

# EViews 3.1 Tutorial

by Manfred W. Keil

to Accompany

## Introduction to Econometrics

by James H. Stock and Mark W. Watson

### 1. EViews: INTRODUCTION

This tutorial will introduce you to a statistical and econometric software package called EViews, or Econometric Views. EViews runs on both the Windows (9x, Me, NT 4.0 or 2000) and Macintosh platforms. 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](http://www.eviews.com). The program comes with two manuals, a *User's Guide* and a *Command and Programming Reference*. Both manuals can be ordered separately (\$40 each, \$75 for the pair) by calling (949) 856-3368 or writing to [sales@eviews.com](mailto:sales@eviews.com). The *User's Guide* is better for first-time users. There is also a cheaper student version (EViews 3.1 Student Version for Windows 9x, 2000, NT 4.0, \$39.95) and EViews Basics available. The difference between the student version and the full version is in the limitation on the size of data sets ("capacity limitation" is 1,000 observations for each series and no more than 10,000 observations for all series) and the availability of some estimation methods such as ARCH (discussed in Chapter 14), GMM, SURE, FIML, 3SLS, FIML and TSLS system estimators (don't worry what these estimation methods stand for at this point), as well as some of the estimators discussed in the appendix to Chapter 9. Furthermore, and perhaps most importantly for you right now, the student version does not allow you to run EViews in "batch mode." 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, you will almost certainly run programs in batch mode.<sup>1</sup>

Econometrics deals with three types of data: cross-sectional data, time series data, and panel data, or longitudinal data (see Chapter 1 of the 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 rates and inflation and unemployment rates; to quarterly data (GDP)

---

<sup>1</sup> All results in this tutorial were computed using EViews 3.1. This was done to allow the tutorial to be used by students who either have bought the EViews 3.1 Student Version and/or use EViews in labs/networks that have not upgraded from the previous version. Some of the features missing from Version 3.1 are the DF-GLS test discussed in Chapter 14 and some statistical features, such as computing summary statistics by quantiles (see, for example, Table 4.1).

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 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 the most commonly used econometrics package for time series analysis in academics, business, and government. It can also be used for cross-section and panel data. EViews allows you to save results within a program and to “retrieve” these results for further calculations. 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 execute the sequence of statements. 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.

EViews is not particularly good at graphing; actually, the graphs look a bit unprofessional. If you need to graph data, it is best to save it in a spreadsheet or ASCII format, which EViews allows you to do, 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 ([manfred.keil@claremontmckenna.edu](mailto:manfred.keil@claremontmckenna.edu)).

## 2. CROSS-SECTIONAL DATA

### Interactive Use

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.



(Your command window may appear smaller, but you can enlarge it using the usual windows methods.)

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*.

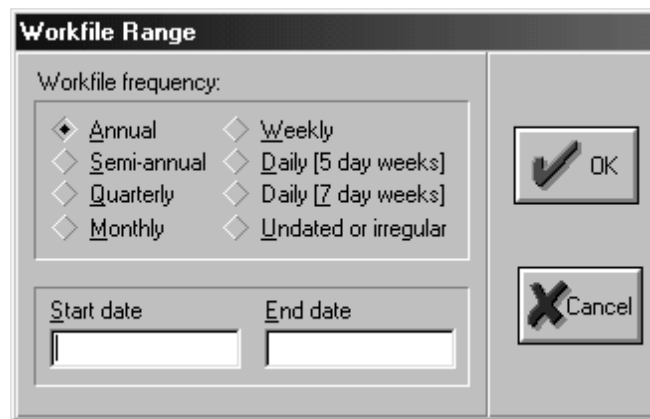
In this tutorial, we will work with two data applications, one cross-sectional (student test scores and student-teacher ratio) and one time series (forecasting inflation).

## Data Input and Simple Data Analysis

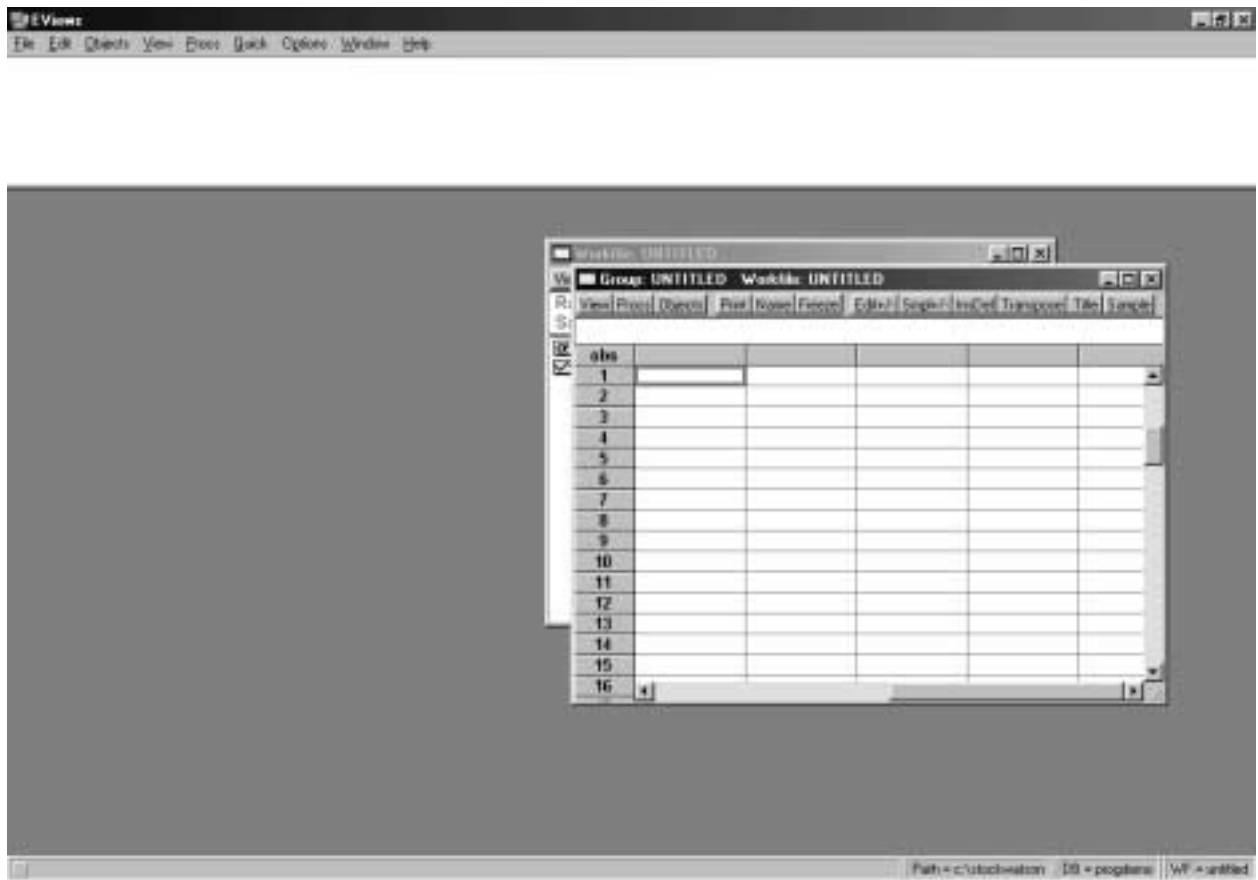
### a) The Easy and Tedious Way: Clicking and Entering

