Presentation of Regression Results

I've put together some information on the "industry standards" on how to report regression results. Every paper uses a slightly different strategy, depending on author's focus. Please review the earlier handout on presenting data and tables, much of that discussion applies to regression tables as well

Regression Tables

- EViews generates a lot of information that you will not use for your analysis. You will probably need to create tables within Word that report this information in a more succinct way.

- Standard information to report on a regression table
  - Dependent variable
  - Explanatory variables
  - Estimates for constant term and coefficient estimates for explanatory variables
  - Standard errors for estimates above
  - Indication of which variables are statistically significant (using *, see below)
  - R-squared
  - Number of observations/sample period

- It is common to report more than one regression in your paper, although you may only discuss one for your presentation. Based on which variables you find to be significant, you may drop some of the explanatory variables to improve the efficiency (precision of the regression's "fit" to the data) of the coefficient estimates.

<u>Cross section data on math scores</u>

Regression Output

EViews output looks like the following

Dependent Variable: MATH91
Method: Least Squares
Date: 04/26/05 Time: 08:29
Sample: 1 407
Included observations: 402

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.025811	0.090037	-0.286669	0.7745
DRUGS	-0.946305	0.437244	-2.164248	0.0310
ENROLL	0.072621	0.077262	0.939934	0.3478
MATH87	0.637418	0.037199	17.13537	0.0000
SES	0.134832	0.039340	3.427364	0.0007
URBAN	0.000486	0.000880	0.552353	0.5810
MALE	0.069283	0.061524	1.126117	0.2608

R-squared	0.550107	Mean dependent var	-0.025206
Adjusted R-squared	0.543273	S.D. dependent var	0.902533
S.E. of regression	0.609946	Akaike info criterion	1.866367
Sum squared resid	146.9535	Schwarz criterion	1.935957
Log likelihood	-368.1397	F-statistic	80.49775
Durbin-Watson stat	1.971175	Prob(F-statistic)	0.000000

An example of what the regression table "should" look like. Note that it should be made clear in the text what the variables are and how each is measured.

Table #1: Regression Results for Student 1991 Math Scores (standard deviations from the mean)

Constant	-0.026
	(0.090)
Drugs	-0.946**
	(0.437)
Enrollment	0.0726
	(0.077)
1987 math score	0.637***
	(0.037)
Socio-economic status	0.135***
	(0.039)
Urban	0.0005
	(0.0009)
Male	0.069
	(0.062)
R-squared	0.550
No. observations	407

Standard errors are reported in parentheses.
*, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

If you want to report results from multiple regressions, you can use the above format. If you clearly label each column, you will be able to refer to this table in your text when comparing regression results and conducting your analysis. For example, the table below reports four different regressions

In the regressions below, the researcher is interested in the individual characteristics of the student (regressions A and B), the characteristics of the classroom (regression C), and the characteristics of the area surrounding the school (regressions D and E)**.**

The adjusted R-squared is reported because it "corrects" for adding more variables to a regression. Adding more explanatory variables always improves the R-squared (more data provide a better fit), the adjusted R-squared accounts for this.

The number of observations should be reported for each regression, if the availability of the data somehow limits the number of observations. If the same number is used in each regression, then reporting the sample once is sufficient.

Table #2: Regression Results for Student 1991 Math Scores (standard deviations from the mean)

	A	B	C	D	E
Constant	-0.044	-0.071	-0.095	0.012	-0.026
	(0.031)	(0.044)	(0.067)	(0.083)	(0.090)
1987 math score	0.648***	0.645***	0.648***	0.641***	0.637***
	(0.037)	(0.037)	(0.037)	(0.038)	(0.037)
Socio-economic status	0.133***	0.134***	0.129***	0.133***	0.135***
	(0.039)	(0.039)	(0.039)	(0.040)	(0.039)
Enrollment			0.050	0.067	0.0726
			(0.059)	(0.078)	(0.077)
Drugs				-0.917**	-0.946**
				(0.437)	(0.437)
Urban				0.0004	0.0005
				(0.0008)	(0.0009)
Male		0.051			0.069
		(0.061)			(0.062)
R-squared	0.540	0.541	0.540	0.549	0.550
Adjusted R-squared	0.537	0.537	0.536	0.543	0.543
No. observations	407				

Standard errors are reported in parentheses.
*, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

In the table above, it is easy for the reader/audience to identify right away which variables are statistically significant (different from zero, so that they have an effect on Math 91 score). Also, the table above shows a natural progression of regression models used by the researcher (see the description on the previous page.

If you are familiar with regression analysis, then you might report other key statistics related to possible heteroskedasticity or autocorrelation in the regression equation. These could be reported in the same area as the R-squared statistics above.

Note that while you may run several regressions, only report the results for those that would be of interest to the reader. Reporting results from hundreds of regressions is not useful – you couldn't possible discuss the results for all them in your research paper.

You presentation stands alone from your research paper in some sense. You should only report the results for one or two regressions during your presentation because this is all that you will have time to discuss. Choose the one or two that you believe will help you to address the research question most effectively with the limited time that you have.

Time series data on inflation and unemployment

This uses the Phillips Curve data from before (see previous handout). Note that the regressions are divided into sub-samples. Notice, the researcher did a total of 6 regressions – one for each sample period (full sample, early sample, and late sample) using two different specifications (one that includes the lagged interest rate as a proxy for expected inflation). The "Raw output" below is for one of those regressions. The table below reports results from all 6 regressions.

Raw output for 1968-2004 regression of a standard Phillips Curve:

Dependent Variable: INFLATION100
Method: Least Squares
Date: 04/26/05 Time: 13:16
Sample: 1968M12 2004M12
Included observations: 433

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.976473	0.618711	3.194503	0.0015
UNEMPLOYMENT	0.467007	0.097761	4.777041	0.0000

R-squared	0.050285	Mean dependent var	4.853794
Adjusted R-squared	0.048081	S.D. dependent var	3.016956
S.E. of regression	2.943533	Akaike info criterion	5.001707
Sum squared resid	3734.352	Schwarz criterion	5.020509
Log likelihood	-1080.870	F-statistic	22.82012
Durbin-Watson stat	0.015753	Prob(F-statistic)	0.000002

Table #1: Estimated Phillips Curves (1948-2004)

Standard Phillips Curve

	Full sample	*1948-1968*	*1969- 2004*
Constant	0.208	0.648***	1.977***
	(0.065)	(0.037)	(0.467)
Unemployment	-0.071	0.645***	0.134***
	(0.044)	(0.037)	(0.039)

Expectations-augmented Phillips Curve

	Full sample	*1948-1968*	*1969- 2004*
Constant	0.208***	0.390**	0.422***
	(0.065)	(0.178)	(0.072)
Unemployment	-0.032**	-0.059*	-0.071***
	(0.011)	(0.033)	(0.012)
Lagged inflation	0.990***	0.937***	1.00***
	(0.006)	(0.017)	(0.006)

Standard errors are reported in parentheses.
*, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.