Spring 2020

Applied Econometrics and Time Series Analysis

Sushmitha Kallur Raju

Table of Contents

1.Introduction

The goal of this project is to study the factors influencing traffic deaths and analyze if the different policies issued are effective in reducing traffic deaths, in particular fatalities caused due to drunk drivers . The dataset has records of the 48 US states collected annually over the years 1982 to 1988 for the traffic fatality rate and some other demographics in each state.

Following are the variables with their definitions:

Variable	Descriptions
state	State ID (FIPS) Code
year	Year
spircons	Per Capita Pure Alcohol Consumption (Annual, Gallons)
unrate	State Unemployment Rate (%)
perinc	Per Capita Personal Income ($)
beertax	Tax on Case of Beer ($)
sobapt	% Southern Baptist
mormon	% Mormon
mlda	Minimum Legal Drinking Age (years)

dry	% Residing in Dry Counties A dry county is a county whose government forbids the sale of any kind of alcoholic beverages. Some prohibit off-premises sale, some prohibit on-premises sale, and some prohibit both.
yngdrv	% of Drivers Aged 15-24
vmiles	Ave. Mile per Driver
jaild	Mandatory Jail Sentence
comserd	Mandatory Community Service
allmort	# of Vehicle Fatalities (#VF)
mrall	Vehicle Fatality Rate (VFR)
allnite	# of Night-time VF (#NVF)
mralln	Night-time VFR (NFVR)
allsvn	# of Single VF (#SVN)
a1517	#VF, 15-17 year olds
mra1517	VFR, 15-17 year olds
a1517n	#NVF, 15-17 year olds

mra1517n	NVFR, 15-17 year olds
a1820	#VF, 18-20 year olds
a1820n	#NVF, 18-20 year olds
mra1820	VFR, 18-20 year olds
mra1820n	NVFR, 18-20 year olds
a2124	#VF, 21-24 year olds
mra2124	VFR, 21-24 year olds
a2124n	#NVF, 21-24 year olds
mra2124n	NVFR, 21-24 year olds
aidall	# of alcohol-involved VF
mraidall	Alcohol-Involved VFR
pop	Population
pop1517	Population, 15-17 year olds

pop1820	Population, 18-20 year olds
pop2124	Population, 21-24 year olds
miles	total vehicle miles (millions)
gspch	GSP Rate of Change This is a measure of economic growth

Problem Statements:

I have tried to analyze the factors affecting the alcohol involved vehicle fatality rate. The main questions that I have tried to find answers for are :

1. Do factors such as unemployment rate, personal income, per capita alcohol consumption, age of the driver etc have a significant effect on alcohol involved vehicle fatality rate?
2. Is making jail sentences mandatory for drunk drivers effective in reducing the alcohol involved vehicle fatality rate? Similarly, introducing policies such as mandatory community service for drunk drivers, beer tax, minimum legal age for drinking or increasing “dry” counties in the state have a negative effect on reducing alcohol involved vehicle fatality rate?

2.Data Munging

The dim() function tells us the dimension of the dataset which is 336 records and 39 columns

```
> dim(car_fatality)
[1] 336  39
```

