Problem set – Binary choice

Note that we will work on the problem individually in class. However, you may want to look it over and have a first try before class in order to be able to get through.

1. You want to model the choice between fishing from a boat and fishing from a pier.

   Read the data into Stata

   ```
   . use http://fmwww.bc.edu/ec-p/data/mus/mus15data
   ```

   and create a binary variable that is defined as follows

   $$y_i = \begin{cases} 1 & \text{if fishing from a charter boat} \\ 0 & \text{if fishing from a pier} \end{cases}$$

   also create your explanatory variable, the log relative price of charter vs pier fishing:

   $$x_i = \ln(p_{charter}/p_{pier}).$$

   you will estimate models of the following form

   $$\Pr(y_i = 1|x_i) = F(\beta_0 + \beta_1 x_i)$$

   (a) Describe the data and calculate some relevant descriptive statistics (think about what you would like to know).

   (b) Estimate your model using the LPM, Probit and Logit, and

        i. Compare the three models on the basis of the slope estimates.

        ii. Compare the three models on the basis of the statistical significance of the slope estimates.

        iii. Compare the three models on the basis of the predicted probabilities.

        iv. Compare the Logit and the Probit model on the basis of log-likelihood.

        v. Compare the three models graphically by plotting the predicted probabilities. What would you ideally like to compare these curves to?

   (c) Obtain and compare the partial effects for the Probit and Logit models in the following ways:

        i. evaluated at the sample average of the log relative price,

        ii. averaged over the complete sample,

        iii. averaged over the individuals that chose pier fishing,

        iv. average over the individuals that chose charter boat fishing.

   (d) Use -xtile- to create a new variable that contains the quintiles of the variable income. Now estimate a LPM and Probit of y_i with dummies for each quintile as RHS-variables. Compare the predicted values and marginal effects from both models. What do you conclude?