

# Estimating heterogeneous choice models with `oglm`

Richard Williams

Department of Sociology, University of Notre Dame, Notre Dame, IN

[Richard.A.Williams.5@ND.Edu](mailto:Richard.A.Williams.5@ND.Edu)

Last revised October 17, 2010 – Forthcoming in *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 author's Stata program `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 special cases of a heterogeneous choice model and alternative parameterizations of 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 (Williams 2009; Keele & Park 2006)

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 (2009) 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. The current paper takes a more applied approach, and illustrates how the author's Stata command `oglm` (Ordinal Generalized Linear Models<sup>1</sup>) 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 and alternative parameterizations of `oglm`'s heterogeneous choice model; yet despite these equivalencies, it is possible to interpret the results of these models in very different ways. The paper further argues that heterogeneous

---

<sup>1</sup> The name is slightly misleading in that `oglm` can also estimate the nonlinear models presented here.

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 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.

$$\begin{aligned} y &= 1 \text{ if } -\infty < y^* < \kappa_1, \\ y &= 2 \text{ if } \kappa_1 < y^* < \kappa_2, \\ y &= 3 \text{ if } \kappa_2 < y^* < \kappa_3, \\ y &= 4 \text{ if } \kappa_3 < y^* < \kappa_4, \\ y &= 5 \text{ if } \kappa_4 < y^* < +\infty \end{aligned}$$

The model for the underlying  $y^*$  can be written as

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

where the  $x$ 's are the explanatory variables, the  $\alpha$ 's are coefficients that give the effect of each  $x$  on  $y^*$ ,  $\varepsilon_i$  is a residual term often assumed to have either a logistic or normal(0, 1) distribution, and  $\sigma$  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  $\pi^2/3$  (as in logit) or 1 (as in probit)<sup>2</sup>. Further, because  $y^*$  is unobserved, we do not actually estimate the  $\alpha$ s. Rather, we estimate parameters called  $\beta$ s. As Allison (1999, citing Amemiya 1985:269) notes, the  $\alpha$ s and the  $\beta$ s are related this way:

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

This now leads us to a potential problem with the ordered logit/probit model. When  $\sigma$  is the same for all cases – residuals are homoskedastic – the ratio between the  $\beta$ s and the  $\alpha$ s is also the same for all cases. But, when  $\sigma$  differs across cases – there is heteroskedasticity – the ratio also differs (Allison 1999). 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.”

---

<sup>2</sup> 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 (i.e.  $\pi^2/3$ ) for logit models and 1.0 for probit.

We will illustrate this first by a series of hypothetical examples. Remember,  $\sigma$  is an adjustment factor for the residual variance. Therefore,  $\sigma$  is fixed at 1 for one group, and the  $\sigma$  for the other group reflects how much greater or smaller that group's residual variance is. In each example, the  $\alpha$ s and  $\sigma$  for group 0 are fixed at 1. For group 1, the values of the  $\alpha$ s and  $\sigma$  are systematically varied. We then see how cross-group comparisons of the  $\beta$ s, i.e. the parameters that are actually estimated in a logistic regression, are affected by differences in residual variability.

Case 1: Underlying alphas are equal, residual variances differ

	Group 0	Group 1
Model using $\alpha$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + 2\varepsilon_i$
Model using $\beta$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = .5x_{i1} + .5x_{i2} + .5x_{i3} + \varepsilon_i$

In Case 1, the underlying  $\alpha$ s all equal 1 in both groups. But, because the residual variance is twice as large for group 1 as it is for group 0, the  $\beta$ s are only 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

	Group 0	Group 1
Model using $\alpha$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = 2x_{i1} + 2x_{i2} + 2x_{i3} + 2\varepsilon_i$
Model using $\beta$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$

In Case 2, the  $\alpha$ s are twice as large in group 1 as in group 0. But, because the residual variances also differ, the  $\beta$ s for the two groups 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

