An Introduction to EViews

The purpose of the computer assignments in BUEC 333 is to give you some experience using econometric software to analyse real-world data. Along the way, you'll become acquainted with the frustrations of obtaining & manipulating data, and you'll get some practice designing and implementing empirical studies. You'll probably also rediscover the age-old maxim that all empirical works takes (at least) three times as long as you expect it to.

To complete the computer assignments, you'll use the EViews software installed on the lab PCs in WMC 2502 and WMC 2506.

Helpful hint: The EViews HELP menu is your friend. All EViews procedures are documented within, as are methods to get data in/out of EViews, etc. Most of the things you'll have to do for the assignments are quite "easy" to do in EViews (meaning you should be able to accomplish it by pointing and clicking, rather than writing a program). So if you're having trouble figuring out how to do something, it's probably because you're not familiar with the specific command that does what you need. There are several places to look when you're stuck – from the HELP menu, you can access:
1. EVIEWS HELP TOPICS – the first place to look if you're stuck, including a Demo that you might find helpful if you're having a hard time getting started (for the Demo, look under User's Guide -> EViews Fundamentals).
2. QUICK HELP REFERENCE – the second place to look, also where you can find some sample data & programs
3. USER'S GUIDE (PDF) – a complete PDF manual, in 2 parts.
4. COMMAND REFERENCE (PDF) – probably more than you need for this course, but contains more detailed information than the User's Guide on various EViews command.

The Basic Idea – how EViews works

EViews is software for econometric analysis. The first step is to get your data into EViews. The second step is to verify that you've actually succeeded in getting the data into EViews without error. The third step is usually to do some exploratory analysis: computing descriptive statistics, generating some plots and histograms to get acquainted with the data, etc. The fourth step is to do the "real" analysis – use the data to answer some question that has been posed to you. This document will help you get through the first 3 steps – we'll deal with the fourth step as the semester progresses.

Step 1: Getting your data into EViews

First you'll need to acquire the data from somewhere. In BUEC 333, that will be from the course website. In the real world it may be from some other website, a CD-ROM, etc. The data will usually be in a format that is foreign to EViews, e.g., an Excel spreadsheet, an ASCII text file, or whatever. Excel files (what we'll mostly use in BUEC 333) have a .xls extension to their filename. ASCII files usually have a .txt extension.

Once you've acquired the data, you need to get it into EViews. EViews organizes objects (objects include data series, graphs, parameter estimates, and lots of other things) into what is called a **workfile**. You begin by creating a new workfile:

FILE -> NEW -> WORKFILE

This opens a dialog box that asks for some characteristics of the data. Most of the time we'll be using UNSTRUCTURED/UNDATED data (select this from the WORKFILE STRUCTURE TYPE drop-down menu). You'll need to enter the number of observations in the data file – with Excel formatted data you can check this by opening the data file in Excel & counting the rows (BUT DON'T FORGET TO CLOSE THE EXCEL FILE, OR YOU'LL GET AN ERROR MESSAGE WHEN YOU TRY TO READ IT INTO EVIEWS!) When you've selected a workfile structure & entered the number of observations, click OK.

A new window will appear – this is your new UNTITLED workfile. You'll see it contains two objects: called C and RESID. For now, ignore these. Across the top of the workfile window there are a series of drop down menus. To get your data into EViews, select:

PROC -> IMPORT -> READ TEXT-LOTUS-EXCEL

This opens a new dialog box that you can use to browse for your data file. When you've done so, click OPEN. This opens a new dialog box where you give information about the data file you're importing. We'll assume it's an Excel file. The usual Excel format has data series (variables) in columns, and each row is an observation: in the DATA ORDER menu, select BY OBSERVATION – SERIES IN COLUMNS. You'll need to tell EViews which cell contains the first data element (the UPPER-LEFT DATA CELL – check this by inspecting the Excel file), and the name of the Excel worksheet that contains the data you want to import. Finally, you'll need to enter the name of each data series (variable) in the NAMES FOR SERIES OR NUMBER IF NAMED IN FILE. **If you're importing an alpha variable (one whose values are WORDS rather than NUMBERS) you need a $ after the variable name.** When you've entered the variable names, click OK to close the dialog box. You'll see a bunch of new objects in your workfile: one for each variable you imported, with the names that you assigned. EViews calls these DATA SERIES, but you can think of them as variables.

