

gologit2 documentation updates

Richard Williams, Department of Sociology, University of Notre Dame

Richard.A.Williams.5@ND.Edu

Last revised March 19, 2016

When using `gologit2` in your work, the suggested citation is

Williams, Richard. 2006. "Generalized Ordered Logit/ Partial Proportional Odds Models for Ordinal Dependent Variables." *The Stata Journal* 6(1):58-82. The published article is available for free at <http://www.stata-journal.com/article.html?article=st0097>

You may also wish to cite

Williams, Richard. 2016. "Understanding and Interpreting Generalized Ordered Logit Models." *Journal of Mathematical Sociology* 40(1):7-20.
<http://dx.doi.org/10.1080/0022250X.2015.1112384>.

Since the 2006 article was written, there have been several enhancements to `gologit2`. A brief summary follows. See the help file for more details.

1. The `gologit2` support page and troubleshooting FAQ can be found at

<http://www3.nd.edu/~rwilliam/gologit2/index.html>

<http://www3.nd.edu/~rwilliam/gologit2/tsfaq.html>

2. `gologit2` now supports factor variables and the `svy:` prefix. It requires Stata 11.2 or higher. If you are condemned to using an older version of Stata, the old version of `gologit2` has been renamed `gologit29`. Both versions are available from SSC. The new version was announced at <http://www.statalist.org/forums/forum/general-stata-discussion/general/296459-major-update-to-gologit2-now-available>

NOTE: You will generally want to use the `gsvy:` prefix rather than `svy:.` `gsvy` has the same syntax as `svy` but provides customized support for `gologit2`'s unique options, such as `autofit`.

NOTE: Long and Freese (2014) discuss `gologit2` on pp. 371-377. Their book was written before `gologit2` supported factor variables. The above announcement shows a simpler way to run the example they give in their book.

3. `gologit2` now supports the `margins` command. Support is especially good in Stata 14+, where `gologit2` supports marginal analysis with multiple outcomes, i.e. you only have to give one `margins` command for all outcomes rather than separate commands for each outcome. For more details see <http://www.statalist.org/forums/forum/general-stata-discussion/general/1294323-new-versions-of-gologit2-and-oglm-available-on-ssc>.

4. If the user considers them more appropriate for their data, probit, complementary log-log, log-log and cauchit links can be used instead of logit. The `link()` function 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. For example, to estimate a goprobit model,

```
. use http://www.indiana.edu/~jslsoc/stata/spex_data/ordwarm2.dta, clear
. gologit2 warm i.yr89 i.male i.white age ed prst, link(p)
```

The following advice 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. The cauchit distribution has tails that are bigger than the normal distribution's, hence the cauchit link may be useful when you have more extreme values in either direction.

Warnings: Programs differ in the names used for these latter two links. Stata's loglog link corresponds to SPSS PLUM's cloglog link; and Stata's cloglog link is called nloglog in SPSS.

5. `gologit2` includes additional diagnostic measures. An oddity of `gologit`/`goprobit` models is that it is possible to get negative predicted probabilities. McCullagh & Nelder discuss this in *Generalized Linear Models*, 2nd edition, 1989, p. 155: "The usefulness of non-parallel regression models is limited to some extent by the fact that the lines must eventually intersect. Negative fitted values are then unavoidable for some values of x , though perhaps not in the observed range. If such intersections occur in a sufficiently remote region of the x -space, this flaw in the model need not be serious." This seems to be a fairly rare occurrence, and when it does occur there are often other problems with the model, e.g. the model is overly complicated and/or there are very small N s for some categories of the dependent variable. `gologit2` will give a warning message whenever any in-sample predicted probabilities are negative. If it is just a few cases, it may not be worth worrying about, but if there are many cases you may wish to modify your model, data, or sample, or use a different statistical technique altogether.