

Using gretl for Principles of Econometrics, 4th Edition

Version 1.041₁

Lee C. Adkins
 Professor of Economics
 Oklahoma State University
 February 1, 2013

Since this manual is based on the examples from Principles of Econometrics, 4th edition (POE4) by Hill et al. (2011), you should also download and install the accompanying data files that go with this book. The file is available at <http://www.learneconometrics.com/gretl/POE4data.exe>.

This is a self-extracting windows file that will install the POE4 data sets onto the c:\Program Files (x86)\ngretl\data directory of your computer's harddrive.¹ If you have installed gretl

2.1 Simple Linear Regression Model

The simple linear regression model is

$$food_exp_t = \beta_1 + \beta_2 income_t + e_t \quad t = 1, 2, \dots, T \quad (2.1)$$

where $food_exp_t$ is your dependent variable, $income_t$ is the independent variable, e_t is random error, and β_1 and β_2 are the parameters you want to estimate. The errors of the model, e_t , have an average value of zero for each value of $income_t$; each has the same variance, σ^2 , and are uncorrelated with one another. The independent variable, $income_t$, has to take on at least two different values in your dataset. If not, you won't be able to estimate a slope! The error assumptions can be summarized as $e_t | income_t \text{ iid } N(0, \sigma^2)$. The expression *iid* stands for **independently and identically distributed** and means that the errors are statistically independent from one another (and therefore uncorrelated) and that each has the same probability distribution. Taking a random sample from a single population accomplishes this.

2.2 Retrieve the Data

The first step is to load the food expenditure and income data into **gretl**. The data file is included in your **gretl** sample files—provided that you have installed the *Principles of Econometrics* data supplement that is available from our website. See section 1.1.1 for details.

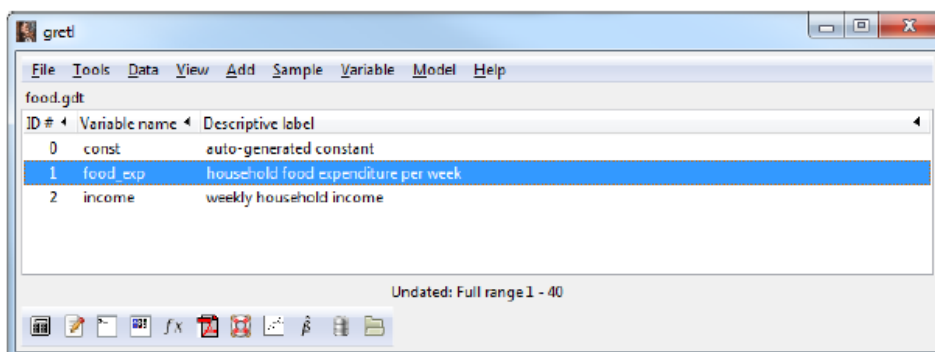


Figure 2.1: The main `gretl` window. The food expenditure data is loaded from `food.gdt` using `File>Open data>sample file` and choosing the food dataset from the sample files that accompany *POE4*.

Load the data from the data file `food.gdt`. Recall, this is accomplished by the commands `File>Open data>Sample file` from the menu bar.¹ Choose `food` from the list. When you bring the file containing the data into `gretl` your window will look like the one in Figure 2.1. Notice that in the `Descriptive label` column contains some information about the variables in the program's memory. For some of the datasets included with this book, it may be blank. These descriptions, when they exist, are used by the graphing program to label your output and to help you keep track of variables that are available for use. Before you graph your output or generate results for a report or paper you may want to label your variables to make the output easier to understand. This can be accomplished by editing the attributes of the variables.

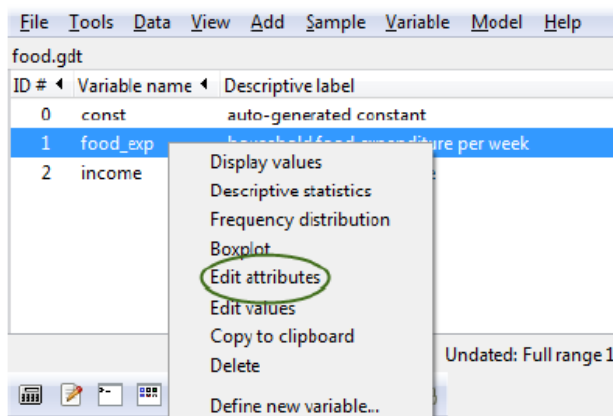


Figure 2.2: Highlight the desired variable and right-click to bring up the pull-down menu shown here. You can also use F2 or keyboard shortcut 'CTRL+e' to bring up the dialog.

¹ Alternately, you could click on the open data button on the toolbar. It's the one that looks like a folder on the far right-hand side.

