Comments on the Homework

Johannes Boehm

February 4, 2010

1 Homework 1, 2

- When you interpret parameter estimates, always pay attention to the units the variables are measured in. Aviod using "A increases by 5 units if B increases by one unit".
- The R^2 is always between 0 and 1. You cannot use it to explain whether the two variables are positively or negatively related. Use the sign of the estimated coefficient instead.
- Avoid saying "there is a weak relationship between A and B". What exactly is a "weak relationship"? Do you mean that the coefficient is low, or do you mean that the fit (i.e. the R^2) is bad? Try to be precise.
- Be careful with the word "test", as it has a special meaning in statistics (econometrics being a part thereof). You should only use it in the context of hypothesis testing.
- When your homework is longer than two, sometimes three pages (excluding computer output), chances are high that you wrote something unnecessary. Try to be precise. In the exam you will have very limited time.

2 Homework 3

- If we already know $\beta_2 > 0$, then there is no reason to test. However, if we can only rule out $\beta_2 \geq 0$, then there is reason to use a one-sided test.
- Maybe I was not really clear during class: the fact that the statistic $t = (b_2 - \beta_2^*)/sd(b_2)$ is t-distributed with $N - k$ degrees of freedom only holds under $H_0 : \beta = \beta_2^*$.

3 Homework 4

- When you describe regression output, write down the regression equation instead of the STATA command.
- SF is father's schooling, SM is mother's schooling. Read the description of the data.

4 Homework 5, 6

- Please attach the computer output to the homework. Otherwise it's hard to check whether what you say makes sense.
- There is some confusion about the interpretation of the p value, which may be partly caused by the interpretation in the book. Probably the easiest definition/interpretation is the following: "The p value is the smallest significance level such that you still reject the null hypothesis". If you *really* want to interpret it as a probability: suppose your test statistic is 5, then you could say "Under the null hypothesis, the probability that the test statistic is greater than 5 [in absolute value] is exactly p". If possible, I would refrain from interpreting the p value as a probability, but rather use the first definition. I will talk to Chris about this and will let you know how he thinks about this.
- If you run a regression with an intercept and a dummy variable, and the intercept cannot be reasonably interpreted (say, because the other regressors are nowhere near zero), then it clearly does not make sense to interpret the intercept for the subsample which is characterized by the dummy variable (i.e. if b_1 is the intercept estimate and δ is the estimate for the coefficient of the dummy, you cannot interpret $b_1 + \delta$ if you cannot interpret b_1). You should rather interpret the coefficient of the dummy on its own.

5 Homework 7

- A statement like "we reject H_0 at all significance levels" cannot be correct. For sufficiently small test size *alpha* the critical value goes to infinity, so for *some* very small test size you will not be able to reject H_0. If you want to emphasize that the t-statistic is very large, say that you reject H_0 at the 0.1 percent level.
- Many of you wrote something like "the test statistic is significant". Usually the word "significant" is only used for the parameters that you test, i.e. "the parameters are jointly significant". I do not really know how important things like these are in the exam.

6 Homework 8

- Some of you confuse estimate and true value, especially in the context of the bias in the case of omitted variable bias. In the first problem, we had

$$E(b_2) = \beta_2 + \beta_3 \frac{\sum(S_i - \bar{S})(EXP_i - E\bar{X}P)}{\sum(S_i - \bar{S})}.$$

  You cannot simply plug in the estimates for the true values β_2, β_3, because you will never know the true values. Sad, but true.

- When you claim that there is omitted variable bias, you have to show that the excluded variable is actually important.

- Having an upward bias is not equivalent to the estimates actually being higher when you leave out the variable. The bias is a theoretical moment (see first point), the difference in the estimates is in some ways a realization of a random variable. See also question 4.

7 Homework 11

- Be specific when you specify H_0 and H_1. Do not write something like "no systematic difference in the estimates" for H_0 in the Hausman exogeneity test (yes, it's probably in the book). What is a systematic difference? Write "no measurement error" instead, that's well defined.