In Chapters 4 to 7 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, but for the purpose of this introduction it will be useful that you become aware of this possibility. As a result, I will use a sub-sample of 20 observations from this data set.

To start, we must establish 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*.



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, so check this field and enter 20 in the “End date” box. You will see a workfile window, which contains two entries. 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 variables, starting with the name, in the grey box to the immediate right of “obs.” (Click on the grey box and enter first “testscr” and “str” in the adjacent one to the right.)

### *Manual Data Entry*

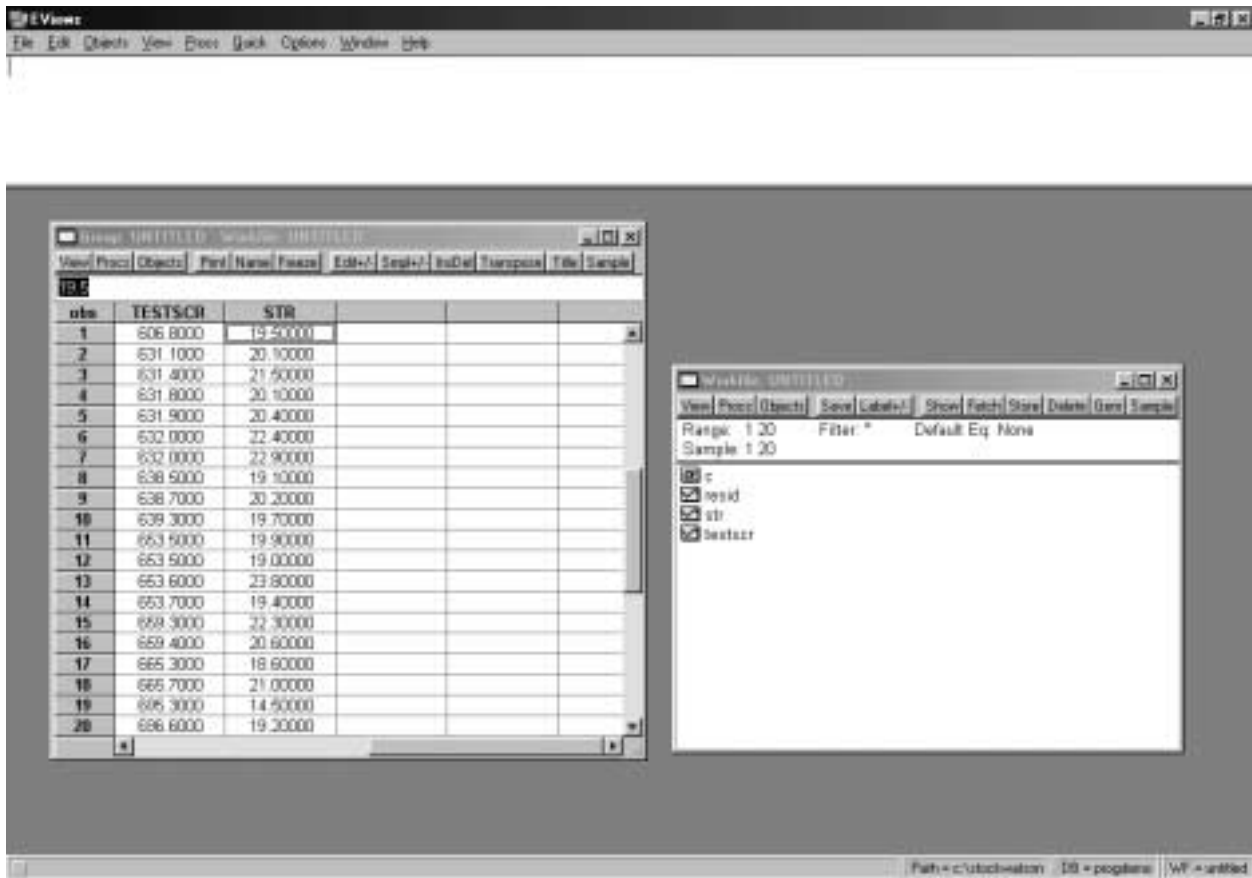
Here are the data to enter. (EViews will add zeros. You will see later how to get rid of these.)

obs	TESTSCR	STR
1	606.8	19.5
2	631.1	20.1
3	631.4	21.5
4	631.8	20.1
5	631.9	20.4
6	632.0	22.4
7	632.0	22.9
8	638.5	19.1
9	638.7	20.2
10	639.3	19.7
11	653.5	19.9
12	653.5	19.0
13	653.6	23.8

14	653.7	19.4
15	659.3	22.3
16	659.4	20.6
17	665.3	18.6
18	665.7	21.0
19	695.3	14.5
20	696.6	19.2

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.

