Comparing Logit and Probit Coefficients Across Groups: Problems, Solutions, and Problems with the Solutions

Richard Williams
Notre Dame Sociology
rwilliam@ND.Edu
http://www.nd.edu/~rwilliam

Meetings of the European Survey Research Association
Ljubljana, Slovenia
July 19, 2013

Comparing Logit and Probit Coefficients across groups

- We often want to compare the effects of variables across groups, e.g. we want to see if the effect of education is the same for men as it is for women
- But many/most researchers do not realize that methods typically used with continuous dependent variables to compare effects across groups may be problematic when the dependent variable is binary or ordinal

- We often think that the observed binary or ordinal variable y is a collapsed version of a latent continuous unobserved variable y*.
- Because y* is unobserved, its metric has to be fixed in some way. This is typically done by scaling y* so that its residual variance is $\pi^2/3 = 3.29$.
- But this creates problems similar to those encountered when analyzing standardized coefficients in OLS
  - unless the residual variance really is the same in both groups (i.e. errors are homoskedastic) the coefficients will be scaled differently and will *not* be comparable.

Case 1: True coefficients are equal, residual variances differ

	Group 0	Group 1
True coefficients	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = x_{i1} + x_{i2} + x_{i3} + 2\varepsilon_i$
Standardized Coefficients	$y_i^* = x_{i1} + x_{i2} + x_{i3} + \varepsilon_i$	$y_i^* = .5x_{i1} + .5x_{i2} + .5x_{i3} + \varepsilon_i$

In Case 1, the true coefficients 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 standardized βs (i.e. the ones reported by most logistic regression programs) are only half as large for group 1 as for group 0. *Naive comparisons of coefficients can indicate differences where none exist*.

Substantive Example: Allison's (1999) model for group comparisons

- Allison (Sociological Methods and Research, 1999) analyzes a data set of 301 male and 177 female biochemists.
- Allison uses logistic regressions to predict the probability of promotion to associate professor.

Table 1: Results of Logit Regressions Predicting Promotion to Associate Professor for Male and Female Biochemists (Adapted from Allison 1999, p. 188)

	Men		*Women*		*Ratio of*	*Chi-Square*
Variable	*Coefficient*	*SE*	*Coefficient*	*SE*	*Coefficients*	*for Difference*
Intercept	-7.6802***	.6814	-5.8420***	.8659	.76	2.78
Duration	1.9089***	.2141	1.4078***	.2573	.74	2.24
Duration squared	-0.1432***	.0186	-0.0956***	.0219	.67	2.74
Undergraduate selectivity	0.2158***	.0614	0.0551	.0717	.25	2.90
Number of articles	*0.0737****	*.0116*	*0.0340***	*.0126*	*.46*	*5.37**
Job prestige	-0.4312***	.1088	-0.3708*	.1560	.86	0.10
Log likelihood	-526.54		-306.19			
Error variance	3.29		3.29			

**p < .05, **p < .01, *** p < .001*

- As his Table 1 shows, the effect of number of articles on promotion is about twice as great for males (.0737) as it is for females (.0340).
- 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 1999:186).
- BUT, Allison warns, women may have more heterogeneous career patterns, and unmeasured variables affecting chances for promotion may be more important for women than for men.

- Allison argued that "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 proposed one way for dealing with group comparisons, but there are others

Solution I: Modify the Model & Make the hetero go away

- Williams (2010) notes that often the appearance of heteroskedasticity is actually caused by other problems in model specification, e.g. variables are omitted, variables should be transformed (e.g. logged), squared terms should be added
  - Williams (2010) shows that the heteroskedasticity issues in Allison's models go away if articles^2 is added to the model

Solution 2: Heterogeneous Choice Models

- Heterogeneous choice/ location-scale models explicitly specify the determinants of heteroskedasticity in an attempt to correct for it.
- In the tenure problem, Allison and Williams both let residual variability differ by gender (but more complicated variance models are also possible)

The Heterogeneous Choice (aka Location-Scale) Model

- Can be used for binary or ordinal models
- Two equations, choice & variance
- Binary case :

$$\Pr(y_i = 1) = g\left(\frac{x_i\beta}{\exp(z_i\gamma)}\right) = g\left(\frac{x_i\beta}{\exp(\ln(\sigma_i))}\right) = g\left(\frac{x_i\beta}{\sigma_i}\right)$$

Problem: Radically different interpretations are possible

- Hauser and Andrew noted that the effects of SES variables on educational attainment declined with each educational transition
- They modeled this via what they called the *logistic response model with proportionality constraints.*
- If the LRPC holds, the effects of variables differ only by a scale factor across each transition (or group), e.g. the model could hold if each SES variable only had half as large an effect on transition 2 as it did on transition 1.