The glimpse() function gives an overview of the dataset. It represents the dataset in a transpose version. It also tells us about the datatype of each column.

```
> glimpse(car_fatality)
Observations: 336
Variables: 39
$ state    <fct> AL, AL, AL, AL, AL, AL, AL, AZ, AZ, AZ, AZ, AZ, A...
$ year     <int> 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1982, 1...
$ spircons <dbl> 1.37, 1.36, 1.32, 1.28, 1.23, 1.18, 1.17, 1.97, 1...
$ unrate   <dbl> 14.4, 13.7, 11.1, 8.9, 9.8, 7.8, 7.2, 9.9, 9.1, 5...
$ perinc   <dbl> 10544.15, 10732.80, 11108.79, 11332.63, 11661.51,...
$ beertax  <dbl> 1.5393795, 1.7889907, 1.7142856, 1.6525424, 1.609...
$ sobapt   <dbl> 30.3557, 30.3336, 30.3115, 30.2895, 30.2674, 30.2...
$ mormon   <dbl> 0.32829, 0.34341, 0.35924, 0.37579, 0.39311, 0.41...
$ mlda     <dbl> 19.00, 19.00, 19.00, 19.67, 21.00, 21.00, 21.00, ...
$ dry      <dbl> 25.0063, 22.9942, 24.0426, 23.6339, 23.4647, 23.7...
$ yngdrv   <dbl> 0.211572, 0.210768, 0.211484, 0.211140, 0.213400,...
$ vmiles   <dbl> 7233.887, 7836.348, 8262.990, 8726.917, 8952.854,...
$ jaild    <int> 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0...
$ comserd  <int> 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0...
$ allmort  <int> 839, 930, 932, 882, 1081, 1110, 1023, 724, 675, 8...
$ mrall    <dbl> 0.000212836, 0.000234848, 0.000233643, 0.00021934...
$ allnite  <int> 146, 154, 165, 146, 172, 181, 139, 131, 112, 149,...
$ mralln   <dbl> 3.70370e-05, 3.88889e-05, 4.13638e-05, 3.63094e-0...
$ allsvn   <int> 99, 98, 94, 98, 119, 114, 89, 76, 60, 81, 75, 85,...
$ a1517    <int> 53, 71, 49, 66, 82, 94, 66, 40, 40, 51, 48, 72, 5...
$ mra1517  <dbl> 0.000253589, 0.000351485, 0.000248731, 0.00033846...
$ a1517n   <int> 9, 8, 7, 9, 10, 11, 8, 7, 7, 8, 11, 19, 16, 14, 5...
$ mra1517n <dbl> 0.000043062, 0.000039604, 0.000035533, 0.00004615...
$ a1820    <int> 99, 108, 103, 100, 120, 127, 105, 81, 83, 118, 10...
$ a1820n   <int> 34, 26, 25, 23, 23, 31, 24, 16, 19, 34, 26, 30, 2...
$ mra1820  <dbl> 0.0004468448, 0.0004928683, 0.0004752587, 0.00046...
$ mra1820n <dbl> 0.0001534618, 0.0001186535, 0.0001153540, 0.00010...
$ a2124    <int> 120, 124, 118, 114, 119, 138, 123, 96, 80, 123, 1...
$ mra2124  <dbl> 0.000413793, 0.000427586, 0.000409722, 0.00040140...
$ a2124n   <int> 32, 35, 34, 45, 29, 30, 25, 36, 17, 33, 30, 25, 3...
$ mra2124n <dbl> 0.000110345, 0.000120690, 0.000118056, 0.00015845...
$ aidall   <dbl> 309.438, 341.834, 304.872, 276.742, 360.716, 368....
$ mraidall <dbl> 0.000078498, 0.000086322, 0.000076428, 0.00006882...
$ pop      <dbl> 3942002, 3960008, 3988992, 4021008, 4049994, 4082...
$ pop1517  <dbl> 208999.6, 202000.1, 197000.0, 194999.7, 203999.9,...
$ pop1820  <dbl> 221553.4, 219125.5, 216724.1, 214349.0, 212000.0,...
$ pop2124  <dbl> 290000.1, 290000.2, 288000.2, 284000.3, 263000.3,...
$ miles    <dbl> 28516, 31032, 32961, 35091, 36259, 37426, 39684, ...
$ gspch    <dbl> -0.022124760, 0.046558253, 0.062797837, 0.0274899...
```

