *parallel lines* assumption. For consistency with other major statistical packages, `oglm` and `gologit2` also use the terminology *parallel lines*, but researchers should realize that others may use different but equivalent phrasings.

white (0 = nonwhite, 1 = white), age (measured in years), ed (years of education), and prst (occupational prestige scale). `ologit` yields the following results.

```
. use http://www.indiana.edu/~jslsoc/stata/spex_data/ordwarm2.dta, clear
(77 & 89 General Social Survey)

. ologit  warm yr89 male white age ed prst, nolog

Ordered logit estimates                           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   .0798988     6.56   0.000     .3673037    .6805013
        male |  -.7332997   .0784827    -9.34   0.000    -.8871229   -.5794766
       white |  -.3911595   .1183808    -3.30   0.001    -.6231815   -.1591374
         age |  -.0216655   .0024683    -8.78   0.000    -.0265032   -.0168278
          ed |   .0671728    .015975     4.20   0.000     .0358624    .0984831
        prst |   .0060727   .0032929     1.84   0.065    -.0003813    .0125267
-------------+----------------------------------------------------------------
       _cut1 |  -2.465362   .2389126          (Ancillary parameters)
       _cut2 |   -.630904   .2333155
       _cut3 |   1.261854   .2340179
------------------------------------------------------------------------------
```

Both Long and Freese (2006) and Williams (2006) use a Brant test to show that the assumptions of the ordered logit model are violated. But, the main problems seem to be with the variables yr89 and male. Williams (2006) shows that a generalized ordered logit model, estimated by `gologit2`, provides a superior fit while introducing only a few additional parameters. `gologit2` relaxes the parallel lines constraint for those variables that violate it (yr89 and male), while maintaining the constraint for others. Williams' paper discusses the model in detail, but his main results can be reproduced with the command

```
. gologit2  warm yr89 male white age ed prst, autofit lrf store(gologit2)
```

The model chi-square for the `gologit2` model is 338.30 with 10 d.f., a significant improvement over the ordered logit model (301.72 with 6 d.f.). At the same time, the `gologit2` model is much more parsimonious than a multinomial logit model, which has a model chi-square of 349.53 but requires 18 degrees of freedom. Williams therefore concludes (p. 58) that "`gologit2` can estimate models that are less restrictive than the parallel lines models estimated by `ologit` (whose assumptions are often violated) but more parsimonious and interpretable than those estimated by a non-ordinal method, such as multinomial logistic regression (i.e. `mlogit`)."

We will now consider whether a heterogeneous choice model might also be a reasonable alternative in this case. Both `gologit2` and the Brant test identified yr89 and male as the variables that violated the assumptions of the ordered logit model, so we include them in the variance equation.[12]

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


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
------------------------------------------------------------------------------
```

The Brant test identified male and yr89 as the most problematic variables for the original ordered logit model, and these variables have significant effects in both the choice and variance equations. Both the `gologit2` and `oglm` models provide a much better fit to the data than does the ordered logit model. From a purely empirical standpoint, cases can be made for both approaches:

```
. lrtest gologit2 oglm, stats force

Likelihood-ratio test                                 LR chi2(2)  =      7.28
(Assumption: oglm nested in gologit2)                 Prob > chi2 =    0.0263

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
        oglm |   2293    -2995.77   -2830.256     11     5682.513    5745.626
    gologit2 |   2293    -2995.77   -2826.618     13     5679.236    5753.825
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The models are not nested, but nonetheless we can note that the `gologit2` model produces a larger model chi-square (338.30 versus 331.03) but at the cost of 2 degrees of freedom. The BIC statistic favors the `oglm` model, while the AIC statistic leans slightly towards the `gologit2` model. Additional analyses (not shown) reveal that the predicted probabilities and marginal

---

[12] Stepwise selection (see example 6) also results in the variables yr89 and male being included in the variance equation.

effects for each model are very similar. Ergo, from a purely empirical standpoint, there is little reason for preferring one model over the other, and either clearly fits better than the ordered logit model.

