Panel Data Analysis in Julia

Christian Kascha

January 12, 2017

Panel Data Analysis in Julia

Purpose of this Talk

- Present basic data analysis steps in Julia.
- Lower the entry costs for you.

The Data Set

- Data on 48 U.S. states for 7 years from 1982 to 1988
- Stock and Watson: *Introduction to Econometrics, 3rd International Edition, Companion Website*

Purpose of the Analysis

Background

- One fourth of fatal crashes in the US involve a driver who was drinking.

Interesting variables

fatality rate number of annual traffic deaths per 10,000 people

beer tax the tax on a case of beer in 1988 dollars

Question

How does the beer tax affect traffic fatalities?

Get the data

Read in the Data

```
using DataFrames
df = readtable("sw-datasets/csv/fatality_mod.csv";
       separator=';')
df[:mrall] *= 10000
```

Let's check that we have done the right thing

```
summary(df)
```

```
"336×43 DataFrames.DataFrame"
```

Checking the data

Get the first six lines

```
head(df[:,1:4])
```

```
6×4 DataFrames.DataFrame
 Row  state  year  spircons  unrate

 1    1      1982  1.37      14.4
 2    1      1983  1.36      13.7
 3    1      1984  1.32      11.1
 4    1      1985  1.28      8.9
 5    1      1986  1.23      9.8
 6    1      1987  1.18      7.8
```

Checking the data

Get some descriptive stats

```
describe(df[:beertax])
```

```
Summary Stats:
Mean:           0.513256
Minimum:        0.043311
1st Quartile:   0.208849
Median:         0.352589
3rd Quartile:   0.651573
Maximum:        2.720764
```

Checking the data

Get some descriptive stats

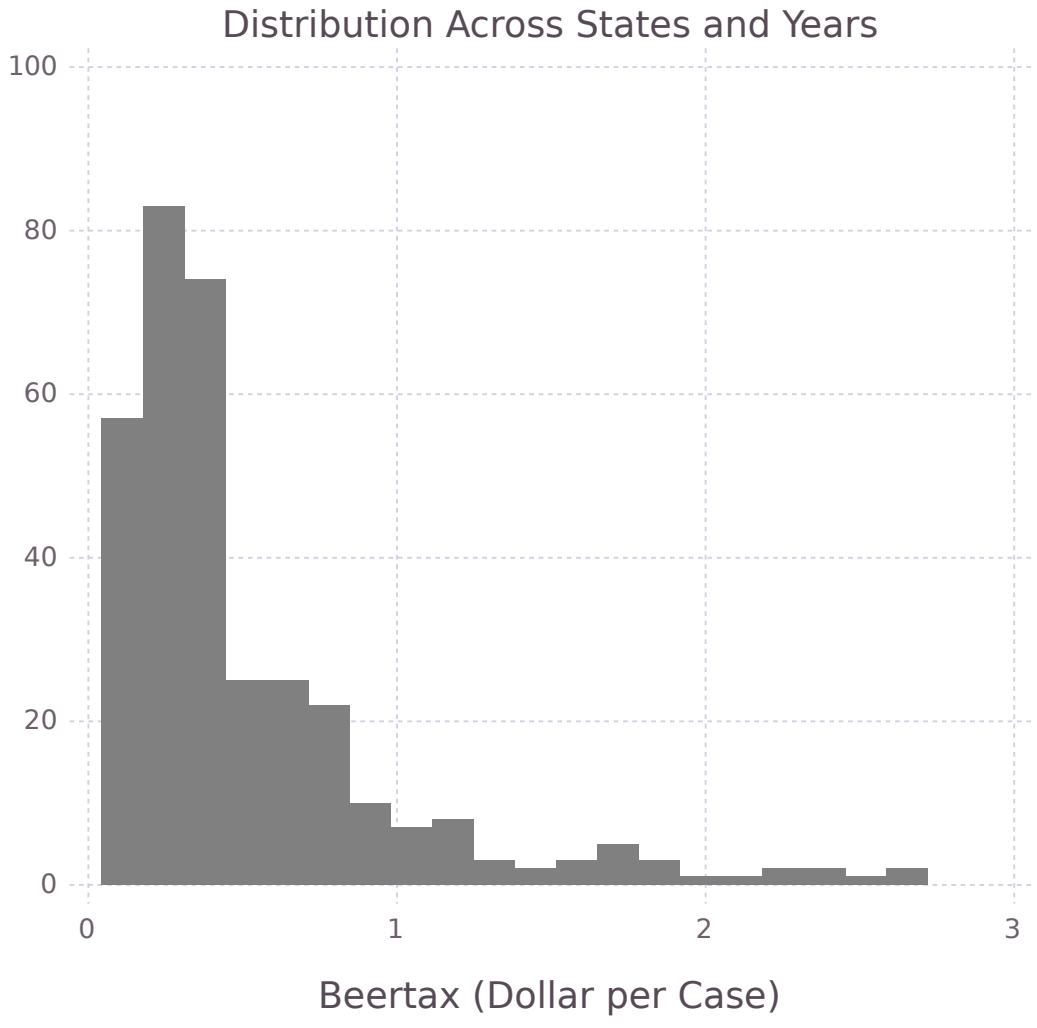
```
describe(df[:mrall])

Summary Stats:
Mean:           2.040444
Minimum:        0.821210
1st Quartile:   1.623710
Median:         1.955955
3rd Quartile:   2.417887
Maximum:        4.217840
```