The sapply() and sum() function helps us find the missing values in the dataset. From the findings, it is clear that jaild (mandatory jail sentence) comserd(mandatory community service) has missing values.

```
> miss.val <- data.frame(miss.val = sapply(car_fatality, function(x) +   sum((is.na(x)))))
> miss.val
          miss.val
state            0
year             0
spircons         0
unrate           0
perinc           0
beertax          0
sobapt           0
mormon           0
mlda             0
dry              0
yngdrv           0
vmiles           0
jaild            1
comserd          1
allmort          0
mrall            0
allnite          0
mralln           0
allsvn           0
a1517            0
mra1517          0
a1517n           0
mra1517n         0
a1820            0
a1820n           0
mra1820          0
mra1820n         0
a2124            0
mra2124          0
a2124n           0
mra2124n         0
aidall           0
mraidall         0
pop              0
pop1517          0
pop1820          0
pop2124          0
miles            0
gspch            0
```

Replacing the missing values with median using impute function:

```
> car_fatality$jaild <- as.numeric( impute(car_fatality$jaild, median))
> car_fatality$comserd <- as.numeric( impute( car_fatality$comserd,median))
```

```
> miss.val
          miss.val
state            0
year             0
spircons         0
unrate           0
perinc           0
beertax          0
sobapt           0
mormon           0
mlda             0
dry              0
yngdrv           0
vmiles           0
jaild            0
comserd          0
allmort          0
mrall            0
allnite          0
mralln           0
allsvn           0
a1517            0
mra1517          0
a1517n           0
mra1517n         0
a1820            0
a1820n           0
mra1820          0
mra1820n         0
a2124            0
mra2124          0
a2124n           0
mra2124n         0
aidall           0
mraidall         0
pop              0
pop1517          0
pop1820          0
pop2124          0
miles            0
gspch            0
```

The car_fatality dataframe is now free of any missing values and is ready for data exploration.

3.Data Exploration

Upon viewing the table, we noticed that variables like population, beertax, VFR, mormon are all on different scales. Therefore, we scaled them as follows:

- beertax/100, unrate/100, sobapt/100, mormon/100, dry/100 : Transformed these variables to a proportion
- log(percapita), log(vmiles): Transformed to logarithmic values
- factor(jaild), factor(comserd): Transformed to categorical variable
- floor(mlda): Rounded off to the nearest integer

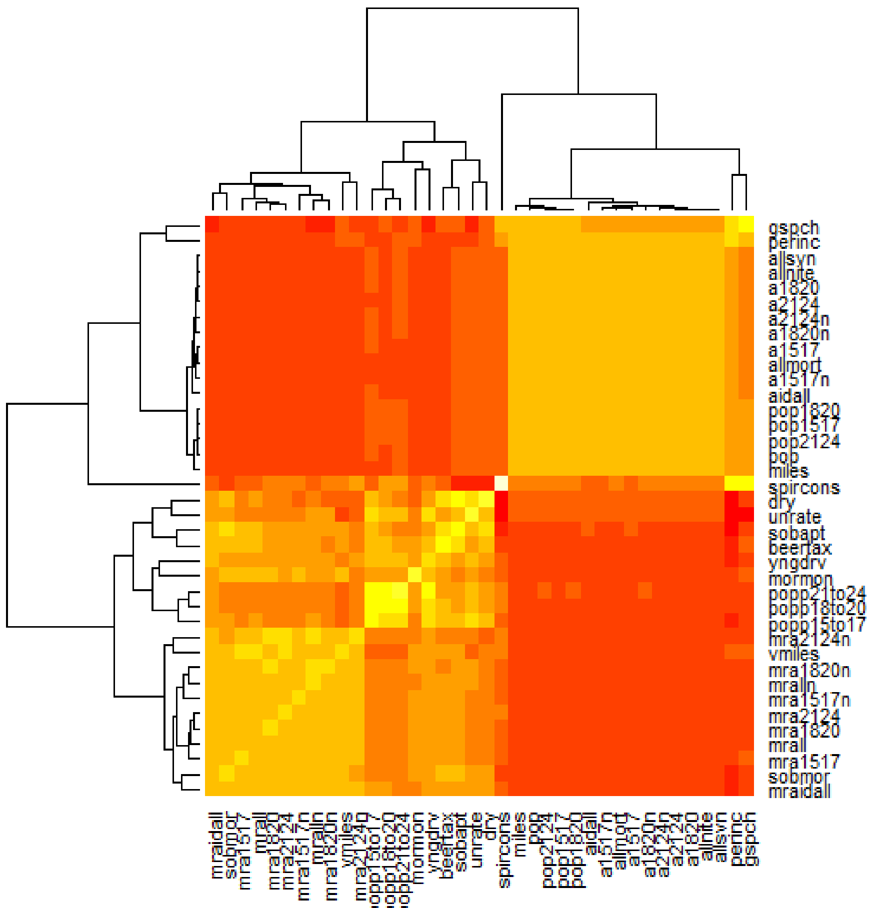
Economic theory interpretation of variables

