

# Panel Data Analysis

## Homework 1

Deadline: January 22<sup>nd</sup>, 2019 at 14:00am

1. A researcher is using a panel data set on  $n = 1000$  workers over  $T = 10$  years (from 2001 to 2010) that contains the workers' earnings, gender, education, and age. Assume that the education variable is changing over time for at least some individuals in the sample (for example, because some individuals return to school). The researcher is interested in the effect of education on earnings. Give some examples of unobserved person-specific variable that are correlated with both education and earnings. Can you think of examples of time-specific variables that might be correlated with education and earnings? How would you control for these person-specific and time-specific effects in a panel data regression?
2. Some U.S. states have enacted laws that allow citizens to carry concealed weapons. These laws are known as "shall-issue" laws because they instruct local authorities to issue a concealed weapons permit to all applicants who are citizens, are mentally competent, and have not been convicted of a felony (some states have some additional restrictions). Proponents argue that, if more people carry concealed weapons, crime will decline because criminals are deterred from attacking other people. Opponents argue that crime will increase because of accidental or spontaneous use of the weapon. In this exercise, you will analyze the effect of concealed weapons laws on violent crimes. You will find a data file, Guns.dta, that contains a balanced panel of data from 50 U.S. states, plus the District of Columbia, for the years 1977-1999.
  - a. What is the average violent crime rate across all the observations in our data. For which state id and year do we see the highest rate of violent crime (vio) in our data?
  - b. Generate the average violent crime rate per year. Which year shows the highest average violent crime rate in the data? (Draw a graph of "average violent crime rate per year" versus "year" save the graph).
  - c. Estimate (1) a regression of  $\log(\text{vio})$  against shall and (2) a regression of  $\log(\text{vio})$  against shall, incarc\_rate, density, avginc, pop, pb1064, pw1064, and pm1029.
    - i . Interpret the coefficient on shall in regression (1). Is this estimate large or small in a "real-world" sense ?

- ii. Does adding the control variables in regression (2) change the estimated effect of a shall-carry law in regression (1) as measured by the statistical significance? As measured by the “real-word” significance of the estimated coefficient? Which regression is more credible and why?
  - iii. Suggest a variable that varies across states but plausibly varies little or not at all over time, and that could cause omitted variable bias in regression(2).
- d. Do the results change when you add fixed state effects? If so, which set of regression results is more credible and why? Are the state effects jointly statistically significant?
- e. You notice from part (b) that there are some time trends in the violent crime variable, i.e. the violent crime rate seems to increase for the first periods in the sample and then decrease for the last periods in the sample. Given this information, add some relevant variables/controls to your specification on part d.