BA-BMECV2502U Econometrics

September 9, 2025

Copenhagen Business School

Department of Economics

Ralf A. Wilke

2025/2026, efterårsemester

Problem Set 3: OLS: topics & Policy Analysis

1. Use the data in OKUN.dta to answer this question. The Okun's law from macroeconomics suggests that the relationship between change in unemployment rate and change in real GDP is negative.

   (a) Estimate the equation

   $$pcrgdp_t = \beta_0 + \beta_1 cunem_t + u_t.$$

   [Okun's law suggests that the coefficient on $cunem$ is -2.] and test the errors for AR(1) serial correlation, without assuming $\{cunem_t; t = 1, 2, ...\}$ is strictly exogenous. What do you conclude?

   (b) Regress the squared residuals, $\hat{u}_t^2$, on $cunem_t$ (this is the Breusch-Pagan test for heteroskedasticity in the simple regression case). What do you conclude?

(c) Obtain the heteroskedasticity-robust and the Newey-West standard error for the OLS estimate $\hat{\beta}_1$. Are they substantially different from the usual OLS standard errors? What do you conclude?

2. When the errors in a regression model have AR(1) serial correlation, why do the OLS standard errors tend to underestimate the sampling variation in $\hat{\beta}_j$? It is always true that the OLS standard errors are too small?

3. (a) Suppose a regression of the OLS residuals on the lagged residuals produces $\hat{\rho} = 0.841$ and se$(\hat{\rho}) = 0.053$. What implications does this have for OLS?

   (b) If you want to use OLS but also want to obtain a valid standard error for the coefficients, what would you do?

4. Use the data in INTDEF.dta for this exercise. A simple equation relating the three-month T-bill rate to the inflation rate (constructed from the Consumer Price Index) is

$$i3_t = \beta_0 + \beta_1 inf_t + u_t.$$

   (a) Estimate this equation by OLS, omitting the first time period for later comparisons. Report the results in the usual form.

   (b) Now, first difference the equation and use:

$$\Delta i3_t = \beta_0 + \beta_1 \Delta inf_t + \Delta u_t.$$

   Estimate this by OLS and compare the estimate of β_1 with the previous estimates.

5. Use the data in JTRAIN98.dta to answer this question. The variable *unem98* is a binary variable indicating whether a worker was unemployed in 1998. It can be used to measure the effectiveness of the job training program in reducing the probability of being unemployed.

   (a) What percentage of workers was unemployed in 1998, after the job training programme? How does this compare with the unemployment rate in 1996?

   (b) Run the simple regression *unem98* on *train*. How do you interpret the coefficient on *train*? Is it statistically significant? Does it make sense to you?

   (c) Add to the regression in part (b) the explanatory variables *earn96*, *educ*, *age*, and *married*. Now interpret the estimated training effect. Why does it differ so much from that in part (b)?

(d) Now perform full regression adjustment by running a regression with a full set of interactions, where all variables (except the training indicator) are centered around their sample means:

$$unem98_i \text{ on } train_i,\ earn96_i,\ educ_i,\ age_i,\ married_i,$$
$$train_i * (earn96_i - \overline{earn}96),\ train_i * (educ_i - \overline{educ}),\ train_i * (age_i - \overline{age}),$$
$$train_i * (married_i - \overline{married}_i).$$

This regression uses all of the data. What happens to the estimated average treatment effects of $train$ compared with part (c). Does its standard error change much?

(e) Are the interaction terms in part (d) jointly significant?

(f) Verify that you obtain exactly the same average treatment effect if you run two separate regressions and use the formula in equation (7.43) in W2020 (or bottom of page 11 of lecture slides). That is, run two separate regressions for the control and treated groups, obtain the fitted values $\widehat{unem98}_i^{(1)}$ and $\widehat{unem98}_i^{(0)}$ for everyone in the sample, and then compute

$$\hat{\tau}_{ura} = n^{-1} \sum_{i=1}^{n} \left[\widehat{unem98}_i^{(1)} - \widehat{unem98}_i^{(0)} \right].$$

Check this with the coefficient on $train$ in part (d). Which approach is more convenient for obtaining a standard error?

These problems have been taken from the textbook "Introductory Econometrics" by J.Wooldridge, 7th edition, 2020.