



EViews Illustrated

for Version 6

Richard Startz

University of Washington

EViews Illustrated

for Version 6

Copyright © 2007 Quantitative Micro Software, LLC

All Rights Reserved

Printed in the United States of America

Disclaimer

The author and Quantitative Micro Software assume no responsibility for any errors that may appear in this book or the EViews program. The user assumes all responsibility for the selection of the program to achieve intended results, and for the installation, use, and results obtained from the program.

Trademarks

Windows, Windows 95/98/2000/NT/Me/XP, and Microsoft Word and Microsoft Excel are trademarks of Microsoft Corporation. PostScript is a trademark of Adobe Corporation. Professional Organization of English Majors is a trademark of Garrison Keillor. All other product names mentioned in this manual may be trademarks or registered trademarks of their respective companies.

Quantitative Micro Software, LLC

4521 Campus Drive, #336, Irvine CA, 92612-2699

Telephone: (949) 856-3368

Fax: (949) 856-2044

web: www.eviews.com

First edition: 2007

Editor: Meredith Startz

Index: Palmer Publishing Services

Chapter 3. Getting the Most from Least Squares

Regression is the king of econometric tools. Regression’s job is to find numerical values for theoretical parameters. In the simplest case this means telling us the slope and intercept of a line drawn through two dimensional data. But EViews tells us lots more than just slope and intercept. In this chapter you’ll see how easy it is to get parameter estimates plus a large variety of auxiliary statistics.

We begin our exploration of EViews’ regression tool with a quick look back at the NYSE volume data that we first saw in the opening chapter. Then we’ll talk about how to instruct EViews to estimate a regression and how to read the information about each estimated coefficient from the EViews output. In addition to regression coefficients, EViews provides a great deal of summary information about each estimated equation. We’ll walk through these items as well. We take a look at EViews’ features for testing hypotheses about regression coefficients and conclude with a quick look at some of EViews’ most important views of regression results.

Regression is a big subject. This chapter focuses on EViews’ most important regression features. We postpone until later chapters various issues, including forecasting ([Chapter 8. “Forecasting”](#)), serial correlation ([Chapter 13. “Serial Correlation—Friend or Foe?”](#)), and heteroskedasticity and nonlinear regression ([Chapter 14. “A Taste of Advanced Estimation”](#)).

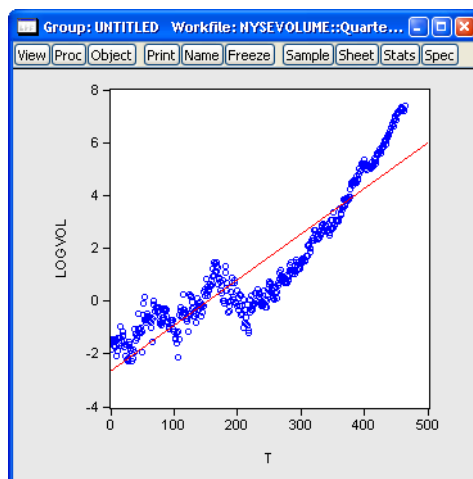
A First Regression

Returning to our earlier examination of trend growth in the volume of stock trades, we start with a scatter diagram of the logarithm of volume plotted against time.

EViews has drawn a straight line—a regression line—through the cloud of points plotted with $\log(\text{volume})$ on the vertical axis and time on the horizontal. The regression line can be written as an algebraic expression:

$$\log(\text{volume}_t) = \alpha + \beta t$$

Using EViews to estimate a regression lets us replace α and β with numbers based on the data in the workfile. In a bit we’ll see that EViews estimates the regression line to be:



$$\log(\text{volume}) = -2.629649 + 0.017278t$$

In other words, the intercept α is estimated to be -2.6 and the slope β is estimated to be 0.017.

Most data points in the scatter plot fall either above or below the regression line. For example, for observation 231 (which happens to be the first quarter of 1938) the actual trading volume was far below the predicted regression line.

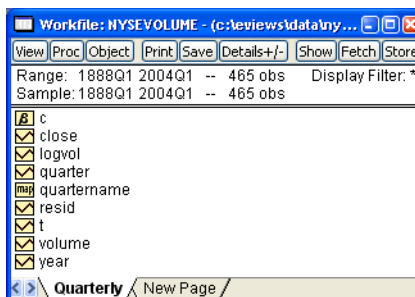
In other words, the regression line contains errors which aren't accounted for in the estimated equation. It's standard to write a regression model to include a term u_t to account for these errors. (Econometrics texts sometimes use the Greek letter epsilon, ϵ , rather than u for the error term.) A complete equation can be written as:

$$\log(\text{volume}_t) = \alpha + \beta t + u_t$$

Regression is a statistical procedure. As such, regression analysis takes uncertainty into account. Along with an estimated value for each parameter (e.g., $\beta = 0.017$) we get:

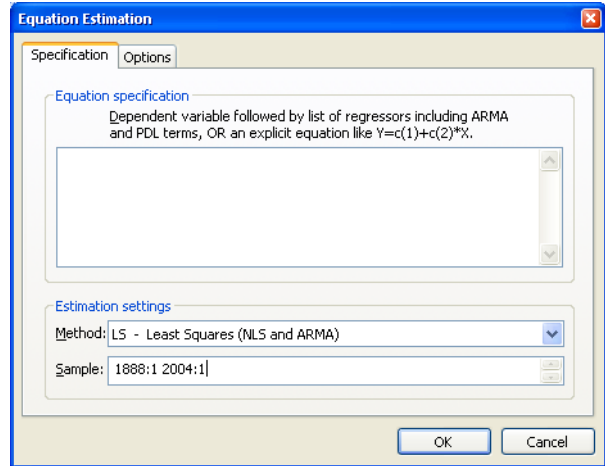
- Measures of the accuracy of each of the estimated parameters and related information for computing hypothesis tests.
- Measures of how well the equation fits the data: How much is explained by the estimated values of α and β and how much remains unexplained.
- Diagnostics to check up on whether assumptions underlying the regression model seem satisfied by the data.