This is what you should see after completing the data entry:



## 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 workfile, click on it, and then click on **Descriptive Statistics** and **Common sample**. You should see the following output:

The screenshot displays the EViews software interface. The main window shows a table of summary statistics for two variables: TESTSCR and STR. The statistics are as follows:

	TESTSCR	STR
Mean	648.4700	20.21000
Median	646.4000	20.10000
Maximum	696.6000	25.60000
Minimum	606.8000	14.50000
Std. Dev.	21.92583	1.948602
Skewness	0.634227	-0.783309
Kurtosis	3.330038	5.134627
Jarque-Beta	1.431582	5.842082
Probability	0.488806	0.063878
Observations	20	20

The interface also shows a toolbar with buttons for 'Show', 'Fetch', 'Store', 'Delete', 'Data', and 'Sample'. The status bar at the bottom indicates the path 'c:\stata\work\10 - popline\wf - untitled'.

If your summary statistics differ, then check the data again. 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. **Number**, for example, allows you to get rid of unnecessary digits after the decimal point, but appears only after you “freeze” the object. Once you freeze an object, you can cut and paste it into your word processing file.

Once you have entered the data, there are various things you can do with it. First, let's get back to the data. Click on **View** and then choose **SpreadSheet**. This allows you to see the data again. 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 “SW20smpl.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*. 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.

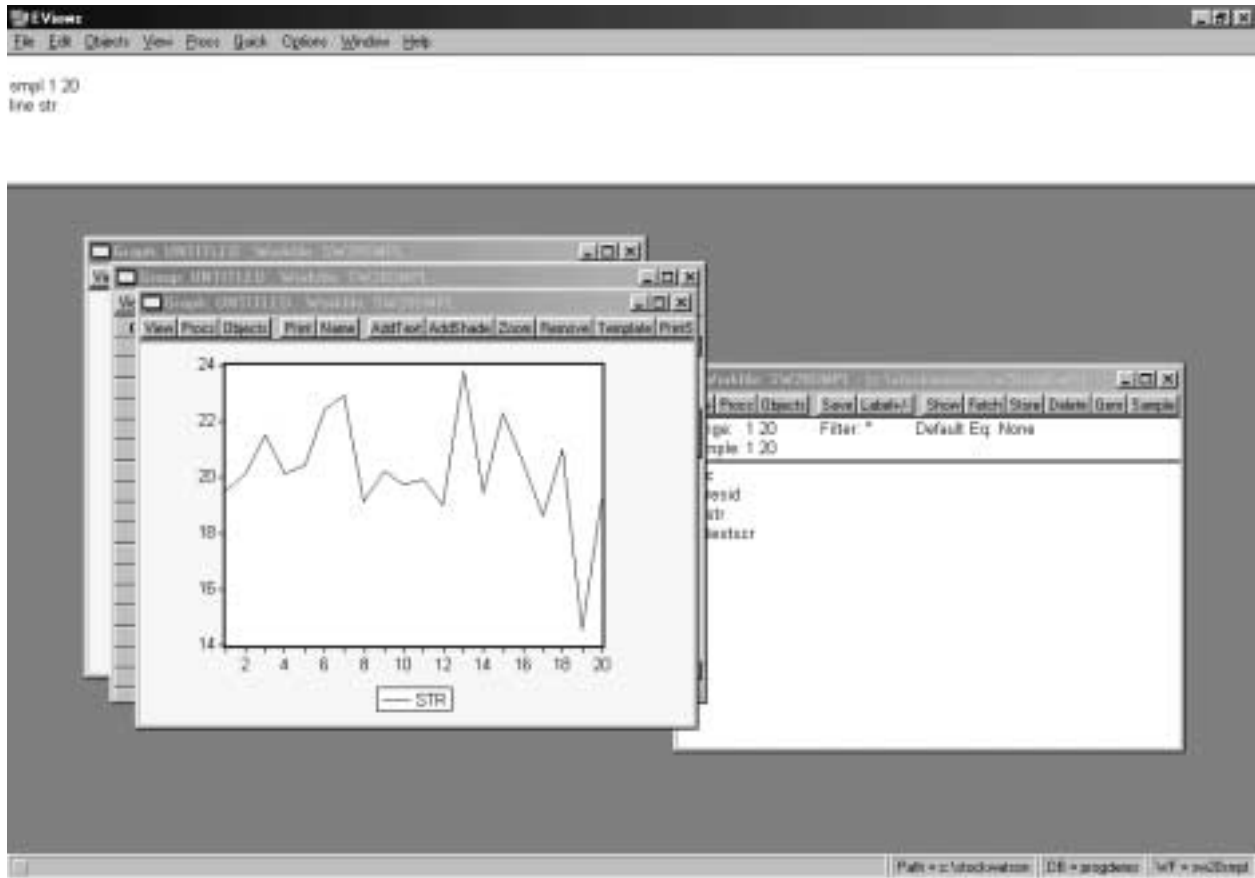
### *Graphical Presentations*

Most often it is a good idea to generate graphs (“pictures”) to get some “feel” for the data. 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, and scatterplots (crossplots), where one variable is graphed against another.

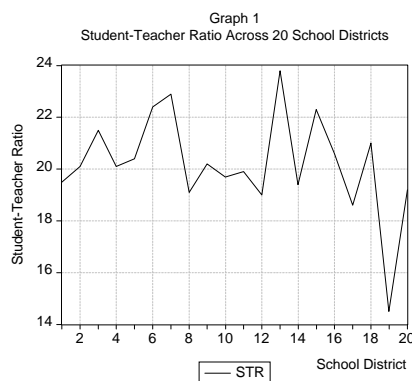
First set the sample to 1-20 either by clicking on the **Sample** button in the workfile toolbar or by entering “smpl 1 20” in the command line. Then type the command “line str” in the command line.<sup>2</sup> In the future, in interactive use you will most often work in the *command window* rather than clicking on buttons.

---

<sup>2</sup> Alternatively the same graph can be generated by marking the variable *str* first and then double clicking on it. In the resulting *Series* window, click on **View /Line Graph**.