From a substantive standpoint, the simplicity of the `oglm` model and the additional insights that are gained by adding only two parameters to the ordered logit model may be highly appealing. The negative coefficients in the variance equation reveal that men were less variable in their attitudes than were women, and that variability in attitudes toward working women declined across time. Both results seem both plausible and substantively interesting. Women, torn between traditional and new roles, may be more divided in their feelings toward working women. Consensus may have increased across time as the notion of women working became more socially acceptable and less divisive.

There is no guarantee that other examples will show an equally tight race between the `gologit2` and `oglm` models, and ultimately theoretical concerns should guide the choice between the two. Nonetheless, this example illustrates that, when the assumptions of the ordered logit model are violated, researchers may want to at least consider the possibility that a heterogeneous choice model is warranted.

3.5 Example 5: A trivial change with seemingly non-trivial implications

We now present a seemingly innocuous change to Allison's model that was presented in example 1. Instead of using the variable female (coded 1 if female, 0 if male) we use male (coded 1 if male, 0 if female). Most people would probably expect that such a trivial change would have no meaningful impact on the model – but the actual results seem to suggest otherwise.

```
. * Example 5:
. * As before, use female in the equations
. quietly oglm  tenure year yearsq select articles prestige female , het(female)
store(oglm_f)

. * Now use male instead
. quietly oglm  tenure year yearsq select articles prestige male , het(male)
store(oglm_m)

. * Do females only logit model again, using oglm
. quietly oglm  tenure year yearsq select articles prestige if female, store(females)

. * Do males only logit model again, using oglm
. quietly oglm  tenure year yearsq select articles prestige if male, store(males)

. esttab oglm_f oglm_m males females, stats(N ll chi2 df_m) mtitle
```

	(1) oglm_f	(2) oglm_m	(3) males	(4) females
tenure				
year	1.910***	1.411***	1.909***	1.408***
	(9.56)	(7.17)	(8.92)	(5.47)
yearsq	-0.140***	-0.103***	-0.143***	-0.0956***
	(-8.24)	(-6.68)	(-7.70)	(-4.36)
select	0.182***	0.134***	0.216***	0.0551
	(3.45)	(3.41)	(3.51)	(0.77)
articles	0.0635***	0.0470***	0.0737***	0.0340**
	(6.22)	(5.80)	(6.37)	(2.69)
prestige	-0.446***	-0.330***	-0.431***	-0.371*
	(-4.60)	(-4.07)	(-3.96)	(-2.38)
female	-0.939*			
	(-2.53)			
male		0.694***		
		(3.69)		
lnsigma				
female	0.302*			
	(2.07)			
male		-0.302*		
		(-2.07)		
cut1				
_cons	7.491***	6.231***	7.680***	5.842***
	(11.36)	(10.04)	(11.27)	(6.75)
N	2797	2797	1741	1056
ll	-836.3	-836.3	-526.5	-306.2
chi2	413.1	413.1	302.4	114.6
df_m	7	7	5	5

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Comparing our earlier model that used female and the new model that substitutes male, the log likelihoods, model chi-squares and degrees of freedom are all the same. In the variance equation, as we would expect, the coefficient for male is opposite in sign to what it was for female. Perhaps surprisingly, however, all the coefficients in the choice equation are different from before, as are the z values. This raises the troubling possibility that seemingly innocuous differences in variable coding could lead to different conclusions about whether or not the effect of a variable is statistically significant.

Why does this occur, and what should be done about it? This is very similar to the situation that occurs when a regression model includes both main effects and interaction effects. For example, if a model includes x1, x2, and x1*x2, then the coefficient for x1 reflects the effect of x1 when x2 equals zero. Further, the t or z value for x1 tests whether the effect of x1 differs from zero when x2 = 0; even if the effect of x1 is insignificant when x2 = 0, it may be significant for other values of x2.

Put another way, we can think of the coefficients in the choice equation as being the coefficients for a group where $\sigma = 1$, and hence the log of $\sigma = 0$. The log of σ will equal 0 when all the variables in the variance equation have a value of zero. Compare the coefficients in the first model, where males are coded 0, with the coefficients in the model for males only, and note how similar they are. Similar points hold for the second model that uses the variable male and females are coded 0, and the last model for females only.

