Crime Diff-in-Diff Analysis

Paul Laskowski

Introduction

The file crime3.Rdata contains a two-year panel of crime data taken from E. Eide (1994) *Economics of Crime: Deterence of the Rational Offender.* Amsterdam: North Holland. The data are for 53 police districts in Norway and include a variable for the clear-up rate, or fraction of recorded crimes solved by the police. The clear-up rate can be interpreted as measuring the probability that a criminal is captured, which is different than the severity of punishment if the capture takes place.

You will use a difference-in-difference analysis to estimate the causal effect that the clear-up rate has on the amount of crime.

The differenced model you will estimate is:

$$clc\hat{r}ime = \beta_0 + \beta_1 cclr\hat{p}rc1 + \beta_2 cclr\hat{p}rc2$$

Here, the dependent variable is the change in the log of number of crimes, while the predictors are change in the clear-up rate, lagged by 1 year and by 2 years.

Question 1

Write down the corresponding non-differenced population model for the log of the number of crimes, lcrime.

a) What assumption is needed for this model to be causal?

b) Explain why a non-differenced regression is unlikely to produce consistent estimates of the model parameters.

Question 2

Do you expect OLS regression to consistently estimate the parameters of the differenced model? Provide your best arguments for one side or the other.

Question 3

Carry out the regression above, and interpret your coefficients.

What further data would you want to have in order to form a more convincing analysis?

More Analysis of the Crime Data

The first part of this activity is a prequel to the previous activity, and the second part is based on the model considered in the previous activity.

1. EDA: make plots of the following combinations of variables (having a linear fit may be helpful so consider using *ggplot2*). For each, comment on the linearity of the relationships and the strength of the linear relationships.

   (a) crime and clrprc1 (and clrprc2)

   (b) lcrime and clrprc1 (and clrprc2)

   (c) clcrime and cclrprc1 (and cclrprc1)

   Suppose you fit a model that uses clrprc1 and clrprc2 to predict lcrime. Even if this model fits the data better, what concerns might you have?

2. Consider the differenced model in the previous activity

   (a) Assess the conditions/assumptions and look for unusual observations. Use the function *vif* to get variance inflation factors and comment on collinearity.

   (b) Is there evidence cclrprc1 and cclrprc1 are jointly significant?

   (c) Is there evidence cclrprc1 and cclrprc1 are individually significant? Note that you can use the function *confint* to get confidence intervals for the model coefficients.

   (d) Is there evidence the coefficient for cclrprc2 is different from -0.10?

   (e) Is there evidence the effects of cclrprc1 and cclrprc2 are the same?