We're re-using the data from [Chapter 1. “A Quick Walk Through”](#) to illustrate the features of EViews' regression procedure. If you want to follow along on the computer, use the workfile “NYSEVOLUME” as shown.

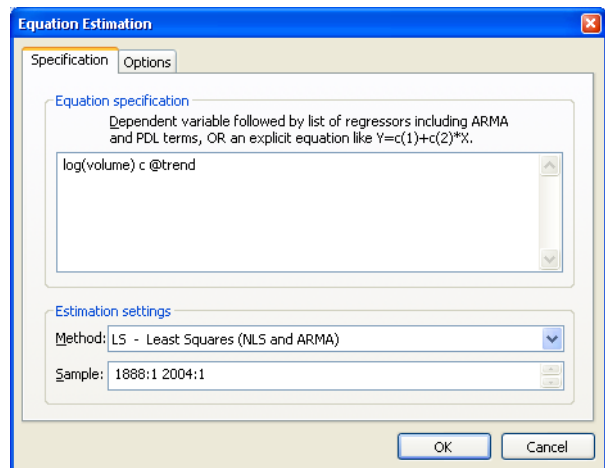


EViews allows you to run a regression either by creating an equation object or by typing commands in the command pane. We'll start with the former approach. Choose the menu command **Object/New Object...** Pick **Equation** in the **New Object** dialog.

The empty equation window pops open with space to fill in the variables you want in the regression.



Regression equations are easily specified in EViews by a list in which the first variable is the *dependent* variable—the variable the regression is to explain, followed by a list of explanatory—or *independent*—variables. Because EViews allows an expression pretty much anywhere a variable is allowed, we can use either variable names or expressions in our regression specification. We want *log volume* for our dependent variable and a time trend for our independent variable. Fill out the equation dialog by entering “log(volume) c @trend”.



Hint: EViews tells one item in a list from another by looking for spaces between items. For this reason, spaces generally aren't allowed inside a single item. If you type:

```
log (volume) c @trend
```

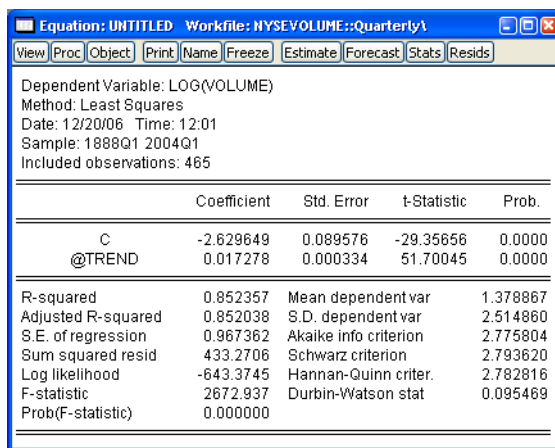
you'll get an error message.

Exception to the previous hint: When a text string is called for in a command, spaces are allowed inside paired quotes.

Reminder: The letter “C” in a regression specification notifies EViews to estimate an intercept—the parameter we called α above.

Hint: Another reminder: `@trend` is an EViews function to generate a time trend, 0, 1, 2,

Our regression results appear below:



	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.629649	0.089576	-29.35656	0.0000
@TREND	0.017278	0.000334	51.70045	0.0000
R-squared	0.852357	Mean dependent var	1.378867	
Adjusted R-squared	0.852038	S.D. dependent var	2.514860	
S.E. of regression	0.967362	Akaike info criterion	2.775804	
Sum squared resid	433.2706	Schwarz criterion	2.793620	
Log likelihood	-643.3745	Hannan-Quinn criter.	2.782816	
F-statistic	2672.937	Durbin-Watson stat	0.095469	
Prob(F-statistic)	0.000000			

The Really Important Regression Results

There are 25 pieces of information displayed for this very simple regression. To sort out all the different goodies, we’ll start by showing a couple of ways that the main results might be presented in a scientific paper. Then we’ll discuss the remaining items one number at a time.

A favorite scientific convention for reporting the results of a single regression is display the estimated equation inline with standard errors placed below estimated coefficients, looking something like:

$$\log(\text{volume}_t) = -2.629649 + 0.017278 \cdot t, \text{ ser} = 0.967362, R^2 = 0.852357$$

(0.089576) (0.000334)

Hint: The dependent variable is also called the *left-hand side* variable and the independent variables are called the *right-hand side* variables. That's because when you write out the regression equation algebraically, as above, convention puts the dependent variable to the left of the equals sign and the independent variables to the right.

The convention for inline reporting works well for a single equation, but becomes unwieldy when you have more than one equation to report. Results from several related regressions might be displayed in a table, looking something like [Table 2](#).

Table 2

	(1)	(2)
Intercept	-2.629649 (0.089576)	-0.106396 (0.045666)
t	0.017278 (0.000334)	-0.000736 (0.000417)
t^2	—	6.63E-06 (1.37E-06)
$\log(\text{volume}(-1))$	—	0.868273 (0.022910)
ser	0.967362	0.289391
R^2	0.852357	0.986826

Column (2)? Don't worry, we'll come back to it later.