A very important implication of the above is that z values and Wald tests should generally NOT be relied on for hypothesis testing involving variables in the choice equation – or at least, if they are used, researchers need to be clear on what hypotheses are being tested. As the examples show, the z values in the choice equation are not invariant across arbitrary changes in the coding of the variance equation variables, e.g. the z value for prestige is -4.60 when female is used in the model but only -4.07 when male is used instead[13]. Particularly in borderline situations, such differences could lead to different conclusions as to whether or not the effect of a variable was statistically significant. Likelihood ratio tests, luckily, do not have this problem. To illustrate this point, we will do LR tests for the effect of prestige, using first female and then male in the models.

```
. * Test prestige under the male versus female models
. * Female is in the model:
. quietly oglm  tenure (year yearsq select articles female), het(female) store(f1)

. quietly oglm  tenure (year yearsq select articles female prestige), het(female)
store(f2)

. lrtest f1 f2, stats

Likelihood-ratio test                                 LR chi2(1)  =     22.34
(Assumption: f1 nested in f2)                         Prob > chi2 =    0.0000

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
          f1 |   2797   -1042.828   -847.4507      7     1708.901    1750.456
          f2 |   2797   -1042.828   -836.2824      8     1688.565    1736.055
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

. * Male is in the model:
. quietly oglm  tenure (year yearsq select articles male), het(male) store(m1)

. quietly oglm  tenure (year yearsq select articles male prestige), het(male)
store(m2)
```

[13] This may be part of the reason that the `hetprob` command in Stata does not allow the use of the `nestreg` prefix, which by default does Wald tests. An additional complication with `nestreg` is that, when Wald tests are used and a variable appears in both the choice and variance equations, both effects will be tested. When using the `nestreg` or `sw` prefix commands with `oglm`, it is strongly recommend that the `lr` (likelihood ratio) option be specified.

```
. lrtest m1 m2, stats

Likelihood-ratio test                                 LR chi2(1)  =     22.34
(Assumption: m1 nested in m2)                         Prob > chi2 =    0.0000

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
          m1 |   2797   -1042.828   -847.4507      7     1708.901    1750.456
          m2 |   2797   -1042.828   -836.2824      8     1688.565    1736.055
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

We see that the LR tests give the same value (22.34) regardless of whether male or female is used in the model.

Another implication of these results is that researchers may want to code the variables in the variance equation so that zero is a substantively meaning value. In the current examples, zero is meaningful in that it stands for one gender or the other. In other cases, however, zero may not even be a value that can occur in the data, e.g. no one may have an IQ score of zero. In such instances, researchers may want to consider centering the variables in the variance equation (i.e. subtract the mean from each case) so that a score of 0 on the log of sigma reflects an “average” person. Or, the zero point might be chosen to represent some other meaningful value, e.g. subtract 12 from years of education so that a score of 0 stands for a high school graduate. Again, this is similar to recommendations that are sometimes made for OLS regression models that include interaction effects. Such changes do not affect the fit of the model, but they may make it easier to interpret results.

3.6 Example 6: Using stepwise selection as a model building and diagnostic device

Stepwise selection procedures are often criticized for their atheoretical nature. But, as this example will show, stepwise selection can help to identify theoretically plausible alternative models that the researcher may wish to consider, and can also be used as a diagnostic device even when the researcher does not want to ultimately present a heterogeneous choice model.

Stepwise selection of variables is easily done in Stata via the use of the `sw` prefix command. With `oglm`, stepwise selection can be used for *either* the choice or variance equation. To do it for the variance equation, the `flip` option can be used to reverse the placement of the choice and variance equations in the command line. The variables in the choice equation can then be specified using the `eq2` option. Using the biochemist data and stepwise selection for the variance equation produces a somewhat different model than the one Allison proposed.

