

Alternatives to Logistic Regression (Brief Overview)

Richard Williams, University of Notre Dame, <https://academicweb.nd.edu/~rwilliam/>

Last revised October 10, 2024

Probit

$$y^* = \alpha + \sum X\beta + \varepsilon, \varepsilon \sim N(0,1)$$

If $y^* \geq 0$, $y = 1$

If $y^* < 0$, $y = 0$

The predicted values in a probit model are like Z-scores. Somebody who has a predicted score of 0 has a 50% chance of success. Somebody with a score of 1 has about an 84% chance of success.

As is the case in the logistic regression model, linear changes in the probit produce nonlinear changes in the probability of success. At the extremes, changes in the probit have little effect on the probability of success. The effects of changes are much greater when the probit is around 0.

Probit	P(Y=1)	Change in P(Y=1)
-4	0.003167%	0.003138%
-3	0.134990%	0.131823%
-2	2.275013%	2.140023%
-1	15.865525%	13.590512%
0	50.000000%	34.134475%
1	84.134475%	34.134475%
2	97.724987%	13.590512%
3	99.865010%	2.140023%
4	99.996833%	0.131823%

The probit and logistic regression models tend to produce very similar predictions. The parameter estimates in a logistic regression tend to be 1.6 to 1.8 times higher than they are in a corresponding probit model.

Long (1997, p. 83) says that the choice between the logit and probit models is largely one of convenience and convention, since the substantive results are generally indistinguishable. But, in some cases, the need to generalize a model may be an issue. For example, multiple-equation systems involving qualitative dependent variables are based on the probit model (e.g. see `biprobit`). For models with nominal dependent variables that have more than 2 categories, the logit model (estimated by `mlogit`) may be preferred because the corresponding probit model (estimated by `mprobit`) is too computationally demanding. For panel data, you can estimate a fixed effects model with logit but not with probit.

The Linear Probability Model (LPM)

We have already talked about the LPM, mostly in a very critical fashion. However, there are various people (mostly Economists I think) who defend the use of the Linear Probability Model, i.e. OLS regression with a binary dependent variable. Here are some discussions on the topic:

<http://statisticalhorizons.com/whats-so-special-about-logit>

<http://statisticalhorizons.com/linear-vs-logistic>

<https://statisticalhorizons.com/in-defense-of-logit-part-2> (this piece best represents my own views)

<http://blogs.worldbank.org/impactevaluations/whether-to-probit-or-to-probe-it-in-defense-of-the-linear-probability-model>

<http://davegiles.blogspot.com/2012/06/another-gripe-about-linear-probability.html>

<http://marcfbellemare.com/wordpress/8951>

<http://prisonrodeo.tumblr.com/post/52055757707>

<http://marcfbellemare.com/wordpress/9024>

Arguments include claims that the LPM is more flexible and easier to interpret; and also that the marginal effects tend to be the same whether you use OLS or logit. However, I think these claims pertain to Average Marginal Effects, which are useful, but can also conceal how effects differ across the range of a variable. In any event I think AMEs can get you most of the same advantages as the LPM while still estimating the model correctly.

[Optional] Other alternatives

The `mlogit` command also fits logistic regression models. You might sometimes prefer it over `logit` because some post-estimation commands are written so as to only work with `mlogit` and not `logit` (e.g. Long and Freese's `mlogtest` command, which we will discuss later.)

Several binary regression models are special cases of the models that can be estimated with my `gologit2` and `oglm` programs, e.g. these programs can estimate the same models that `logit`, `probit` and `cloglog` do. They can also estimate several other models not otherwise available in Stata. We will discuss these programs much more later.

As Long and Freese (2006, p. 181) note, Stata can estimate other types of binomial models. `blogit` and `bprobit` fit logit and probit models on grouped (“blocked”) data. The `hetprob` command allows the assumed variance of the errors in the probit model to vary as a function of the independent variables; we’ll talk more about such models later. `hetprob`’s heteroskedastic probit model is another special case of the models that can be estimated by `oglm`.

Some other alternatives are discussed in the appendices.

Appendix A: Complementary log-log models [OPTIONAL]

The Stata command `cloglog` fits maximum likelihood complementary log-log models. The model can be written as

$$\Pr(Y = 1 | X) = 1 - \exp\{-\exp(X\beta)\}$$

Or as

$$\log\{-\log(1 - p_i)\} = \alpha + \sum X\beta$$

According to the Stata reference manual and Powers and Xie (2000), complementary log-log analysis is an alternative to logit and probit analysis, but it is unlike these other estimators in that the transformation is not symmetric about 0, i.e. it is skewed to the right. Typically this model is used when the positive (or negative) outcome is rare.

As with other models, changes in the values of XB produce nonlinear changes in the probability of success.

cloglog	Pr(Y=1)
-3	4.85680%
-2.5	7.88063%
-2	12.65770%
-1.5	19.99893%
-1	30.77994%
-0.5	45.47608%
0	63.21206%
0.5	80.77044%
1	93.40120%
1.5	98.86857%
2	99.93820%
2.5	99.99949%
3	100.00000%

According to Powers and Xie (2000, pp. 83-84) the complementary log-log model is derived from the assumption that the error distribution (or distribution of the latent variable) follows a standard extreme value distribution. They further say that, aside from its use as an alternative probability model, this model has been used extensively to model grouped survival data. For individual level data, the parameters have a similar interpretation to those from the logistic regression model.