Histogram of Beertax

Code

```
using Gadfly
myTheme = Theme(default_color=colorant"grey")

p = plot(df, x="beertax", Geom.histogram(bincount=20),
 Guide.title("Distribution Across States and Years"),
 Guide.xlabel("Beertax (Dollar per Case)"), myTheme);

img = PDF("beertax_hist.pdf", 12cm, 12cm)
draw(img, p)
```

Histogram of Beertax



Scatter Plot - All States and Years

Code

```
p = plot(df, y="mrall", x="beertax", Geom.point,
 Guide.title("Distribution Across States and Years"),
 Guide.xlabel("Beertax (Dollars per Case)"),
 Guide.ylabel("Fatalities (per 10'000)"), myTheme)

img = PDF("scatterplot.pdf", 12cm, 12cm)

draw(img, p)
```

Scatter Plot - All States and Years

Using 1982 Data for an Ordinary Regression

First try

```
df_1982 = df[(df[:year] .== 1982),:];

a, b = linreg(df_1982[:beertax], df_1982[:mrall])
```

```
(2.0103812626061535,0.14846034326952268)
```

Higher beer taxes lead to more fatalities???

By the way

Get the source code with `edit(funktionname)` like so: `edit(linreg)`.

What if . . . . ?

Some doubts

- Suppose that states that have a more problematic drinking culture are more likely to adopt a higher beer tax?
- Could we observe "drinking culture"?

