Econometrics project

Prof. Patrick Waelbroeck

Academic year 2024-2025

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: 0-25%, 26%-74%, 75%-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*. Interpret all coefficients.
11. Run a regression of *lprice* on a *constant*, *crime*, *lnox*, *rooms*, *lproptax*. Interpret all coefficients.
12. In the specification of question 9, test the hypothesis H_0: $\beta_{nox} = 0$ vs. H_1: $\beta_{nox} \neq 0$ at the 1% level using the p-value of the test
13. In the specification of question 9, test the hypothesis H_0: $\beta_{crime} = \beta_{proptax}$ at the 10% level
14. In the specification of question 9, test the hypothesis H_0: $\beta_{nox} = 0$, $\beta_{proptax} = 0$ at the 10% level
15. In the specification of question 9, test the hypothesis H_0: $\beta_{nox} = -500$, $\beta_{proptax} = -100$ at the 10% level using the p-value of the test
16. In the specification of question 9, test the hypothesis that all coefficients are the same for observations with low levels of *nox* vs. medium and high levels of *nox*.
17. Repeat the test of question 16 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

18. Explain the problem of heteroskedasticity with an example of the course.
19. In the specification of question 9, 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.

20. In the specification of question 10, test the hypothesis of no heteroskedasticity of linear form
21. In the specification of question 11, test the hypothesis of no heteroskedasticity of linear form
22. Comment on the differences between your results of questions 20,21, 22.
23. Using the specification of question 9, identify the most significant variable causing heteroskedasticity using the student statistics and run a WLS regression with the identified variable as weight. Compare the standards errors with those of question 9. Comment on your results.

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.

24. Define strict and weak stationarity.
25. Explain ergodicity and state the ergodic theorem. Illustrate with an example.
26. Why do we need both stationarity and ergodicity?
27. Explain “spurious regression”.
28. Make all time series stationary by computing the difference between the original variable and a moving average of order 2x10. Give the formula for the exact weights.
29. Using the original dataset, test the unit root hypothesis for all variables.
30. Transform all variables so that they are stationary using either your answers to questions 28 or to question 29.
31. Explain the difference between ACF and PACF.
32. Plot and comment on the ACF and PACF of all variables.
33. Explain the principle of parsimony and its relationship with Ockham’s razor using the theory of information criterion.
34. Explain the problem of auto-correlation of the errors.
35. Using only stationary variables, run a regression of GDP on constant, unemployment and inflation and test the hypothesis of no-autocorrelation of errors.
36. Regardless of your answer to question 35, correct auto-correlation with GLS. Test again for the presence of auto-correlation. Comment on your results.
37. For all variables, construct their lag 1 and lag 2 variables.
38. 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?
39. State and test the no-Granger causality hypothesis of unemployment on GDP at the 1% level
40. Divide the sample in two groups: 1900-1960 and 1961-2000. Test the stability of coefficients between the two periods.
41. Test the structural breakpoint using a trim ratio of 30% at the 1% level
42. Divide the sample into 3 periods of equal length. Test that the coefficients of the second and the third periods are equal. Formulate the null hypothesis and interpret your results.