# EViews 6 Tutorial

by Manfred W. Keil

to Accompany

**Introduction to Econometrics**

by James H. Stock and Mark W. Watson

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. EViews: Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>2. Cross-Sectional Data</strong></td>
<td>3</td>
</tr>
<tr>
<td>Interactive Use: Data Input and Simple Data Analysis</td>
<td></td>
</tr>
<tr>
<td>a) The Easy and Tedious Way: Manual Data Entry</td>
<td>4</td>
</tr>
<tr>
<td>b) Summary Statistics</td>
<td>7</td>
</tr>
<tr>
<td>c) Graphical Presentations</td>
<td>8</td>
</tr>
<tr>
<td>d) Simple Regression</td>
<td>11</td>
</tr>
<tr>
<td>e) Entering Data from a Spreadsheet</td>
<td>14</td>
</tr>
<tr>
<td>f) Importing Data Files directly into EViews</td>
<td>15</td>
</tr>
<tr>
<td>g) Multiple Regression Model</td>
<td>16</td>
</tr>
<tr>
<td>h) Data Transformations</td>
<td>19</td>
</tr>
<tr>
<td><strong>Batch Files</strong></td>
<td>21</td>
</tr>
<tr>
<td><strong>3. Time Series Data</strong></td>
<td>23</td>
</tr>
<tr>
<td><strong>4. Summary of Frequently Used EViews Commands</strong></td>
<td>25</td>
</tr>
<tr>
<td><strong>5. Final Note</strong></td>
<td>27</td>
</tr>
</tbody>
</table>
1. EVIEWS: INTRODUCTION

This tutorial will introduce you to a statistical and econometric software package called EViews. The most current professional version is EViews 7.1, but the student version still runs on EViews 6. Both EViews 6 and 7.1 are sufficiently similar that those who have access to EViews 7 can comfortably use this tutorial for the more advanced version. The difference is only apparent in more advanced techniques that you, as a first time user, will not encounter in a course of econometrics (or at least not in the beginning of the course).

EViews runs on the Windows (2000, 2003, XP, Vista, Server 2008, or Windows 7), but not on a Mac platform (unless you can run Windows on your Mac). It is produced by Quantitative Micro Software (QMS) in Irvine, California. You can read about various product information at the firm’s Web site, www.eviews.com. EViews 7 comes with four manuals, a User’s Guide (2 books), a Command Reference, and an Object Reference book. The manuals can only be ordered as a set, but they can be ordered without purchasing the EViews program; the cost is $140 for the set. However, these manuals can be accessed within the program through the Help function. You can order EViews by calling (949) 856-3368 or writing to sales@eviews.com. In addition, if you purchase the Student Version ($39.95), you will receive PDF copies of the User’s Guide and Command Reference manuals for EViews 6 plus some additional material to get started. The User’s Guide is better for first-time users. In some cases, you may have received the Student Version bundled with a textbook.

The difference between the student version and the full version is in the limitation on the size of data sets (“capacity limitation” is 1,500 observations for each series and no more than 15,000 observations for all series; students can work with larger data sets but will then not be able to save/import the workfile) and the availability of some features such as advanced seasonal adjustment methods (X11 and X12). Furthermore, and perhaps most importantly for you right now, the student version does not allow you to run EViews in “batch mode.” Instead you can only use the interactive use. This tutorial will explain the difference between interactive use and batch mode below. Once you have gone through the first series of commands in interactive mode, you will almost certainly want to run programs in batch mode.

Econometrics deals with three types of data: cross-sectional data, time series data, and panel (longitudinal) data (see Chapter 1 of the Stock and Watson (2011) textbook). In a time series you observe the behavior of a single entity over multiple time periods. This can range from high frequency data such as financial data (hours, days); to data observed at somewhat lower (monthly) frequencies, such as industrial production, inflation, and unemployment rates; to quarterly data (GDP) or annual (historical) data. In a cross-section you analyze data from multiple entities at a single point in time. One big difference between time series and cross-sectional analysis is that the order of the observation numbers does not matter in cross-sections. With time series, you would lose some of the most interesting features of the data if you shuffled the observations. Finally, panel data can be viewed as a combination of time series and cross-sectional data, since multiple entities are observed at multiple time periods. EViews allows you to work with all three types of data.
EViews is most commonly used for time series analysis in academics, business, and government, but you can work with it easily when you have cross-sections and/or panel data. EViews allows you to save results within a program and to “retrieve” these results for further calculations later. Remember how you calculated confidence intervals in statistics say for a population mean? Basically you needed the sample mean, the standard deviation, and some value from a statistical table. In EViews you can calculate the mean and standard deviation of a sample and then temporarily “store” these. You then work with these numbers in a standard formula for confidence intervals. In addition, EViews provides the required numbers from the relevant distribution (normal, \( \chi^2 \), \( F \), etc.).

While EViews is truly interactive, you can also run a program as a “batch” job, i.e., you write a sequence of commands and then execute the program in one go. In the good old days the equivalent was to submit a “batch” of cards, each containing a single command, to a technician, who would use a card reader to enter these into the computer, and the computer would then execute the sequence of statements. (You stored this batch of cards typically in a filing cabinet, and the deck was referred to as a “file.”) While you will work at first in interactive mode by clicking on buttons, you will very soon discover the advantage of running your regressions in batch mode. This method allows you to see the history of commands, and you can also analyze where exactly things went wrong if there are problems with any of your commands. This tutorial will initially explain the interactive use of EViews, since it is more intuitive. However, we will switch as soon as it makes sense into the batch mode.¹

While EViews produces graphs and charts, these can often be improved upon by saving the data used in these graphs in a spreadsheet or ASCII format, and then to import the data into Excel (or another spreadsheet program you prefer). Even better, since EViews works in a Windows format, it allows you to cut and paste the data into any other Windows-based program.