To do this, first highlight the variable whose attributes you want to edit, right-click and the menu shown in (see Figure 2.2) appears. Select **Edit attributes** to open a dialog box (Figure 2.3) where you can change the variable's name, assign variable descriptions and display names. Describe and label the variable `food_exp` as 'Food Expenditure' and `income` as 'Weekly Income (\$100)'. The dialog can also be opened using F2 from the main `gretl` window or using the keyboard

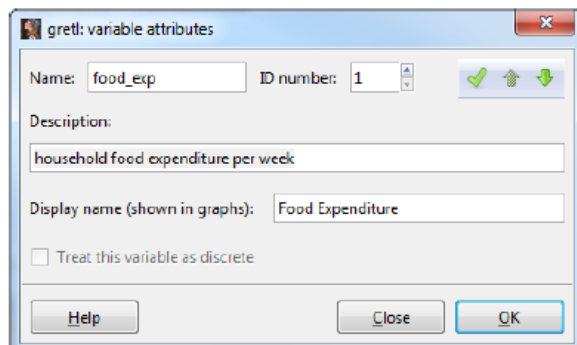


Figure 2.3: Variable edit dialog box

shortcut, 'E.' Finally, the `setinfo` command can be used to set the description and the label used

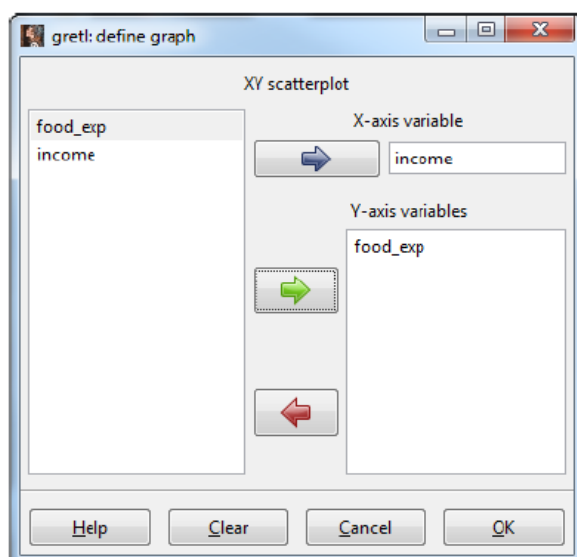


Figure 2.4: Use the dialog to plot of the food expenditure against Weekly Income in graphs.

In the following example a script is generated that opens the `food.gdt` dataset, and adds variable descriptions, and assigns a label to be used in subsequent graphs.


```

open "@gretl\dir\data\poe\food.gdt"
setinfo food_exp -d "household food expenditure per week" \
  -n "Food Expenditure"
setinfo income -d "weekly household income" -n "Weekly Income"
labels

```

The `-d` flag is given followed by a string in double quotes. It is used to set the descriptive label. The `-n` flag is used similarly to set the variable's name in graphs. Notice that in the first and last uses of `setinfo` in the example that I have issued the continuation command (`\`) since these commands are too long to fit on a single line. If you issue the `labels` command, `gretl` will respond by printing the descriptions to the screen.

2.3 Graph the Data

To generate a graph of the food expenditure data that resembles the one in Figure 2.6 of *POE*, you can use the  button on the `gretl` toolbar (third button from the right). Clicking this button brings up a dialog to plot the two variables against one another. Figure 2.4 shows this dialog where x is placed on the x-axis and y on the y-axis. The result appears in Figure 2.5. Notice that the labels applied above now appear on the axes of the graph.

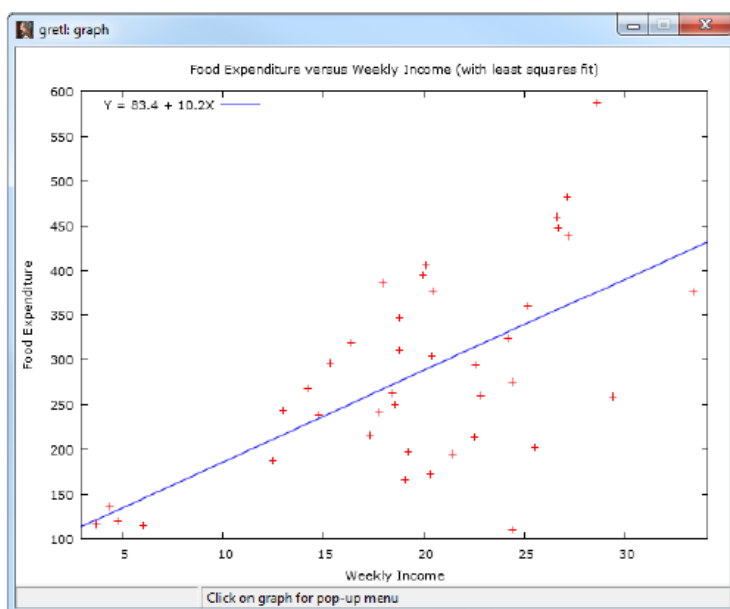


Figure 2.5: XY plot of the food expenditure data

Figure 2.5 plots food expenditures on the y axis and Weekly Income on the x . **Gretl**, by default, also plots the fitted regression line. The benefits of assigning labels to the variables becomes more obvious. Both X- and Y-axes are informatively labeled and the graph title is changed as well. More on this later.