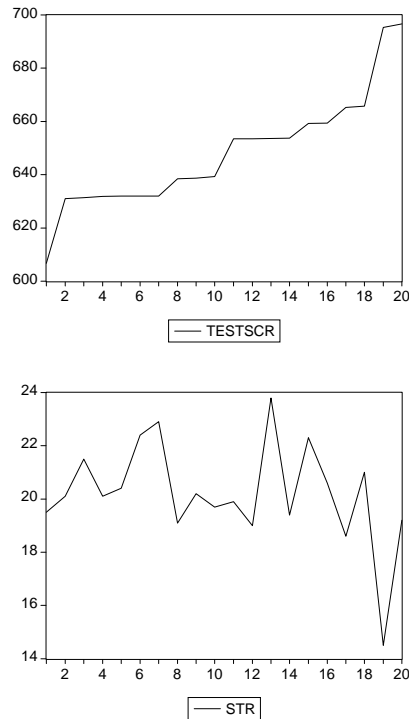


After the graph appears, double click on the graph and alter it until it looks like the one below. Some of the alterations can be made in the resulting dialog box; others, such as text inserted, title of the graph, etc., have to be edited in.



Because in general we are interested either in causal relationships between variables or in the ability of one variable to 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. In this case, “m” means “display

multiple graphs.” Use the line command to generate the graph below.<sup>3</sup>

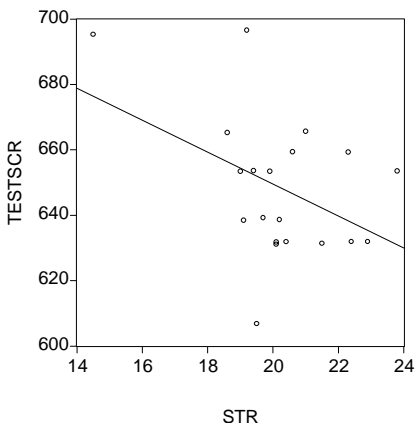


To get an even better idea about the relationship, you can display a two-dimensional relationship in a scatterplot. The command is *scat series1 series2*, where *series1* is used for the horizontal axis.<sup>4</sup> Adding the option (*r*) fits a linear (regression) line through the points.

---

<sup>3</sup> Pushing buttons is relegated to footnotes from here on. You should work with commands now. If you have to, mark *testscr* and *str* this time, select **View** and then **Graph** in the Group Window (or in one graph below the other: select **View** and then **Multiple Graphs**).

<sup>4</sup> Double click on the graph that displays the test score and the student-teacher ratio in the same graph. The dialog box opens. In the upper-left corner choose **Scatter Graph** under **Graph Type**. Experiment with the options, including the **Freeze** button, until you can replicate the graph above.



### *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. For the 20 school districts in our sample, we seem to have confirmed the existence of a negative relationship between large classes and poor student performance. We even 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 may want to predict what the effect of reduction by one in the student-teacher ratio would be.

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

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

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,  $\beta_1$  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. Working with the *command* window,<sup>5</sup> type

```
ls(h) testscr c str
```

where the “h” in parentheses indicates that you are using heteroskedasticity-robust standard errors. The output appears as follows:

**Dependent Variable: TESTSCR**

**Method: Least Squares**

**Date: xx/xx/xx Time: xx:xx**

**Sample: 1 20**

**Included observations: 20**

**White Heteroskedasticity-Consistent Standard Errors & Covariance**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	747.0976	47.19888	15.82872	0.0000
STR	-4.880141	2.284288	-2.136395	0.0466
<b>R-squared</b>	<b>0.188297</b>	<b>Mean dependent var</b>	<b>648.4700</b>	
<b>Adjusted R-squared</b>	<b>0.143203</b>	<b>S.D. dependent var</b>	<b>21.92583</b>	
<b>S.E. of regression</b>	<b>20.29528</b>	<b>Akaike info criterion</b>	<b>8.953294</b>	
<b>Sum squared resid</b>	<b>7414.174</b>	<b>Schwarz criterion</b>	<b>9.052867</b>	
<b>Log likelihood</b>	<b>-87.53294</b>	<b>F-statistic</b>	<b>4.175609</b>	
<b>Durbin-Watson stat</b>	<b>0.484121</b>	<b>Prob(F-statistic)</b>	<b>0.055929</b>	

According to these results, lowering the student-teacher ratio by one student per class

<sup>5</sup> If you are working in a Group Window, possibly by having invoked the **Show** option, then click on **Procs**. 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 **Objects** and the **New Object** and finally **Equation**. The same dialog box will open.

results in an increase of almost five points, on average, in the districtwide test score. Using the notation of your textbook, you should display the results as follows:

$$\widehat{TestScore} = 747.1 - 4.88 \times STR$$

(47.2) (2.28)

Note that the result for the 20 chosen school districts is more than twice as strong as for the sample of all 420 school districts. However, this is a rather small sample and the standard error of the estimator is relatively large.

