First test in Introductory Econometrics — Tentative answers

Robert M. Kunst

November 8, 2017

1. Consider the Gauss-Markov assumptions for the linear regression model:[8 points]

   (a) Please provide the five assumptions that are required in the *multiple* linear regression model for the BLUE property of the OLS estimator.[**A: Obvious**]

   (b) What does 'BLUE' stand for? Is it conceivable that an estimator with a smaller variance can be constructed, even if the five assumptions hold?[**A: first part, obvious; second part, yes, there can be biased estimators with smaller variance, and there can be unbiased nonlinear estimators with smaller variance, with the special issue that the latter case is not possible with a Gaussian distribution.**]

   (c) Why do we often assume a sixth property MLR.6? For what do we need that one? OLS is BLUE anyway.[**A: for inference, for the exact distribution in restriction tests**]

   (d) Which of the five Gauss-Markov assumptions is not necessary for the unbiasedness of OLS?[**A: homoskedasticity**]

2. Problems in empirical applications.[6 points]

   (a) A macro-economist estimates an investment equation. She regresses aggregate equipment investment on a short-run interest rate (bill rate) and on a long-run interest rate (bond rate). Correlation between the two interest rates is positive and strong, approximately 0.8. MLR.1, MLR.2, MLR.4, MLR.5 are fulfilled. Will OLS be BLUE?[**A: yes. Non-zero correlation among regressors does not imply multicollinearity in the sense**

of MLR.3; even if it is 0.99 it will not invalidate Gauss-Markov conditions, and OLS will be BLUE. Of course, it may cause large variances and low significance of t–values, but this was not the question here.]

(b) The same macro-economist adds a third regressor to the equation, the (arithmetic) average of the short and the long rate. What is the problem of this specification?[**A: Now we have a violation of MLR.3, and the OLS estimate cannot be calculated**]

(c) Just for the sake of an experiment, instead of averaging the two interest rates, she now forms the geometric average by multiplying them and taking the square root. Will she encounter any estimation problem this time, with the three regressors short rate, long rate, and geometric average?[**A: In this case, the third regressor is a non-linear function of the other regressors, and MLR.3 will hold**]

3. Restriction tests in multiple regression.[5 points]

(a) A micro-economist regresses the monthly salary (wage) of 500 individuals on several indicators, such as the years of education x_1, a gender dummy x_2, an indicator for the ethnic group x_3, mother's salary or wage x_4, father's salary or wage x_5. The regression software yields coefficient estimates and standard errors, but no t–statistics. How can the economist calculate t–statistics? What are these t–statistics good for?[**A: t–values or t–statistics can be calculated by dividing coefficient estimates by the corresponding standard errors, in symbols $t_j = \hat{\beta}_j/\hat{\sigma}_j$. They serve for testing restriction hypotheses $H_0 : \beta_j = 0$ in t–tests.**]

(b) It turns out that t–statistics on β_2 and β_3 are small in absolute values, and the economist considers eliminating the variables x_2 and x_3. What is the correct way to formulate this restriction as a null hypothesis, written in the parameters of the model? What is the distribution of the corresponding restriction statistic under the null?[**A: $H_0 : \beta_2 = 0, \beta_3 = 0$ or compactly $H_0 : \beta_2 = \beta_3 = 0$. The null distribution of the corresponding restriction test, and F–test, is an F distribution, here an F(2,494) distribution. Note that hypotheses such as $H_0 : \hat{\beta}_2 = 0$ are regarded as serious errors and yield no points**]

(c) It turns out that the p–value for the test described in (b) is 0.23.

Does the test reject at a significance level of 5%?[**A: No. Only if the p–value were less than 0.05, the test would reject.**]

4. Understanding basic concepts of econometrics.[6 points]

   (a) What is the difference between residuals and errors?[**A: Residuals $\hat{u}_i$ are observed random variables defined as $y_i - \hat{\beta}_0 - \hat{\beta}_1 y_{1,i} - \ldots - \hat{\beta}_k y_{k,i} = y_i - \hat{y}_i$. Errors u_i are unobserved random variables defined as $y_i - \beta_0 - \beta_1 y_{1,i} - \ldots - \beta_k y_{k,i}$.**]

   (b) Why can errors never be correlated with parameters?[**A: Parameters are (typically unobserved) constants. Errors are random variables. A random variable cannot be correlated with a real constant.**]

   (c) Why is $H_0 : \hat{\beta} = 0$ an invalid null hypothesis?[**A: Hypotheses should be formulated about the values of parameters, not about estimates. For example, $\hat{\beta}_0 = 0$ is an uninteresting null, as it can be and usually is rejected by a single observation of the intercept estimate.**]