Models compared

$$\Pr(y_i = 1) = g\left(\frac{x_i\beta}{\exp(z_i\gamma)}\right) = g\left(\frac{x_i\beta}{\exp(\ln(\sigma_i))}\right) = g\left(\frac{x_i\beta}{\sigma_i}\right)$$

$$\log_e\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{j0} + \lambda_j \sum_k \beta_k X_{ijk}$$

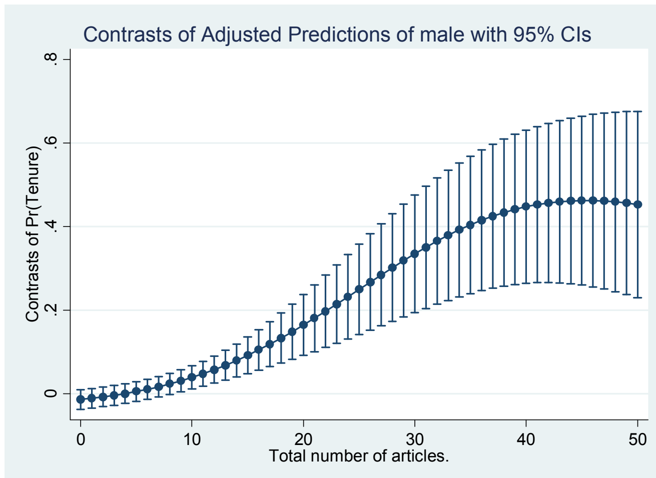
- Williams (2010) showed that, even though the rationales behind the models are totally different, heterogeneous choice models produce identical fits to the LRPC models estimated by Hauser and Andrew

- Indeed, when the models are both applied to Allison's tenure data, the estimated coefficients are exactly identical or can be easily converted from one parameterization to the other

- But, the theoretical concerns that motivate the models lead to radically different interpretations of the results.
  - 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, because every variable has a smaller effect on women than it does men
  - Somebody looking at these exact same numbers from the standpoint of the hetero choice model would conclude there is no inequality; effects of variables are the same for both men and women and only appear different because differences in residual variability cause coefficients to get scaled differently

Solution III: Compare Predicted Probabilities across groups

- Long (2009) proposes a different analytical approach that he says avoids the problems with the previous approaches.
- Long estimates models that allow for, say, every variable to interact with gender. He then creates graphs like the following that plot differences in predicted probabilities of tenure for men and women



- This simple example shows that the predicted probabilities of tenure for men and women differ little for those with small numbers of articles
- But, the differences become greater as the number of articles increases. For example, a women with 40 articles is predicted to be 45 percent less likely to get tenure than a man with 40 articles.

Critique of Long

- Once differences in predicted probabilities are discovered, policy makers may decide that some sort of corrective action should be considered, i.e. the graphs will show you whether there is a reason to be concerned in the first place

- At the same time, Long's approach may be frustrating because it doesn't try to explain why the differences exist. i.e. is it because the effects of variables differ across groups or is it because of differences in residual variability?

- From a policy standpoint, we would like to know what is causing these observed differences in predicted probabilities
  - If it is because women are rewarded less for each article they write, we may want to examine if women's work is not being evaluated fairly
  - If it is because of differences in residual variability, we may want to further examine why that is. For example, if family obligations create more career hurdles for women then they do men, how can we make the workplace more family-friendly?
  - But if we do not know what is causing the differences, we aren't even sure where to start if we want to eliminate them.

- But, as we have seen, when we try to explain group differences, the coefficients can be interpreted in radically different ways.
- Given such ambiguity, some might argue that you should settle for description and not strive for explanation (at least not with the current data).
- Others might argue that you should go with the model that you think makes most theoretical sense, while acknowledging that alternative interpretations of the results are possible.

Conclusions

- Researchers need to be aware that comparisons of effects across groups are much more difficult with logit and ordered logit models than with OLS
- But unfortunately the proposed ways for dealing with these issues have problems of their own
- At this point, it is probably fair to say that the descriptions of the problems with group comparisons may be better, or at least more clear-cut, than the various proposed solutions.

Selected References

- Allison, Paul. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research* 28(2): 186-208.
- 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 September 27, 2011 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=609104)
- 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. 2009. Group comparisons in logit and probit using predicted probabilities. Working Paper, June 25, 2009. Retrieved September 27, 2011 (http://www.indiana.edu/~jslsoc/files_research/groupdif/groupwithprobabilities/groups-with-prob-2009-06-25.pdf)
- Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd Edition. College Station, Texas: Stata Press.
- Williams, Richard. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients across Groups. *Sociological Methods & Research* 37(4): 531-559. A pre-publication version is available at http://www.nd.edu/~rwilliam/oglm/RW_Hetero_Choice.pdf.
- Williams, Richard. 2010. Fitting Heterogeneous Choice Models with oglm. *The Stata Journal* 10(4):540-567. A pre-publication version is available at http://www.nd.edu/~rwilliam/oglm/oglm_Stata.pdf.

For more information, see:

http://www3.nd.edu/~rwilliam