Finally, there is a warning about the limitations of this tutorial. The purpose is to help you gain an initial understanding of how to work with EViews. I hope that the tutorial looks less daunting than the manuals. However, it cannot replace the accompanying manuals, which you will have to consult for more detailed questions (alternatively use “Help” in the program). Feel free to provide me with feedback of how we can improve the tutorial for future generations of students (mkeil@cmc.edu). Colleagues of mine and I have decided to set up a “Wiki” run by students but supervised by faculty at my academic institution. We have found that the “wisdom of crowds” often produces very valuable information for those who follow. This is, of course, just a suggestion. Finally you may want to think about working with statistical software as learning a new language: practicing it routinely will result in improvement. If you set it aside for too long, you will only remember the most important lines but will forget the important details. Another danger of tutorials like this is that you simply follow the instructions and when you are done, you don’t remember the commands. It is therefore a good idea to keep a separate sheet and to write

¹ As mentioned above, the very reasonably priced student version does not run batch files. However, even if you purchased the student version, the academic version may be available to you at your college/university, or you may decide to upgrade on your own.
down commands and examples of them if you think you will use them later on. I will give you short exercises so that you can practice the commands on your own.

2. CROSS-SECTIONAL DATA

**Interactive Use: Data Input and Simple Data Analysis**

Let’s get started. Click on the EViews icon to begin your session. What you see next is the EViews window, with the *title bar* at the top, the *command window* immediately below and the *status line* at the very bottom (ignore my path, etc., below).

![EViews Window](image)

The results of your various operations will be displayed between the command window and the status line in the so-called *work area*. In interactive use, EViews allows you to execute commands either by clicking on command buttons or by typing the equivalent command into the *command window*.

In this tutorial, we will work with two data applications, two cross-sectional (California Test Score Data Set used in chapters 4-9; Current Population Survey Data Set used in Chapters 3 and 8), and one time series (U.S. Macro Quarterly Data Set used in Chapter 14 and 16).
a) The Easy and Tedious Way: Manual Data Entry

In Chapters 4 to 9 you will work with the California Test Score Data Set. These are cross-sectional data, referred to in EViews as “undated or irregular” data. There are 420 observations from K-6 and K-8 school districts for the years 1998 and 1999. You will not want to enter a large amount of data manually, unless you have collected data by yourself (something that economists are doing more and more). The alternative is to enter the data into a spreadsheet (Excel) and then to cut and paste the data (see below). However, for the purpose of this introduction it will be useful that you become aware of entering, and editing, data. Here I will use a sub-sample of 10 observations from the California Test Score Data Set.

To start, we must create a workfile in EViews. Click on the File pull-down menu, and then on New and Workfile. As is common in Windows programs, you will see a dialog box.

![Workfile Create dialog box](image)

This particular dialog box asks you for the start and end dates of your data set, and for the type of data you are entering. We are working with undated or irregular data (cross sectional data), so use the pull-down menu for Workfile Structure Type and select Unstructured/Undated. Then enter 10 in the Observations box. While you are at it, enter “SW10smpl” into the “WF” field.
You will see a workfile window, which contains two entries ($c$, $resid$). Do not worry about these for the moment. To enter the data into a format similar to the spreadsheets you have become familiar with, click on **Quick** in the title bar, and then on **Empty Group (Edit Series)**.
Next enter the data for two variables (two columns). Here are the 10 observations to enter.
(EViews will add zeros. You will see later how to get rid of these.)

<table>
<thead>
<tr>
<th>obs</th>
<th>TESTSCR</th>
<th>STR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>606.8</td>
<td>19.5</td>
</tr>
<tr>
<td>2</td>
<td>631.1</td>
<td>20.1</td>
</tr>
<tr>
<td>3</td>
<td>631.4</td>
<td>21.5</td>
</tr>
<tr>
<td>4</td>
<td>631.8</td>
<td>20.1</td>
</tr>
<tr>
<td>5</td>
<td>631.9</td>
<td>20.4</td>
</tr>
<tr>
<td>6</td>
<td>632.0</td>
<td>22.4</td>
</tr>
<tr>
<td>7</td>
<td>632.0</td>
<td>22.9</td>
</tr>
<tr>
<td>8</td>
<td>638.5</td>
<td>19.1</td>
</tr>
<tr>
<td>9</td>
<td>638.7</td>
<td>20.2</td>
</tr>
<tr>
<td>10</td>
<td>639.3</td>
<td>19.7</td>
</tr>
</tbody>
</table>

Once you have entered the data, close the object (you will be asked “Delete Untitled GROUP ?”
Click on “Yes.”) You will be able to (re)name the variables. Click on SER01, then rightclick and chose “Rename…” and enter “testscr.” Do the same to change SER02 to “str.”

Entering data in this way is very tedious, and you will make data input errors frequently. You will see below how to enter data directly from a spreadsheet or an ASCII file, which are the most common forms of data you will receive in the future. Also, you noticed when you entered the test score (testscr) first and then the student-teacher ratio (str) that you were automatically moved into the test score column after entering each student-teacher data point. This is an unfortunate feature, but there is no alternative unless you enter all the data by observation.