Hint: Good scientific practice is to report only digits that are meaningful when displaying a number. We've printed far too many digits in both the inline display and in [Table 2](#) so as to make it easy for you to match up the displayed numbers with the EViews output. From now on we'll be better behaved.

EViews regression output is divided into three panels. The top panel summarizes the input to the regression, the middle panel gives information about each regression coefficient, and the bottom panel provides summary statistics about the whole regression equation.

The most important elements of EViews regression output are the estimated regression coefficients and the statistics associated with each coefficient. We begin by linking up the num-

bers in the inline display—or equivalently column (1) of Table 2—with the EViews output shown earlier.

The names of the independent variables in the regression appear in the first column (labeled “Variable”) in the EViews output, with the estimated regression coefficients appearing one column over to the right (labeled “Coefficient”). In econometrics texts, regression coefficients are commonly denoted with a Greek letter such as α or β or, occasionally, with a Roman b . In contrast, EViews presents you with the variable names; for example, “@TREND” rather than “ β ”.

The third EViews column, labeled “Std. Error,” gives the standard error associated with each regression coefficient. In the scientific reporting displays above, we’ve reported the standard error in parentheses directly below the associated coefficient. The standard error is a measure of uncertainty about the true value of the regression coefficient.

The standard error of the regression, abbreviated “*ser*,” is the estimated standard deviation of the error terms, u_t . In the inline display, “*ser* = 0.967362” appears to the right of the regression equation proper. EViews labels the *ser* as “S.E. of regression,” reporting its value in the left column in the lower summary block.

Note that the third column of EViews regression output reports the standard error of the estimated coefficients while the summary block below reports the standard error of the regression. Don’t confuse the two.

The final statistic in our scientific display is R^2 . R^2 measures the overall fit of the regression line, in the sense of measuring how close the points are to the estimated regression line in the scatter plot. EViews computes R^2 as the fraction of the variance of the dependent variable explained by the regression. (See the *User’s Guide* for the precise definition.) Loosely, $R^2 = 1$ means the regression fit the data perfectly and $R^2 = 0$ means the regression is no better than guessing the sample mean.

Hint: EViews will report a negative R^2 for a model which fits worse than a model consisting only of the sample mean.

The Pretty Important (But Not So Important As the Last Section’s) Regression Results

We’re usually most interested in the regression coefficients and the statistical information provided for each one, so let’s continue along with the middle panel.

t-Tests and Stuff

All the stuff about individual coefficients is reported in the middle panel, a copy of which we've yanked out to examine on its own.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.629649	0.089576	-29.35656	0.0000
@TREND	0.017278	0.000334	51.70045	0.0000

The column headed “t-Statistic” reports, not surprisingly, the *t*-statistic. Specifically, this is the *t*-statistic for the hypothesis that the coefficient in the same row equals zero. (It's computed as the ratio of the estimated coefficient to its standard error: *e.g.*, $51.7 = 0.017/0.00033$.)

Given that there are many potentially interesting hypotheses, why does EViews devote an entire column to testing that specific coefficients equal zero? The hypothesis that a coefficient equals zero is special, because if the coefficient does equal zero then the attached coefficient drops out of the equation. In other words, $\log(\text{volume}_t) = \alpha + 0 \times t + u_t$ is really the same as $\log(\text{volume}_t) = \alpha + u_t$, with the time trend not mattering at all.

Foreshadowing hint: EViews automatically computes the test statistic against the hypothesis that a coefficient equals zero. We'll get to testing other coefficients in a minute, but if you want to leap ahead, look at the equation window menu **View/Coefficient Tests...**

If the *t*-statistic reported in column four is larger than the *critical value* you choose for the test, the estimated coefficient is said to be “statistically significant.” The critical value you pick depends primarily on the risk you're willing to take of mistakenly rejecting the null hypothesis (the technical term is the “size” of the test), and secondarily on the degrees of freedom for the test. The larger the risk you're willing to take, the smaller the critical value, and the more likely you are to find the coefficient “significant.”

Hint: EViews doesn't compute the degrees of freedom for you. That's probably because the computation is so easy it's not worth using scarce screen real estate. Degrees of freedom equals the number of observations (reported in the top panel on the output screen) less the number of parameters estimated (the number of rows in the middle panel.) In our example, $df = 465 - 2 = 463$.

The textbook approach to hypothesis testing proceeds thusly:

1. Pick a size (the probability of mistakenly rejecting), say five percent.
2. Look up the critical value in a *t*-table for the specified size and degrees of freedom.

3. Compare the critical value to the t -statistic reported in column four. Find the variable to be “significant” if the t -statistic is greater than the critical value.

EViews lets you turn the process inside out by using the “ p -value” reported in the right-most column, under the heading “Prob.” EViews has worked the problem backwards and figured out what size would give you a critical value that would just match the t -statistic reported in column three. So if you are interested in a five percent test, you can reject if and only if the reported p -value is less than 0.05. Since the p -value is zero in our example, we’d reject the hypothesis of no trend at any size you’d like.

Obviously, that last sentence can’t be literally true. EViews only reports p -values to four decimal places because no one ever cares about smaller probabilities. The p -value isn’t literally 0.0000, but it’s close enough for all practical purposes.

