LABOR FORCE PARTICIPATION DATA SET

STATA COMMANDS

c.doughertylse.ac.uk

January 2002

This document gives explicit instructions for doing the LFP regression exercises using Stata version 7. First, a few general comments. As far as the Windows environment is concerned, Stata is only semi-housetrained and it can be an effort to find out how to do things with it. There is a built-in help resource, but it could be improved. The manuals accompanying Stata version 7 are an improvement on previous ones, but likewise there is still scope for further improvement. Often the quickest way to find out how to do something is by getting an experienced user to show you. But it is worth making the effort because Stata is powerful and easy to use, once you know what you are doing.

Files

As with all statistical packages, you will be working with two types of file: a data file and output files. The LFP data file, `LFP.dta`, is described at the beginning of the *EAEF* manual (the extension `.dta` indicates that it is a data file). As you do the exercises, the file will increase in size because you will occasionally add new variables to it, but you will never need to change its name. When you have added a new variable and wish to save the expanded file, do so with the existing name and overwrite the previous version. You cannot print this file directly because it is in a special format that no printer can recognize. If you wish to print out some of the data, you have to do this from within Stata.

Output files are described in Stata as log files. You will create many of them, one for each Stata session. Assuming that you wish to preserve the output for each session, you should give each file a different name. To open a log file, click on the button with a scroll on the Stata menu bar. This is the fourth from the left. You are given a choice of type of log file at the time of opening it. The default is a "formatted log file" with extension `.smcl`. Avoid this, and choose instead a plain log file with extension `.log`. Plain `.log` files are ASCII (text) files and can be imported with no fuss into any word-processing package. Incidentally, having specified the name of your `.log` file and the directory (folder) in which it is to be located, it is natural to look for a tab with "open" or "OK" to click. There isn't one. You have to click on the button marked "save" to open the file. To close the file at the end of your session, click on the scroll icon again and choose the close option.

Stata windows

When you launch Stata, you will see four Windows: a command window, a results window, a variables window, and a review window.

Command window: This one-line window is where you type in your instructions. You can save on typing in two ways. Instead of typing the name of a variable, you can click on its name in the variables window. Second, when you need to give a command that is similar to a previous one, you can do this by editing the previous command rather than starting from scratch. You do this by pressing the Page Up key as often as necessary. It is easy to make mistakes when entering Stata commands and this will keep you sane.

Results window: You would only use the results window if you do not intend to save your output and have not opened a log window. However, even if you have opened a log window, you need to be able to see the bottom of the results window. The reason is that if a command gives rise to more output than can be shown at once in the results window, the first window-full will be shown and the instruction --more-- in blue letters will appear at the bottom of the results window. To see another window-full, press the space-bar, and keep doing this until all the output has been displayed. You will not be able to issue any more commands until you have done this. If the bottom of the results file is not visible, you will not be able to see the --more-- instruction and you will think that Stata has hung up on you.

Variables window: This contains a list of the variables in the data set.

Review window: This simply lists your most recent commands.

Common commands

Here are a few commands that will be useful in the exercises:

- `reg` followed by a list of variable names. The first variable is regressed on the rest.
- `sum` followed by a list of variable names. This produces a table giving the mean, standard deviation, maximum and minimum for each variable listed.
- `tab` followed by one variable name. This produces a frequency distribution for the variable
- `tab` followed by two variable names. This produces a cross-tabulation with the first-named variable providing the rows and the second-named one providing the columns
- `gen` followed by an equation. This creates a new variable defined as the dependent variable of the equation.

Adding an `if` expression at the end of a command, for example `if y>10`, makes it selective as indicated. Most `if` conditions are straightforward, but there is one that is not: a condition which uses an = sign, like `if y==10`, must repeat the = sign as shown. This is to distinguish between the use of = in equations defining variables and its use in tests for equality.

EXERCISES

You must download the LFP data set in order to do this exercise. Instructions are given at the beginning of the *EAEF* manual. You should download it only once

Exercise 1 Probit analysis

What factors affect having a job?

Use probit analysis to regress *WORKING* on *AGE*, *S*, *CHILDL06*, *CHILDL16*, *MARRIED*, *ETHBLACK*, and *ETHHISP* for males and females separately. Perform appropriate statistical tests and derive the marginal effects.

```
probit WORKING S AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP
if MALE==0
```

Exercise 2 Tobit analysis

What factors affect number of hours worked?

Regress *HOURS* on *AGE*, *S*, *CHILDL06*, *CHILDL16*, *MARRIED*, *ETHBLACK*, and *ETHHISP* for females (1) using OLS and all the observations, (2) using OLS and only the observations for which *HOURS* > 0, and (3) using tobit analysis with lower bound 0. Compare the results.

```
reg HOURS AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP if
MALE==0 & HOURS > -1
```

```
reg HOURS AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP if
MALE==0 & HOURS > 0
```

```
tobit HOURS AGE CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP if
MALE==0 & HOURS > -1, ll(0)
```

Exercise 3 Tobit analysis

What factors affect number of hours worked?

In Exercise 2 it would have been desirable to have included *EARNINGS* as an explanatory variable. Why is this not possible?

(No Stata analysis required)

Exercise 4 Sample selection bias

Is there evidence that sample selection bias affects the fitting of earnings functions?

Regress LGEARN on *S*, *ASVABC*, *ETHBLACK*, and *ETHHISP* for males and females separately, using *S*, *AGE*, *CHILDL06*, *CHILDL16*, *MARRIED*, *ETHBLACK*, and *ETHHISP* as selection variables. Compare the results with those of the corresponding OLS regressions.

```
heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE
CHILDL06 CHILDL16 MARRIED ETHBLACK ETHHISP)
```

```
reg LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0
```

Exercise 5 Sample selection bias

Is there evidence that being a single mother is a special deterrent to working?

Investigate whether having a child aged less than 6 is likely to be an especially powerful deterrent to working if the mother is unmarried by downloading the LFP94 data set from the website and repeating the regressions in this section adding an interactive dummy variable *MARL06* defined as the product of *MARRIED* and *CHILDL06* to the selection part of the model.

```
g MARL06 = MARRIED*CHILDL06
```

```
heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE
CHILDL06 CHILDL16 MARRIED MARL06 ETHBLACK ETHHISP)
```

Exercise 6 Sample selection bias

Effect of including a selection variable in the regression specification

It has been asserted that including in the regression specification a variable that is part of the selection process may give rise to an apparently significant effect. Evaluate this by adding *CHILDL06* to an otherwise standard earnings function for females, regressing *LGEARN* on *S*, *ASVABC*, *ETHBLACK*, *ETHHISP*, and *CHILDL06*.

```
reg LGEARN S ASVABC ETHBLACK ETHHISP CHILDL06 if MALE==0
```