In general, you can look at variables in your workfile by typing in the command

$$Show \ varname1, \ varname2, \ ...$$

where varnamei refers to a variable that exists in your workfile. Try it here by typing

$$show \ testscr \ str$$

You should see the following:
**b) Summary Statistics**

For the moment, let’s just see if we are working with the same data set. Locate the View button at the upper-left corner of the Group window, click on it, and then click on **Descriptive Stats** and **Common sample**. You should see the following output (instead of using Prnt Scrn on my computer, I pressed the Freeze button in EViews. This allows me to copy and paste output into another Windows based program, a feature that will come in handy down the road when you want to display some of your output):

Sample: 1 10

<table>
<thead>
<tr>
<th></th>
<th>TESTSCR</th>
<th>STR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>631.3500</td>
<td>20.59000</td>
</tr>
<tr>
<td>Median</td>
<td>631.9500</td>
<td>20.15000</td>
</tr>
<tr>
<td>Maximum</td>
<td>639.3000</td>
<td>22.90000</td>
</tr>
<tr>
<td>Minimum</td>
<td>606.8000</td>
<td>19.10000</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>9.264418</td>
<td>1.260908</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.992947</td>
<td>0.782889</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>6.247292</td>
<td>2.295517</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>11.01344</td>
<td>1.228314</td>
</tr>
<tr>
<td>Probability</td>
<td>0.004059</td>
<td>0.541097</td>
</tr>
<tr>
<td>Sum</td>
<td>6313.500</td>
<td>205.9000</td>
</tr>
<tr>
<td>Sum Sq. Dev.</td>
<td>772.4650</td>
<td>14.30900</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
The summary statistics are explained in Chapter 2 of your textbook (for example, Kurtosis is defined in equation (2.15) on page 25 in Stock and Watson (2011).

If your summary statistics differ, then check the data again. (To return to the data observations, either click on View and then choose Spread Sheet, or simply click on the Sheet button). Once you have located the data problem, click the Edit+/- button on the workfile toolbar, move to the observation in question, enter the correct value, and press Enter. You may want to explore some of the other toolbar buttons to see their functions. CellFmt, for example, allows you to get rid of unnecessary digits after the decimal point, but appears only after you “Freeze” the object and click Edit +/-.

Once you have entered the data, there are various things you can do with it. You may want to keep a hard copy of what you just entered. If so, click on the Print button.

In general, it is a good idea to save the data and your work frequently in some form. Many of us have learned through painful experiences how easy it is to lose hours of work by not backing up data/results in some fashion. There are two ways to save data in EViews. One is to save an entire workfile (Save), and the other is to store individual series (Store).

Press the Save button in the workfile toolbar, or click on File and then SaveAs in the main menu. Follow the usual Windows format for saving files (drives, directories, file type, etc.). If you save workfiles in EViews readable format, then you should use the extension “.WF1.” Once you have saved a workfile, you can call it up the next time you intend to use it by clicking on File and then Open. Try these operations by saving the current workfile under the name “SW10smpl.wf1.”

Alternatively, you may want to just save a few series of the current workfile. The reason is that sometimes you use some of these original series, or transformations of these series, in a different workfile. Let’s save the test score and student-teacher series. First mark the two series in the workfile by clicking on testscr, then hold down the control button and click on str. (Make sure that you are doing this in the Workfile window, not in the Group View window.) After that, press the Store button in the workfile toolbar. Once again, a dialog box will pop up. Store the two data series in the EViews subdirectory with the extension “.db.” Next time you need to retrieve these two series, you can simply click on the Fetch button in the workfile toolbar.

c) Graphical Presentations

Most often it is a good idea to generate graphs (“pictures”) to get some “feel” for the data. You will be able to detect outliers which may be the result of data entry errors or you will be able to see if the data “makes sense.” Although EViews offers many graphing options, there are two that you will use most often: line graphs, where one or more variables are plotted across entities (these will become more important in time series analysis when you are plotting over time), and scatterplots (crossplots), where one variable is graphed against another.
First set the sample to 1-10 either by clicking on the Sample button in the workfile toolbar and typing in “1 10” under Sample range pairs or by entering “smpl 1 10” in the command window. (The command window is the white box located directly under the main menu and is where you will type all commands.) Then type, in the next line, the command “freeze(graph_str) str.line” in the command line. This will create the line graph and give it a name (graph_str here, but other names, such as graph_1 or mygraph, can be chosen instead). Think of freezing an object as taking a photograph of it and giving it a name. This allows you to locate it easily in your photo album later. You can also edit the photograph later. Most importantly, you can cut and paste it into your word-processing program. “graph_str” now appears in the workfile window. Double click on it to see the graph you just created. In the future and when in interactive use, you will most often work in the command window rather than clicking on buttons.

After the graph appears, either double click on the graph or click on the Options button, and alter it until it looks like the one below. Some of the alterations can be made in the resulting dialog box; others, such as text inserted, title of the graph, etc., have to be edited in.

![Graph 1: Student-Teacher Ratio Across 10 School Districts](image)

Typically we are interested either in causal relationships between variables or in the ability of one variable to predict (later, in time series, forecast) another, it is a good idea to plot two variables together. Commands, such as line, can often be modified by an option in parentheses.

---

2 Make sure you press enter after each command line for the program to register it. If you make a typo, you can go back and correct your mistake, but be sure to press enter again for the computer to read the command. When typing commands, be careful where you put spaces, because the computer is sensitive to these and “freeze(graph)” is not the same as “freeze (graph)”. Oftentimes, the underscore “_” is used in a variable name in place of an actual space.

