Interpreting Regression Results: a basic framework

Nicola Limodio*

December 13, 2012

This scheme is meant to be illustrative and does not apply to all contexts, neither for problem sets or exam questions. Following this scheme does not provide you with a one-fit-all solution for econometrics interpretation; this course and in general understanding this subject goes well beyond “cookbook recipes” and you need to study from the textbook, solving problem sets and studying past exams.

However here I report the basic results that need to be noticed in interpreting the output of a regression and that may help you to structure logically your comments.

Suppose we face the following econometric model

$$y_i = b_1 + b_2 x_{1i} + b_3 x_{2i} + \epsilon_i$$

Significance: first and foremost we need to check whether the whole model has any explanatory power, this is easily spot through the $F$ test. Recall that this test is based on the following null hypothesis

$$H_0: b_2 = b_3 = 0 \quad H_1: b_2 \neq 0 \text{ or } b_3 \neq 0 \text{ or } b_2 = b_3 \neq 0$$

first we check that the $F$ statistic permits to reject this test at a certain level (ie 5%), then we move to a lower level (ie 1%) and if still the test holds we proceed and check for 0.1% (this kind of progressive testing should always be kept in mind for $F$, $t$ tests and other general tests). Now we can move on and verify whether each coefficient is statistically different from zero (you need to know how to define the $F$ statistic in terms of $R^2$). Stata will report for each coefficient its magnitude, standard errors, $t$ statistic and the $p$-value associated with the above $t$ statistic. Recall that for each coefficient the null hypothesis is

$$H_0: b_2 = 0 \quad H_1: b_2 \neq 0$$

at this point you need to discuss again the $t$ statistic and apply the progressive testing that was previously discussed.

Sign: quickly report the sign of each coefficient and whether it confirms our ex-ante expectation.

Magnitude: first of all define the coefficient (is it a marginal effect? an elasticity? a semi-elasticity?), the move on discussing its size and especially keep in mind the question and why this coefficient may be informative. For more complex models in which non-linear, interactive terms or dummy variables appear, it may be useful to re-arrange terms in order to gain a clearer insight. For example assume now we face this model

$$y_i = b_1 + b_2 x_{1i} + b_3 x_{2i} + b_4 x_{2i}^2 + \epsilon_i$$

---

*Special thanks to Stephan Maurer for providing useful feedback.
what does $b_4$ mean? In principle it represents the change in $y_i$ given that all other variables but $x_{2i}^2$ stay constant; but clearly as $x_{2i}^2$ changes also $x_{2i}$ does. Therefore we need to be a little careful in this part. If the previous model is rearranged as follows

$$y_i = b_1 + b_2 x_{1i} + (b_3 + b_4 x_{2i}) x_{2i} + \epsilon_i$$

then it is clear that $b_3 + 2b_4 x_{2i}$ represents the marginal effect of $x_{2i}$ on $y_i$ and that this depends on the level of $x_{2i}$, while $b_4$ reports whether the changes in $x_{2i}$ are increasing ($b_4 > 0$), constant ($b_4 = 0$) or decreasing ($b_4 > 0$) in $x_{2i}$.

Explanation: how can we explain the results of the regression? Do these follow an economic model, sociological theory or lab experiment? Which forces may lie behind this result? The regression we saw in class analyzed

$$S_i = b_1 + b_2 AB_i + b_3 MALE_i + \epsilon_i$$

how educational attainment (years of schooling $S_i$) depends on an ability index ($AB_i$) and a dummy indicating sex ($MALE_i$ takes value 1 when individual $i$ is male and 0 for female). From the regressions in our dataset it turned out that $b_3$ was significant, negative and close to one, hence males seem to have on average almost one year of schooling less than females. As many of you argued this is puzzling. An interpretation of this result may be clear if we believe that male educational attainment depends on two forces: a positive one that pushes guys to stay in school more (ie. my PhD cohort is made of 18 out of 20 males) and a negative one that pushes guys to drop out of school earlier (ie. a job can be found more easily and with a higher wage than girls). If we believe that these are the forces behind male educational attainment, then this regression implies that the second (negative) force is much stronger and results in a negative $b_3$ coefficient.

This is a simple and general framework for analysing regressions, during your problem sets or exam questions you may be asked specifically to develop at a deeper level some of these arguments (ie. construct the $F$ or $t$ test, discuss $R^2$, etc...), hence this scheme is non-exhaustive and you should not rely on it for a pass grade. This document contains only some essential pieces of information that I have obsessively reported during classes and that you should have already mastered massively... but in case you have not done it yet, take some time for this.