Probit and Logit (quick overview)

Federico Nutarelli

The order of the topics covered here is, more or less, the same argument covered in the R script.

Overview of Logit and Probit

First of all, Probit (Logit) regression is used for binary depenents:

$$y = \begin{cases} 0 \ \textit{if no} \\ 1 \ \textit{if yes} \end{cases} .$$

Examples for instance in labor market is whether an individual participates in the labor market or not.
In general, binary outcome models estimate the probability that $y = 1$ as a function of the independent variables:

$$Pr[y = 1|x] = F(x'\beta) \tag{1}$$

In the Probit model, $F(x'\beta)$ is the cdf of the standard normal distribution, i.e.:

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz$$

The Logit model is the cdf of the logistic distribution:

$$F(x'\beta) = \Lambda(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}$$

Notice that $x'\beta$ is a quantile z such that the Probit coefficient β is the change in z (the quantile) associated with a one unit change in x. Basically, if x changes of one unit, how does z changes? We assume that the effect of a change of x on z is linear. Notice however that the link between the dependent and z is **not** linear because $\Phi(.)$ is not linear in x. Since the dependent variable is a nonlinear function of the regressors, the coefficient on x has no simple interpretation.
So, **how do we interpret the coefficients of the probit?** In general, you cannot interpret the coefficients from the output of a probit regression; not in any standard way, at least.

Marginal effects

If you are asked, how to interpret the results of a probit (logit) **directly**, what you usually would like is to assess whether an increase in x increases/decreases the **likelihood** that $y = 1$ (makes that outcome more/less likely). In other words, an increase in x makes the outcome of 1 **more or less likely**. We interpret the **sign** of the coefficient but **not** the magnitude. The magnitude cannot be interpreted using the coefficient because different models have different scales of coefficients.

When estimating probit and logit models, it is common to report the marginal effects after reporting the coefficients. The marginal effects reflect the change in the probability of $y = 1$ given a 1 unit change in an independent variable x. In general, in fact, you need to interpret the **marginal effects** of the regressors, that is, how much the (conditional) probability of the outcome variable changes when you change the value of a regressor, holding all other regressors constant at some values. This is different from the linear regression case where you are directly interpreting the estimated coefficients. This is so because **in the linear regression case, the regression coefficients are the marginal effects**.

In the probit regression, there is an additional step of computation required to get the marginal effects once you have computed the probit regression fit. This is easy to see:

- Linear regression case: $E[Y|X] = \beta_0 + \sum_i \beta_k X_{ki}$, whose marginal effects is β_k;
- Probit model: $E[Y|X] = \Phi(\beta_0 + \sum_i \beta_k X_{ki})$, whose marginal effects are $\beta_k \cdot \Phi(\beta_0 + \sum_i \beta_k X_{ki})$, **which is not the same as the regression coefficient!**. For short, marginal effects are: $\Phi(x'\beta)\beta_j$.
- Logit model: $E[Y|X] = \Lambda(\beta_0 + \sum_i \beta_k X_{ki})$, whose marginal effects are (for short) $\Lambda(x'\beta)[1 - \Lambda(x'\beta)]\beta_k = \frac{e^{x'\beta}}{1+e^{x'\beta}} \cdot \frac{1}{1+e^{x'\beta}} \approx \Lambda(x'\beta) \cdot \beta_j$. We will use the latter in R.

Problem: how can we compute the quantity represented by the marginal effects and what are the choices of the other regressors that should enter this formula? Thankfully, R provides this computation after a probit regression, and provides some defaults of the choices of the other regressors (there is no universal agreement on these defaults).

Specifically, the R procedure goes like this: The predicted probability that $y = 1$ given $x_1, x_2, \ldots, x_k$ can be calculated in two steps:

1) Compute $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$, obtaining $\hat{z}$
2) Look up $\Phi(\hat{z})$ by calling pnorm().

β_j is the effect on z of a one unit change in regressor x_j, holding constant all other $k1$ regressors.

In R, Probit models can be estimated using the function glm() from the package stats. Using the argument "family" we specify that we want to use a Probit link function.

Odds ratios

In general, given a probability of success p (so the probability of failure is $(1-p)$), the odds ratio measures the probability that $y = 1$ relative to the probability that $y = 0$.
Now, in the Logit model, for instance, we have seen that:

$$p = \frac{e^{(x'\beta)}}{1 + e^{(x'\beta)}}$$

, hence:

$$\frac{p}{(1-p)} = e^{x'\beta}$$

In order to simplify, we can simply take the log of the odds ratio:

$$log\frac{p}{(1-p)} = x'\beta$$

An odds ratio of 2, for instance, simply means that the outcome $y = 1$ is twice as more likely than the outcome $y = 0$. Though odds ratio are important and reported as default in many popular softwares (e.g. STATA), in economics, it is more popular to report the marginal effects.

McFadden R^2

Why do we need an alternative R^2? Remember that the R-squared equals $\frac{SSRegression}{SSTotal}$, which mathematically must produce a value between 0 and 100%.$SStotal$ is the total sum of squares: $(y_i - \overline{y})^2$ and $SSregression$ is the explained sum of squares $(f_i - \overline{y})$. In nonlinear regression, $SSRegression + SSError$ do not equal $SSTotal$! This completely invalidates R-squared for nonlinear models, and it no longer has to be between 0 and 100%.
So we need an alternative R^2 which is given by:

$$1 - L_{ur}/L_r$$

. The latter compares the unrestricted (i.e. estimated on all parameters, or in our case, all Xs) log-likelihood L_{ur} for the model we are estimating and the restricted (i.e. a log-likelihood without any X) L_r with only an intercept.
If the independent variables have no explanatory power, the restricted model will be the same as the unrestricted model and the R-squared will be 0.

Prediction

Prediction in general **linear** models focuses mainly on predicting the values of the conditional mean:

$$E[Y|X_1 = x_1, \ldots X_p = x_p] = f(\eta) = f(\beta_0 + \beta_1 x_1 + \ldots \beta_p x_p)$$

using $\hat{\eta} = \hat{\beta}_0+\hat{\beta}_1 x_1+\ldots\hat{\beta}_p x_p$ and not on predicting the conditional response (i.e. $\hat{Y}|X$). The reason is that confidence intervals, the main difference between both kinds of prediction (of conditional mean and of conditional response), depend heavily on the family we are considering for the response. For instance, if we think about it, for the logistic function the CI **of the conditional response** (not the ones of the coefficients, which need to be computed!) can be only $\{0, 1\}$. To provide you an intuition, $\hat{Y}$ is a probability so stays between 0 and 1,, while we know that CI are constructed by adding and subtracting 1.96 which would make the CI for $\hat{Y}$ larger than necessary, thus useless.
To stay with the same notation, for the logistic model, the prediction of the conditional response follows immediately from $\Lambda(\eta)$:

$$\hat{Y}|(X_1 = x_1, \ldots X_p = x_p) = \begin{cases} 1 \ \textit{with probability } \Lambda(\eta) \\ 0 \ \textit{with probability } 1-\Lambda(\eta) \end{cases} .$$

As a consequence, we can predict Y as 1 if $\Lambda(\eta) > 1$ and 0 otherwise (Y takes indeed only two values).
From this we can understand easily that **type="link"** returns the $\hat{\eta}$ directly, i.e. $\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots\hat{\beta}_p x_p$ which you might have heard as log-odds in the logistic, while **type="response"** returns the probabilities in the logistic, i.e. $\Lambda(\eta)$.
$E[Y|X]$ can be reached as an average of the latter.