3 Alternatively the same graph can be generated by marking the variable testscr first and then double clicking on it. In the resulting Series window, click on View/Graph/Line&Symbol. You can then freeze the graph by clicking on the Freeze button.
In this case, “m” means “display multiple graphs.” Use the line command to generate the graph below. This will require you first to define or create a “group” by giving it a name (here size_perform but others, such as mygroup are possible). Next you tell the program which series form the group, here str and testscr. Then “freeze” the graph as before.

The line commands are

```plaintext
group size_perform str testscr
freeze(two_series_plot) size_perform.line(m)
```

You should see the following two graphs in your EViews display (I used copy/paste here).

To get an even better idea about the relationship, you can display a two-dimensional relationship in a scatterplot (see p. 92 of your Stock and Watson (2011) textbook). The command is

```plaintext
4 Pushing buttons is relegated to footnotes from here on. You should work with commands now. If you have to, mark testscr and str, opening the two variables as a group, then select View/Multiple Graphs/Line).
```
where size_perform refers to the name of a previously created group.\textsuperscript{5}

(Not to worry about the positive slope here. Remember, this is a sample, and a very small one at that. After all, you may get 10 heads in 10 flips of a coin.)

d) Simple Regression

There is a commonly held belief among many parents that lower student-teacher ratios will result in better student performance. Consequently, in California, for example, all K-3 classes now have a maximum student-teacher ratio of 20 ("Class Size Reduction Act" – CSR).

For the 10 school districts in our sample, we seem to have found a positive relationship between larger classes and poor student performance. Not to worry – we will soon work with all 420 observations from the California School Data Set, and we will then find the negative relationship you have seen in the textbook – for now, we are more concerned about learning techniques in EViews.

\textsuperscript{5} Alternatively select View/Graph/Scatter/Details: Fit Lines: Regression Line. Choose None in the resulting Global Fit Options Box.
In the previous section, we included a regression line in the scatterplot, something that you should have encountered towards the end of your statistics course. However, the graph of the regression line does not allow you to make exact quantitative statements about the relationship; you want to know the exact values of the slope and the intercept. For example, in general applications, you may want to predict the effect of increase by one in the explanatory variable (here the student-teacher ratio) on the dependent variable (here the test scores).

To answer the questions relating to the more precise nature of the relationship between large classes and poor student performance, you need to estimate the regression intercept and slope. A regression line is little else than fitting a line through the observations in the scatterplot according to some principle. You could, for example, draw a line from the test score for the lowest student-teacher ratio to the test score for the highest student-teacher ratio, ignoring all the observations in between. Or you could sort the data by student-teacher ratio and split the sample in half so that the observations with the lowest ten student-teacher ratios are in one set, and the observations with the highest ten student-teacher ratios are in the other set. For each of the two sets you could calculate the average student-teacher ratio and the corresponding average test score, and then connect the two resulting points. Or you could just eyeball the relationship. Some of these principles have better properties than others to infer the true underlying (population) relationship from the given sample. The principle of ordinary least squares (OLS), for example, will give you desirable properties under certain restrictive assumptions that are discussed in Chapter 4 of the Stock/Watson textbook.

Back to computing. If the dependent variable, $Y$, is only determined by a single explanatory variable $X$ in a linear fashion of the type

$$Y_i = \beta_0 + \beta_1 X_i + u_i i=1,2, ..., N$$

with “$u$” representing the error, or random disturbance, not accounted for by the linear equation, then the task is to find some value for $\beta_0$ and $\beta_1$. If you had values for these coefficients, then $\beta_1$ describes the effect of a unit increase in $X$ on $Y$. Often a regression line is a linear approximation to an underlying relationship and the intercept $\beta_0$ only has a useful meaning if observations around $X=0$ occur in the data. As we have seen in the scatterplot above, there are no observations around the student-teacher ratio of zero, and it is therefore better not to interpret the numerical value of the intercept at all. Your professor most likely will give you a serious penalty in the exam for interpreting the intercept here because with no students present, there is no score to record. (What would be the function of the teacher in that case?)

There are various ways to estimate the regression line. The command for regressing a variable $Y$ on a constant (intercept) and another variable $X$ is:

```
ls Y c X
```
where “ls” stands for least squares. Here, working with the *command* window,\(^6\) type

\[
ls(h) \text{ testscr } c \text{ str}
\]

where the “\(h\)” in parentheses indicates that you are using heteroskedasticity-robust standard errors (“\(c\)” stands for the intercept).

The output appears as follows:

```
Dependent Variable: TESTSCR
Method: Least Squares
Date: 01/18/11   Time: 19:52
Sample: 1 10
Included observations: 10
White Heteroskedasticity-Consistent Standard Errors & Covariance

                                 Coefficient  Std. Error  t-Statistic  Prob.
C                           618.8527   51.06075   12.11993       0.0000
STR                     0.606961   2.333492 0.260108       0.8013

R-squared       0.006824
Mean dependent var  631.3500
Adjusted R-squared  -0.117323
S.D. dependent var  9.264418
S.E. of regression   9.792813
Akaike info criterion  7.578031
Sum squared resid    767.1935
Schwarz criterion    7.638548
Hannan-Quinn criter. 7.511644
F-statistic          0.054969
Durbin-Watson stat   0.820522
                    0.853391
```