Hint: t -statistics and p -values are different ways of looking at the same issue. A t -statistic of 2 corresponds (approximately) to a p -value of 0.05. In the old days you’d make the translation by looking at a “ t -table” in the back of a statistics book. EViews just saves you some trouble by giving both t - and p -.

Not-really-about-EViews-digression: Saying a coefficient is “significant” means there is statistical evidence that the coefficient differs from zero. That’s not the same as saying the coefficient is “large” or that the variable is “important.” “Large” and “important” depend on the substantive issue you’re working on, not on statistics. For example, our estimate is that NYSE volume rises about one and one-half percent each quarter. We’re very sure that the increase differs from zero—a statement about statistical significance, not importance.

Consider two different views about what’s “large.” If you were planning a quarter ahead, it’s hard to imagine that you need to worry about a change as small as one and one-half percent. On the other hand, one and one-half percent per quarter starts to add up over time. The estimated coefficient predicts volume will double each decade, so the estimated increase is certainly large enough to be important for long-run planning.

More Practical Advice On Reporting Results

Now you know the principles of how to read EViews’ output in order to test whether a coefficient equals zero. Let’s be less coy about common practice. When the p -value is under 0.05, econometricians say the variable is “significant” and when it’s above 0.05 they say it’s “insignificant.” (Sometimes a variable with a p -value between 0.10 and 0.05 is said to be “weakly significant” and one with a p -value less than 0.01 is “strongly significant.”) This practice may or may not be wise, but wise or not it’s what most people do.

We talked above about scientific conventions for reporting results and showed how to report results both inline and in a display table. In both cases standard errors appear in parentheses below the associated coefficient estimates. “Standard errors in parentheses” is really the first of two-and-a-half reporting conventions used in the statistical literature. The second convention places the t -statistics in the parentheses instead of standard errors. For example, we could have reported the results from EViews inline as

$$\log(\text{volume}_t) = -2.629649 + 0.017278 \cdot t, \quad \text{ser} = 0.967362, \quad R^2 = 0.852357$$
$$(-29.35656) \quad (51.70045)$$

Both conventions are in wide use. There’s no way for the reader to know which one you’re using—so you have to tell them. Include a comment or footnote: “Standard errors in parentheses” or “ t -statistics in parentheses.”

Fifty percent of economists report standard errors and fifty percent report t -statistics. The remainder report p -values, which is the final convention you’ll want to know about.

Where Did This Output Come From Again?

The top panel of regression output, shown on the right, summarizes the setting for the regression.

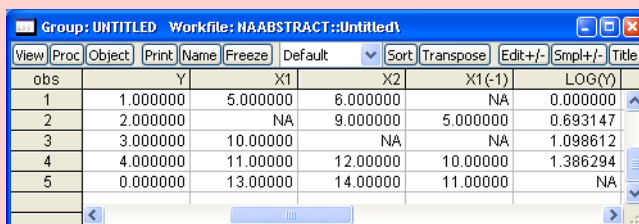
```
Dependent Variable: LOG(VOLUME)
Method: Least Squares
Date: 12/20/06 Time: 12:01
Sample: 1888Q1 2004Q1
Included observations: 465
```

The last line, “Included observations,” is obviously useful. It tells you how much data you have! And the next to last line identifies the sample to remind you which observations you’re using.

Hint: EViews automatically excludes all observations in which any variable in the specification is NA (not available). The technical term for this exclusion rule is “list-wise deletion.”

Big (Digression)

Hint: Automatic exclusion of NA observations can sometimes have surprising side effects. We'll use the data abstract at the right as an example.



obs	Y	X1	X2	X1(-1)	LOG(Y)
1	1.000000	5.000000	6.000000	NA	0.000000
2	2.000000	NA	9.000000	5.000000	0.693147
3	3.000000	10.000000	NA	NA	1.098612
4	4.000000	11.000000	12.000000	10.000000	1.386294
5	0.000000	13.000000	14.000000	11.000000	NA

Data are missing from observation 2 for X1 and from observation 3 for X2. A regression of Y on X1 would use observations 1, 3, 4, and 5. A regression of Y on X2 would use observations 1, 2, 4, and 5. A regression of Y on both X1 and X2 would use observations 1, 4, and 5. Notice that the fifth observation on Y is zero, which is perfectly valid, but that the fifth observation on log(Y) is NA. Since the logarithm of zero is undefined EViews inserts NA whenever it's asked to take the log of zero. A regression of log(Y) on both X1 and X2 would use only observations 1 and 4.

The variable, X1(-1), giving the previous period's values of X1, is missing both the first and third observation. The first value of X1(-1) is NA because the data from the observation before observation 1 doesn't exist. (There *is* no observation before the first one, eh?) The third observation is NA because it's the second observation for X1, and that one is NA. So while a regression of Y on X1 would use observations 1, 3, 4, and 5, a regression of Y on X1(-1) would use observations 2, 4, and 5.

Moral: When there's missing data, changing the variables specified in a regression can also inadvertently change the sample.

What's the use of the top three lines? It's nice to know the date and time, but EViews is rather ungainly to use as a wristwatch. More seriously, the top three lines are there so that when you look at the output you can remember what you were doing.

Dependent Variable: LOG(VOLUME)
Method: Least Squares
Date: 12/20/06 Time: 12:01
Sample: 1888Q1 2004Q1
Included observations: 465

"Dependent Variable" just reminds you what the regression was explaining—LOG(VOLUME) in this case.

"Method" reminds us which statistical procedure produced the output. EViews has dozens of statistical procedures built-in. The default procedure for estimating the parameters of an equation is "least squares."