```
. sw, pe(.01) lr: oglm  tenure female year yearsq select  articles prestige,
eq2(female  year yearsq select  articles prestige ) flip store(sw1)
LR test               begin with empty model
p = 0.0000 <  0.0100  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     -.759368   -.0764838
        year |   2.108752   .2486633     8.48   0.000     1.621381    2.596123
      yearsq |  -.1542213   .0208579    -7.39   0.000    -.1951019   -.1133406
      select |   .1744644   .0598623     2.91   0.004     .0571364    .2917924
    articles |   .0628407   .0157851     3.98   0.000     .0319026    .0937789
    prestige |  -.6118689   .1307262    -4.68   0.000    -.8680877   -.3556502
-------------+----------------------------------------------------------------
lnsigma      |
    articles |    .030149   .0091448     3.30   0.001     .0122256    .0480724
-------------+----------------------------------------------------------------
       /cut1 |   7.959556   .7637106    10.42   0.000      6.46271    9.456401
------------------------------------------------------------------------------
```

As the above shows, in Allison's Biochemist data, the only variable that enters into the variance equation using `oglm`'s stepwise selection procedure is number of articles. A very plausible argument can be made for this: there may be little residual variability among those with few articles (with most getting denied tenure) but there may be much more variability among those with more articles (having many articles may be a necessary but not sufficient condition for tenure). Hence, while heteroskedasticity may be a problem with these data, it may not be for the reasons first thought.

It is important to realize, however, that apparent problems with heteroskedasticity in a model may actually reflect other problems with the model specification. Relevant variables may be omitted from the model; subgroup differences may be being ignored; and variables may need to be transformed in some way, e.g. logged or squared. In the present example, the number of articles ranges from 0 to 73. It may be that, at some point, additional articles have less effect or even a negative effect on the likelihood of getting tenure (e.g. somebody might have a lot of articles but they aren't that good.) One simple way to address such a possibility is to add articles^2 to the model[14].

```
. gen articles2 = articles^2
. oglm  tenure female year yearsq select   articles articles2 prestige, het(articles) store(sw2)
```

[14] We thank Maarten Buis for suggesting that we consider adding terms for nonlinear effects to the model.

```
Heteroskedastic Ordered Logistic Regression       Number of obs   =       2797
                                                  LR chi2(8)      =     439.77
                                                  Prob > chi2     =     0.0000
Log likelihood = -822.94311                       Pseudo R2       =     0.2109

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
tenure       |
      female |  -.3470778   .1470054    -2.36   0.018    -.6352031   -.0589526
        year |   1.764339   .2233366     7.90   0.000     1.326608    2.202071
      yearsq |  -.1282567   .0182644    -7.02   0.000    -.1640544   -.0924591
      select |   .1631087   .0503776     3.24   0.001     .0643704    .2618471
    articles |   .1481165   .0246791     6.00   0.000     .0997464    .1964866
   articles2 |   -.002716   .0008273    -3.28   0.001    -.0043374   -.0010945
    prestige |  -.4909742   .1124811    -4.36   0.000    -.7114332   -.2705152
-------------+----------------------------------------------------------------
lnsigma      |
    articles |   .0081942   .0095091     0.86   0.389    -.0104432    .0268316
-------------+----------------------------------------------------------------
       /cut1 |   7.375548   .6803437    10.84   0.000     6.042099    8.708997
------------------------------------------------------------------------------

. lrtest sw1 sw2, stats

Likelihood-ratio test                                  LR chi2(1)  =     11.74
(Assumption: m3 nested in m4)                          Prob > chi2 =    0.0006

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
          m3 |   2797   -1042.828   -828.8122      8     1673.624    1721.115
          m4 |   2797   -1042.828   -822.9431      9     1663.886    1717.313
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

As we see, adding articles^2 significantly improves fit and makes the coefficient in the variance equation insignificant. Hence, even if the researcher does not want to use stepwise selection as a model-building device or does not want to present a heterogeneous choice model, he or she may still wish to use stepwise selection to diagnose potential problems in the model which can then be addressed in other ways. Of course, researchers can also use theoretical reasons to identify those variables that might raise concerns about heteroskedasticity and specify the models themselves.