- **spircons**: If per capita consumption of alcohol is more then it indicates the states VFR also increases.
- **unrate**: If the unemployment rate in the state is high, then people tend to get more frustrated and depressed. The alcohol consumption increases and hence the VFR increases.
- **perinc**: If the per capita personal income increases, then economy of the state will be good. This indicates that the person will have better lifestyle and good cars with good safety features. Hence alcohol involved VFR decreases.
- **beertax**: If beertax increases, then less would buy beer. Hence it can be said that beertax and VFR are negatively correlated.
- **sobapt**: From online research, it is stated that Southern Baptist convention confirms biblical warnings that use of alcohol leads to physical, mental and emotional damage. From this it can be stated that sobapt and alcohol involved VFR are negatively related.
- **mormon**: From online research, the mormon religion has restrictions on alcohol beverages ,including beer. This means there is a negative correlation between mormon, and alcohol involved VFR.
- **mlda**: If the minimum legal driving age increases, then alcohol involved VFR decreases. The probability of accidents among young intoxicated drivers is more.
- **dry**: If the percentage of dry counties in a state is high then the number alcohol consumed VFR is low as the drivers will not be intoxicated and will be more responsible while driving.
- **yngdrv**: If the percentage of drivers aged 15-24 increases, then the probability of alcohol involved VFR increases due to less experience. Hence there is a positive correlation.
- **vmiles**: If average miles per driver increases, then the driver becomes more exhausted and hence the alcohol involved VFR will increase.
- **pop**, **pop1517**, **pop1820**, **pop2124**: These variables are positively correlated with alcohol involved VFR.
- **miles**: There is a positive correlation with alcohol involved VFR.
- **gspch**: If the gsp which is the measure of economic growth increases then more people can afford to buy cars and hence the alcohol involved VFR increases.

Correlation matrix

Considering only the numeric variables and eliminating the factor variables, we created the correlation matrix and heatmap to know the relation between the variables. The heatmap seems to be compliant with the economic theory stated above. For further analysis, we have ignored the highly correlated variables to avoid multicollinearity problem.

```
sapply(carpp,class)

sapply(carpp,is.factor)

correlation_car<-cor(carpp[sapply(carpp, function(x) !is.factor(x))])
correlation_car

heatmap(cor(correlation_car),Rowv= FALSE, Colv= FALSE, dendogram="none",cellnote= round(cor(correlation_car),2),
         notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))
```