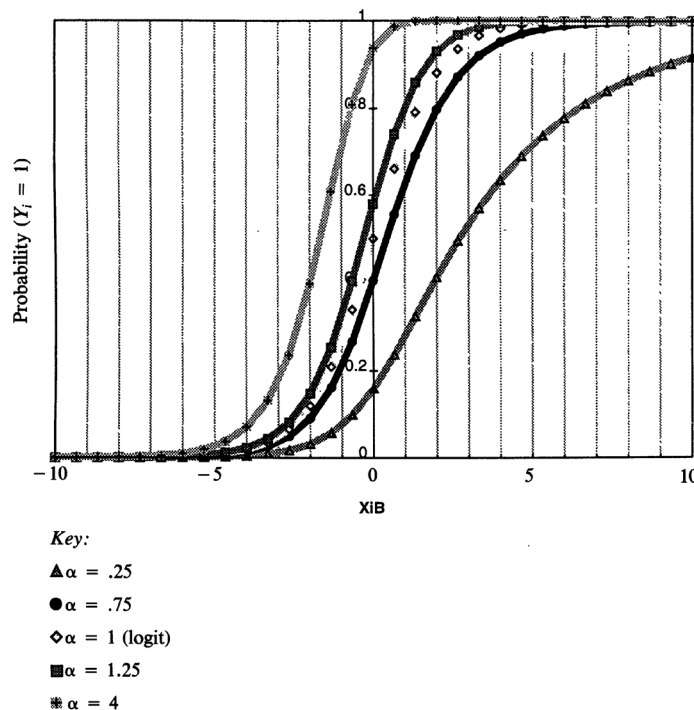
I'll admit that I do not have a very good understanding of the complementary log-log model. Powers and Xie give an example of where it produces results very similar to logistic regression, and the few examples I tried produced similar results as well. But, it might make more of a difference when the probability of success is very small or very large.

Appendix B: Scobit (Skewed Logit) [OPTIONAL]

I include this just on the wild chance that you come across an article that uses this method or for some reason you want to use it yourself. Other than in the article that introduced the method, I have never seen it used. Nor have I had much luck even getting it to work the few times I have tried it.

The article described below is cited in the Stata reference manual and is available from JSTOR. Basically, the author, Jonathan Nagler, argues that, with both logit and probit, it is assumed that individuals with a .5 probability of success are most sensitive to changes in the independent variables, e.g. a 1 unit change in X will have a greater effect on someone who has a .5 probability of success than it will have on someone with a .3 or .7 probability of success. Nagler argues that this isn't necessarily the case, e.g. somebody with a .4 probability of success may be more affected by a 1 unit change in X than somebody who has a .5 probability of success. If so, the distribution is “skewed” - things aren't symmetric about .5. He shows various diagrams which involve modifications of the shape of the logistic distribution, where the changes are determined by a parameter called α , e.g.

Figure 2.A. Cumulative Distribution for Scobit
[Y - Axis = $1 - F(-X_i\beta)$]



($\alpha = .25$ is the rightmost curve, $\alpha = 4$ is the leftmost.) He says that, if individuals with high initial probabilities of choosing alternative 1 are most sensitive to stimulus, α should be high. If individuals with low initial probabilities of choosing alternative 1 are most sensitive to stimulus, α should be low. Regular logistic regression is basically a special case of the scobit model, where $\alpha = 1$. Stata provides a means for testing whether scobit is necessary or if it is ok to stick with logistic (i.e. it provides a test of whether α significantly differs from 1).

It appears that scobit models can be tough to estimate. I tried a few, and often got fatal errors and/or it took very long to get a solution. Further, the Stata online help warns that

In many cases, the ancillary parameter alpha can be highly collinear with the other parameters in the model (i.e., the coefficients). This will lead to a singular variance-covariance matrix (i.e., missing values for one or more standard errors). In these cases, other models (e.g., logit, probit, or cloglog) may be more appropriate for the data.

The log likelihood for scobit is often poorly approximated by a quadratic function. Because of this, standard-error estimates can sometimes be very poor, and Wald tests can lead to incorrect conclusions. Hence, likelihood-ratio tests are recommended for all statistical inference; see help lrtest.

Scobit: An Alternative Estimator to Logit and Probit

Jonathan Nagler

American Journal of Political Science, Vol. 38, No. 1. (Feb., 1994), pp. 230-255.

Stable URL: <http://links.jstor.org/sici?sici=0092-5853%28199402%2938%3A1%3C230%3ASAAETL%3E2.0.CO%3B2-K>

Abstract

Logit and probit, the two most common techniques for estimation of models with a dichotomous dependent variable, impose the assumption that individuals with a probability of .5 of choosing either of two alternatives are most sensitive to changes in independent variables. This assumption is imposed by the estimation technique because both the logistic and normal density functions are symmetric about zero. Rather than let methodology dictate substantive assumptions, I propose an alternative distribution for the disturbances to the normal or logistic distribution. The resulting estimator developed here, scobit (or skewed-logit), is shown to be appropriate where individuals with any initial probability of choosing either of two alternatives are most sensitive to changes in independent variables. I then demonstrate that voters with initial probability of voting of less than .5 are most sensitive to changes in independent variables. And I examine whether individuals with low levels of education or high levels of education are most sensitive to changes in voting laws with respect to their probability of voting.