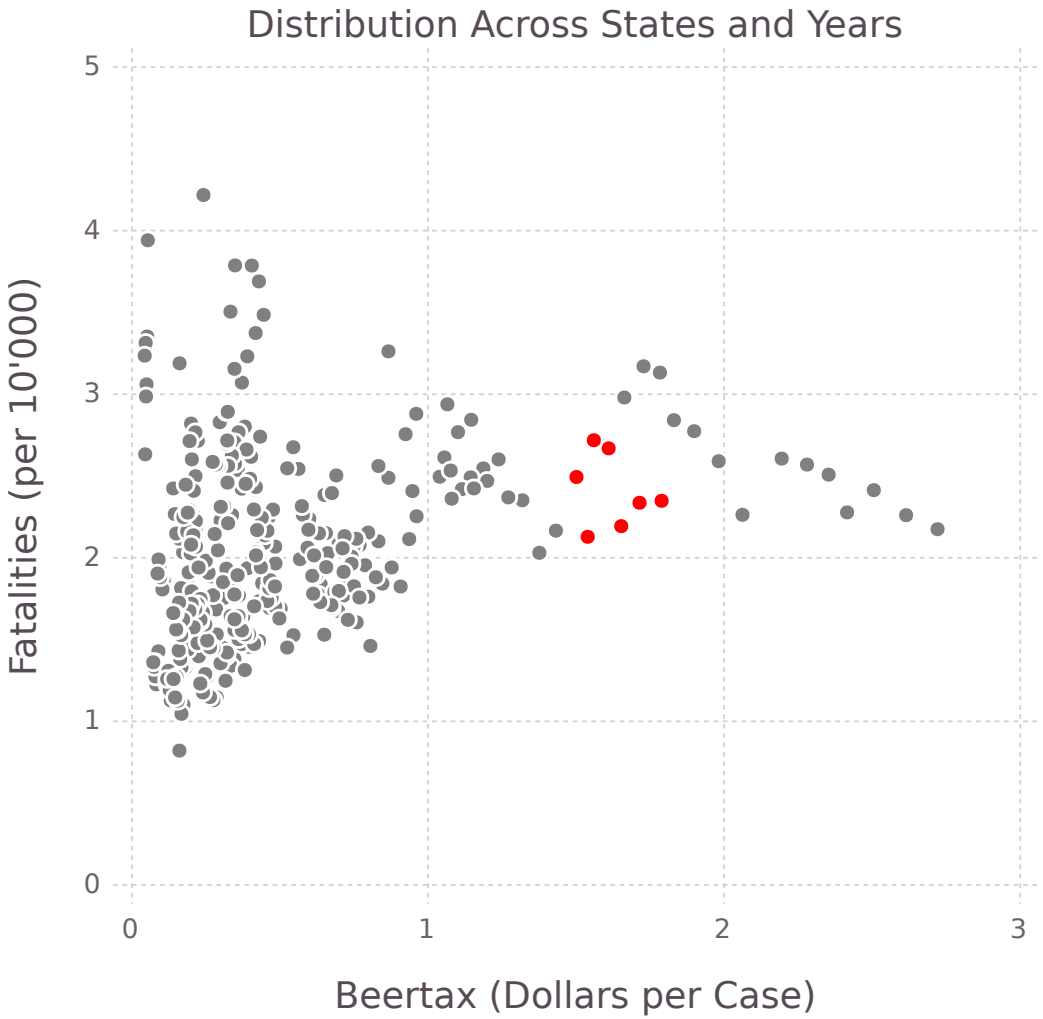
Omitted variable bias in Pictures - Code

Let's look at some single states here:

```
myTheme2 = Theme(default_color=colorant"red")

df_state1 = df[(df[:state] .== 1),:]
p = plot(layer(df, y="mrall", x="beertax", Geom.point,
       order=1, myTheme),
 layer(df_state1, y="mrall", x="beertax", Geom.point,
       order=2, myTheme2),
 Guide.title("Distribution Across States and Years"),
 Guide.xlabel("Beertax (Dollars per Case)"),
 Guide.ylabel("Fatalities (per 10'000)"))

img = PDF("scatterplot_state1.pdf", 12cm, 12cm)

draw(img, p)
```

Omitted variable bias in Pictures

Omitted variable bias in Pictures



Omitted variable bias in Pictures

Omitted variable bias in Pictures

Panel Data Analysis with Fixed Effects

- Panel data analysis allows us to take unobserved state differences that are constant over the years into account by analysing all states and years jointly.
- In Julia, the package that provides the necessary routines is called FixedEffectsModels.

Panel Data Analysis Syntax

Model

The econometric model looks like this

$$\text{fatalities}_{st} = \alpha_s + \beta \text{beertax}_{st} + u_{st}$$

Syntax

```
using FixedEffectModels

df[:stateFixedEffect] = pool(df[:state]);
panel_reg_1 = reg(mrall ~ beertax |> stateFixedEffect, df)
```

Panel Data Analysis Results

```
                          Fixed Effect Model
=============================================================
Number of obs:                336  Degrees of freedom:           49
R2:                         0.905  R2 within:                 0.041
F-Statistic:              12.1904  p-value:                   0.000
Iterations:                     1  Converged:                    true
=============================================================
          Estimate Std.Error  t value Pr(>|t|) Lower 95% Upper 95%
-------------------------------------------------------------
beertax  -0.655874   0.18785 -3.49148    0.001  -1.02561 -0.286135
=============================================================
```

Taking state - specific differences into account, we obtain an estimate with a negative sign.

Thank You!

Questions?