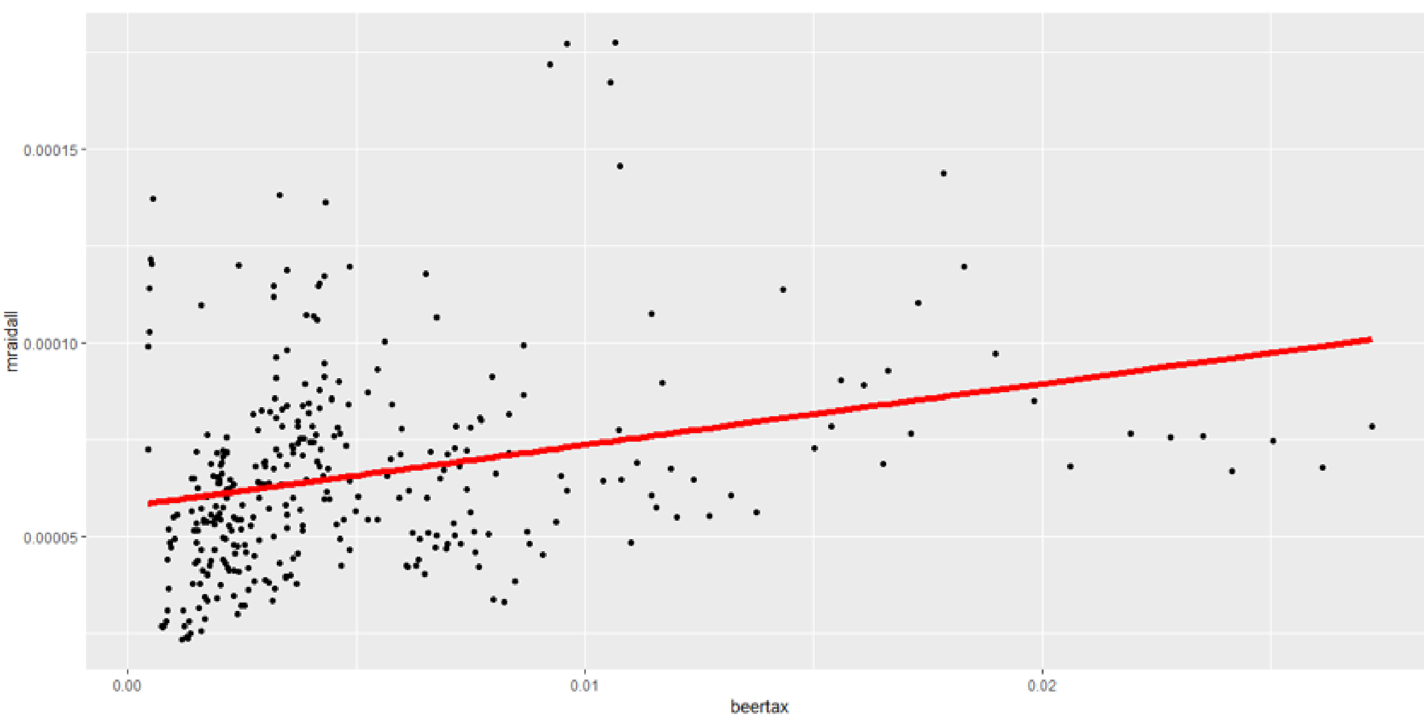


Regression analysis

To understand the topic on “How drunk driving can affect the accident facility?” , we ran a regression plot between beertax vs alcohol involved VFR (mraidall).

```
> coeftest(pl,vcov.=vcovHC)

t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 5.7815e-05 1.9085e-06 30.2935 < 2.2e-16 ***
beertax     1.5810e-03 3.0433e-04  5.1952 3.566e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In the below plot, I observe that there is a positive correlation between the beertax and alcohol involved VFR(mraidall) which is against the assumption of the economic theory stated above. This can be due to simultaneous causality bias and unobserved heterogeneity. There is a chance that taxes could have been imposed on those states where more people drink while driving. Cultural attitude towards driving and driving could have been an omitted variable where the beertax variable has picked up and could have caused an upward bias.



In order to overcome this issue of omitted variable bias, it is important to consider suitable instrument variables. The instrument variables should mainly satisfy the following conditions:

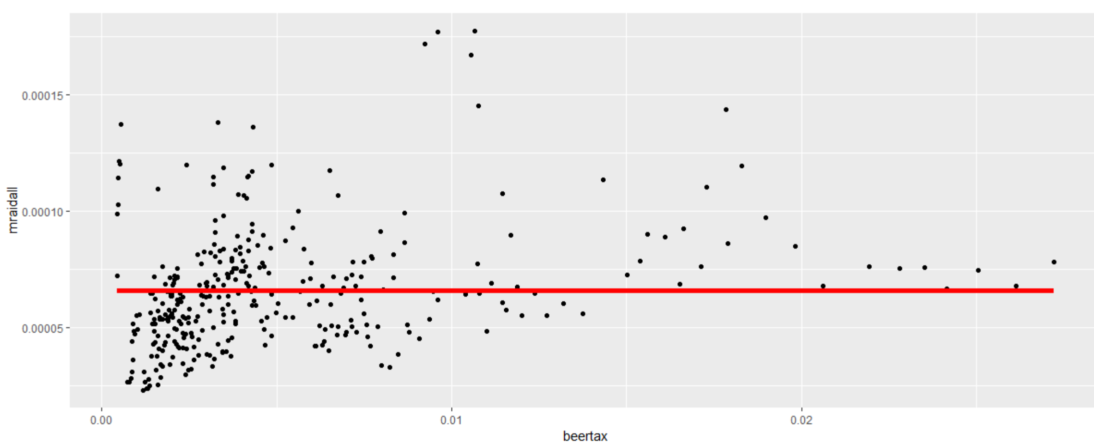
- It should not have direct effect on the dependent variable i.e. alcohol involved VFR and should not be involved as an explanatory variable
- It should not be correlated to the regression error term
- It should strongly be correlated to the independent variable i.e. beertax the endogenous explanatory variable

In the above case, the variables mormon and sobapt can be used as instrument variables. Firstly, they do not have direct effect on alcohol involved VFR as the religious groups are against alcohol consumption and hence the fatality is not correlated to these variables. Secondly, they are not correlated to error terms. Thirdly, these variables are correlated to the beertax variable as they can force the government to take stringent steps to ban alcohol consumption and hence impose more beer tax.

Instrument variables

Considering mormon and sobapt as instrument variables, we performed the following regression analysis

```
> p2<-plm(mraidall~beertax|mormon+sobapt,data=carpp)
> summary(p2)
Oneway (individual) effect Within Model
Instrumental variable estimation
   (Balestra-Varadharajan-Krishnakumar's transformation)

Call:
plm(formula = mraidall ~ beertax | mormon + sobapt, data = carpp)

Balanced Panel: n = 48, T = 7, N = 336

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-9.3544e-05 -5.8990e-06 -2.6722e-07  4.5895e-06  4.7727e-05

Coefficients:
          Estimate  Std. Error t-value Pr(>|t|)
beertax -0.00014202  0.00263138  -0.054    0.957