#### *a) 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. This is what you should see in EViews:

EVIEWS - Group: UNTITLED Workfile: UNTITLED

File Edit Objects View Proc Quick Options Window Help

View Proc Objects Panel Name Prices Edit(s) Single Initial Comments Tab Sample

obs	SER01	SER02	SER03	SER04	SER05	SER06	SER07	SER08	SER09	SER10	SER11
1	195.0000	10.90000	0.510200	2.040800	67.00000	690.8000	0.343500	6394.911	17.86991	22.69000	0.00000
2	240.0000	11.19000	95.41670	47.91670	101.0000	661.2000	0.420833	5099.361	21.52996	9.824000	4.560000
3	1950.000	62.90000	95.03230	76.32260	169.0000	649.6000	0.100032	5901.966	18.69723	8.979000	30.00000
4	243.0000	14.00000	36.47540	77.04920	95.00000	647.7000	0.348794	7101.631	17.36714	8.070000	0.00000
5	1336.000	71.90000	33.10860	70.42700	171.0000	640.9500	0.128090	6235.960	18.67133	9.000333	13.66000
6	137.0000	6.400000	12.31890	86.99660	25.00000	605.5500	0.182482	6990.147	21.40635	10.41900	12.40000
7	196.0000	10.00000	12.80320	94.62370	28.00000	606.7500	0.143690	5253.331	19.50000	6.577000	68.71000
8	968.0000	42.50000	19.80630	100.0000	66.00000	609.0000	0.074324	4665.746	20.89412	8.174000	46.96000
9	379.0000	19.00000	32.19000	93.13980	35.00000	612.5000	0.092348	5395.548	19.94737	7.399500	30.00000
10	2247.000	108.0000	79.98420	87.31640	0.000000	612.6500	0.000000	5036.211	20.80696	11.61333	40.27000
11	446.0000	21.00000	19.60990	85.87440	86.00000	615.7500	0.150825	4647.662	21.23809	8.931000	52.91000
12	967.0000	47.00000	71.71310	96.60560	56.00000	616.3000	0.066738	5447.346	21.00000	7.389500	54.60000
13	103.0000	5.000000	22.42990	98.13080	25.00000	616.3000	0.242718	6667.149	20.60000	5.339500	42.71000
14	487.0000	24.34000	24.62940	77.14640	0.000000	616.3000	0.000000	4816.613	20.06822	8.279000	20.53000
15	649.0000	36.00000	14.83790	76.27120	31.00000	616.4500	0.047786	5621.456	18.02776	9.630000	80.12000
16	652.0000	42.07000	24.21420	94.29670	80.00000	617.3500	0.093997	6026.360	20.25196	7.464000	49.41000
17	491.0000	26.92000	11.20160	87.79970	100.0000	618.0500	0.203666	6723.230	16.97707	6.216000	66.50000
18	421.0000	25.90000	8.951100	77.90970	90.00000	618.3000	0.118795	6689.666	16.50900	7.764000	60.96000
19	6980.000	303.0300	21.28240	94.97120	960.0000	619.8000	0.138635	6064.616	22.70402	7.023000	77.00000
20	2688.000	136.0000	23.43790	93.22620	139.0000	620.3000	0.061711	5433.593	19.91111	5.699000	49.81000
21	440.0000	24.00000	34.77270	100.0000	69.00000	620.5000	0.156816	5725.563	18.33333	7.941000	40.66000
22	475.0000	21.00000	21.64990	91.54640	53.00000	621.4000	0.111579	4642.105	22.61905	9.630000	16.21000
23	2538.000	130.5000	18.91110	70.81670	168.0000	621.7500	0.066698	5107.066	19.44826	7.409500	45.07000
24	476.0000	19.00000	43.86590	100.0000	0.000000	622.0500	0.000000	4699.662	25.05263	9.630000	39.00000
25	2367.000	114.0000	16.80100	90.62370	216.0000	622.6000	0.091642	4655.464	20.67544	8.019000	76.66000
26	1588.000	85.00000	22.40720	85.14720	196.0000	623.1000	0.124685	5415.153	18.66235	8.523000	40.41000
27	7306.000	319.8000	17.00190	89.03490	742.0000	623.2000	0.101580	4897.672	22.84653	7.983181	73.72000
28	2601.000	136.0000	16.07110	92.19630	269.0000	623.4500	0.103422	5223.912	19.26667	7.306000	70.01000
29	647.0000	64.00000	16.29280	90.20070	67.00000	623.6000	0.079103	5139.166	19.25000	8.934000	56.96000
30	452.0000	22.00000	14.48880	81.02960	95.00000	624.1500	0.121681	4614.262	20.54646	8.564000	11.06000
31	4142.000	201.0000	35.56290	81.59660	969.0000	624.9500	0.137373	5342.233	20.62697	8.613000	80.41000
32	2102.000	99.79000	15.21990	80.20490	224.0000	624.9500	0.106666	5347.460	21.07260	12.40900	63.17000
33	10012.00	464.9000	29.76390	91.69340	721.0000	625.3000	0.072014	5036.459	21.52681	8.126616	66.17000
34	2486.000	126.0000	12.69200	95.09300	202.0000	625.9500	0.081190	5117.142	19.90400	11.43100	53.41000
35	*										

File c:\school\ev... 10 x program WF = edited

Next you need to rename Ser01 to Ser13 with the names as they appeared in the original Excel worksheet. The names were as follows:

enrl\_tot teachers calw\_pct meal\_pct computer testscr comp\_stu expn\_stu str avginc el\_pct read\_scr math\_scr

In the EVIEWS program, click on *Ser01*, type *enrl\_tot*, and press enter. You will be asked if you want to replace SER01 with ENRL\_TOT. Say yes and continue. EVIEWS is ready to receive the second variable name (*teachers*), and so on. Write in all the variable names. When you are done, you are ready to save the workfile. Name it SWcaschool.wf1.

Views: [Group: UNTITLED] [Worksheet: UNTITLED]

File Edit Objects View Tools Quick Options Window Help

View [Data Objects] [Field Names] [Fields] [Edit] [Setup] [InDel] [Functions] [Title] [Samples]