According to these results, lowering the student-teacher ratio by one student per class results in an decrease of 0.6 points, on average, in the districtwide test score. Using the notation of your textbook, you should display the results as follows:

\[
\bar{\text{TestScore}} = 618.9.1 + 0.61 \times \text{STR}, \quad R^2 = 0.007, \quad \text{SER} = 9.8
\]

\[
(51.1) \quad (2.33)
\]

\(^6\) If you are working in a Group Window, possibly by having invoked the Show option, then click on Proc. Next press Make Equation, and a dialog box will open. If EViews has not suggested a regression of the test score on the student-teacher ratio plus a constant (“\(C\); this letter is reserved in EViews for the constant – actually a vector of ones – and you are not allowed to give another variable this name), then type in the variable names in that order (EViews takes the first variable as the dependent variable; it does not matter if you place the constant before the explanatory variable or after). Alternatively, start in the Main menu and click on Object and the New Object and finally Equation. The same dialog box will open.
Note that the result for the 10 chosen school districts is quite different from the sample of all 420 school districts. However, this is a rather small sample and the regression $R^2$ is quite low. As a matter of fact, in Chapter 5 of your textbook, you will learn that the above slope is not statistically significant.

e) Entering Data from a Spreadsheet

So far you entered data manually. Most often you will work with larger data sets that are external to the EViews program, i.e., they will not be included in, or be part of, the program itself. This makes sense as data sets either become very large or are generated by another program, such as a spreadsheet.

Stock and Watson present the California test score data set in Chapter 4 of the textbook. Locate the corresponding Excel file caschool.xls and open it. Next, following the procedures discussed previously, open a new EViews workfile with 420 observations, and use the Quick/Edit Group (Empty Series) procedure. Return to the Excel file and mark F2:R421. Next, using the “copy” and “paste” commands common to Windows programs, move the data block to EViews. You presumably are familiar with this procedure. Make sure to select the grey box to the immediate right of “obs” before pasting (this will highlight that column). Next “rename” the 13 variables according to the name in the cells F1:R1.

This is what you should see in EViews:

14
When you are done, you are ready to save the workfile. Name it *caschool.wfl*.

You can now reproduce Equation (4.7) from the textbook. Use the regression command you previously learned to generate the following output (“Freeze” the output and use the CellFmt button to adjust the number of digits after the decimal point).

![Regression Output](image)

(You can find the standard errors and the $t$-statistic on p. 129 of the Stock and Watson (2011) textbook. The regression $R^2$, sum of squared residuals ($SSR$), and standard error of the regression ($SER$) are presented in Section 4.3.)

**f) Importing Data Files directly into EViews**

Even though the *cut and paste* method seemed straightforward enough, there is a second, more direct way to import data into EViews from Excel, which does not involve copying and pasting data points.

Start again with a new workfile in EViews. Next press **Proc /Import /Read Text-Lotus-Excel**. A dialog box will open, and you will first have to specify the location where your data file (caschool.xls) resides. After you double click on the file, another dialog box opens.  

---

*There is a slight problem here if you use EViews 6, rather than EViews 7. EViews 6 only recognizes .xls Excel files, not .xlsx. In that case, open caschool.xlsx and save it as an “Excel 97-2003 Workbook” which will produce the needed caschool.xls file. You can then proceed as indicated in the paragraph.*
Unfortunately, you do have to provide names for the series though, so copy the names of the variables from cell F1 (enrl_tot) to R1 (math_scr) in your Excel file and paste them into this field. Finally, EViews has suggested that the first data is in cell B2. Change this to F2, the first data point corresponding to enrl_tot. Before you click OK, you need to close the Excel file. The following window is what the dialog box should look like before hitting the return button. Note that EViews also allows you to import other types of data files, e.g. STATA files, although this may be a bit more complicated.

EViews will show that the data exist in the Workfile Window. You may want to check that the data were properly retrieved by typing the command “show testscr str” or running the same regression as before.

You can also save data in ASCII, spreadsheet, STATA, SPSS, and other formats by clicking on File/ Save As and then looking at the various options in “Save as type.”

g) Multiple Regression Model

Economic theory most often suggests that the behavior of a certain variable is influenced not only by another single variable, but by a multitude of factors. The demand for a product depends not only on the price of the product but also on the price of other goods, income, taste, etc. Similarly, the Phillips curve suggests that inflation depends not only on the unemployment rate, but also on inflationary expectation and possibly supply shocks, etc.
An extension of the simple regression model is the multiple regression model, which incorporates more than one regressor (see Equation (6.7) in the textbook on page 189).

\[ Y_i = \beta_{0i} + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \ldots + \beta_{ki}X_{ki} + u_i, \quad i = 1, \ldots, n. \]

To estimate the coefficients of the multiple regression model, you proceed in a similar way as in the simple regression model. The difference is that you now need to list the additional explanatory variables. In general, the command is:

\[ \text{ls}\{\text{options}\} Y c X_1 X_2 \ldots X_k. \]

where \(\text{options}\) can be omitted (this is the default and gives you homoskedasticity-only standard errors) and can be replaced by various possible entries.

