# Third and final test in Introductory Econometrics

Robert M. Kunst

January 26, 2016

1. You run a cross-section regression using OLS, but you suspect that heteroskedasticity might be a problem. Your data set has 100 observations and 2 regressor variables apart from the intercept. [12 points]

   (a) You have run several tests for heteroskedasticity, and they reject homoskedasticity. Now, you consider two options: first, weighted least squares using estimates $\hat{\sigma}_i^2$ for the unknown error variances $\sigma_i^2$; second, OLS with standard errors corrected according to White-Eicker. Which of the two approaches promises (at least asymptotically) efficient estimation of the coefficients $\beta_j$?

   (b) Assume again that only OLS and also good variance estimates $\hat{\sigma}_i^2$ for the error variances are available. Describe how you can construct the weighted least squares (WLS) estimator.

   (c) Now assume you are stubborn and you neither correct via White-Eicker nor use WLS, but you stick to your original OLS. Will the OLS coefficient estimates be unbiased and consistent, assuming all other Gauss-Markov conditions hold (except homoskedasticity)?

2. You run a time-series regression on the model $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \beta_3 y_{t-1} + u_t$ using OLS for $n = 100$ observations. [12 points]

   (a) You want to check autocorrelation in the errors, and you decide to use the Breusch-Godfrey test for two lags. Your software does not contain any automatic command for such a test, but you can of course run OLS and you can also save residuals. Describe how you would implement the Breusch-Godfrey test here: what is the dependent variable of the auxiliary regression, what are the regressors of that auxiliary regression, how do you calculate your test statistic, what is its asymptotic distribution under the null?

   (b) The Breusch-Godfrey test rejects. Will your OLS estimator be consistent?

   (c) What would you consider doing now that the Breusch-Godfrey test has rejected, in order to obtain consistent estimation? Try and give an answer in words.

   (d) Somebody tells you that the Durbin-Watson test statistic on the regression $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \beta_3 y_{t-1} + u_t$ is quite nice and close to 2. Can you take this argument seriously? Is this regression static or dynamic?

3. A pharmaceutical manufacturer is interested in whether her drug really decreases blood pressure. 1000 persons are selected randomly from the entire population of a country, and some data for these persons is collected, particularly variables $P_i$ (the blood pressure of person $i$) and $D_i$ (the amount of monthly intake of the drug of concern by $i$; data on the usage of other drugs is not available). The regression $P_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 A_i + u_i$ is estimated by OLS, where $X_i$ denotes the gender of $i$ (dummy with 1 for females and 0 for males) and $A_i$ the age of $i$ in years.[12 points]

   (a) You find that $\hat{\beta}_1$ is actually positive, i.e. the drug appears to increase blood pressure rather than decrease it. Do you have a good explanation for this apparently counterintuitive result? What effects could be hidden in the errors $u$?

   (b) You suspect that $D$ and $u$ may be correlated, and you consider instrumental variable (IV) estimation. A potential instrument is $E$, a dummy variable with 1 in the eastern parts of the country and 0 in the west. This $E$ is positively correlated with $D$, as a competing pharmaceutics producer is very actively marketing in the west. Which (two) conditions must this instrument fulfill to be valid?

   (c) You find that gender and $D$ are positively correlated, as the women of this country enjoy taking drugs more than men. Does this correlation make your IV estimator inconsistent?

4. For the following tests, write down the null and alternative hypotheses. You can use words or symbols, but unusual notation must be explained. [9 points]:

   (a) The Durbin-Watson test;

   (b) The Breusch-Pagan test;

   (c) The $F$–test that corresponds to the second and third regression coefficient in a regression with 5 regressors.