Total Sum of Squares:    5.39e-08
Residual Sum of Squares: 5.3883e-08
R-Squared:      0.00079951
Adj. R-Squared: -0.16631
F-statistic: 0.0894031 on 1 and 287 DF, p-value: 0.76515
```

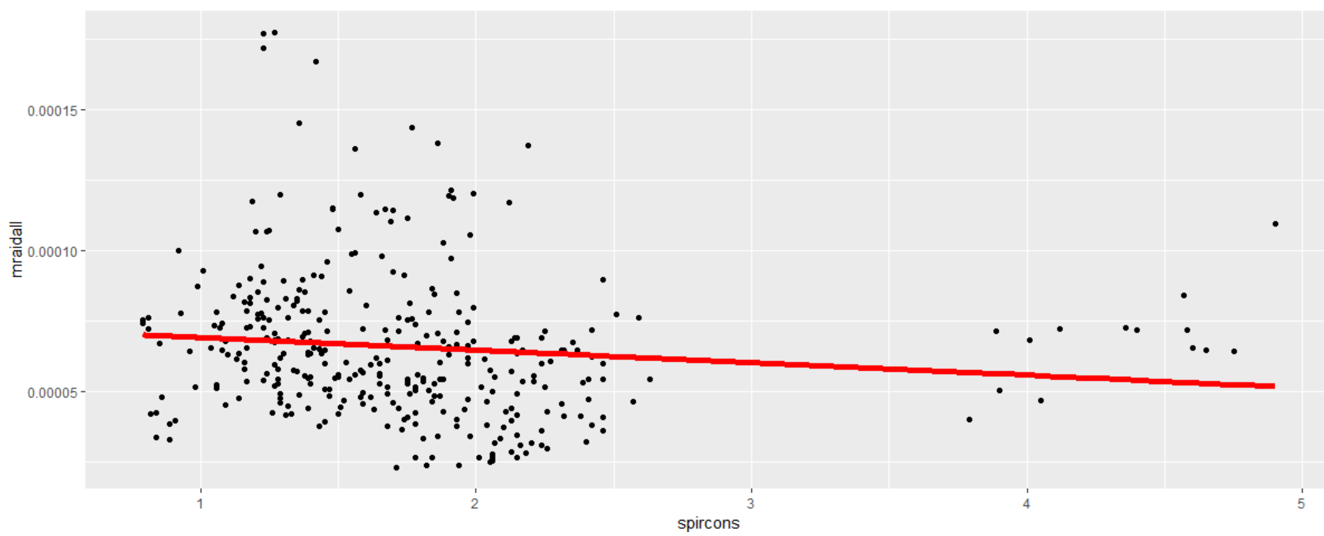


It can be observed from the above output that there is a negative correlation between the beer tax and alcohol involved VFR which is complaint with our economic theory. However, on observing the p-value(0.957) we can conclude that beertax is an insignificant variable as its p-value is above the significance level(0.05).The null hypothesis, which states that beer tax has no effect on the alcohol involved VFR cannot be rejected. Therefore, to overcome the endogeneity problem within this model, I have to further my analysis to choose a right model.

Variables of interest for model selection:

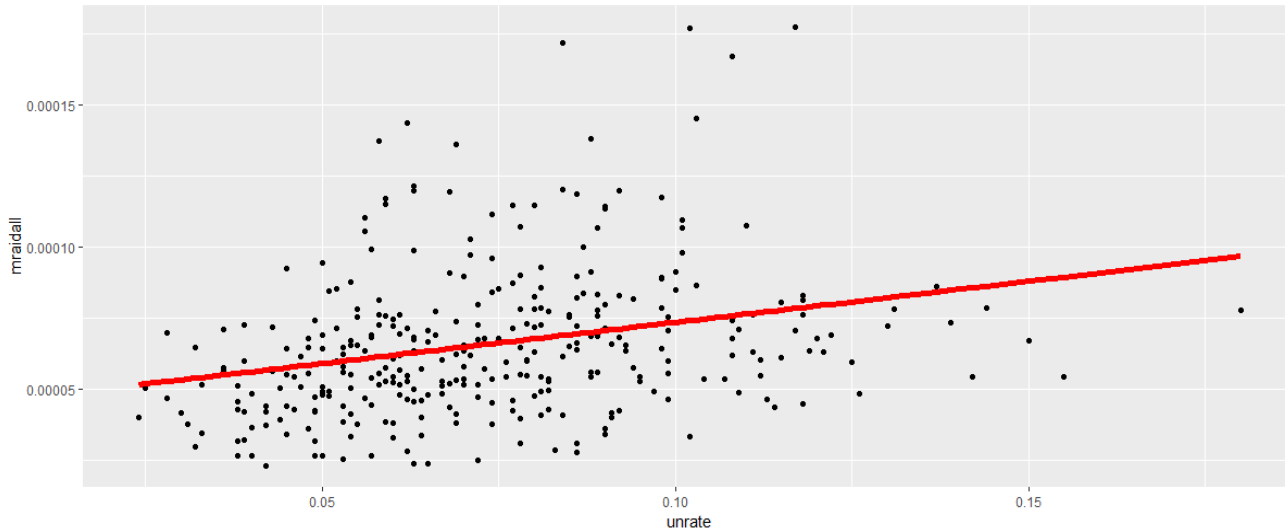
- **spircons** - negatively correlated to mraidall which is against the economic theory

