Econometrics project MODS202

Prof. Patrick Waelbroeck

Academic year 2023-2024

PART 1 - CROSS-SECTION DATA

This part uses the dataset HPRICE2.RAW described in HPRICE2.DES.

1. State the fundamental hypothesis under which the Ordinary Least Squares (OLS) estimators are unbiased.
2. Show that under this assumption the OLS estimators are indeed unbiased.
3. Explain the sample selection bias with an example from the course.
4. Explain the omitted variable bias with an example from the course
5. Explain the problem of multicollinearity. Is it a problem in this dataset?
6. Create three categories of *nox* levels (low, medium, high), corresponding to the following percentiles: 0-39%, 40%-60%, 61%-100%
7. Compute for each category of *nox* level the average median price and comment on your results
8. Produce a scatter plot with the variable *price* on the y-axis and the variable *nox* on the x-axis. Is this a ceteris paribus effect?
9. Run a regression of *price* on a *constant*, *crime*, *nox*, *rooms*, *proptax*. Comment on the histogram of the residuals. Interpret all coefficients.
10. Run a regression of *lprice* on a *constant*, *crime*, *nox*, *rooms*, *proptax*. Comment on the histogram of the residuals. Interpret all coefficients.
11. Run a regression of *lprice* on a *constant*, *crime*, *lnox*, *rooms*, *lproptax*. Comment on the histogram of the residuals. Interpret all coefficients.
12. In the specification of question 10, test the hypothesis H_0: $\beta_{nox} > 0$ vs. H_1: $\beta_{nox} < 0$ at the 10% level
13. In the specification of question 10, test the hypothesis H_0: $\beta_{nox} = 0$ vs. H_1: $\beta_{nox} \neq 0$ at the 10% level using the p-value of the test
14. In the specification of question 10, test the hypothesis H_0: $\beta_{crime} = \beta_{proptax}$ at the 1% level
15. In the specification of question 10, test the hypothesis H_0: $\beta_{nox} = 0$, $\beta_{proptax} = 0$ at the 1% level
16. In the specification of question 10, test the hypothesis H_0: $\beta_{nox} = -500$, $\beta_{proptax} = -100$ at the 1% level using the p-value of the test
17. In the specification of question 10, test the hypothesis H_0: $\beta_{nox} + \beta_{proptax} = -1000$ at the 1% level using the p-value of the test
18. In the specification of question 10, test the hypothesis that all coefficients are the same for observations with low levels of *nox* vs. medium and high levels of *nox*.
19. Repeat the test of question 18 but now assuming that only the coefficients of *nox* and *proptax* can change between the two groups of observations. State and test H_0.

PART 2 - HETEROSKEDASTICITY

20. Explain the problem of heteroskedasticity with an example of the course.
21. Suppose that Eu u′ = $\sigma^2\Omega$. Show that the GLS estimator is the best linear unbiased estimator.
22. In the specification of question 10, test the hypothesis of no heteroskedasticity of linear form, i.e. in the regression of u^2 on *constant*, *crime*, *nox*, *rooms*, *proptax*, test H_0: δ_{crime}, δ_{nox}, δ_{room}, $\delta_{proptax}$ = 0, where the coefficients δ_k (k = crime, nox, rooms, proptax) are associated with the corresponding explanatory variables.
23. In the specification of question 11, test the hypothesis of no heteroskedasticity of linear form
24. In the specification of question 12, test the hypothesis of no heteroskedasticity of linear form
25. Comment on the differences between your results of questions 22,23, 24.
26. Regardless of the results of the test of question 22, identify the most significant variable causing heteroskedasticity using the student statistics and run a WLS regression with the identified variable as weight.

PART 3 - TIME SERIES DATA

This part uses the threecenturies_v2.3 datasets. Import Real GDP at market prices, unemployment rate and consumer price inflation for the period 1900-2000 in Python from the A1 worksheet.

27. Define strict and weak stationarity.
28. Explain ergodicity and state the ergodic theorem. Illustrate with an example.
29. Why do we need both stationarity and ergodicity?
30. Explain “spurious regression”.
31. Define a moving average and explain the trade-off involved in the choice of the size of the window and of whether to center or not the moving average.
32. Make all time series stationary by computing the difference between the original variable and a moving average of order 5.
33. Compute the difference between the resulting time series of question 32 and a moving average of order 10. Give the exact formula for the weights after the two transformations.
34. Using the original dataset, test the unit root hypothesis for all variables.
35. Transform all variables so that they are stationary using either your answers to questions 32-33 or to question 34.
36. Explain the difference between ACF and PACF.
37. Plot and comment on the ACF and PACF of all variables.
38. Explain the principle of parsimony and its relationship with Ockham’s razor using the theory of information criterion.
39. Explain the problem of auto-correlation of the errors.
40. Using only stationary variables, run a regression of GDP on constant, unemployment and inflation and test the hypothesis of no-autocorrelation of errors.
41. Regardless of your answer to question 40, correct auto-correlation with GLS. Comment on the difference with question 40.
42. For all variables, construct their lag 1 and lag 2 variables.
43. Run a regression of GDP on constant, lag 1 unemployment, lag 2 unemployment, lag 1 inflation, lag 2 inflation. What is the number of observations and why?
44. State and test the no-Granger causality hypothesis of inflation on GDP at the 10% level
45. Divide the sample in two groups: 1900-1950 and 1951-2000. Test the stability of coefficients between the two periods.

46. Test the structural breakpoint using a trim ratio of 25% at the 10% level