See if you can reproduce the following regression output, which corresponds to Column 5 in Table 7.1 of the textbook (page 238). The option used below is \((h)\) to produce heteroskedasticity-robust standard error (EViews refers to these as “White Heteroskedasticity-Consistent Standard Errors”).

\[ \text{Dependent Variable: TESTSCR} \]
\[ \text{Method: Least Squares} \]
\[ \text{Date: 01/20/11  Time: 11:56} \]
\[ \text{Sample: 1 420} \]
\[ \text{Included observations: 420} \]
\[ \text{White Heteroskedasticity-Consistent Standard Errors & Covariance} \]

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>700.39</td>
<td>126.48</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-1.01</td>
<td>-3.77</td>
<td>0.0002</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.13</td>
<td>-3.58</td>
<td>0.0004</td>
</tr>
<tr>
<td>MEAL_PCT</td>
<td>-0.53</td>
<td>-13.87</td>
<td>0.0000</td>
</tr>
<tr>
<td>CALW_PCT</td>
<td>-0.05</td>
<td>-0.82</td>
<td>0.4150</td>
</tr>
</tbody>
</table>

R-squared 0.775 Mean dependent var 654.157
Adjusted R-squared 0.773 S.D. dependent var 19.053
S.E. of regression 9.084 Akaike info criterion 7.263
Sum squared resid 34247.463 Schwarz criterion 7.311
Log likelihood -1520.188 Hannan-Quinn criter. 7.282
F-statistic 357.054 Durbin-Watson stat 1.430
Prob(F-statistic) 0.000
The interpretation of the coefficients is equivalent to that of a controlled science experiment: it indicates the effect of a unit change in the relevant variable on the dependent variable, *holding all other factors constant* (“ceteris paribus”).

Section 7.2 of the Stock and Watson (2011) textbook discusses the *F*-statistic for testing restrictions involving multiple coefficients. To test whether all of the above coefficients are zero with the exception of the intercept, click on **View/Coefficient Tests/Wald-Coefficient Restrictions**. The regression coefficients are stored in a vector $c(1)$ to $c(k+1)$, where the number in parentheses indicates the order of appearance in the regression output. Thus in the example $c(1)$ is the intercept or constant term, $c(2)$ is the coefficient on STR, and so forth. To execute the above test, enter the following and press enter:

The computer will generate the following output:

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Value</th>
<th>df</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F</em>-statistic</td>
<td>361.6835</td>
<td>(4, 415)</td>
<td>0.0000</td>
</tr>
<tr>
<td>Chi-square</td>
<td>1446.734</td>
<td>4</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Null Hypothesis Summary:

<table>
<thead>
<tr>
<th>Normalized Restriction ($= 0$)</th>
<th>Value</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(2)</td>
<td>-1.014353</td>
<td>0.268861</td>
</tr>
<tr>
<td>C(3)</td>
<td>-0.129022</td>
<td>0.036258</td>
</tr>
<tr>
<td>C(4)</td>
<td>-0.528619</td>
<td>0.038117</td>
</tr>
<tr>
<td>C(5)</td>
<td>-0.047854</td>
<td>0.058654</td>
</tr>
</tbody>
</table>

Restrictions are linear in coefficients.

See if you can generate the *F*-statistic of 5.43 following Equation (7.6) in the Stock and Watson.
(2011) text and listed at the bottom of page 223 (restrict the coefficients of $STR$ and $Expn$ to be zero.8

h) Data Transformations

So far, we have only used data in regressions that already existed in some file that we either created or used. Almost always, you will be required to transform some of the raw data that you received before you run a regression. In EViews you transform variables by using the “genr” (as in generate) command. For example, Chapter 8 of the Stock/Watson textbook introduces the polynomial regression model, logarithms, and interactions between variables. Let us reproduce Equations (8.2), (8.11), (8.18), and (8.37) here. The following commands generate the necessary variables9:

```plaintext
genr avginc2=avginc^2
genr avginc3=avginc^3
genr lavginc=log(avginc)
genr ltestscr=log(testscr)
genr strpctel=str*el_pct
```

8 A word of caution here. In the above table, the $F$-statistic is 361.6835. In the regression output above, the same listed $F$-statistic is 357.054, even though it tests for the same restrictions, namely that all slope coefficients are zero. Note that the latter statistic is the homoskedasticity-only $F$-statistic, even though the equation was estimated using heteroskedasticity-robust standard errors.

9 For example, I have generated a variable called “avginc2”, and assigned it to be the square of the previously defined variable “avginc.” Note that I am generating variable names that are self-explanatory. They could have been called “variable1,” “variable2,” “variable3,” etc. but it is a good idea to create variable names that you can remember.
Next run the four regressions using the same technique as for multiple regression analysis. Finally save your workfile again and exit the workfile.

**Exercise**

One of the problems with the type of tutorial you are working on is that you just follow instructions without internalizing them. A typical student will finish the tutorial with few problems but then little is retained. If I asked you to retrieve a data set and to run a few regressions, for example, would you be able to do that? Or would you say “how do I do this?”

Let’s see how much you understood. Go to the Stock and Watson website for the 3rd edition ([http://www.pearsonhighered.com/stock_watson](http://www.pearsonhighered.com/stock_watson)). Go to Student Resources in the Companion Web Site, and download the CPS data set for Chapter 8 (Data Sets for Replicating Empirical Results: CPS Data Used in Chapter 8). Then replicate the results for columns (1) from Table 8.1 on page 284 of the Stock and Watson (2011) textbook. Why do you think your results differ from those listed in the table? What if you found a way to restrict your sample to only include individuals who are at least 30 but not older than 64? To find a way to restrict your sample, look for Help and the *smpl if* command. Then restricting your sample to those individuals in that age group, replicate columns (1) to (3). For column (4), define potential experience as (age – Years of education – 6).
Batch Files

*You will skip this section if you only have the Student Version, since this feature is not available for your version.*

So far, you have either clicked on buttons in EViews or used the “Command Window” to type executable statements. But what if you wanted to keep a permanent record of all the transformations you made, regressions you tried, graphs you created, etc.? In that case, you would need to create a “program” that consists of line commands similar to those that you used in the “Command Window” previously. After having created such a program, you can then execute (“run”) it and view the output afterwards (if you did not make any errors). Batch files can also include loops and conditional branching.

