# Panel Data

Antonio Pacifico
Federica Romei

May 8, 2012

1. Guns is a balanced panel of data on 50 US states, plus the District of Columbia (for a total of 51 states), by year for 1977 to 1999. A data frame containing 1,173 observations on 13 variables:

   - *state*: factor indicating state.
   - *year*: factor indicating year.
   - *violent*: violent crime rate (incidents per 100,000 members of the population).
   - *murder*: murder rate (incidents per 100,000).
   - *robbery*: robbery rate (incidents per 100,000).
   - *prisoners*: incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year).
   - $afam$: percent of state population that is African-American, ages 10 to 64.
   - *cauc*: percent of state population that is Caucasian, ages 10 to 64.
   - *male*: percent of state population that is male, ages 10 to 29.
   - *population*: state population, in millions of people.
   - *income*: real per capita personal income in the state (US dollars).
   - *density*: population per square mile of land area, divided by 1,000.
   - *shall*: Does the state have a shall carry law in effect in that year?

Some U.S. states have enacted lows 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 , mentally competent, and have not been convicted of a felony ( some state have 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.

Estimate (1) a regression of $ln(violent)$ against $shall$ and a (2) regression of $ln(violent)$ against $shall$, $prisoners$, $density$, $income$, $population$, $afam$, $cauc$ and $male$.

(a) Interpret the coefficient on shall in regression (1) and (2). Is this estimate large or small in "real-world" sense?

Does adding the control variables in regression (2) changes the estimated effect of a shall-carry law in regression (1), as measured by statistical significance? As measured by "real-world" significance of the estimated coefficient?

**Solution:** If you regress $ln(violent)$ on $shall$ you will have this output:

```
 Linear regression                                      Number of obs =    1173
                                                        F(  1,  1171) =   86.86
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.0866
                                                        Root MSE      =  .61735

------------------------------------------------------------------------------
             |               Robust
    lviolent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.4429646   .0475283    -9.32   0.000    -.5362148   -.3497144
       _cons |   6.134919   .0193039   317.81   0.000     6.097045    6.172793
------------------------------------------------------------------------------
```

It seems that State who enforces Shall Law have 44% violent crime less than the State who doesn't enforce this law. Then according to this regression it should be good enforce the law allowing the citizens to carry guns. If we run second regression we will have this output:

```
Linear regression                                       Number of obs =    1173
                                                        F(  8,  1164) =   95.67
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.5643
                                                        Root MSE      =  .42769

------------------------------------------------------------------------------
             |               Robust
    lviolent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.3683869   .0347879   -10.59   0.000     -.436641   -.3001329
   prisoners |   .0016126   .0001807     8.92   0.000     .0012581    .0019672
     density |   .0266885   .0143494     1.86   0.063    -.0014651     .054842
      income |   .0012051   .0072778     0.17   0.869     -.013074    .0154842
  population |   .0427098   .0031466    13.57   0.000     .0365361    .0488836
        afam |   .0808526   .0199924     4.04   0.000     .0416274    .1200778
        cauc |   .0312005   .0097271     3.21   0.001      .012116    .0502851
        male |   .0088709   .0120604     0.74   0.462    -.0147917    .0325334
       _cons |   2.981738   .6090198     4.90   0.000     1.786839    4.176638
------------------------------------------------------------------------------
```

Again the enforcement of the law seems to decrease a lot the percentage of violent crime. Indeed the States that enforced "Shall law" seems to have 36% less violent crime than State that doesn' t enforce.
The effect seems to be huge from a real world perspective and seems to encourage the enforcement of Shall law.

(b) Suggest a variable that varies across states but plausibility varies little- or not at all- over time, and could cause omitted variable bias in the regression (2)

**Solution:** One possible problem could be the presence of State specific characteristic that influences both the adoption of "Shall law" and the rate of violent crime. There could be some States governed by Conservative Party that are very concern about Security. This could increase the probability to adopt Shall law and to fight violent crime but this doesn't mean that Shall law is good to decrease violent crime.

(c) Do the results change when you add fixed state effect? If so, which set of regression results is more credible, and why?

**Solution:** In order to add fixed state effect you have to write

```
areg lviolent shall prisoners density income population afam cauc male,
                                absorb (state)
```

```
Linear regression, absorbing indicators                Number of obs =    1173
                                                       F(  8,  1114) =   38.77
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9411
                                                       Adj R-squared =  0.9380
                                                       Root MSE      =  .16072

------------------------------------------------------------------------------
    lviolent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.0461415   .0188668    -2.45   0.015      -.08316    -.009123
   prisoners |   -.000071   .0000936    -0.76   0.448    -.0002547    .0001126
     density |  -.1722901   .0850362    -2.03   0.043    -.3391392   -.0054409
      income |  -.0092037   .0059083    -1.56   0.120    -.0207963    .0023889
  population |   .0115247   .0087239     1.32   0.187    -.0055924    .0286417
        afam |   .1042804   .0177564     5.87   0.000     .0694407    .1391201
        cauc |   .0408611   .0050745     8.05   0.000     .0309044    .0508177
        male |  -.0502725   .0064037    -7.85   0.000    -.0628373   -.0377078
       _cons |   3.866017   .3847716    10.05   0.000     3.111058    4.620975
-------------+----------------------------------------------------------------
       state |       F(50, 1114) =    142.570   0.000          (51 categories)
```