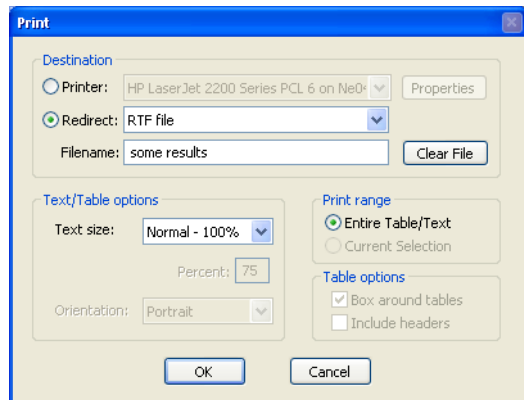
The third line just reports the date and time EViews estimated the regression. It's surprising how handy that information can be a couple of months into a project, when you've forgotten in what order you were doing things.

Since we're talking about looking at output at a later date, this is a good time to digress on ways to save output for later. You can:

- Hit the **Name** button to save the equation in the workfile. The equation will appear in the workfile window marked with the **E** icon. Then save the workfile.

Hint: Before saving the file, switch to the equation's label view and write a note to remind yourself why you're using this equation.

- Hit the **Print** button.
- Spend output to a Rich Text Format (RTF) file, which can then be read directly by most word processors. Select **Redirect:** in the **Print** dialog and enter a file name in the **Filename:** field. As shown, you'll end up with results stored in the file "some results.rtf".
- Right-click and choose **Select non-empty cells**, or hit **Ctrl-A**—it's the same thing. Copy and then paste into a word processor.



Freeze it

If you have output that you want to make sure won't ever change, even if you change the equation specification, hit **Freeze**. Freezing the equation makes a copy of the current view in the form of a table which is detached from the equation object. (The original equation is unaffected.) You can then **Name** this frozen table so that it will be saved in the workfile. See [Chapter 17. "Odds and Ends."](#)

Summary Regression Statistics

The bottom panel of the regression provides 12 summary statistics about the regression. We'll go over these statistics briefly, but leave technical details to your favorite econometrics text or the *User's Guide*.

R-squared	0.852357	Mean dependent var	1.378867
Adjusted R-squared	0.852038	S.D. dependent var	2.514860
S.E. of regression	0.967362	Akaike info criterion	2.775804
Sum squared resid	433.2706	Schwarz criterion	2.793620
Log likelihood	-643.3745	Hannan-Quinn criter.	2.782816
F-statistic	2672.937	Durbin-Watson stat	0.095469
Prob(F-statistic)	0.000000		

We've already talked about the two most important numbers, "R-squared" and "S.E. of regression." Our regression accounts for 85 percent of the variance in the dependent variable and the estimated standard deviation of the error term is 0.97. Four other elements, "Sum squared residuals," "Log likelihood," "Akaike info criterion," and "Schwarz criterion" are used for making statistical comparisons between two different regressions. This means that they don't really help us learn anything about the regression we're working on; rather, these statistics are useful for deciding if one model is better than another. For the record, the sum of squared residuals is used in computing F -tests, the log likelihood is used for computing likelihood ratio tests, and the Akaike and Schwarz criteria are used in Bayesian model comparison.

The next two numbers, "Mean dependent var" and "S.D. dependent var," report the sample mean and standard deviation of the left hand side variable. These are the same numbers you'd get by asking for descriptive statistics on the left hand side variables, so long as you were using the sample used in the regression. (Remember: EViews will drop observations from the estimation sample if any of the left-hand side or right-hand side variables are NA—*i.e.*, missing.) The standard deviation of the dependent variable is much larger than the standard error of the regression, so our regression has explained most of the variance in $\log(\text{volume})$ —which is exactly the story we got from looking at the R-squared.

Why use valuable screen space on numbers you could get elsewhere? Primarily as a safety check. A quick glance at the mean of the dependent variable guards against forgetting that you changed the units of measurement or that the sample used is somehow different from what you were expecting.

"Adjusted R-squared" makes an adjustment to the plain-old R^2 to take account of the number of right hand side variables in the regression. R^2 measures what fraction of the variation in the left hand side variable is explained by the regression. When you add another right hand side variable to a regression, R^2 always rises. (This is a numerical property of least squares.) The adjusted R^2 , sometimes written \bar{R}^2 , subtracts a small penalty for each additional variable added.

"F-statistic" and "Prob(F-statistic)" come as a pair and are used to test the hypothesis that none of the explanatory variables actually explain anything. Put more formally, the " F -sta-

tistic” computes the standard F -test of the joint hypothesis that all the coefficients, except the intercept, equal zero. “Prob(F-statistic)” displays the p -value corresponding to the reported F -statistic. In this example, there is essentially no chance at all that the coefficients of the right-hand side variables all equal zero.

Parallel construction notice: The fourth and fifth columns in EViews regression output report the t -statistic and corresponding p -value for the hypothesis that the individual coefficient in the row equals zero. The F -statistic in the summary area is doing exactly the same test for all the coefficients (except the intercept) together.

This example has only one such coefficient, so the t -statistic and the F -statistic test exactly the same hypothesis. Not coincidentally, the reported p -values are identical and the F is exactly the square of the t , $2672 = 51.7^2$.

Our final summary statistic is the “Durbin-Watson,” the classic test statistic for serial correlation. A Durbin-Watson close to 2.0 is consistent with no serial correlation, while a number closer to 0 means there probably is serial correlation. The “DW,” as the statistic is known, of 0.095 in this example is a very strong indicator of serial correlation.