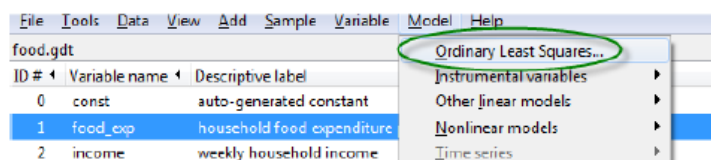


Figure 2.6: From the menu bar, select **Model>Ordinary Least Squares** to open the least squares dialog box

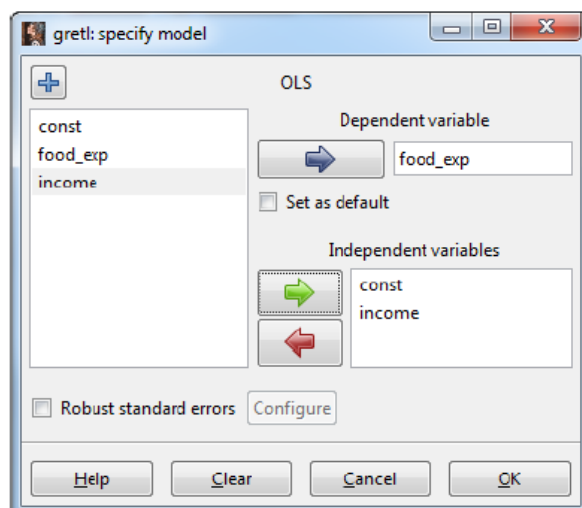


Figure 2.7: The Specify Model dialog box opens when you select **Model>Ordinary least squares**


2.4 Estimate the Food Expenditure Relationship

Now you are ready to use **gretl** to estimate the parameters of the food expenditure equation.

$$food_exp_t = \beta_1 + \beta_2 income_t + e_t \quad t = 1, 2, \dots, T \quad (2.2)$$

From the menu bar, select **Model>Ordinary Least Squares** from the pull-down menu (see Figure 2.6) to open the dialog box shown in Figure 2.7. From this dialog you'll need to tell **gretl** which variable to use as the dependent variable and which is the independent variable. Notice that by default, **gretl** assumes that you want to estimate an intercept (β_1) and includes a constant as

an independent variable by placing the variable `const` in the list by default. To include x as an independent variable, highlight it with the cursor and click the ‘Add->’ button.

The `gretl` console (see section 1.3.1) provides an easy way to run a regression. The `gretl` console is opened by clicking the console button on the toolbar, . The resulting console window is shown in Figure 2.8.

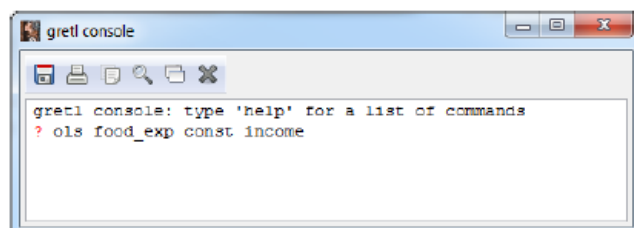


Figure 2.8: The Gretl console window. From this window you can type in `gretl` commands directly and perform analyses very quickly—if you know the proper `gretl` commands.

At the question mark in the console simply type

```
ols y const x
```

to estimate your regression function. The syntax is very simple, `ols` tells `gretl` that you want to estimate a linear function using ordinary least squares. The first variable listed will be your dependent variable and any that follow, the independent variables. These names must match the ones used in your data set. Since ours in the food expenditure example are named, `y` and `x`, respectively, these are the names used here. Don’t forget to estimate an intercept by adding a constant (`const`) to the list of regressors. Also, don’t forget that `gretl` is case sensitive so that `x` and `X` are different entities.

This yields window shown in Figure 2.9 below. The results are summarized in Table 2.1.

An equivalent way to present results, especially in very small models like the simple linear regression, is to use equation form. In this format, the `gretl` results are:

$$\widehat{\text{food_exp}} = 83.4160 + 10.2096 \text{ income}$$

$$T = 40 \quad \bar{R}^2 = 0.3688 \quad F(1, 38) = 23.789 \quad \hat{\sigma} = 89.517$$

(standard errors in parentheses)

Finally, notice in the main `gretl` window (Figure 1.3) that the first column has a heading called

Table 2.1: OLS estimates using the 40 observations 1–40.

