EXERCISE SHEET 11: HORNUNG (2015): RAILROADS AND GROWTH IN PRUSSIA

Main learning goals: 1. More on instrumental variables

2. Some basics on propensity score matching
3. Implement and understand logistic regressions
4. Interpretations of marginal effects
5. Estimating treatment effects with matching estimators

Hand-in: As usual: do-file and the corresponding log-file (saved in .log-format) containing the following steps (numbered in the do-file).

1. Open the dataset `hornung-rail-cross-section.dta` and make yourself familiar with the variables and the data structure. Generate the population growth rates needed for the study.
2. Replicate the summary statistics of table 2 of the paper by using the `tabstat` command. Use `estpost ttest` for the t-tests. Briefly interpret your findings. Do your results for the t-test change if (a) you exclude the observations with area changes before 1872 or (b) constrain on the same sample by using the `listwise` option.
3. Set up a global macro with the control variables. Then, replicate table 3 with the `tabstat` and `estpost ttest` commands. What can we learn from this table? What are potential concerns/problems?
4. Estimate the OLS regression (1) for the cross-section. Make sure that you constrain the sample as specified in the paper. How could reverse causality, unobserved heterogeneity, and omitted variable bias affect the estimates and weaken the identification?
5. Briefly discuss the role of the instrument. Estimate the IV regressions for first (panel B) and second stage (panel C) of table 5, column (1) and (2). Is the instrument valid? What can we learn by comparing OLS and IV results? Are you convinced by the way how Erik Hornung defends the exclusion restriction? Do your results change if you use the straight line corridor of the least cost path?
6. Re-estimate column (1) and (2) of table 5 for the sample of cities that never changed their area size over the sample period. Furthermore, make sure that you include the same observations for both regressions. Do the results change qualitatively?
7. What is the role of propensity score matching in this context? Replicate table 3. Briefly take a detour to discuss the logit model as example for non-linear models (since we have not seen them in the course so far) in comparison to the usual linear probability model. To do so, estimate the probability of receiving a railroad access based on the ex ante controls from table 3. Use the `margins` command to estimate the marginal effect of population growth at the means for both models and briefly discuss.
8. Use `psmatch2` to implement the propensity score matching with a kernel to replicate panel B of table 7. Make sure that you impose the right constraints. What is the role of the common support? Use `psgraph` to plot it.
9. Compute the weighted averages for panel B of table 7 and check how to compute the differences. What is the relation between matching and unobserved heterogeneity?

Further comments: Submit 3 files (.do, .log, .pdf(with graphs)) to Ilias. Usual procedures apply.