

## EXERCISE: Models with Limited Dependent Variables

### A Logistic Model of the Decision to Smoke or not to Smoke

Within the document *Smoking and Drinking Amongst Adults 2008*, which is a component of the *General Lifestyle Survey 2008*, you will find *Table 1.3 Percentage who have never smoked cigarettes regularly by sex and age: 1974 to 2008*.

You are asked to use the **gretl** program to fit a logistic function to describe the probability of having been a smoker as a function of age. You may choose to make a comparison of two probability schedules that vary either according to sex at a specific date or that vary between dates for men or for women.

For the independent variable, you may take the mid points of the age intervals that are given in table 1.3. Observe that, whereas the argument of the normal distribution ranges over the interval  $(-\infty, \infty)$ , ages are bounded by zero. You are asked to accommodate this feature in the model via the appropriate transformation of the independent variable and to determine what difference this makes to the analysis.

The **gretl** function *Logistic regression* carries out an OLS regression using the logistic transformation  $\log\{y/(\gamma - y)\}$  of the dependent variable  $y$ . The dependent variable must be strictly positive. If it is a decimal fraction, between 0 and 1, the default is to use a  $\gamma$  value (the asymptotic maximum of the dependent variable) of 1. If the dependent variable is a percentage, between 0 and 100, the default  $\gamma$  is 100.

If you wish to set a different maximum, use the optional *y<sub>max</sub>=value* syntax following the list of regressors. The supplied value must be greater than all of the observed values of the dependent variable.

### A Binary Response Model of Labor Market Participation.

Use the data in the file `participation.data` to construct a model of Swiss labour force participatoion. The data were taken from Michael Gerfin (1996), "Parametric and Semiparametric Estimation of the Binary Response Model of Labour Market Participation," *Journal of Applied Econometrics*, Vol. 11, No. 3, 1996, pp. 321–340.

There are 872 observations.

Column is the observation number.

Column 2 is LFP, a labour force participation dummy.

Column 3 is LNNLINC, the log of nonlabour income.

Column 4 is AGE, age in years divided by 10.

Column 5 is EDUC, years of formal education.

Column 6 is NYC, the number of young children (younger than 7).

Column 7 is NOC, the number of older children.

Column 8 is FOREIGN, a dummy, equal to 1 if obs is not Swiss.