OLS, using observations 1–40				
Dependent variable: food_exp				
	Coefficient	Std. Error	<i>t</i> -ratio	p-value
const	83.4160	43.4102	1.9216	0.0622
income	10.2096	2.09326	4.8774	0.0000
Mean dependent var	283.5735	S.D. dependent var	112.6752	
Sum squared resid	304505.2	S.E. of regression	89.51700	
R^2	0.385002	Adjusted R^2	0.368818	
$F(1, 38)$	23.78884	P-value(F)	0.000019	
Log-likelihood	−235.5088	Akaike criterion	475.0176	
Schwarz criterion	478.3954	Hannan–Quinn	476.2389	

ID #. An ID # is assigned to each variable in memory and you can use the ID # instead of its variable name in your programs. For instance, the following two lines yield identical results:

```

1  ols food_exp const income
2  ols 1 0 2

```

One (1) is the ID number for `food_exp` and two (2) is the ID number of `income`. The constant has ID zero (0). If you tend to use long and descriptive variable names (recommended, by the way), using the ID number can save you a lot of typing (and some mistakes).

2.4.1 Elasticity

Elasticity is an important concept in economics. It measures how responsiveness one variable is to changes in another. Mathematically, the concept of elasticity is fairly simple:

$$\varepsilon = \frac{\text{percentage change in } y}{\text{percentage change in } x} = \frac{\Delta y/y}{\Delta x/x} \quad (2.3)$$

In terms of the regression function, we are interested in the elasticity of average food expenditures with respect to changes in income:

$$\varepsilon = \frac{\Delta E(y)/E(y)}{\Delta x/x} = \beta_2 \frac{x}{E(y)}. \quad (2.4)$$

$E(y)$ and x are usually replaced by their sample means and β_2 by its estimate. The mean of `food_exp` and `income` can be obtained by using the cursor to highlight both variables, use the

View>Summary statistics from the menu bar as shown in Figure 2.10, and the computation can be done by hand. However, you can make this even easier by using the `gretl` language to do all of the computations—no calculator needed! Simply open up a new script and type in:

```
1  ols food_exp const income --quiet
2  scalar elast=$coeff(income)*mean(income)/mean(food_exp)
```

This yields the output shown in figure 2.11.

Following a least squares regression, `Gretl` stores the least squares estimates of the constant and the slope in variables called `$coeff(const)` and `$coeff(income)`, respectively. In addition, it uses `mean(income)` and `mean(food_exp)` to compute the mean of the variables `income` and `food_exp`. The `--quiet` option is convenient when you don't want or need the output from the regression printed to the screen. The result from this computation appears below in Figure 2.12.

2.4.2 Prediction

Similarly, `gretl` can be used to produce predictions. The predicted food expenditure of an average household having weekly income of \$2000 is:

$$\widehat{food_exp}_t = 83.42 + 10.21 income_t = 83.42 + 10.21(20) = 287.61 \quad (2.5)$$

Remember, `income` is measured in \$100, so 20 in the above expression represents $20 * \$100 = \$2,000$. The `gretl` script is:

```
scalar yhat = $coeff(const) + $coeff(income)*20
```

which yields the desired result.

2.4.3 Estimating Variance

In section 2.7 of *POE4*, you are given expressions for the variances of the least squares estimators of the intercept and slope as well as their covariance. These estimators require that you estimate the overall variance of the model's errors, σ^2 . `Gretl` does not explicitly report the estimator, $\hat{\sigma}^2$, but rather, its square root, $\hat{\sigma}$. `Gretl` calls this "S.E. of regression" which you can see from the output is 89.517. Thus, $89.517^2 = 8013.29$. `Gretl` also reports the sum of squared residuals, equal to 304505.2, from which you can calculate the estimate. Dividing the sum of squared residuals by the estimator's degrees of freedom yields $\hat{\sigma}^2 = 304505/38 = 8013.29$.

The estimated variances and covariance of the least squares estimator can be obtained once the model is estimated by least squares by selecting the **Analysis>Coefficient covariance matrix** command from the pull-down menu of the **model** window as shown in Figure 2.13. The result is:

Covariance matrix of regression coefficients:

const	income	
1884.44	-85.9032	const
	4.38175	income

So, estimated variances of the least squares estimator of the intercept and slope are 1884.44 and 4.38175, respectively. The least squares standard errors are simply the square roots of these numbers. The estimated covariance between the slope and intercept -85.9032 .

You can also obtain the variance-covariance matrix by specifying the `--vcv` option when estimating a regression model. For the food expenditure example use:

```
ols food_exp const income --vcv
```

to estimate the model using least squares and to print the variance covariance matrix to the results window.