obs	ENR1_TOT	TEACHERS	CALW_PCT	MEAL_PCT	COMPUTER	TESTSCR	COMP_STU	EXPN_STU	STR	AVGINC	EL_PCT
1	125.0000	10.90000	0.510200	2.040800	67.00000	690.8000	0.343590	6384.911	17.88991	22.69000	0.0000
2	240.0000	11.19000	95.41670	47.91670	101.0000	661.2000	0.420833	5099.391	21.52496	9.824000	4.5800
3	1550.000	82.90000	55.03230	76.32260	169.0000	643.6000	0.309032	5501.955	16.60723	8.978000	30.00
4	243.0000	14.00000	36.47540	77.04920	85.00000	647.7000	0.345794	7101.831	17.35714	8.978000	0.0000
5	1335.000	71.50000	33.10860	78.42700	171.0000	640.9500	0.128900	5235.968	18.67133	9.080333	13.96
6	137.0000	6.400000	12.31880	86.95650	25.00000	605.9500	0.182482	5880.147	21.40625	10.41500	12.40
7	195.0000	10.00000	12.90320	94.62370	29.00000	606.7500	0.143680	5253.331	19.50000	6.577000	68.71
8	898.0000	42.50000	19.80630	100.0000	66.00000	609.0000	0.074324	4965.746	20.89412	8.174000	46.96
9	379.0000	19.00000	32.19000	93.13960	35.00000	612.5000	0.092348	5395.548	19.94737	7.395000	30.00
10	2247.000	108.0000	78.99430	67.31640	0.000000	612.6500	0.000000	5036.211	20.80696	11.61333	40.20
11	446.0000	21.00000	15.62690	85.87440	86.00000	615.7500	0.192625	4647.690	21.23809	8.931000	52.91
12	987.0000	47.00000	71.71310	98.60560	56.00000	616.3000	0.056739	5447.345	21.00000	7.389500	54.60
13	103.0000	5.000000	22.42990	98.13080	25.00000	616.3000	0.242718	6967.149	20.60000	5.339000	42.71
14	487.0000	24.34000	24.60940	77.14940	0.000000	616.3000	0.000000	4818.613	20.00022	6.279000	20.50
15	649.0000	36.00000	14.83790	76.27120	31.00000	616.4500	0.047786	5621.456	18.02778	9.630000	80.10
16	652.0000	42.07000	24.21420	84.29570	80.00000	617.3500	0.093897	6026.360	20.25196	7.454000	49.41
17	491.0000	28.92000	11.20160	87.79970	100.0000	618.0500	0.203666	6723.238	16.97787	6.216000	85.50
18	421.0000	26.50000	8.651100	77.90970	90.00000	618.3000	0.118705	5689.886	16.50990	7.764000	58.90
19	6880.000	303.0300	21.28240	94.97120	960.0000	619.8000	0.139535	5064.616	22.70402	7.023000	77.00
20	3688.000	135.0000	23.43750	93.22920	138.0000	620.3000	0.051711	5433.593	19.91111	5.699000	49.81
21	440.0000	24.00000	34.77270	100.0000	89.00000	620.5000	0.156818	5725.563	18.33333	7.941000	40.66
22	475.0000	21.00000	21.64990	91.64640	63.00000	621.4000	0.111679	4542.106	22.61805	9.630000	16.21
23	2538.000	130.5000	18.91110	70.81670	168.0000	621.7500	0.066588	5107.086	19.44828	7.405000	45.00
24	475.0000	19.00000	43.85590	100.0000	0.000000	622.0500	0.000000	4689.662	25.05263	9.630000	39.00
25	2367.000	114.0000	15.80100	80.62370	216.0000	622.6000	0.091642	4695.464	20.67544	8.019000	76.68
26	1588.000	85.00000	22.40720	89.14720	198.0000	623.1000	0.124895	5415.153	16.65235	8.523000	40.41
27	7306.000	319.8000	17.00190	88.03490	742.0000	623.2000	0.101560	4997.872	22.84553	7.983181	73.70
28	2601.000	135.0000	15.07110	82.19530	268.0000	623.4500	0.103422	5223.912	19.26667	7.309000	70.01
29	647.0000	44.00000	15.29280	80.20070	67.00000	623.6000	0.079103	5139.166	19.25000	8.934000	55.96
30	452.0000	22.00000	14.49880	81.02950	65.00000	624.1500	0.121681	4614.262	20.54545	8.554000	11.00
31	4142.000	201.0000	35.56290	81.50650	569.0000	624.5500	0.137373	5342.293	20.60697	6.613000	80.40
32	2102.000	99.79000	15.31990	90.28490	224.0000	624.9500	0.106685	5347.458	21.07289	12.40900	63.10
33	10012.00	464.9000	29.76390	91.59340	721.0000	625.3000	0.072014	5036.459	21.53681	8.126616	65.10
34											

Path: c:\stata\stata\ D8 = popline\ WF = added

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 adjust the number of digits after the decimal point).

**Dependent Variable: TESTSCR**

**Method: Least Squares**

**Sample: 1 420**

**Included observations: 420**

**White Heteroskedasticity-Consistent Standard Errors & Covariance**

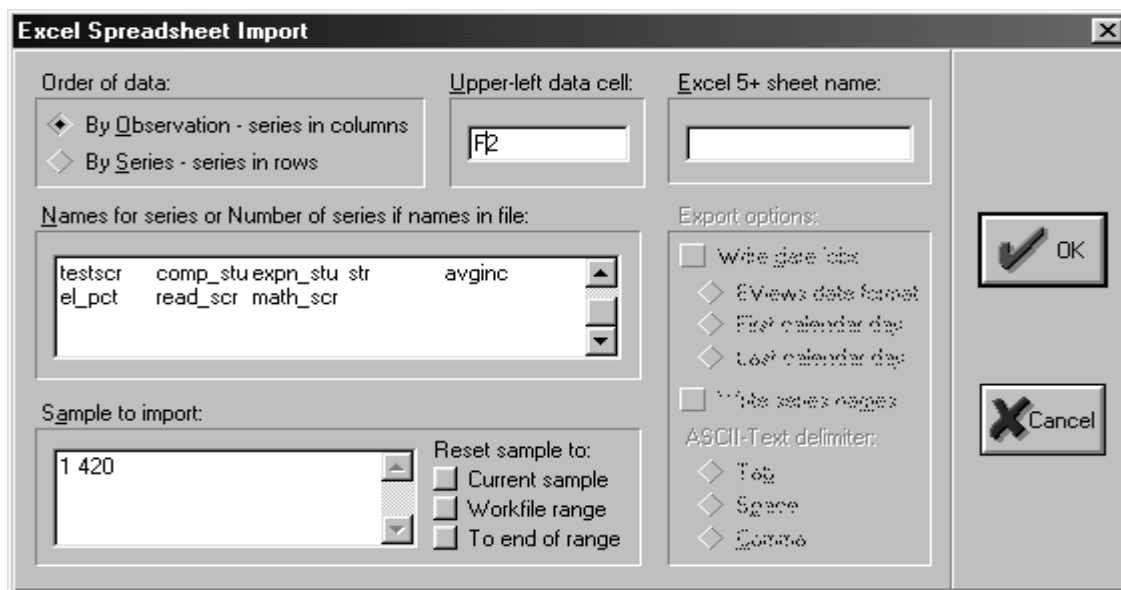
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	698.9	10.36	67.44	0.00
STR	-2.28	0.52	-4.39	0.00

R-squared	0.05	Mean dependent var	654.16
Adjusted R-squared	0.05	S.D. dependent var	19.05
S.E. of regression	18.58	Akaike info criterion	8.69
Sum squared resid	144315.48	Schwarz criterion	8.71
Log likelihood	-1822.25	F-statistic	22.58
Durbin-Watson stat	0.13	Prob(F-statistic)	0.00

(You can find the standard errors and the  $t$ -statistic on p. 114 of the Stock/Watson textbook. The regression  $R^2$ , sum of squared residuals ( $SSR$ ), and standard error of the regression ( $SER$ ) are presented in Section 4.8.)