EViews has extensive facilities both for testing for the presence of serial correlation and for correcting regressions when serial correlation exists. We’ll look at the Durbin-Watson, as well as other tests for serial correlation and correction methods, later in the book. (See [Chapter 13. “Serial Correlation—Friend or Foe?”](#)).

A Multiple Regression Is Simple Too

Traditionally, when teaching about regression, the simple regression is introduced first and then “multiple regression” is presented as a more advanced and more complicated technique. A simple regression uses an intercept and one explanatory variable on the right to explain the dependent variable. A multiple regression uses one or more explanatory variables. So a simple regression is just a special case of a multiple regression. In learning about a simple regression in this chapter you’ve learned all there is to know about multiple regression too.

Well, almost. The main addition with a multiple regression is that there are added right hand-side variables and therefore added rows of coefficients, standard errors, *etc.* The model we’ve used so far explains the log of NYSE volume as a linear function of time. Let’s add two more variables, time-squared and lagged log(volume), hoping that time and time-squared will improve our ability to match the long-run trend and that lagged values of the dependent variable will help out with the short run.

In the last example, we entered the specification in the **Equation Estimation** dialog. I find it much easier to type the regression command directly into the command pane, although the

method you use is strictly a matter of taste. The regression command is `ls` followed by the dependent variable, followed by a list of independent variables (using the special symbol “C” to signal EViews to include an intercept.) In this case, type:

```
ls log(volume) c @trend @trend^2 log(volume(-1))
```

and EViews brings up the multiple regression output shown to the right.

You already knew some of the numbers in this regression because they appeared in the second column in Table 1 on page 63. When you specify a multiple regression, EViews gives one row in the output for each independent variable.

The screenshot shows the EViews regression output window for the equation: UNTITLED, Workfile: HYSEVOLUME::Quarterly1. The dependent variable is LOG(VOLUME). The method used is Least Squares. The date is 12/20/06 and the time is 12:18. The sample is (adjusted): 1888Q2 2004Q1, with 464 observations included after adjustments.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.106396	0.045666	-2.329866	0.0202
@TREND	-0.000736	0.000417	-1.764606	0.0783
@TREND^2	6.63E-06	1.37E-06	4.829663	0.0000
LOG(VOLUME(-1))	0.868273	0.022910	37.89886	0.0000


R-squared	0.986826	Mean dependent var	1.385802
Adjusted R-squared	0.986740	S.D. dependent var	2.513120
S.E. of regression	0.289391	Akaike info criterion	0.366505
Sum squared resid	38.52359	Schwarz criterion	0.402193
Log likelihood	-81.02909	Hannan-Quinn criter.	0.380553
F-statistic	11485.70	Durbin-Watson stat	2.342018
Prob(F-statistic)	0.000000		

Hint: Most regression specifications include an intercept. Be sure to include “C” in the list of independent variables unless you’re sure you don’t want an intercept.

Hint: Did you notice that EViews reports one fewer observation in this regression than in the last, and that EViews changed the first date in the sample from the first to the second quarter of 1888? This is because the first data we can use for lagged volume, from second quarter 1888, is the (non-lagged) volume value from the first quarter. We can’t compute lagged volume in the first quarter because that would require data from the last quarter of 1887, which is before the beginning of our workfile range.

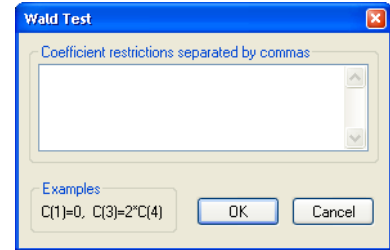
Hypothesis Testing

We’ve already seen how to test that a single coefficient equals zero. Just use the reported *t*-statistic. For example, the *t*-statistic for lagged log(volume) is 37.89 with 460 degrees of freedom (464 observations minus 4 estimated coefficients). With EViews it’s nearly as easy to test much more complex hypotheses.

Click the  button and choose **Coefficient Tests/Wald – Coefficient Restrictions...** to bring up the dialog shown to the right.

In order to whip the **Wald Test** dialog into shape you need to know three things:

- EViews names coefficients C(1), C(2), C(3), *etc.*, numbering them in the order they appear in the regression. As an example, the coefficient on LOG(VOLUME(-1)) is C(4).
- You specify a hypothesis as an equation restricting the values of the coefficients in the regression. To test that the coefficient on LOG(VOLUME(-1)) equals zero, specify “C(4) = 0”.
- If a hypothesis involves multiple restrictions, you enter multiple coefficient equations separated by commas.



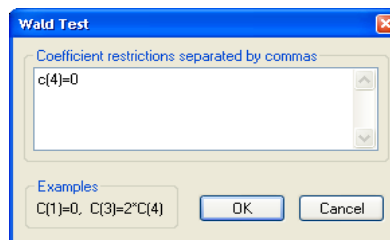
Let's work through some examples, starting with the one we already know the answer to: Is the coefficient on LOG(VOLUME(-1)) significantly different from zero?