On the course website you'll find the hockey data used for the in-class demo (NHL1988-99.XLS) and variable names in the text file VARNAMES.TXT. I suggest you try importing these data into EViews at this point.

Step 2: Make sure the data have imported properly

ALWAYS ALWAYS ALWAYS check your data before you start using them. To do so, select all the data series you just imported (usual Windows selection rules apply – use SHIFT+CLICK and/or CTRL+CLICK to select multiple objects) except C and RESID, RIGHT-CLICK the selected objects, and then OPEN -> AS GROUP. This opens the

selected data series in the default (spreadsheet) view. From here, you can browse the data to ensure they've been imported correctly (e.g., compare the imported data series to the original Excel file).

If the data appear to have been imported correctly, you can close the GROUP window and proceed to step 3. **I RECOMMEND THAT YOU SAVE THE WORKFILE AT THIS POINT (FILE -> SAVE)** so you can avoid going through the whole importing step the next time you use these data (just use FILE -> OPEN in future).

If your data were not imported correctly, you'll need to figure out what you've done wrong and return to step 1.

Step 3: Exploratory analysis

EViews gives you lots of ways to explore your data with just a few mouse clicks. You can open a single data series by double-clicking it, or open multiple series using OPEN -> AS GROUP like we did in step 2. You can print on-screen information using the PRINT menu if you're asked to include it in an assignment. Here are some examples of ways to view a single data series:

VIEW -> DESCRIPTIVE STATISTICS & TESTS -> STATS TABLE
Displays sample statistics like the sample mean, sample standard deviation (STD. DEV.) etc.

VIEW -> DESCRIPTIVE STATISTICS -> HISTOGRAM AND STATS
Displays a histogram of a data series and some sample statistics. This is an easy way to view the shape of a variable's marginal probability distribution function (pdf).

VIEW -> DESCRIPTIVE STATISTICS -> STATS BY CLASSIFICATION
Displays conditional statistics (i.e., conditional means and variances, etc.). You'll need to name a data series to classify the data by (the conditioning variable).

VIEW -> GRAPH -> DISTRIBUTION -> EMPIRICAL CDF
You'll find this useful for viewing the shape of a variable's cdf

VIEW -> DISTRIBUTION -> KERNEL DENSITY GRAPHS
This displays a "smoothed" histogram (the default options should be fine), which approximates the variable's pdf

VIEW -> ONE-WAY TABULATION
Displays the marginal distribution (pdf and cdf) of a discrete random variable in table form. For a continuous random variable, EViews will split the data into ranges.

There are some other ways to view groups of random variables:

VIEW -> GRAPH -> SCATTER

Displays a scatter plot of two random variables. This is a useful way to examine their joint distribution.

VIEW -> DESCRIPTIVE STATS -> COMMON SAMPLE
Is like the STATS TABLE above, but for multiple variables.

VIEW -> COVARIANCE ANALYSIS
Produces a table of sample correlations and/or covariances between the selected variables.

VIEW -> N-WAY TABULATION
Displays the joint distribution (pdf and cdf) of two discrete random variables in a table. By default, it gives counts in each cell. If you select the ROW% and COLUMN% options, it will display percentages of the data in each cell as well, and the margins of the table give the marginal distribution. If the variables are continuous, EViews will split the data into ranges.