	Group 0	Group 1
Model using $\alpha$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = 2x_{i1} + 2x_{i2} + 2x_{i3} + 3\varepsilon_i$
Model using $\beta$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = \frac{2}{3}x_{i1} + \frac{2}{3}x_{i2} + \frac{2}{3}x_{i3} + \varepsilon_i$

In Case 3, the  $\alpha$ s are again twice as large in group 1 as in group 0. But, because of the large differences in residual variances, the  $\beta$ s 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  $\beta$ s 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 (Duncan 1975). 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,  $\sigma$  can differ across cases, hence correcting for heteroskedasticity. The heterogeneous choice model accomplishes this by simultaneously estimating two equations: one for the determinants of the outcome, or choice, and another for the determinants of the residual variance. The choice equation can be written as

$$y_i^* = \sum_k x_{ik} \beta_k + \varepsilon_i \quad (1a)$$

The location/ choice equation gives the value of the underlying latent variable. In the above,  $x$  is a vector of  $k$  values for the  $i$ th observation. The  $x$ 's are the explanatory variables and are said to be the determinants of the choice, or outcome. The  $\beta$ s show how the  $x$ s affect the choice.

The variance equation can be written as

$$\sigma_i = \exp\left(\sum_j z_{ij} \gamma_j\right) \quad (1b)$$

The scale/ variance equation indicates how the underlying latent variable is scaled for each case, i.e. it reflects differences in residual variability that, if left unaccounted for, would cause values to be scaled differently across cases. In the above,  $z$  is a vector of  $j$  values for the  $i$ th 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,  $\sigma_i = 1$ . The  $\gamma$ s show how the the  $z$ 's affect the variance (or more specifically, the log of  $\sigma$ ; estimating the log of  $\sigma$  guarantees that  $\sigma$  itself will always have a positive value).

For an ordered variable  $y$  with  $M$  categories coded 1 to  $M$ , the full heterogeneous choice model (using logit link) can then be written as<sup>3</sup>

$$P(y_i > m) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k - \kappa_m}{\exp\left(\sum_j z_{ij} \gamma_j\right)} \right) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k - \kappa_m}{\sigma_i} \right), m = 1, 2, \dots, M - 1, \quad (1c)$$

---

<sup>3</sup> 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.

where

$$\begin{aligned}\text{invlogit}(x) &= \text{inverse logit function of } x = \exp(x) / [1 + \exp(x)], \\ \exp\left(\sum_j z_{ij}\gamma_j\right) &= \exp(\ln(\sigma_i)) = \sigma_i, \\ \kappa_0 &= -\infty \text{ and } \kappa_M = \infty\end{aligned}$$

The full model shows how the choice and variance equations are combined to come up with the probability for any given response, e.g. you can compute the probability that a person with a given set of characteristics will “Strongly Agree” or “Disagree” with a statement. In the above formula, the  $\kappa$ s 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 ( $\beta_0$ ) rather than subtract a cutpoint. The end result is the same because the cutpoint and constant are opposite in sign. The logit link function is used here, but others are possible, such as probit, complementary log-log, log-log and cauchit.

When  $\sigma_i = 1$  for all cases and links `logit` or `probit` are used, the heterogeneous 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 `hetprobit` (except that `hetprobit` uses an intercept rather than a cutpoint.) As we will see, while less obvious, various other models that have appeared in the literature are also special cases of heterogeneous choice models.

### 3 The `oglm` command

#### 3.1 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 redisplay 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`.

### 3.2 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 “Insigma”, 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.

### 3.3 Other standard Stata options supported by `oglm`

`robust cluster level`

### 3.4 Options available when replaying results

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

### 3.5 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.

## 4 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, 2007) `esttab` program (all available from SSC) be installed. The first two 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 third 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 (b) why likelihood ratio tests, when possible, are often 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.



## 4.1 Example 1: Allison's Model of Group Comparisons

Allison (1999) analyzes a data set of 301 male and 177 female biochemists<sup>4</sup>. 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 (2009) provides an extended discussion of the strengths and weaknesses of Allison's proposed strategy, some of which we will expand on later. The Appendix of Allison's paper presents the Stata code that is needed to estimate his models<sup>5</sup>. We begin by summarizing Allison's discussion and then show how his results can be replicated using `oglm`.