Hint: We know the results of this test already, because EViews computed the appropriate test statistic for us in its standard regression output.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.106396	0.045666	-2.329866	0.0202
@TREND	-0.000736	0.000417	-1.764606	0.0783
@TREND*2	6.63E-06	1.37E-06	4.829663	0.0000
LOG(VOLUME(-1))	0.868273	0.022910	37.89886	0.0000

R-squared	0.986826	Mean dependent var	1.385802
Adjusted R-squared	0.986740	S.D. dependent var	2.513120
S.E. of regression	0.289391	Akaike info criterion	0.366505
Sum squared resid	38.52359	Schwarz criterion	0.402193
Log likelihood	-81.02909	Hannan-Quinn criter.	0.380553
F-statistic	11485.70	Durbin-Watson stat	2.342018
Prob(F-statistic)	0.000000		

Complete the **Wald Test** dialog with $C(4) = 0$.



EViews gives the test results as shown to the right.

EViews reports an F -statistic rather than a t -statistic because the F - applies for both single and multiple restrictions. When the hypothesis involves a single restriction, as is the case here, whether you prefer the t - or the F - is a matter of taste and you can compute the t -statistic as the square root of the F -. For example, $37.89 = \sqrt{1436}$.

Test Statistic	Value	df	Probability
F-statistic	1436.324	(1, 460)	0.0000
Chi-square	1436.324	1	0.0000

Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
C(4)	0.868273	0.022910	

Restrictions are linear in coefficients.

Hint: The p -value reported by EViews is computed for a two-tailed test. If you're interested in a one-tailed test, you'll have to look up the critical value for yourself.

Suppose we wanted to test whether the coefficient on $\text{LOG}(\text{VOLUME}(-1))$ equaled one rather than zero. Enter " $c(4) = 1$ " to find the new test statistic.

So this hypothesis is also easily rejected.

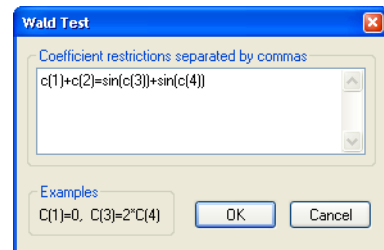
Test Statistic	Value	df	Probability
F-statistic	33.05886	(1, 460)	0.0000
Chi-square	33.05886	1	0.0000

Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
-1 + C(4)	-0.131727	0.022910	

Restrictions are linear in coefficients.

Econometric theory warning: If you’ve studied the advanced topic in econometric theory called the “unit root problem” you know that standard theory doesn’t apply in this test (although the issue is harmless for this particular set of data). Take this as a reminder that you and EViews are a team, but you’re the brains of the outfit. EViews will obediently do as it’s told. It’s up to you to choose the proper procedure.

EViews is happy to test a hypothesis involving multiple coefficients and nonlinear restrictions. To test that the sum of the first two coefficients equals the product of the sines of the second two coefficients (and to emphasize that EViews is perfectly happy to test a hypothesis that is complete nonsense) enter “ $c(1) + c(2) = \sin(c(3)) + \sin(c(4))$ ”.



Not only is the hypothesis nonsense, apparently it’s not true.

Test Statistic	Value	df	Probability
F-statistic	458.5520	(1, 460)	0.0000
Chi-square	458.5520	1	0.0000

Normalized Restriction (= 0)	Value	Std. Err.
C(1) + C(2) - @SIN(C(3)) - @SI...	-0.870353	0.040644

Delta method computed using analytic derivatives.

A good example of a hypothesis involving multiple restrictions is the hypothesis that there is no time trend, so the coefficients on both t and t^2 equal zero. Here’s the **Wald Test** view after entering “ $c(2) = 0, c(3) = 0$ ”.

The hypothesis is rejected. Note that EViews correctly reports 2 degrees of freedom for the test statistic.

Test Statistic	Value	df	Probability
F-statistic	16.81854	(2, 460)	0.0000
Chi-square	33.63707	2	0.0000

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.000736	0.000417
C(3)	8.63E-06	1.37E-06

Restrictions are linear in coefficients.

Representing

The **Representations** view, shown at the right, doesn't tell you anything you don't already know, but it provides useful reminders of the command used to generate the regression, the interpretation of the coefficient labels C(1), C(2), *etc.*, and the form of the equation written out with the estimated coefficients.

```

Equation: UNTITLED Workfile: NYSEVOLUME::Quarterly1
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Estimation Command:
=====
LS LOG(VOLUME) C @TREND @TREND^2 LOG(VOLUME(-1))

Estimation Equation:
=====
LOG(VOLUME) = C(1) + C(2)*@TREND + C(3)*@TREND^2 + C(4)*LOG(VOLUME(-1))

Substituted Coefficients:
=====
LOG(VOLUME) = -0.106395655155 - 0.000736156481983*@TREND +
6.63104154813e-06*@TREND^2 + 0.868273200859*LOG(VOLUME(-1))

```

Hint: Okay, okay. Maybe you didn't really need the representations view as a reminder. The real value of this view is that you can copy the equation from this view and then paste it into your word processor, or into an EViews batch program, or even into Excel, where with a little judicious editing you can turn the equation into an Excel formula.

What's Left After You've Gotten the Most Out of Least Squares

Our regression equation does a pretty good job of explaining $\log(\text{volume})$, but the explanation isn't perfect. What remains—the difference between the left-hand side variable and the value predicted by the right-hand side—is called the *residual*. EViews provides several tools to examine and use the residuals.