Creating new variables
You're not limited to working with the set of variables that you import – you can use these to create new ones. We use the GENR menu in the WORKFILE window for this. Clicking GENR opens a dialog box. You create a new variable by entering an equation that defines it. The equations can be mathematical or logical. Here are some examples.

Creating variables with mathematical equations:
In the hockey data on the course website, there is a variable called AGE. Suppose we want to create a new variable (call it AGE_SQUARED) that is the square of AGE. In the ENTER EQUATION box, type:

AGE_SQUARED = AGE*AGE

And click OK. Another example: a new variable called LN_SALARY that is the natural logarithm of the SALARY variable plus 1:

LN_SALARY = LOG(1 + SALARY)

(We add 1 in case SALARY = 0. Why?)

Pretty much any mathematical operation will work – multiplication, addition, taking logarithms, raising values to an exponent, etc. You can see some examples of mathematical expressions in Chapter 6 of the EView User's Guide (the .pdf file you can access from inside EViews).

Creating variables with logical expressions
Logical expressions are mathematical statements that are either true or false. For example, if AGE = 25, then the logical expression AGE > 30 is false, and the logical expression AGE < 30 is true. These type of expressions are very useful for manipulating

data, creating new variables, etc. For now, we'll focus on using logical expressions to create new variables that take value 0 or 1. We call these **dummy variables** (or **indicator variables**).

At their most basic level, computers are just big calculators. All they do is store and manipulate numbers. In fact, at their most basic level they only store and manipulate ones and zeros. When a computer encounters a logical expression, it evaluates whether it is true or false. It then stores the value 1 if the expression is true or the value 0 if it is false. This is useful for creating dummy variables. Here's an example. Recall the hockey data on the course website. There is a variable called AGE that measures a player's age in 1999. The logical expression AGE >25 will be true for some players and false for others. We can use this fact to create a dummy variable (call it OVER_25) that equals one if a player is over 25 in 1999, and equals zero if the player is 25 or younger:

OVER_25 = (AGE > 25)

When we create this variable, EViews evaluates the logical expression AGE > 25 for each observation. When it is true, it returns the value 1 and the variable OVER_25 equals 1 for this observation. When it is false, EViews returns the value 0, and the variable OVER_25 equals 0 for this observation. We can use this to compute, for example, the proportion of players that are older than 25 in 1999: create the variable OVER_25 as above, and then compute the sample mean of OVER_25 (open the series OVER_25 and look at its stats table). I suggest you try this example to make sure you understand it – it will probably help if you open the AGE variable and the OVER_25 variables together and browse them as a group.

We can use more complicated expressions. For example, we can create a dummy variable called UNDER_25_AND_RICH that equals 1 if AGE < 25 and SALARY > 1,000,000:

UNDER_25_AND_RICH = (AGE < 25 and SALARY > 1000000)

Note this variable equals 0 if AGE ≥25 **or** SALARY ≤ 1000000. Notice also that EViews understands the word "and" the same way you or I do. EViews also understands the word "or":

NOT_25 = (AGE < 25 or AGE > 25)

A Note on Missing Data

Sometimes we encounter missing values when working with real-world data. For example, in the NHL hockey data on the course website, the SALARY variable is missing for some players. It simply wasn't available when I put the data set together. When EViews reads a missing data value, it assigns the value "NA". Subsequent operations on missing values propagate the "NA" value. For instance, whenever SALARY = NA for a player, UNDER_25_AND_RICH = NA also. This is usually ok, but it's important to keep in mind that a variable like UNDER_25_AND_RICH won't be

available for every player, and statistics computed using this variable will only be based on the fraction of the sample for which SALARY is not missing. Here's a useful trick for working with missing data. You can create a dummy variable that equals 1 if SALARY is missing and 0 otherwise:

SALARY_MISSING = (SALARY = NA)

This variable is useful for computing things like the proportion of the sample where SALARY is missing, examining whether observations with SALARY missing are similar to those without (e.g., compare conditional means of the other variables given SALARY_MISSING), etc.