List of Homework for

The Economics of European Regions: Theory, Empirics, and Policy

Deadline: February 22, 2020

Davide Fiaschi[1] Angela Parenti[2]

January 21, 2020

[1]Dipartimento di Economia e Management, University of Pisa, Via Ridolfi 10, Pisa. *E-mail address:* davide.fiaschi@unipi.it.

[2]Dipartimento di Economia e Management, University of Pisa, Via Ridolfi 10, Pisa. *E-mail address:* angela.parenti@unipi.it.

The couples of student randomly matched and number of assignment for each couple are:

1. Zsófia Hegedüs

2. Gonul Aksan

3. Dal Torrione and Derayati

4. Bini and Lo Conte

The economics of EU regions

We expected a paper of **about 30 pages** and **a zip file with R codes** used to produce pictures, tables and estimates reported in the paper.

1. The dynamics of GDP per capita in EU regions in the period 1992:2006

- Download data at https://eer.ec.unipi.it/teaching-materials/ following the link to "Link to Dropbox directory with datasets".
- Build a panel of NUTS-2 regions for the period 1992:2006 taking any variables useful for explaining the growth of **GDP per capita** according to the Solovian model of growth.
- Estimate the **distribution dynamics in 1992:2006** (annual density, Markov matrix in discrete time and stochastic kernel). Discuss the results in terms of absolute, conditional and σ-convergence.
- Discuss the possible presence of **spatial dependence** on the base of Moran I and LISA using an appropriate geographical spatial matrix (based on contiguity or k-nearest neighbour). Discuss the results in light of theoretical spatial models.
- Estimate a **growth equation for GDP per capita** of EU regions using parametric and semiparametric methods by checking for the presence of endogeneity. Discuss the results in light of theoretical models.
- Discuss the **policy implications** of analysis.

2. The dynamics of GDP per capita in EU regions in the period 2006:2015

- Download data at https://eer.ec.unipi.it/teaching-materials/ following the link to "Link to Dropbox directory with datasets".
- Build a panel of NUTS 2 regions for the period 2006:2015 taking any variables useful for explaining the growth of **GDP per capita** according to the Solovian model of growth.
- Estimate the **distribution dynamics in 2006:2015** (annual density, Markov matrix in discrete time and stochastic kernel). Discuss the results in terms of absolute, conditional and σ-convergence.
- Discuss the possible presence of **spatial dependence** on the base of Moran I and LISA using an appropriate geographical spatial matrix (based on contiguity or k-nearest neighbour). Discuss the results in light of theoretical spatial models.
- Estimate a **growth equation for GDP per capita** of EU regions using parametric and semiparametric methods by checking for the presence of endogeneity. Discuss the results in light of theoretical models.
- Discuss the **policy implications** of analysis.

3. Commuting in EU regions

- Download data at https://eer.ec.unipi.it/teaching-materials/ following the link to "Link to Dropbox directory with datasets" and at Eurostat website: `https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_r_lfe2ecomm&lang=en`
- Build a panel of NUTS 2 regions taking any variables useful for explaining commuting in EU regions. Gap in the availability of data should drive the choice of the length of the period of analysis and the size of the sample, together with the possibility to interpolate data.
- Estimate the **distribution dynamics** (density, Markov matrix in discrete time and stochastic kernel). Discuss the results.
- Discuss the possible presence of **spatial dependence** in commuting on the base of Moran I and LISA using an appropriate geographical spatial matrix (based on contiguity or k-nearest neighbour). Discuss the results in light of theoretical spatial models.
- Estimate the **determinants of commuting** in EU regions using parametric and semiparametric methods and checking for the presence of endogeneity. Discuss the results in light of theoretical models.
- Discuss the **policy implications** of analysis.

4. Labour costs in EU regions

- Download data at https://eer.ec.unipi.it/teaching-materials/ following the link to "Link to Dropbox directory with datasets" and at Eurostat website: `https://ec.europa.eu/eurostat/data/database`
- Build a panel of NUTS 2 regions from 2004 to 2012 taking any variables useful for explaining labour costs in EU regions, also at sectoral level.
- Estimate the **distribution dynamics** (density, Markov matrix in discrete time and stochastic kernel), also at sectoral level. Discuss the results.
- Discuss the possible presence of **spatial dependence** in labour costs on the base of Moran I and LISA using an appropriate geographical spatial matrix (based on contiguity or k-nearest neighbour). Discuss the results in light of theoretical spatial models.
- Estimate the **determinants of labour costs** (also at sectoral level) in EU regions using parametric and semiparametric methods and checking for the presence of endogeneity. Discuss the results in light of theoretical models.
- Discuss the **policy implications** of analysis.

Exam: Causality

We expected a paper of **about 10 pages** and **a zip file with R codes** used to produce pictures, tables and estimates reported in the paper.

1. Assessing the impact of Job training provided under the Job Training Partnership Act

The U.S. Department of Labor began planning for an experimental evaluation of the training provided under the Job Training Partnership Act in 1986. The National JTPA study is the largest randomized training evaluation ever undertaken in the United States, collecting data on roughly 20,000 persons in 16 different sites around the country.

Between November 1987 and September 1989, eligible persons who applied for JTPA services were assigned to one of three different service groups– classroom training, on-the-job training and some combination of the two. (Those with the weakest basic skills were assigned to classroom training.) During the application and assessment process, staff members explained to applicants that not all of them would be served and that a lottery would be conducted to determine who would participate. Therefore, it was only after assessment and assignment to service strategy group, that staff members telephoned a random assignment clerk at the central office to determine which applicants would be eligible to receive the treatment. One-third were assigned to a control group and prevented from receiving JTPA services.

