

Pol Sci 630: Problem Set 13 - Autocorrelation, Fixed, Random, Mixed Effects, and Causal Inference Techniques

Prepared by: Jan Vogler (jan.vogler@duke.edu)

Due Date: Tuesday, December 1st, 2015, 10 AM

It is absolutely essential that you show all your work, including intermediary steps, and comment on your R code to earn full credit. Showing all steps and commenting on code them will also be required in future problem sets.

Please use a **single** PDF file created through knitr to submit your answers. knitr allows you to combine R code and L^AT_EX code in one document, meaning that you can include both the answers to R programming and math problems. Also submit the source code that generates the PDF file (i.e. either .Rnw or .Rmd files)

Make sure that the PDF files you submit do not include any references to your identity. The grading will happen anonymously. You can submit your answer at the following website: <http://ps630-f15.herokuapp.com/>

R Programming

Problem 1: Autocorrelation (4 points)

Do the following in R:

Load the *LDC_IO_replication* data.

a) Estimate a linear regression with the following characteristics and show the results:

1. Dependent variable: net changes in FDI as percentage of GNP

2. Independent variables: tariff Level, GDP per capita, Polity score
3. Please use lagged versions of the independent variables (lagged by 1 year). This includes a lagged version of tariff level that is not included in the original dataset.
4. Please use country fixed effects.

b) Create three lags of your dependent variable and estimate a new regression that includes those lags as independent variables. Compare the results of your two regressions.

c) Now, in addition to the three lags, add a variable to capture a time trend to your model and interpret the results. Considering the results, is there a time trend?

d) Finally, for your last model, including both three lags of your dependent variable and the time trend, do you find any autocorrelation between the residuals at time t and time $t-1$.

Problem 2: Fixed-, Random-, and Mixed-Effects Models (4 points)

Do the following in R:

a) Estimate a **random** effects model with the following characteristics. (Note: random effects here refers to the intercept only)

1. Dependent variable: net changes in FDI as percentage of GNP
2. Independent variables: tariff Level, GDP per capita, Polity score
3. Please use lagged versions of the independent variables (lagged by 1 year).
4. Please assume that the intercept is randomly distributed (but not any of the coefficients).

b) Interpret the results of the regression you estimated.

c) Show the density distribution of the intercept graphically.

Estimate a **mixed** effects model with the following characteristics. (Note: Mixed effects refers to the inclusion of some coefficients as randomly distributed — for details see below).

1. Dependent variable: net changes in FDI as percentage of GNP
 2. Independent variables: tariff Level, GDP per capita, Polity score
 3. Please use lagged versions of the independent variables (lagged by 1 year).
 4. Assume that the intercept and the coefficient of tariff level are randomly distributed across units.
 5. Assume that the coefficients of the other independent variables is fixed across units.
- d) Interpret the results of the regression you estimated.
- e) Show the distribution of the coefficient of tariff level graphically.

Problem 3: Regression Discontinuity Designs (4 points)

Do the following in R:

a) Load the *SocialSecurity* data that is available on the course website and display the summary statistics.

This artificial dataset contains observations of two variables, namely the vote share that the “liberal party” has received and expenditures for social security.¹ The liberal party is in a two-party system (with the other party being the “conservative party”) in which gaining 50% of the vote share means election victory. The goal of the analysis is to estimate the effect of liberal-party election victories on social security expenditures via RDD.

b) Turn the data frame into a regression discontinuity object and display the average values of social security expenditures for different bins of liberal party vote shares.

¹Assume that each data point represents one election period in one state.

- c) Estimate two parametric regressions of the effects of liberal-party election victory on social security expenditures. One regression should be a regular OLS expression without higher-order polynomials. The second regression should include polynomials of up to the 3rd order. Choose a bandwidth of 15. Interpret the results of your regressions — is there a discontinuity in the data at a vote share of 50%?
- d) Estimate a nonparametric regression of the effects of liberal-party election victory on social security expenditures. Use the IK procedure introduced in the tutorial to estimate the optimal bandwidth.
- e) For the non-parametric regression, show a regression sensitivity plot (for bandwidths of 5 to 20) and interpret it.
- f) Conduct a regression sensitivity test and interpret the results.
- g) Conduct a McCrary density test and interpret the results.

Statistical Theory: Differences-in-differences

Problem 4 (4 points)

Answer the following questions:

- a) In the tutorial it was shown that a simple differences-in-differences estimate using four groups is the following:

$$\delta = (\bar{x}_{t2} - \bar{y}_{t2}) - (\bar{x}_{t1} - \bar{y}_{t1})$$

Where x is the treatment group and y is the control group and $t1$ and $t2$ indicate time points 1 and 2 respectively.

Assuming that the parallel trends assumption is valid, how does the above calculation allow us to make a causal inference about the effect of the treatment of x ?

- b) What happens if the parallel trends assumption is violated?

c) Describe one way how the parallel trends assumption can be relaxed. What needs to be true to relax it?