In this case we are running a regression with $n-1$ dummy variable, where $n$ is the number of the States. We try to clean our regression from the characteristic of the States. As you can see the coefficient of *shall* drop dramatically. Indeed States that enforces "Shall law" have 4% violent crime less than States that don't enforce. Moreover the coefficient is not more significant at 1% level.
The regression that is more significant is this one because, as you can see, Adjusted $R^2$ and $R^2$ are incredibly high.

(d) Do the results change when you add fixed time effects? If so, which set of regression is more credible and why?

**Solution:** Another problem can arise from specific time features. Imagine that in a specific year crime drop and many States adopt Shall law. It can happen that in specific year more citizens are concern with Security and decide to vote at National Level for Conservative Party and push for more severe law against crime.
To take in consideration this effect we have to add time fixed effect building $T-1$

dummy variables, where $T$ is the number of the years in the dataset. To do this write on Stata:

```
tab year, generate (year_)
```

then

```
areg lviolent shall prisoners density income population afam cauc male
                        year_*, absorb (state)
```

and you will have this output (actually we put a shorter version):

```
Linear regression, absorbing indicators                Number of obs =    1173
                                                       F( 30,  1092) =   26.14
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9562
                                                       Adj R-squared =  0.9530
                                                       Root MSE      =  .14003

------------------------------------------------------------------------------
    lviolent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -.0279935   .0171578    -1.63   0.103    -.0616596    .0056725
   prisoners |    .000076   .0000903     0.84   0.400    -.0001012    .0002531
     density |   -.091555   .0762821    -1.20   0.230    -.2412312    .0581212
      income |   .0009587   .0064349     0.15   0.882    -.0116676    .0135849
  population |  -.0047544   .0078675    -0.60   0.546    -.0201916    .0106827
       afam |   .0291862    .022692     1.29   0.199    -.0153387    .0737111
       cauc |   .0092501   .0078617     1.18   0.240    -.0061756    .0246759
       male |   .0733254   .0156139     4.70   0.000     .0426887     .103962
------------------------------------------------------------------------------
```

The coefficient of Shall is not significantly different from zero at 10% significant level. This means that it seems to be not relation between the Shall law and the rate of violent crime. Moreover everything seems to be explained by State and Time specific factor.

(e) Repeat the analysis using $ln(robbery)$ and $ln(murder)$ in place of $ln(violent)$.

**Solution:** We write only the Stata code:

```
areg lrobbery shall prisoners density income population afam cauc male
                        year_*, absorb (state)
areg lmurder shall prisoners density income population afam cauc male
                        year_*, absorb (state)
```

(f) In your view, what are the most important remaining threats to the internal validity of

this regression analysis?

**Solution:** We can have a lot of problem. First of all we cannot take in consideration State-Time specific features that can affect our regression. Secondly, it could be that some State adopt the "Shall law" to fight the increase in violent crime, then the causality can go on the other way round.

(g) Based on your analysis, what are the conclusions would you draw about the effects of concealed-weapon law on these criminal rates?

**Solution:** Up to you. :D

2. It's extremely important to know what are the determinant of the wage. Assume that your boss give to you dataset wage2.dta. It contains 935 observation on the following variables:

   - $wage$: monthly earnings
   - $hours$: average weekly hours
   - $IQ$: IQ score
   - $KWW$: knowledge of world work score
   - $educ$: years of education
   - $exper$: years of work experience
   - $tenure$: years with current employer
   - $age$: age in years
   - $married$: =1 if married
   - $black$: =1 if black
   - $south$: =1 if live in south
   - $urban$: =1 if live in SMSA
   - $sibs$: number of siblings
   - $brthord$: birth order
   - $meduc$: mother's education
   - $feduc$: father's education
   - $lwage$: natural log of wage

   (a) Your boss is very lazy and doesn't want a regression with too many regressors. He asks to you to regress $log(wage)$ on max 6 variables BUT there should be an interaction term, variable to the square and a dummy. You should be able to comment properly your output. [1]

   **Solution:** There is not a right answer but remember that in this kind of work you have to:

   - Start from certain assumption. This means that you have to justify a priori the choice of certain variables respect to others. Don't come to the whiteboard saying "i pick this and this randomly". Right answer could be i decide to choose this variable because i think it's significant as determinant of wage for this reasons..
   - Do the regression properly.

[1] If we won't be able to correct in class we will correct next session.

- Understand the results and be able to comment both from a statistic point of view and from a logic point of view.
- Speak to the class not to me. You are explaining something to other people! Don't be shy!