4 Other features of oglm

`oglm` has several other features that may make it useful to researchers. `oglm` supports multiple link functions, including logit (the default), probit, complementary log-log, log-log and cauchit. Several special cases of ordinal generalized linear models can also be estimated by oglm, including the parallel lines models of `ologit` and `oprobit` (where error variances are assumed to be homoskedastic), the heteroskedastic probit model of `hetprob` (where the dependent variable must be a dichotomy and the only link allowed is probit), the binomial generalized linear models of `logit`, `probit` and `cloglog` (which also assume homoskedasticity), as well as similar models that are not otherwise estimated by Stata. This makes oglm particularly useful for testing whether constraints on a model (e.g. homoskedastic errors) are justified, or for determining whether one link function is more appropriate for the data than are others.

Other features of oglm include support for linear constraints, making it possible, for example, to impose and test the constraint that the effects of x1 and x2 are equal. oglm works with several prefix commands, including by, nestreg, xi, svy and sw. Its predict command includes the ability to compute estimated probabilities. The actual values taken on by the dependent variable are irrelevant except that larger values are assumed to correspond to "higher" outcomes. Up to 20 outcomes are allowed. oglm was inspired by the SPSS PLUM routine but differs somewhat in its terminology, labeling of links, and the variables that are allowed when modeling heteroskedasticity.

5 The oglm command

5.1 Syntax

oglm supports many standard Stata options, which work the same way as they do with other Stata commands. Several other options are unique to or fine-tuned for oglm. The complete syntax is

```
oglm depvar [indepvars] [weight] [if exp] [in range] [,
        link(logit/probit/cloglog/loglog/cauchit) force lrforce store(name)
        constraints(clist) robust cluster(varname) level(#) or irr rrr eform hr log
        hetero(varlist) scale(varlist) eq2(varlist) hc ls flip maximize_options ]
```

oglm shares the features of all estimation commands; see help est. oglm typed without arguments redisplays previous results. The following options may be given when redisplaying results:

```
store or irr rrr hr eform level(#)
```

by, svy, nestreg, stepwise, xi and possibly other prefix commands are allowed; see help prefix.

fweights, iweights, and pweights are allowed; see help weights.

5.2 Options unique to or fine-tuned for oglm

link(*link*) specifies the link function to be used. The legal values are link(*logit*), link(*probit*), link(*cloglog*), link(*loglog*) and link(*cauchit*) which can be abbreviated as link(*l*), link(*p*), link(*c*), link(*ll*) and link(*ca*). link(*logit*) is the default if the option is omitted.

Users should keep in mind that programs differ in the names used for some links. Stata's loglog link corresponds to SPSS PLUM's cloglog link; and Stata's cloglog link is called nloglog in

SPSS. The following advice for choosing an appropriate link function is adapted from Norusis (2005, p. 84): Probit and logit models are reasonable choices when the changes in the cumulative probabilities are gradual. If there are abrupt changes, other link functions should be used. The log-log link may be a good model when the cumulative probabilities increase from 0 fairly slowly and then rapidly approach 1. If the opposite is true, namely that the cumulative probability for lower scores is high and the approach to 1 is slow, the complementary log-log link may describe the data.

`hetero(`*`varlist`*`)`, `scale(`*`varlist`*`)` and `eq2(`*`varlist`*`)` are synonyms (use only one of them) and can be used to specify the variables believed to affect heteroskedasticity in heterogeneous choice/ location-scale models. In such models the model chi-square statistic is a test of whether any of the choice/location parameters or the heteroskedasticity/scale parameters differ from zero; this differs from `hetprob`, where the model chi-square only tests the choice/location parameters. The more neutral-sounding `eq2(`*`varlist`*`)` alternative is provided because it may be less confusing when using the `flip` option.

`flip` causes the command-line placement of the location and scale variables to be reversed, i.e. what would normally be the choice/location variables will instead be the variance/scale variables, and vice- versa. This is primarily useful if you want to use the `sw` or `nestreg` prefix commands to do stepwise selection or hierarchical entry of the heteroskedasticity/scale variables. (Just be sure to keep straight which set of variables is which.) If you do this, use the likelihood ratio test options of `nestreg` or `sw`, because the default Wald tests may be wrong otherwise.