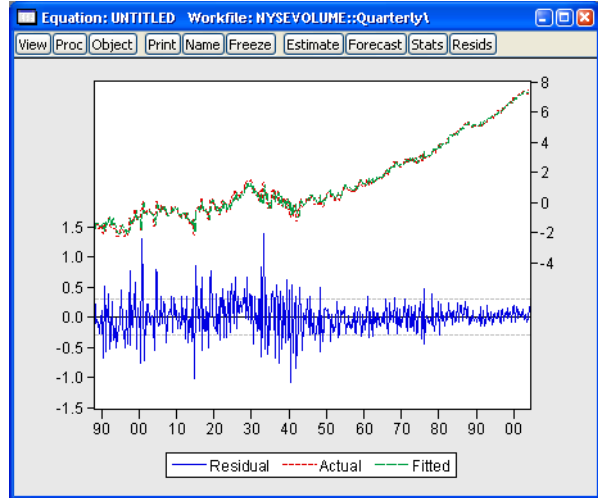
Peeking at the Residuals

The View **Actual, Fitted, Residual** provides several different ways to look at the residuals.

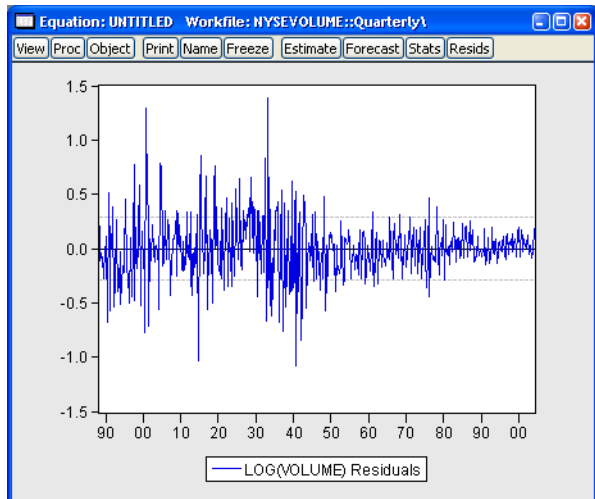
Actual, Fitted, Residual Table
 Actual, Fitted, Residual Graph
 Residual Graph
 Standardized Residual Graph

Usually the best view to look at first is **Actual, Fitted, Residual/Actual, Fitted, Residual Graph** as illustrated by the graph shown here.

Three series are displayed. The residuals are plotted against the left vertical axis and both the actual ($\log(\text{volume})$) and fitted (predicted $\log(\text{volume})$) series are plotted against the vertical axis on the right. As it happens, because our fit is quite good and because we have so many observations, the fitted values nearly cover up the actual values on the graph. But from the residuals it's easy to see two facts: our model fits better in the later part of the sample than in the earlier years—the residuals become smaller in absolute value—and there are a very small number of data points for which the fit is really terrible.

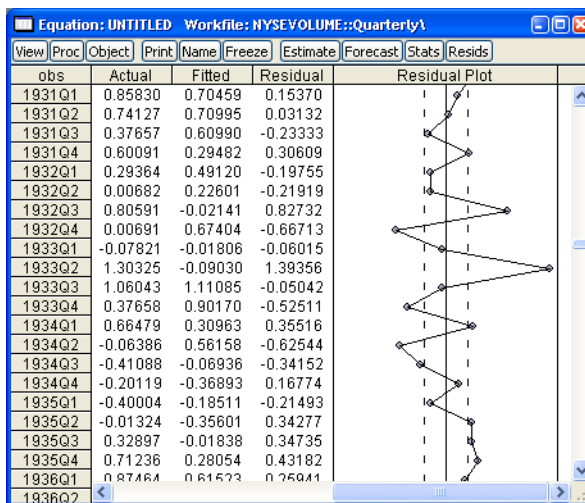


Points with really big positive or negative residuals are called *outliers*. In the plot to the right we see a small number of spikes which are much, much larger than the typical residual. We can get a close up on the residuals by choosing **Actual, Fitted, Residual/Residual Graph**.



It might be interesting to look more carefully at specific numbers. Choose **Actual, Fitted, Residual/Actual, Fitted, Residual Table** for a look that includes numerical values.

You can see enormous residuals in the second quarter for 1933. The actual value looks out of line with the surrounding values. Perhaps this was a really unusual quarter on the NYSE, or maybe someone even wrote down the wrong numbers when putting the data together!



Grabbing the Residuals

Since there is one residual for each observation, you might want to put the residuals in a series for later analysis.

Fine. All done.

Without you doing anything, EViews stuffs the residuals into the special series resid after each estimation. You can use RESID just like any other series.

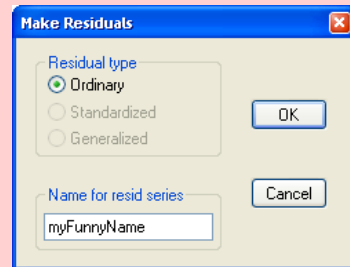
Resid Hint 1: That was a very slight fib. EViews won't let you include RESID as a series in an estimation command because the act of estimation changes the values stored in RESID.

Resid Hint 2: EViews replaces the values in RESID with new residuals after each estimation. If you want to keep a set, copy them into a new series as in:

```
series rememberresids = resid
```

before estimating anything else.

Resid Hint 3: You can store the residuals from an equation in a series with any name you like by using **Proc/Make residual series...** from the equation window.



Quick Review

To estimate a multiple regression, use the `ls` command followed first by the dependent variable and then by a list of independent variables. An equation window opens with estimated coefficients, information about the uncertainty attached to each estimate, and a set of summary statistics for the regression as a whole. Various other views make it easy to work with the residuals and to test hypotheses about the estimated coefficients.

In later chapters we turn to more advanced uses of least squares. Nonlinear estimation is covered, as are methods of dealing with serial correlation. And, predictably, we'll spend some time talking about forecasting.