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 a program presented in the appendix of his paper to estimate a single model for men and women that includes a new parameter he calls  $\delta$ . In this model, the coefficients for men and women are constrained to be equal. The  $\delta$  parameter adjusts for the differences in residual variability between men and women. Allison's model can be written as

---

<sup>4</sup> The data were originally collected by J. Scott Long (Long, Allison and McGinnis 1993) and are available on his website.

<sup>5</sup> The do file included with this paper includes the code needed to replicate Allison's analysis using his own programs.

$$P(y_i = 1) = \text{invlogit} \left( \left( \sum_k x_{ik} \beta_k + \beta_0 \right) * (1 + \delta G_i) \right) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k + \beta_0}{1 / (1 + \delta G_i)} \right) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k + \beta_0}{\sigma_i} \right) \quad (2)$$

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  $\sigma$  for group 0 equals 1 and the  $\sigma$  for group 1 equals  $1/(1 + \delta)$ . The value of  $\delta$  in Allison's model is  $-.26$ , meaning that the standard deviation of the disturbance variance for men (group 0) is 26 percent lower than the standard deviation for women (group 1), i.e. women are more variable in their career histories which causes the estimated coefficients in the female model to be smaller. To the model with  $\delta$  Allison then adds an interaction term for gender \* articles. This interaction term is insignificant. Allison therefore concludes "The apparent difference in the coefficients for article counts in Table 1 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."

Allison used specialized code to estimate his model. However, as Williams (2009) 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 presented in equation 1C simplifies to

$$P(y_i = 1) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k - \kappa}{\exp(G_i \gamma)} \right) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k - \kappa}{\exp(\ln(\sigma_i))} \right) = \text{invlogit} \left( \frac{\sum_k x_{ik} \beta_k - \kappa}{\sigma_i} \right) \quad (3)$$

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 ( $\beta_0$ ) is added in the numerator while in the heterogeneous choice model a cutpoint ( $\kappa$ ) is subtracted. This is a trivial difference because one number is the negative of the other. In both models the numerator is divided by  $\sigma_i$ . The main difference is how the two methods arrive at their estimate of  $\sigma_i$ . Neither method estimates  $\sigma_i$  directly, but  $\sigma_i$  is easily computed from the numbers they do estimate. The heterogeneous choice model estimates the log of  $\sigma_i$ , which guarantees that  $\sigma_i$  will be a positive number. Under Allison's approach,  $\delta$  is estimated, where  $\delta$  is the difference between the values of  $\sigma$  in the two groups. Not surprisingly, then, `oglm` can easily reproduce the estimates from Allison's model. The `het(female)` option tells `oglm` to include female in the variance equation, thus allowing residual variability to differ by gender.

```
. 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)
. * Allison's Table 1 - men only
. quietly logit tenure female year yearsq select articles prestige if female==0
. quietly estimates store male
. * Allison's Table 1 - females only
```

```

. quietly logit tenure female year yearsq select articles prestige if female==1
. quietly estimates store female
. * oglm replication of Allison's delta models from his Table 2
. quietly oglm tenure year yearsq select articles prestige female, het(female)
store(oglm1)
. * Compute Allison's delta
. display (1 - exp(.3022305))/ exp(.3022305)
-.26083233
. quietly oglm tenure year yearsq select articles prestige female f_articles,
het(female) store(oglm2)
. * Compute Allison's delta
. display (1 - exp(.1774193))/ exp(.1774193)
-.16257142
. esttab male female oglm1 oglm2, stats(N ll) mtitle