### 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 **Procs / Import Data/Read Text-Lotus-Excel**. A dialog box will open, and you will first have to specify the location where your data file (SWcaschool.xls) resides. After you double click on the file, another dialog box opens. The data in your Excel file are ordered by observation, so you do not have to change the automatic setting at the top. 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. These have to be pasted into the empty 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 **Show testscr str** or running the same regression as above.

Finally, you can also save data in ASCII or spreadsheet format by clicking on **Procs** and then **Export Data**.

## 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 expectations, productivity growth, the change in price mark-ups, etc.

An extension of the simple regression model is the multiple regression model, which incorporates more than one regressor (see Equation (5.7) in the textbook).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n.$$

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. See if you can reproduce the following regression output, which corresponds to Column 5 in Table 5.2 of the textbook.

**Dependent Variable: TESTSCR**

**Method: Least Squares**

**Date: xx/xx/xx Time: xx:xx**

**Sample: 1 420**

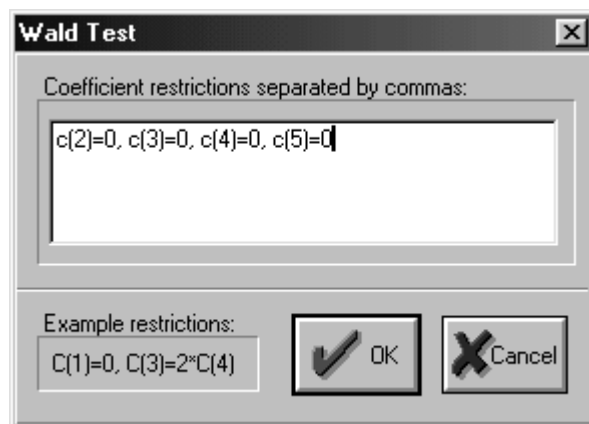
**Included observations: 420**

**White Heteroskedasticity-Consistent Standard Errors & Covariance**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	700.4	5.5	126.48	0.000
STR	-1.01	0.27	-3.77	0.000
EL_PCT	-0.130	0.036	-3.58	0.000
MEAL_PCT	-0.529	0.038	-13.87	0.000
CALW_PCT	-0.048	0.059	-0.82	0.415
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	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 5.8 of your 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,  $c(1)$  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:

**Wald Test:**  
**Equation: Untitled**

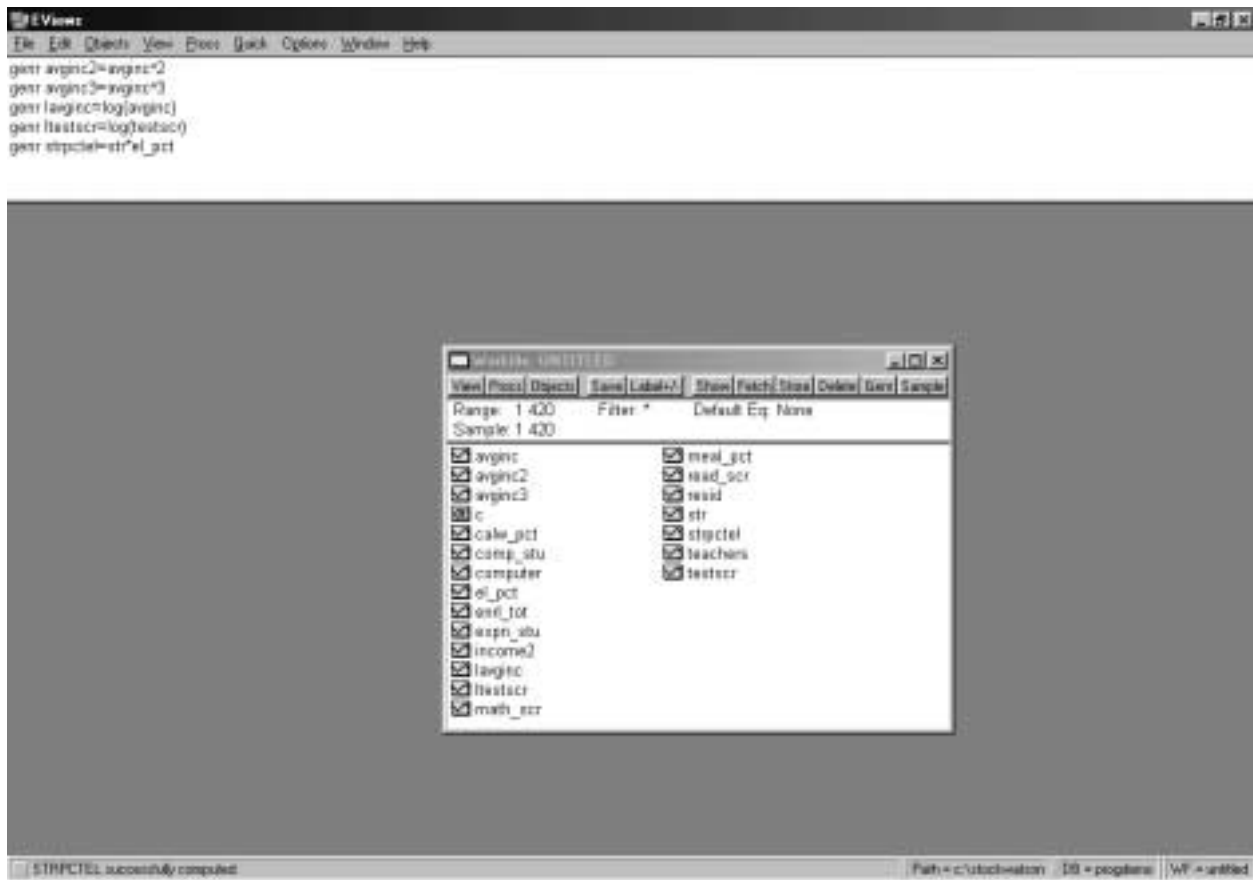
**Null Hypothesis: C(2)=0**  
**C(3)=0**  
**C(4)=0**  
**C(5)=0**

<b>F-statistic</b>	<b>361.6835</b>	<b>Probability</b>	<b>0.000000</b>
<b>Chi-square</b>	<b>1446.734</b>	<b>Probability</b>	<b>0.000000</b>

See if you can generate the  $F$ -statistic of 5.43 following Equation (5.40) in the text.

