

## EXERCISE: Analysis of Panel Data using GRETL

**The Data From Greene.** The `gretl` program is already linked to a data file `greene14_1.gdt`, which is designed to illustrate the analysis of panel data. The data file can be loaded via the menu commands `<File><Open Data><Sample file>`. These will bring the `<<gretl: data files>>` window to the fore. Click on the tab labelled *Greene* and select the airline cost data. The data can be viewed via the `<Data><Select all><Display values>` menu commands.

**The Structure of the Data.** The data comprise observations on 6 airlines taken annually for 15 years. The joint indices for the airline, i.e. the unit, and for the year of observation are given in the first column of the display in the form of `unit:year`. Thus, for example, `1:13` denotes first airline and the 13th year. These indices inform `gretl` of the panel structure the data.

The second and third third columns, which are titled `unit` and `year`, repeat this information in a way that is redundant. The remaining elements in each row correspond to the observations on the dependent variable `C:cost` and the explanatory variables `Q:output`, `PF:fuel price` and `F:load factor`.

**A Command Script.** A command script for estimating the panel data models is already available. This is accessed via the menu sequence `<File> <Data files> <Practice files>`. A window titled `<<gretl: practice files>>` will be revealed from which `greene14_1 panel methods`, `Hausman test` should be selected.

The script indicates that the logarithms of some the variables are taken before an ordinary least-squares regression is run via the command `ols 1_C 0 1_Q 1_PF LF`. The zero in this sequence stands for the vector of units associated with a global constant term. The result of the command may be described as a pooled regression.

The command `hausman` tests the simple pooled model against the principal alternatives, which are the fixed effects and random effects models. The fixed effects model allows the intercept of the regression to vary across the cross-sectional units. An  $F$ -test is reported for the null hypotheses that the intercepts do not differ. The random effects model decomposes the residual variance into two parts, one part specific to the cross-sectional unit and the other specific to the particular observation.

To run the script, click on the run icon, which is the third icon at the top of the window that takes the form of a cogwheel. This will regenerate the results of the regression analysis.

**Importing Panel Data.** An array of ASCII data in a comma-separated or space-separated format can be imported into `gretl` and, given the appropriate instructions, it will interpreted as panel data. Consider the following matrix

array:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 16.2 & 31.5 & 42.7 \\ 6.5 & 87.5 & 78.6 \\ \hline 8.7 & 32.4 & 65.7 \\ 43.6 & 4.8 & 38.2 \\ \hline 5.4 & 55.7 & 75.7 \\ 78.4 & 24.2 & 29.3 \end{bmatrix}$$

Here, the generic submatrix  $X_j$ , of order  $2 \times 3$ , represents observations from 2 dates of three variables affecting the  $j$ th unit.

This array of data, without the matrix braces and the horizontal dividing lines, can be imported into **gretl** via the menu sequence  $\langle File \rangle \langle Open Data \rangle \langle Import \rangle \langle CSV \rangle$ . You will be confronted with the  $\langle \langle \text{gretl: data delimiter} \rangle \rangle$  dialogue box, from which you should select the comma option. Search for the **GTRIAL.CSV** file and open it.

The  $\langle \langle \text{gretl: open data} \rangle \rangle$  dialogue box will appear. Next, in response to the question “Do you want to give the data a time-series or a panel interpretation” answer [Yes]. Within the  $\langle \langle \text{Data structure wizard} \rangle \rangle$ , select *Panel*.

When confronted with a choice between *stacked time series* or *stacked cross section section*, choose the former. Select 3 as the number of cross-section units 2 as the number of periods. (Observe that, in fact, **gretl** proposes 2 cross-sectional units 3 time periods. This is an intelligent proposal in view of the fact that the number of observations  $T$  should exceed the number of variables  $k$ , but it conflicts with our previous specification.)

The data can now be viewed via  $\langle Data \rangle \langle Select all \rangle \langle Display values \rangle$ . The following tableau will be revealed:

	$v_1$	$v_2$	$v_3$
1 : 1	16.2	31.5	42.7
1 : 2	6.5	87.5	78.6
2 : 1	8.7	32.4	65.7
2 : 2	43.6	4.8	38.2
3 : 1	5.4	55.7	75.7
3 : 2	78.4	24.2	29.3

Having mastered these techniques, you are in a position to exploit the panel data facilities of **gretl**.