Do the analysis only for MEN

- Download the folder "Bloom&Al_1997.zip" from the website of the course. Carefully read the paper "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study", Bloom et al. (1997).
- The assignment to treatment and control group were truly random. Discuss with the use of simple equations how in this setting it is possible to define and estimate the average treatment effect of treated.
- Test whether you see any differences between treatment and control in their baseline characteristics, such as race, previous work experience, marital status, age, etc.
- What is your estimate of the impact of the JTPA program on workers' earnings during 30 months after random assignment?
- Generate an estimate of the program effect, after controlling for race, ethnicity, marital status, age and previous earnings using regression analysis. Does the estimated program effect change?

2. Analysis of a natural experiment: same-sex siblings, fertility and women's labor supply

Suppose you are interested in estimating the effect of fertility on labor supply. In particular, you want to know how much a woman's labor supply falls when she has an additional kid. You have US census data from 1980 for a random sample of 30,000 married women aged 21-35 with two or more kids.

- Download the folder "Angrist&Evans_1998.zip" from the website of the course. Carefully read the paper "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", Angrist and Evans (1998).

- The weeks worked is the outcome variable (Y), while whether the woman has more than 2 kids is the treatment (D). Assume a constant treatment effect, $A1 : \beta_{Di} = \beta_D$, and run the OLS regression of Y on D, W_1, W_2, W_3, W_4, W_5, W_6. Interpret the estimated coefficient on D.

- U are all the other factors, besides D and the W_s above, that also affect Y. $E(U|D, W, ...W) = E(U|W, ...W)$ and $E(U|D, W, ...W) = 0$ are two alternative precise mathematical statements of the identifying assumption for the OLS estimate on D to represent the causal effect of fertility on labor supply. Discuss the plausibility of these assumption in this application.

- Use Z as an instrument for D to estimate β_D. Interpret your estimates.

- Assess empirically whether *samesex* is a weak instrument.

3. Difference-in-difference in non-experimental settings

LaLonde (1986) shows how non-experimental estimates generated through standard econometric analysis compare to the experimental ideal.

- Carefully read the paper of Lalonde (1986),"Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 1986, vol. 76, issue 4, 604-20,;
- Download the data "LaLonde.RData" which includes the NSW sample, as well as two non-experimental samples: one based on the Current Population Survey (CPS) and one on the Michigan Panel of Income Dynamics (PSID).
- Data from 1974 and 1975 are pre-treatment, and data from 1978 are post-treatment. The first part of the question uses only experimental data. Also, use data from 1978 only.
- Investigate whether the data is consistent with randomization of the treatment.
- Estimate the effect using the experimental sample.
- Now move to observational data.Use pre-treatment data, too. Now use the sample consisting in the treated from the NSW sample and the comparison individuals from the CPS sample.
- Now you can construct a difference-in-differences estimate. How do these results compare to the experimental results?
- Get the means of each variable. Test if these means differ between treatment and control group.
- What is the assumption that must be made to interpret the difference in difference estimate as a causal effect? Why is this important?
- Compare the difference in the pre-training incomes by constructing a difference between $re74$ and $re75$. Do the same comparison of means that you did for the difference-in-difference estimation, what do you find?

4. Assessing the effect of class size on test scores

The twelfth century rabbinic scholar Maimonides proposed a maximum class size of 40. This same maximum induces a nonlinear and nonmonotonic relationship between grade enrollment and class size in Israeli public schools today. Maimonides' rule of 40 is used here to construct instrumental variables estimates of effects of class size on test scores. The resulting identification strategy can be viewed as an application of Donald Campbell's regression-discontinuity design to the class-size question. The estimates show that reducing class size induces a significant and substantial increase in test scores for fourth and fifth graders, although not for third graders.

- Carefully read the paper of Joshua D. Angrist, Victor Lavy. (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *The Quarterly Journal of Economics*, 114(2), 533-575;
- Download the data "AngristLavyData.RData" (to understand what variables in the data file mean you need to look at tables in the paper).
- Estimate the effect of class size on math scores using OLS without any controls, and then by adding the percentage of disadvantaged students in the class and enrollment as controls. Interpret your results.
- Create a subsample of schools with enrollment between 20 and 60 students. Generate a (predicted) large class dummy based on the first discontinuity at 40 students.
- Use OLS to estimate the effect of being in a large class on math scores assuming that you have a sharp RDD around this discontinuity. Control for the percentage of disadvantaged students in the class and a linear trend in enrollment.
- Use Local Linear Regression to get a point estimate of the effect of being in a large class on math scores assuming you have a sharp RDD. Finally, use a nonparametric bootstrap to estimate the standard error on your RDD point estimate. Compare these results to the estimates you obtained with OLS
- Estimate the effect of class size on math scores using fuzzy RDD. Control for the percentage of disadvantaged students in the class and a linear trend in enrollment.
- If the RDD is valid, then the coefficient of interest should not change significantly whether you include or exclude covariates. Check whether this is the case.
- Explore the validity of the design and the robustness of the results:
  - Manipulation: Plot the distribution of the assignment variable

- Misspecification: Present a graph using binned local averages of class size and math score against enrollment. Use bins of width 20 and make sure that the bins do not cover the discontinuities. Can you see the discontinuity in class size and math scores?
- Placebo check: Conduct the RD analysis where your outcome is percentage disadvantaged pupils.