### *Data Transformations*

You will be frequently 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 6 of the Stock/Watson textbook introduces the polynomial regression model, logarithms, and interactions between variables. Let us reproduce Equations (6.11), (6.23), and (6.37) here. The following commands generate the necessary variables:



Next run the three regressions using the same technique as for multiple regression analysis. Finally save your workfile again and exit the workfile.

### **Batch Files**

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).

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. These flags are useful if you want to remember later what you were doing or if you want others to understand your program. Then “save [it] as” “Tutorialch4.prg” in your directory and click the “Run” button. Make sure that the workfile “Swcaschool.wf1” resides in the same subdirectory. EViews will execute your commands.

```

' Stock and Watson
' chapters 4-7
'
' Chapter 4
'
open Swcaschool.wf1
'
' DIST_CODE      DISTRICT CODE
' READ_SCR      AVG READING SCORE
' MATH_SCR      AVG MATH SCORE
'
' Summary statistics for str and testscr in Table 4.1
' 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
'
' Equation 4.11
' Here is an example of how to use OLS in the program.
' You first define the equation and then use the by now familiar ls command.
'
equation eq4_11.ls(h) testscr c str
'
' Equation 4.33
' Below a binary variable is defined first by setting it to zero for the entire sample
' then to set it to one for observations (smpl stands for sample) where the student-teacher ratio is
' less than 20.
'
genr dsize=0
smpl if str<20
genr dsize=1
smpl 1 420
equation eq4_33.ls(h) testscr c dsize

```

In the “Swcaschool.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.

### 3. TIME SERIES DATA

Let's leave the cross-sectional data and move on to Chapter 12 in the textbook. This is where EViews is superior to other regression packages, such as SAS, SPSS, or STATA, which are more suitable for cross-sectional analysis.

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,

$$Inf_t = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} * 100 = \left( \frac{CPI_t}{CPI_{t-1}} - 1 \right) * 100$$

To create past values of variables, you generate a lag by adding a “(-1)” after the 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  $X(-12)$ .<sup>6</sup>

Type or copy and paste the following lines of code into a new program and run it. 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 unemployment rate and the *future* change in the inflation rate, we also need to calculate the change in the inflation rate, or  $\Delta Inf_t = Inf_t - Inf_{t-1}$ . In the program below, this variable is called “*dinf*”. You need the workfile “ch12\_plus\_14.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.

---

<sup>6</sup> In mathematics, a lag is defined (loosely) through the use of a “lag-operator”  $L$ , where  $L^i x_t = x_{t-i}$ . Similarly, the “difference operator”  $\Delta = (1 - L)$ , so that  $\Delta x_t = x_t - x_{t-1}$ . See Appendix 12.3 of the textbook for more details.

```

' Stock and Watson
' chapter 12_plus_14
'
' Chapter 12
'
open ch12_plus_14.wf1
'
' 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
'
genr lpunew=log(punew)
genr inf=400*(lpunew-lpunew(-1))
genr dinf=inf-inf(-1)
genr yeardinf=inf-inf(-4)
'
' Figure 12.1
'
smpl 1960:1 1999:4
group Fig12_1 inf lhur
Fig12_1.line(m)
freeze(InflatUR) Fig12_1
'
'
' Equation 12.13
'
smpl 1962:1 1999:4
'
' EViews allows you to use (-i) in an equation on variables
' that you did not have to generate previously
'
equation eq12_13.ls(h) dinf c dinf(-1) dinf(-2) dinf(-3) dinf(-4)
'
' Figure 12.13
'
group Fig12_3 lhur(-4) yeardinf
Fig12_3.scats(r)
freeze(Phillips) Fig12_3
'
' Equation 12.17
'
equation eq12_17.ls(h) dinf c dinf(-1) dinf(-2) dinf(-3) dinf(-4) lhur(-1) lhur(-2) lhur(-3) lhur(-4)

```

#### 4. SUMMARY OF FREQUENTLY USED EViews COMMANDS

The command 'genr' *creates new variables and modifies existing variables.*

Examples:

```
genr expn=expn_stu/1000
```

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

```
genr avginc2=avginc^2  
genr lavginc=log(avginc)
```

create the square and log of average income, respectively.

Note that commands of the type

```
genr testscr = 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:

```
genr dinf=inf-inf(-1)  
genr yeardinf=inf-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

```
smpl 1 420  
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 4-7. 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.

```
smpl if str<=20
```

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

The most frequently used *statistical operations* involve running regressions ('ls'), establishing the correlation between variables ('cor'), and graphing variables ('line'). 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:

```
equation eq4_7.ls(h) testscr c str  
equation eqtab5_2_5.ls(h) testscr c str el_pct meal_pct calw_pct  
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. 'ls' then instructs EViews to use OLS estimation for the equation. The dependent variable appears first, followed by the regressors, where 'c' is used for the intercept ('c' is a reserved name in EViews, meaning that you cannot use it to generate a variable called '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 'cor' and 'line' command. Examples:

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

Here the variables str (student-teacher ratio) and testscr (testscore) are assigned to a group called 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, inf (inflation) and lhur (unemployment rate) are assigned to a group called Fig12\_1 and are then plotted (where 'm' is an option that allows for the display of multiple graphs).

```
group Fig12_1 inf lhur  
Fig12_1.line(m)
```

## **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. As mentioned before, this tutorial is not intended to replace them.

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.