`hc` and `ls` affect how the equations are labeled. If `hc` is used, then, consistent with the literature on heterogeneous choice, the equations are labeled “choice” and “variance”. If `ls` is used, the equations are labeled “location” and “scale”, which is consistent with SPSS PLUM and other published literature. If neither option is specified, then the scale/heteroskedasticity equation is labeled “lnsigma”, which is consistent with other Stata programs such as `hetprob`.

`force` can be used to force `oglm` to issue only warning messages in some situations when it would normally give a fatal error. By default, the dependent variable can have a maximum of 20 categories. A variable with more categories than that is probably a mistaken entry by the user, e.g. a continuous variable has been specified rather than an ordinal one. But, if the dependent variable really is ordinal with more than 20 categories, `force` will let `oglm` analyze it (although other practical limitations, such as small sample sizes within categories, may keep it from coming up with a final solution.) Obviously, you should only use `force` when you are confident that you are not making a mistake. `trustme` can be used as a synonym for `force`.

`lrforce` forces Stata to report a Likelihood Ratio Statistic under certain conditions when it ordinarily would not. Some types of constraints can make a Likelihood Ratio chi-square test invalid. Hence, to be safe, Stata reports a Wald statistic whenever constraints are used. But, for many common sorts of constraints (e.g. constraining the effects of two variables to be equal) an LR chi- square statistic is probably appropriate. Note that the `lrforce` option will be ignored when robust standard errors are specified either directly or indirectly, e.g. via use of the `robust` or `svy` options. Use this option with caution.

`store(`*`name`*`)` causes the command `estimates store` *`name`* to be executed when `oglm` finishes. This is useful for when you wish to estimate a series of models and want to save the results. See `help estimates`. The store option may not work correctly when the `svy` prefix is used.

`log` displays the iteration log. By default it is suppressed.

`or` reports the estimated coefficients transformed to relative odds ratios, i.e., exp(b) rather than b; see [R] `ologit` for a description of this concept. Options `rrr`, `eform`, `irr` and `hr` produce identical results (labeled differently) and can also be used. It is up to the user to decide whether the exp(b) transformation makes sense given the link function used, e.g. it probably doesn't make sense when using the probit link.

`constraints(`*`clist`*`)` specifies the linear constraints to be applied during estimation. The default is to perform unconstrained estimation. Constraints are defined with the constraint command. constraints(1) specifies that the model is to be constrained according to constraint 1; constraints(1-4) specifies constraints 1 through 4; constraints(1-4,8) specifies 1 through 4 and 8.

5.3 Other standard Stata options supported by oglm

`robust cluster level`

5.5 Options available when replaying results

`store or irr rrr hr eform level(#)`

5.6 Options available for the predict command

`pr`, the default, calculates the predicted probabilities. If you do not also specify the `outcome()` option, you must specify k new variables, where k is the number of categories of the dependent variable. Say that you fitted a model by typing `oglm result x1 x2`, and result takes on three values. Then you could type `predict p1 p2 p3` to obtain all three predicted probabilities. If you specify the `outcome()` option, you must specify one new variable. Say that result takes on the values 1, 2, and 3. Typing `predict p1, outcome(1)` would produce the same p1.

`xb` calculates the linear prediction. You specify one new variable, for example, `predict linear, xb`. The linear prediction is defined, ignoring the contribution of the estimated cutpoints.

`sigma` calculates the standard deviation, also known as the scale. You specify one new variable, for example, `predict sigma, s`. If the model does not include an equation for heteroskedasticity then the predicted sigma value is missing for all cases.

`stdp` calculates the standard error of the linear prediction. You specify one new variable, for example, `predict se, stdp`.

`outcome(`*`outcome`*`)` specifies for which outcome the predicted probabilities are to be calculated. `outcome()` should contain either a single value of the dependent variable or one of #1, #2, ..., with #1 meaning the first category of the dependent variable, #2 the second category, etc.