```

	(1) male	(2) female	(3) oglm1	(4) oglm2
-----				
main				
year	1.909*** (8.92)	1.408*** (5.47)	1.910*** (9.56)	1.838*** (9.06)
yearsq	-0.143*** (-7.70)	-0.0956*** (-4.36)	-0.140*** (-8.24)	-0.134*** (-7.89)
select	0.216*** (3.51)	0.0551 (0.77)	0.182*** (3.45)	0.170** (3.29)
articles	0.0737*** (6.37)	0.0340** (2.69)	0.0635*** (6.22)	0.0720*** (6.31)
prestige	-0.431*** (-3.96)	-0.371* (-2.38)	-0.446*** (-4.60)	-0.420*** (-4.37)
female			-0.939* (-2.53)	-0.378 (-0.84)
f_articles				-0.0305 (-1.63)
_cons	-7.680*** (-11.27)	-5.842*** (-6.75)		
-----				
lnsigma				
female			0.302* (2.07)	0.177 (1.09)
-----				
cut1				
_cons			7.491*** (11.36)	7.365*** (11.25)
-----				
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

The models labeled oglm1 and oglm2 correspond to the delta models in Allison's Table 2. 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  $\delta$  while `oglm` reports  $\gamma$ , which in this case is  $\ln(\sigma_{\text{Group1}})$ . These results are algebraically equivalent:  $\delta = (1 - \exp(\gamma))/\exp(\gamma) = (1 - \sigma_{\text{Group1}})/\sigma_{\text{Group1}}$ . The code above shows how delta can easily be computed using Stata.

The `oglm1` model says that the standard deviation of the residuals is  $\exp(\gamma) = \exp(.302) = 1.35$  times larger for women than men, while Allison's model using delta makes the equivalent statement that the standard deviation for men is 26 percent smaller than it is for women. In the `oglm2` model, the standard deviation is  $\exp(\gamma) = \exp(.177) = 1.194$  times larger for women, which is the same as saying that the standard deviation for men is 16.25 percent 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. This is not just a matter of convenience: ordinal variables are generally preferable because they contain more information about the underlying latent variable<sup>6</sup>. The variance equation is not limited to a single binary variable, hence increasing the ability of the researcher to estimate a properly specified model. `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.

## 4.2 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 (aka OCG II) survey data (Blau et al 1983; Inter-University Consortium for Political and Social Research 2010). 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 4<sup>th</sup> transition the dependent variable would be coded 0. The person would have no records for the 5<sup>th</sup> and 6<sup>th</sup> transitions because you cannot make those transitions if you haven't made the 4<sup>th</sup>. 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 2<sup>nd</sup> transition and is now eligible to complete the 3<sup>rd</sup>; otherwise `trans3` is coded 0).

---

<sup>6</sup> Williams (2009) discusses in more detail the limitations of binary dependent variables and the advantages offered by ordinal measures.

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).

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

The  $\lambda_j$  introduce proportional increases or decreases in the  $\beta_k$  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  $\lambda_j$  proportionality factor for each of the  $J = 6$  transitions ( $\lambda_1$  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, providing a parsimonious and substantively interesting model.

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,  $\lambda_j$ , 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 empirically indistinguishable. In the heterogeneous choice model (equations 1C and 3), the  $X\beta$ 's are divided by  $\sigma$ , while in the LRPC (equation 4) the  $X\beta$ 's are multiplied by  $\lambda$ s. Since multiplication is simply the inverse of division, it is not surprising that Hauser and Andrew's LRPC results can be easily reproduced using `oglm`<sup>7</sup>. 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.

---

<sup>7</sup> The fit of the LRPC model is presented in Table 5, Model 4 of Hauser and Andrew's (2006) paper. The do files included with this paper show how to exactly reproduce Hauser and Andrew's original results and show the simple algebraic manipulations that convert their parameterization into `oglm`'s.

```
. use lrpc, clear
(Hauser & Andrew, Sociological Methodology 2006 pp. 1-26, modified OCG II data)
. 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
-----						

Equivalencies between the LRPC and heterogeneous choice models are immediately apparent. Hauser and Andrew's LRPC program produces a log likelihood of -33529.654, as does oglm. The coefficients in Hauser and Andrew's betas equation have exact counterparts in oglm's choice equation. Simple algebraic manipulations can yield the other parameters reported by Hauser and Andrews, e.g. the LRPC's lambdas are the reciprocals of the heterogeneous choice model's sigmas.

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

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_{k=1}^{k'} \beta_k X_{ijk} + \sum_{k'+1}^K \beta_{jk} X_{ijk}, \quad j = 1, 2, \dots, 6 \quad (5)$$

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. 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).
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 sibstt19 ln_inc_trunc edhifaom edhimoom, het(trans2
trans3 trans4 trans5 trans6) store(m6)
```

Having noted these equivalences, it is important to realize that the substantive implications and rationales that motivate the models are very different. The LRPC and LRPPC say that effects differ across transitions by scale factors. The 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<sup>8</sup>. 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.