To create a program, click on **File** and then **New** and **Program**. This opens the “Program” box. Type in, or cut and paste, the following commands exactly as they appear below. Use ‘ whenever needed at the beginning of the line to indicate that you have added a comment. Comments are useful if you want to remember later what you were doing or if you want others to understand your program. They do not affect the actual execution of the program. Then save it as “Tutorialch4.prg” in your directory and click the “Run” button.

Make sure to replace the command

```plaintext
open c:\StockandWatsons\caschool.wf1
```

with the relevant directory where you saved the California School Data Set (caschool.wf1) on your computer or the college/university drive.

(To save time, you may want to omit some of the comments below.)

```
*****************************************************************************************
' Stock and Watson
' chapter 4 (EViews 6.0 Version)
' caschool.wf1 is the California School Data Set
' Chapter 4: Linear Regression with One Regressor
*****************************************************************************************
open c:\StockandWatsons\caschool.wf1
*****************************************************************************************
'TESTSCR     AVG TEST SCORE =(READ_SCR+MATH_SCR)/2
' STR         STUDENT TEACHER RATIO (TEACHERS/ENRL_TOT)
'Summary statistics for str and testscr in Table 4.1
```

21
' Define a group and give the table a name (tab4_1).

```
group tab4_1 str testscr
tab4_1.stats
```

' Correlation between str and testscr
' Again, define a group first and then give it a name (cor_str_testscr)

```
group cor_str_testscr str testscr
cor_str_testscr.cor
```

' Figure 4.2: Scatterplot of Test Score and Student-Teacher Ratio

```
group Fig4_2 str testscr
Fig4_2.scat linefit
```

' Equation 4.11
' Here is an example how to use OLS in the program.
' You first define the equation and then use the ls command.

```
equation eq4_11.ls(h) testscr c str
```

' crossplot Figure 4.3 with regression line

```
group Fig4_3 str testscr
Fig4_3.linefit
```

' End of Chapter 4

' Equation 5.18 (chapter 5)

' Below a binary variable is defined first by setting it to zero for the entire sample
' then to set it to one for observations where the student-teacher ratio is less than 20.

```
genr dsize=0
smpl if str<20
ngen dsize=1
smpl 1 420
equation eq4_33.ls(h) testscr c dsize
```

22
In the “caschool.wf1” workfile, you can now click on the equations or tables you have generated. “Eq4_11,” for example, has reproduced Equation (4.11) from your textbook. This is identical to the regression generated above.

A summary of frequently used EViews commands is given at the end of the tutorial. As an exercise, start generating the equations, graphs, and tables of chapters 5-9 in Stock and Watson (2011).

3. TIME SERIES DATA

Let’s leave the cross-sectional data and move on to Chapter 14 in the textbook.

In the time series chapters of the textbook, you will use past values or lags of variables to forecast the dependent variable or for data transformations. We refer to “(t-1)” as a lag (similarly, “(t+1)” is a lead). Imagine you had entered the data for the CPI, but you wanted to forecast the inflation rate or the change in the inflation rate. The next step is therefore to transform the raw data. Specifically,

$$\text{Inf}_t = \frac{\text{CPI}_t - \text{CPI}_{t-1}}{\text{CPI}_{t-1}} \times 100 = \left(\frac{\text{CPI}_t}{\text{CPI}_{t-1}} - 1\right) \times 100$$

To create past values of variables, you generate a lag by adding a “(-1)” after its variable name in the “genr” statement. In a spreadsheet, this amounts to copying an entire data series and pasting it into a new column one observation down: the first observation becomes the second observation, etc. The procedure generalizes to higher lags: X_{t-12} is X(-12).\(^{10}\)

Type the following lines of code into a new program and run it (you can ignore the comments – they are there mainly for you). See if you understand the code in terms of generating the inflation rate and its change. Because the short-run Phillips curve suggests a negative relationship between the change in the inflation rate (perhaps the inflation rate and the expected inflation rate, where the latter is generate through static expectations) and the unemployment, we also need to calculate the change in the inflation rate, or \(\Delta \text{Inf}_t = \text{Inf}_t - \text{Inf}_{t-1}\). In the program below, this variable is called “dinf.” You need the workfile “macro_3e.wf1,” which you can download from the Web site.

After running the program, click on some of the figures and equations to view the output.

\(^{10}\) In mathematics, a lag is defined (loosely) through the use of a “lag-operator” \(L\), where \(L x_t = x_{t-1}\). Similarly, the “difference operator” \(\Delta = (1 - L)\), so that \(\Delta x_t = x_t - x_{t-1}\). See Appendix 14.3 of the textbook for more details.
Stock and Watson

Chapter 14 (some results)

open c:\stockandwatson\macro_3e.wf1

' Note: you must replace the above command with the appropriate
path where you have stored the workfile

' LHUR   Unemployment Rate U.S.
PUNEW   Consumer Price Index U.S.
FYFF   Federal Funds Interest Rate U.S.
FYGM3   3-Month Treasury Bill Interest Rate U.S.
FYGT1   1-Year Treasury Bond Interest Rate
EXRUK   Dollar-Pound Exchange Rate
GDP_JP   Real GDP Japan

smpl 1959:1 2005:4

generate the annualized inflation rate and its change

genn lpunew=log(punew)
genr inf=400*(lpunew-lpunew(-1))
genr dinf=inf-inf(-1)
genr yeardinf=inf-inf(-4)

