KTH Teknikvetenskap
Harald Lang

Tentamensskrivning, 10/6-2009, kl 14.00–19.00.
SF2951 Ekonometri
Hjälpmedel: miniräknare.

Number the pages, and write your name on each sheet.

---

1. Assume you have estimated the regression equation

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + e$$

and you have got the point estimates $\hat{\beta} = (\hat{\beta}_0\ \hat{\beta}_1\ \hat{\beta}_2\ \hat{\beta}_3\ \hat{\beta}_4)'$ and estimated covariance matrix V.

Now you want to test the hypothesis $\beta_1 = \beta_2 = \beta_3$. Describe how you would do this. Describe the test variable (in terms of $\hat{\beta}$ and V) and how you determine for which values of the test variable you reject the hypothesis on a given level p.

2. Assume that you want to estimate the equation

$$y = x'\beta + e, \quad P(e < 0) = 0.25$$

(this means that $x'\beta$ is the first quartile of y conditional on x.) Describe the estimation method for this problem.

3. Assume that you want to estimate the regression equation

$$y_i = \beta_0 + x_{1,i}\beta_1 + x_{2,i}\beta_2 + e_i, \quad i = 1, \dots, n \tag{1}$$

but you suspect that it will suffer from heteroskedasticity. You therefore pre-multiply with the variable u_i which you think is approximately inversely proportional to the variance of e_i. Hence you have the new regression equation

$$u_i y_i = u_i\beta_0 + u_i x_{1,i}\beta_1 + u_i x_{2,i}\beta_2 + \varepsilon_i \tag{2}$$

where ε_i are the error terms in this new equation. We know that if we estimate (1) with OLS, then the sum of the estimated residuals $\sum_i \hat{e}_i = 0$. Is this true also for the estimated residuals in (2), i.e., is $\sum_i \hat{\varepsilon}_i = 0$ necessarily (i.e., not just by chance?) Motivate your answer!

4. Explain the meaning of the conceps
– endogeneity
– multicolinearity
– self selection bias
and give simple examples of each.

5. We are interested in how working experience in years (exper) influences the wage (w). We therefore run two regression equations on the same data:

$$\ln(w) = \alpha_0 + (\text{educ})\alpha_1 + (\text{exper})\alpha_2 + e \tag{1}$$
$$\ln(w) = \beta_0 + (\text{age})\beta_1 + (\text{exper})\beta_2 + \varepsilon \tag{2}$$

("educ" is education in years, "age" is the age of the person in years.) Explain the different interpretations of α_2 and β_2. Which of these coefficients do you expect to be the largest? Motivate!

Answers

1. Define
$$R = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$
and
$$U = (R\hat{\beta})'(RVR')^{-1}(R\hat{\beta})$$
Reject the hypothesis if $U > c$ where c is chosen such that the probability that a $\chi^2(2)$-variable is $> c$ equals p.

2. Let $\hat{e}_i,\ i = 1, \ldots, n$ denote the estimated residuals. The estimation method is
$$\text{minimise} \sum_{i=1}^{n} |\hat{e}_i| h(\hat{e}_i)$$
where
$$h(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 3 & \text{if } x < 0 \end{cases}.$$

3. No. The "normal equations" says that $X'\hat{e} = 0$. In the regression equation (1) the first covariate is the cosntant 1, hence the first normal equation is $\sum \hat{e}_i = 0$. In the equation (2), however, there is no such covariate. The first normal equation is $\sum_i u_i \hat{\varepsilon}_i = 0$.

4. As for "endogeneity" and "multicolinearity", see Hansen's course notes. These concepts are also treated in many of the exercises on the course web page. The problem of "self selection bias" is extensively treated in my comments to Hansen's notes (also available on the course web page.)

5. I expect $\alpha_2 > \beta_2$. The interpretation of α_2 is the (relative) increase in wage due to one more year of working experience for a fixed education level. The interpretation of β_2 is the (relative) increase in wage due to one more year of working experience for a fixed age of the person. This means that one more year of working experience in this case often implies one year less education, and since also education is expected to have a positive impact on wage, it is reasonable to expect that β_2 is less than α_2.