### 4.3 Example 3: 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<sup>9</sup>. He shows that generalized ordered logit models are one way of dealing with the problem. We will now illustrate that heterogeneous choice models may also be attractive alternatives.

---

<sup>8</sup> Using Hauser and Andrew's published code, we also estimated an LRPC model with Allison's biochemist data. The similarities were striking and obvious: other than the intercepts, which the two programs parameterize differently, the coefficient estimates were identical. Most critically, Allison's  $\delta$ , 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 or more groups. 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. 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.

<sup>9</sup> 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.

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), 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
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
warm						
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
(Ancillary parameters)						
_cut1	-2.465362	.2389126				
_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`).<sup>10</sup>

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.<sup>11</sup>

```
. 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 variables `male` and `yr89` have significant effects in both the choice and variance equations. 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 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.

<sup>10</sup> Both the Brant test and `gologit2`'s `autofit` option rely on purely empirical means to identify violations of a model's assumptions. It would be better, of course, if researchers had strong theories about when and where the model's assumptions will be violated, but we suspect this is rarely the case. Given that the alternatives are often to estimate a model whose assumptions are known to be violated (e.g. `ologit`) or to estimate a model that has far more parameters than are necessary (e.g. `mlogit`) the sort of middle ground taken by a program like `gologit2` may be the best choice. Williams (2006) argues that, when theory about the nature of violations is lacking, the use of more stringent significance levels when testing helps to avoid capitalizing on chance.

<sup>11</sup> Stepwise selection (see example 5) also results in the variables `yr89` and `male` being included in the variance equation.

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 either approach:

```
. 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 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. However, from a substantive standpoint, the simplicity of the `oglm` model and the insights about differences in variability across time and gender that are gained by adding only two parameters to the ordered logit model may be highly appealing.

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.

#### 4.4 Example 4: A trivial change with seemingly non-trivial implications

In many types of analyses, it often makes little difference whether z tests or Wald tests or likelihood ratio chi-square tests are used to test hypotheses about individual coefficients. It is important to realize that this is often NOT the case with heterogeneous choice models. In particular, seemingly trivial changes in the coding of variables used in the variance equation can change the hypotheses that z tests or Wald tests of coefficients in the choice equation address. In brief, z tests of individual coefficients in the choice equation are conditional on the coding of the variables in the variance equation, while likelihood ratio tests are not.