`scores` calculates equation-level score variables.

6 Support for oglm

Richard Williams
Department of Sociology
University of Notre Dame
Richard.A.Williams.5@ND.Edu
http://www.nd.edu/~rwilliam/oglm/

7 Acknowledgements

The documentation and source code for several Stata commands (e.g. `ologit_p`) were major aids in developing the oglm documentation and in adding support for the `predict` command. Much of the code is adapted from *Maximum Likelihood Estimation with Stata, Third Edition*, by William Gould, Jeffrey Pitblado and William Sribney. SPSS's PLUM routine helped to inspire `oglm` and provided a means for double-checking the accuracy of the program. Joseph Hilbe, Mike Lacy, Maarten Buis, Glenn Hoetker and Rory Wolfe provided stimulating comments on this paper and/or on the development of `oglm`. Jeff Pitblado assisted with several difficult programming issues. J. Scott Long, Robert Hauser and Megan Andrew provided access to the data sets used in these analyses.

8 References

Allison, Paul. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research* 28(2): 186-208.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Hauser, Robert M. and Megan Andrew. 2006. Another Look at the Stratification of Educational Transitions: The Logistic Response Model with Partial Proportionality Constraints. *Sociological Methodology* 36(1):1-26.

Hoetker, Glenn. 2004. Confounded Coefficients: Extending Recent Advances in the Accurate Comparison of Logit and Probit Coefficients Across Groups. Working Paper, October 22, 2004. Retrieved March 21, 2006 (http://www.business.uiuc.edu/ghoetker/documents/Hoetker_comp_logit.pdf)

Jann, Ben. 2005. Making regression tables from stored estimates. *Stata Journal* 5: 288–308.

Long, J. Scott, Paul D. Allison, and Robert McGinnis. 1993. “Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity.” *American Sociological Review* 58:703-722.

Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata, 2nd Edition*. College Station, Texas: Stata Press.

Mare, Robert D. 1980. Social Background and School Continuation Decisions. *Journal of the American Statistical Association* 75:293–305.

_____. 2006. Response: Statistical Models of Educational Stratification—Hauser and Andrew's Models for School Transitions.” *Sociological Methodology* 36(1):27-37.

Norusis, Marija. 2005. *SPSS 13.0 Advanced Statistical Procedures Companion*. Upper Saddle River, New Jersey: Prentice Hall. See especially the chapter on SPSS PLUM, available on the web at http://www.norusis.com/pdf/ASPC_v13.pdf .

SAS Institute Inc. 2004. *SAS/STAT 9.1 User’s Guide.* Cary, NC: SAS Institute Inc.

Williams, Richard. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82.

_____. Forthcoming. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients across Groups. *Sociological Methods & Research.* A pre-publication version can be found at http://www.nd.edu/~rwilliam/oglm/RW_Hetero_Choice.pdf . Last accessed July 18, 2008.

Wolfe, Rory and William Gould. 1998. An approximate likelihood-ratio test for ordinal response models. *Stata Technical Bulletin* 42: 24-27. In *Stata Technical Bulletin Reprints*, vol 7, 199-204. College Station, TX: Stata Press.

Yatchew, Adonis and Zvi Griliches. Specification Error in Probit Models. 1985. *The Review of Economics and Statistics* 67(1):134-139.

About the Author

Richard Williams is Associate Professor and a former Chairman of the Department of Sociology at the University of Notre Dame. His teaching and research interests include Methods and Statistics, Demography, and Urban Sociology. His work has appeared in the *American Sociological Review*, *Social Forces*, *Stata Journal*, *Social Problems*, *Demography*, *Sociology of Education*, *Journal of Urban Affairs*, *Cityscape*, *Journal of Marriage and the Family*, and *Sociological Methods and Research*. His recent research, which has been funded by grants from the Department of Housing and Urban Development and the National Science Foundation, focuses on the causes and consequences of inequality in American home ownership. He is a frequent contributor to Statalist.