Lesson 4
Limited Dependent Variables

4.1: Logit/Probit Models

Normally, independent variables are dummy variables. But in some cases, it may be the dependent variables that are dummies. For instance, consider whether a person is employed or not. We can define this as a dummy:

Y = 1 if employed, = 0 otherwise.

In such cases the OLD Method is inapplicable and we have to use Logit/Probit Models.

4.2: The Model

We have three files (kug, pdv and ocp) containing data for three mines. Respondents were asked whether they had suffered from any disease. If so, they were assigned a value of 1, and 0 otherwise. We collected data on each disease (given by names), and disease type (grp1…grp7). However, we will work with two other variables: ILL and ILLM.

ILL = 1 means that the person is suffering from any of the diseases studied; = 0 otherwise.

ILLM = 1 means that the person is suffering from diseases common to mine workers (like back ache, asthma, etc.); = 0 otherwise.

Our hypothesis is that a person is more likely to be affected by a disease if he is a mineworker. Whether he is a mineworker or not depend upon the value of another variable MW. This is 1 for mineworkers; = 0 otherwise. To test the hypothesis we will have to estimate the model:

$$ILL = \alpha + \beta\ MW.$$

4.3: Logit Model

Open the file kug and create an intercept them int.

Click on Univariate\LogitProbit Models.

Type ILL INT MW in the Estimation Window and click Start.

Choose the Logit Option.

Note that R^2 cannot be estimated, so we get the pseudo R^2. Click Close and examine the options provided.

4.4: Probit Model

The Probit Model is estimated as above – only you should choose the Probit Model option. You can also choose the probit model while backtracking from the Logit Model.

Estimate ILL and ILLM for each mine using both logit and Probit Models. You can also estimate Models using GRP1 … GRP7 or any individual disease.

Do you think that the hypothesis that mineworkers face greater occupational health hazards is borne out by the results? The value of the pseudo R^2 is also very low – about 0.2. Why do you think this is low?

4.5: Pooled Model

Pool the data for the three mines and create a new file named pool. Run the Logit and probit Models on pooled data.

4.6: Comparison of Coefficients

Enter the values of the coefficients for all the three mines in an Excel sheet. Divide the values of the intercept and slope coefficients from the Logit Model with that from

the Probit Model. Round off to first place of decimal. You will see that the result comes to 1.7 for nearly all the values.

Note

The exercises are based on a ppaer by Debasish Sarkar et al (Journal of Health Management). The results of the exercise do not tally with that of the paper. One reason is that the authors used LIMDEP to estimate the regression. Even if they used Microfit the results would have differed from yours. Compare carefully the results and guess why the difference occurs. Your results should, however, tally with that given on the Table below.

Results:-

Statistics	KUG	OCP	PDV
Logit Model			
Intercept	-1.7370	-1.5305	-2.6955
MW	3.2410	3.1728	4.44384
Pseudo-R^2	0.24831	.24108	.44991
Probit Model			
Intercept	-1.0377	-0.92334	-1.5281
MW	1.9462	1.9090	2.5691
Pseudo-R^2	0.24831	.24108	.44991