' Figure 14.1

smpl 1960:1 1999:4
group Fig14_1 inf lhur
Fig14_1.line(m)
freeze(InflatUR) Fig14_1

' Equation 14.13

smpl 1962:1 2004:4

EViews allows you to use (-i) in an equation on variables
that you had not generated previously

equation eq14_13.ls(h) dinf c dinf(-1 to -4)

' Figure 14.13

group Fig14_3 lhur(-4) yeardinf
Fig14_3.scat linefit
freeze(Phillips) Fig14_3

' Equation 14.17

equation eq14_17.ls(h) dinf c dinf(-1 to -4) lhur(-1 to -4)
4. SUMMARY OF FREQUENTLY USED EVIEW COMMANDS

The command ‘genr’ creates new variables and modifies existing variables.

Examples:

\[
\text{genr expn} = \frac{\text{expn}_{\text{stu}}}{1000}
\]

generates the expenditure variable used in the textbook by dividing the original data by 1,000.

\[
\text{genr avginc}^2 = \text{avginc}^2
\]
\[
\text{genr lavginc} = \log(\text{avginc})
\]

create the square and log of average income, respectively.

Note that commands of the type

\[
\text{genr testscr} = \frac{\text{testscr}}{100}
\]

simply modify an existing variable.

The most frequently used operators are + (addition), - (subtraction), * (multiplication), / (division), ^ (exponentiation). Log(x) calculates the natural logarithm of x (see the above example) and exp(x) computes the exponent of x.

When working with time series data, lags are frequently used. EViews allows you to create these simply by entering (-i) immediately after the variable name:

\[
\text{genr dinf} = \text{inf} - \text{inf}(-1)
\]
\[
\text{genr yeardinf} = \text{inf} - \text{inf}(-4)
\]

The first command generates the quarterly change in the inflation rate (assuming that you work with quarterly data), while the second generates the annual change in the inflation rate.

The sample range is set through the ‘smpl’ command. The command is of the type: smpl n1 n2, where n1 and n2 are the beginning and end dates (first and last observations) for which EViews will execute the commands that follow.

Examples are

\[
\text{smpl 1 420}
\]
\[
\text{smpl 1959:1 2001:4}
\]

In the first case, EViews is instructed to use all 420 observations of the California Test Score Data Set used in Chapters 4-9. The second example restricts the sample to the first quarter of
1959 to the last quarter of 2001.

Note that you can work with a subsample by using relational operators.

\[ \text{smpl if str} \geq 20 \]

only looks at observations with a student-teacher ratio of less than 20.

The most frequently used \textit{statistical operations} involve running regressions (‘\textit{ls}’), establishing the correlation between variables (‘\textit{cor}’), and graphing variables (‘\textit{line}’).

\[ \text{EViews creates results by storing them in so-called objects. Initially, you will use the ‘equation’ object and the ‘group’ object most often, as in the following examples:} \]

\[ \text{equation eq4_7.ls(h) testscr c str} \]
\[ \text{equation eqtab5_2_5.ls(h) testscr c str el_pct meal_pct calw_pct} \]
\[ \text{equation eq12_7.ls(h) dinf c dinf(-1)} \]

In each case, an equation object is declared first and a name is assigned to it. ‘\textit{ls}’ then instructs EViews to use OLS estimation for the equation. The dependent variable appears first, followed by the regressors, where ‘\textit{c}’ is used for the intercept (‘\textit{c}’ is a reserved name in EViews, meaning that you cannot use it to generate a variable called ‘\textit{c}’).

To create a line graph or to view the correlation between variables, you first must assign the variables to a group and name this group. Next, you execute the correlation and graphing through the ‘\textit{cor}’ and ‘\textit{line}’ command. Examples:

\[ \text{group cor_str_testscr str testscr} \]
\[ \text{cor_str_testscr.cor} \]

Here the variables \textit{str} (student-teacher ratio) and \textit{testscr} (test score) are assigned to a group called \textit{cor_str_testscr} (the name was chosen to indicate what the group was used for, but we could have named it almost anything alternatively), and EViews is then instructed to calculate the correlation between the variables in the group. The group can contain more than two variables.

In the following example, \textit{inf} (inflation) and \textit{lhur} (unemployment rate) are assigned to a group called Fig12_1 and are then plotted (where ‘\textit{m}’ is an option that allows for the display of multiple graphs).

\[ \text{group Fig12_1 inf lhur} \]
\[ \text{Fig12_1.line(m)} \]

Another topic you may come across is statistical distribution functions. EViews provides functions that provide access to the density or probability functions, cumulative distribution, quantile functions, and random number generators for a number of standard statistical
distributions. For example, the following command would tell you that $\chi^2 \Pr(Y \leq 7.78) = 0.90$:

```
genr result=@cchisq(7.78,4)
```

For a complete table and descriptions of distribution functions you can use, search for “statistical distribution functions” under the Help section.

5. FINAL NOTE

For a complete list of commands, consult the EViews Command and Programming Reference or the User’s Guide. Alternatively, use the “Help” command inside EViews. Under the Find tab in EViews Help Topics, you can search for whatever you are looking for. As mentioned before, this tutorial is not intended to replace the Reference or User’s Guide. The best way to learn how to use the program is to spend some time exploring and playing with it.

EViews replication batch files for all the results in the Stock/Watson textbook are available from the Web site. You are invited to download these and study them.