Third and final test in Introductory Econometrics

Robert M. Kunst

January 28, 2015

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

   (a) You decide to run the White test for heteroskedasticity. Assume your software does not offer a direct command of that type and you have to do it yourself. How do you construct the dependent variable of the corresponding auxiliary regression? What are the regressors in that auxiliary regression? How do you construct your test statistic from the auxiliary regression? What are the degrees of freedom for the asymptotic χ^2 distribution of the test statistic under the null (answer is an integer number, general answers such as $n-k$ are not valid)?

   (b) Suppose the White test rejects. You consider two options: first, weighted least squares using appropriate estimates $\hat{\sigma}_i^2$ for the unknown σ_i^2; second, OLS with standard errors corrected according to White-Eicker. Which of the two approaches promises efficient estimation of the coefficients β_j?

2. You run a time-series regression in a static model of the type $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + u_t$ using OLS for $n = 100$ observations. [12 points]

   (a) You calculate the Durbin-Watson statistic, it has the value of 1.51. For $k' = 2$ and $n = 100$, you look up the lower bound for the 5% significance point, which turns out to be $d_L = 1.634$. What is your conclusion regarding autocorrelation in the errors?

   (b) You re-specify your model as the dynamic regression $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \beta_3 y_{t-1} + u_t$, and now the Durbin-Watson statistic has the value 1.75. The upper bound for the 5% significance point for $k' = 3$ and $n = 100$ turns out to be $d_U = 1.736$. Can you conclude that there is no significant autocorrelation in the errors?

   (c) You now decide to run the Breusch-Godfrey test. Specifically, you run a regression of OLS residuals from the model in (b) on x_t, z_t, y_{t-1}, and on three lags of the residuals $\hat{u}_{t-1}, \hat{u}_{t-2}, \hat{u}_{t-3}$. How do you construct the test statistic from the regression output? What is its (asymptotic) distribution under the null?

   (d) The Breusch-Godfrey test rejects at 5%. Does this suggest that the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ is consistent for the model in (b)? Would it have been consistent in the static regression of point (a)?

3. You wish to investigate the effects of wearing woollen socks on seasonal infections. The dependent variable y measures the number of days in sick leave because of flu over the year 2011, the dummy SOX is 1 for woollen and 0 for nylon socks, additionally you have a gender dummy SEX and a covariate C for the amount of vitamin C intake during 2011. Data are collected from 1000 persons all over Europe. The naive regression $y_i = \beta_0 + \beta_1 SOX_i + \beta_2 SEX_i + \beta_3 C_i + u_i$ is estimated by OLS. [12 points]

   (a) Someone points out that the average outside temperature $TEMP$ may be an important omitted variable and that woollen socks are more likely to prevent catching a cold in Finland than in Malta. True, gender and temperature may not be correlated. If C and $TEMP$ are correlated, will OLS still be a consistent estimator?

   (b) Why is $TEMP$ itself an invalid instrument for C? Suppose we consider using $PHARM$, the weekly opening hours of pharmacies, as an instrument for C. Which two conditions must such an instrument fulfil?

   (c) It turns out that SOX and SEX are correlated, as females tend to prefer nylon socks. Does this fact affect the consistency of your IV estimation, with $PHARM$ assumed as a valid instrument for C, and SOX and SEX assumed exogenous?

4. For the following tests, write down the null and alternative hypotheses [9 points]:

   (a) The Breusch-Godfrey test;

   (b) The Hausman test for endogeneity;

   (c) The t–test that corresponds to the regression intercept.