ECON-UB 251

Assignment 3, Fall 2022

The learning goals of this assignments are:

1. familiarize with the application of panel data models to economic problems
2. familiarize with the application of binary dependent variable (classification) models
3. developing data wrangling skills

I *strongly* prefer that you complete the assignment in Rmarkdown and a sample template is provided in Brightspace. You can knit the document to Word or PDF (or export to PDF the Word document). Set the `echo` option to `TRUE` so that I can see the code you are using to conduct the analysis. You can discuss the assignment with other students, but each student should submit his/her original work.

Submit in `Brightspace` by 2pm on **Thursday November 17th, 2022**.

1. **Panel Data**

The right to carry a gun is a debated issue in the United States. Supporters of gun rights argue that more guns are likely to deter crime, while opponents believe that it might cause more crime given their wide availability. To investigate this issue we use a panel dataset of 51 US states (including DC) in the period 1977-2014 that includes the following variables:

Variable	Definition
stateid	numerical (integer) state ID (Alabama = 1, Alaska = 2, etc.)
year	4-digit year (1977, 1978,..., 2010)
state	state name (string) ("Alabama", etc.)
vio	violent crime rate (incidents per 100,000 members of the population)
bur	burglary crime rate (incidents per 100,000)
mur	murder rate (incidents per 100,000)
aga	aggravated assault rate (incidents per 100,000)
pap	rape rate (incidents per 100,000)
auto	motor vehicle crime rate (incidents per 100,000)
shall	fraction of year state has a shall-issue law
density	population per square mile of land area, divided by 1000
rpcpi	real per capita personal income in the state
rpcui	real per capita unemployment insurance payments in the state
rpcim	real per capita government income support ("welfare") payments in the state
pbmAABB	percent of state population that is black male, ages AA to BB
pwmAABB	percent of state population that is white male, ages AA to BB

The dataset includes several measures of violence (`vio`,`bur`, `mur`, `auto`) while the variable of interest is `shall` which represents the fraction of the year that a state had a *shall-issue* law in place (that is, the law requires a license to carry a concealed handgun subject to specific criteria determined in the law).

1.1 [**10%**] Save the data file that is available in *Brightspace*, load it in your rmarkdown document, and answer the following questions:

- Use `ggplot2` to make a scatter plot of `density` and `vio` with the color of the dots varying by state (add also `+ theme(legend.position = "none")` to eliminate the legend). Discuss the results and identify the state that stands out relative to the rest.
- Do the same scatter plot for `log(density)` vs `log(vio)` and discuss whether it could be useful to transform the variables

1.2 [10%] Plot the relationship between `shall` and `vio` and discuss the results

1.3 [10%] Use the package `plm` (installation needed) to define the dataset as a `plm` panel dataset. Estimate the *pooled regression model*

$$\log(vio_{it}) = \alpha + \beta_1 shall_{it} + u_{it}$$

- Interpret the coefficient estimate of `shall` and discuss its significance. Which of the views discussed earlier does the estimate seem to support?

1.4 [10%] Estimate the panel data model with state fixed effects, that is,

$$\log(vio_{it}) = \alpha_i + \beta_1 shall_{it} + u_{it}$$

- Discuss the estimate of β_1 and its significance
- Did the coefficient estimate of `shall` change significantly relatively to the pooled estimate? why or why not?

1.5 [10%] Estimate the panel data model with state *and* time fixed effects, that is,

$$\log(vio_{it}) = \alpha_i + \lambda_t + \beta_1 shall_{it} + u_{it}$$

- Discuss the estimate of β_1 and its significance
- Did the coefficient estimate of `shall` change significantly relatively to the pooled and state fixed effect estimate? why or why not?

1.6 [10%] Add the `rpcpi + rpcui + rpcim + density + pbm1019 + pbm2029 + pwm1019 + pwm2029` regressors to the state and state/time fixed effect models.

- Discuss the change in the estimate of β_1 and its significance relative to the model with only `shall`
- We estimated 5 models (pooled, FE with/without controls, TFE with/without controls): what is your conclusion about the relationship between `shall-issue` laws and crime?

2. **Binary Dependent Variables**

Why are some Americans unemployed? In this exercise we consider a sample of Americans that were working in April 2008 and were later interviewed again in April 2009 about their employment status (employed, unemployed, or out of the labor force). This period corresponds to part of the Great Recession that represents a particularly severe contraction in many parts of the US economy. In addition to the employment status, we observe demographic information about the individuals (such as `race`, `married`, employment sector, education level, and `female`). See the variable description below:

Variable Name	Description
Variables from the 2009 Survey	
employed	indicator =1 if employed in 2009
unemployed	indicator =1 if unemployed in 2009
Variables from the 2008 Survey	
age	age
female	indicator =1 if female
married	indicator =1 if
race	= 1 if white (only) = 2 if black (only) = 3 not white (only) or black (only)

Variable Name	Description
union	indicator =1 if a member of a union
ne_states	indicator =1 if from a northeastern state
so_states	indicator =1 if from a southern state
ce_states	indicator =1 if from a central state
we_states	indicator =1 if from a western state
private	indicator =1 if employed in a private firm
government	indicator =1 if employed by the government
self	indicator =1 if self-employed
educ_lths	indicator =1 if highest level of education is less than a high school graduate
educ_hs	indicator =1 if highest level of education is a high school graduate
educ_somecol	indicator =1 if highest level of education is some college
educ_aa	indicator =1 if highest level of education is AA degree
educ_ba	indicator =1 if highest level of education is BA or BS degree
educ_adv	indicator =1 if highest level of education is advanced degree
earnwke	average weekly earnings

2.1 [10%] Produce a Table with the percentage of each employment status in April 2009 for the 5,412 individuals that were employed in April 2008. Discuss the results.

- NB: there are 435 individuals that appear neither employed or unemployed in 2009 (that is, both variables are equal to zero); filter these observations out in this and the following questions

2.2 [10%] Regress the `Employed` variable on `age` and the square of `age` using the Linear Probability Model (LPM)

- Is age a determinant factor of the probability to be employed in 2009?
- Is there evidence of a nonlinear effect of age of the probability to employment?
- Use the estimated coefficients of `age` and its square to plot the effect on the probability to be employed

2.3 [10%] Estimate the following logit models:

1. `employed` on `age` and its square
2. Add the following demographics: `earnwke`, `race`, `married`, `female`, `ne_states`, `so_states`, `ce_states`, `we_states`, `educ_lths`, `educ_hs`, `educ_somecol`, `educ_aa`, `educ_bac`, `educ_adv`

- Estimate the models and discuss the characteristics of workers that were most affected by the Great Recession

2.4 [10%] Model the probability of being unemployed (instead of being employed as in the previous question); estimate the logit model with the demographic variables for the `unemployed` variable and discuss the characteristics of workers that significantly affected the probability of unemployment