To illustrate this, 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.

```
. * 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*** (9.56)	1.411*** (7.17)	1.909*** (8.92)	1.408*** (5.47)
yearsq	-0.140*** (-8.24)	-0.103*** (-6.68)	-0.143*** (-7.70)	-0.0956*** (-4.36)
select	0.182*** (3.45)	0.134*** (3.41)	0.216*** (3.51)	0.0551 (0.77)
articles	0.0635*** (6.22)	0.0470*** (5.80)	0.0737*** (6.37)	0.0340** (2.69)
prestige	-0.446*** (-4.60)	-0.330*** (-4.07)	-0.431*** (-3.96)	-0.371* (-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*** (11.36)	6.231*** (10.04)	7.680*** (11.27)	5.842*** (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 the first two models, as we would expect the log likelihoods, model chi-squares and degrees of freedom are all the same. Also as we would expect, in the variance equations, the coefficient for male is opposite in sign to what it is for female. Perhaps surprisingly, however, all the coefficients in the choice equations are different, as are the z values. Note too that the coefficients in the first model, where males are coded 0, are similar to the coefficients in the

males-only model 3. The same is true for the second model that uses the variable male and females are coded 0, and the last model for females only.

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  $x_1$ ,  $x_2$ , and  $x_1*x_2$ , then the coefficient for  $x_1$  reflects the effect of  $x_1$  when  $x_2$  equals zero. Further, the t or z value for  $x_1$  tests whether the effect of  $x_1$  differs from zero when  $x_2 = 0$ ; even if the effect of  $x_1$  is insignificant when  $x_2 = 0$ , it may be significant for other values of  $x_2$ .

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  $\sigma$  will equal 0 when all the variables in the variance equation have a value of zero. The reported z values in the choice equation, then, are tests of whether or not the effect of a variable differs from zero for a group that has a value of zero for all variables in the variance equation. That is, the tests are conditional on the values of the variables in the variance equation, and a different set of values would yield different conditional tests. The z values are NOT global tests of whether the inclusion of a variable does or does not significantly improve overall model fit.

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<sup>12</sup>. Particularly in borderline situations, such differences could lead to different conclusions as to whether or not the effect of a variable was statistically significant.

Luckily, likelihood ratio tests of individual coefficients do NOT have this problem. They can test whether the inclusion of a variable in the choice equation does or does not significantly improve model fit, and are not conditional on the coding of the variables in the variance equation. 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)
```

---

<sup>12</sup> 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 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)
```

```
. 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 meaningful 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. The coefficients in the choice equation will then tell you the effects of variables on 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.

#### 4.5 Example 5: 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	<b>3.30</b>	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)<sup>13</sup>. One simple way to address such a possibility is to add `articles^2` to the model:

<sup>13</sup> We thank Maarten Buis for suggesting that we consider adding terms for nonlinear effects to the model.

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

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

Log likelihood = -822.94311
```

	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	<b>0.86</b>	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<sup>14</sup>. 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.

## 5 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.

<sup>14</sup> A reviewer suggested that “rather than adding a squared term for productivity, either the square root of articles or the  $\ln(articles + .5)$  are commonly used.” Inclusion of either of these terms also caused the variance coefficient to become insignificant. However, the overall fit of the model was better with  $articles^2$ .

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 `hetprobit` (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`. `oglm` does not currently support factor variables and may or may not support other features that were added to Stata after version 9. 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 and labeling of links.

## 6 Support for `oglm`

Richard Williams  
Department of Sociology  
University of Notre Dame  
[Richard.A.Williams.5@ND.Edu](mailto: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. The 1973 Occupational Changes in a Generation (aka OCG II) data that Hauser and Andrew modified (Blau et al 1983) is made available by the Inter-University Consortium for Political and Social Research (2010). Brian Miller assisted with the analysis.



## 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.
- Blau, Peter M., Otis Dudley Duncan, David L. Featherman, and Robert M. Hauser. 1983. Occupational Changes in a Generation, 1962 And 1973 [Computer file]. Madison, WI: University of Wisconsin [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994. doi:10.3886/ICPSR06162
- Duncan, Otis Dudley. 1975. *Introduction to Structural Equation Models*. Academic Press: New York.
- 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](http://www.business.uiuc.edu/ghoetker/documents/Hoetker_comp_logit.pdf) )
- Inter-University Consortium for Political and Social Research. 2010. Occupational Changes in a Generation, 1962 and 1973. Retrieved October 17, 2010  
(<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/06162>)
- Jann, Ben. 2005. Making regression tables from stored estimates. *The Stata Journal* 5(3): 288–308.
- \_\_\_\_\_. 2007. Making regression tables simplified. *The Stata Journal* 7(2): 227-244.
- Keele, Luke and David K. Park. 2006. Difficult Choices: An Evaluation of Heterogeneous Choice Models. Working Paper, March 3, 2006. Retrieved March 21, 2006  
(<http://www.nd.edu/~rwilliam/oglm/ljk-021706.pdf> )
- 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](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.
- \_\_\_\_\_. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients across Groups. *Sociological Methods & Research* 37(4): 531-559.
- 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*.