```
> p3<-lm(mraidall~spircons,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ spircons, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-4.275e-05 -1.701e-05 -3.724e-06  1.158e-05  1.091e-04 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.373e-05  3.885e-06  18.980   <2e-16 ***
spircons    -4.449e-06  2.064e-06  -2.155   0.0319 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.583e-05 on 334 degrees of freedom
Multiple R-squared:  0.01371,	Adjusted R-squared:  0.01076 
F-statistic: 4.644 on 1 and 334 DF,  p-value: 0.03187
```



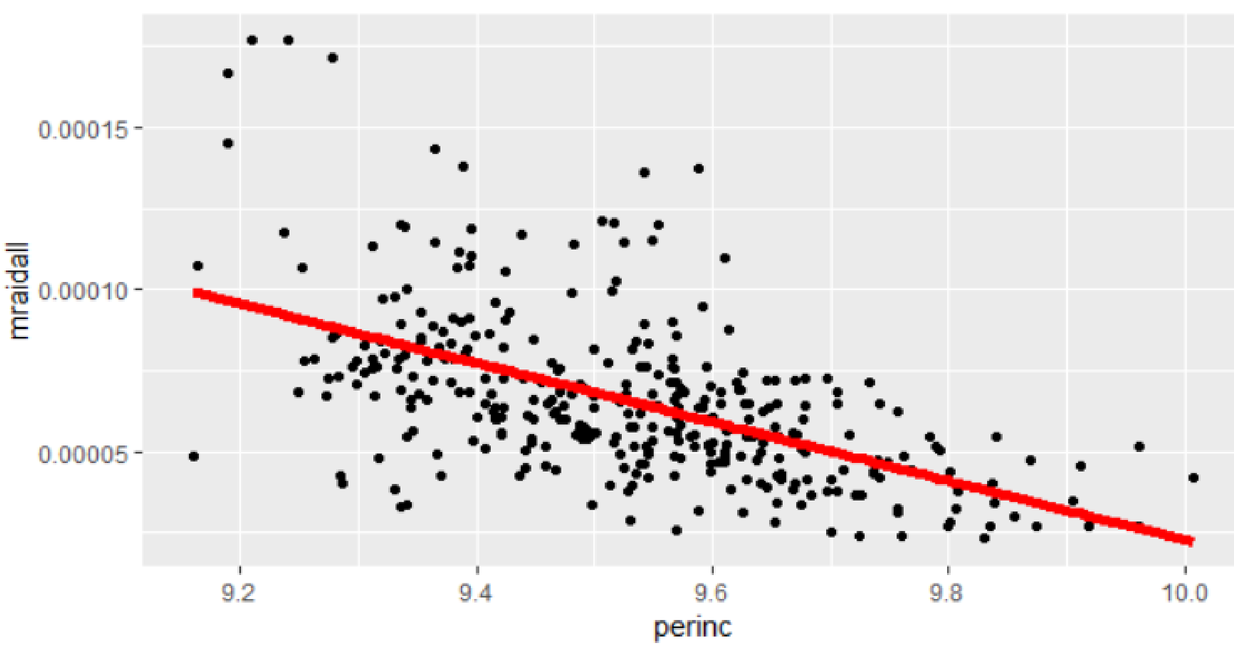
- **beertax** - As seen above, there is a positive correlation with mraidall which is against the economic theory.
- **unrate** - It is positively correlated to mraidall which is same as the economic theory.

```
> p3<-lm(mraidall~unrate,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ unrate, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-4.152e-05 -1.589e-05 -3.875e-06  1.063e-05  1.028e-04 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.469e-05  4.181e-06  10.688  < 2e-16 ***
unrate      2.891e-04  5.381e-05   5.374 1.45e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.495e-05 on 334 degrees of freedom
Multiple R-squared:  0.07957,	Adjusted R-squared:  0.07682 
F-statistic: 28.88 on 1 and 334 DF,  p-value: 1.451e-07
```



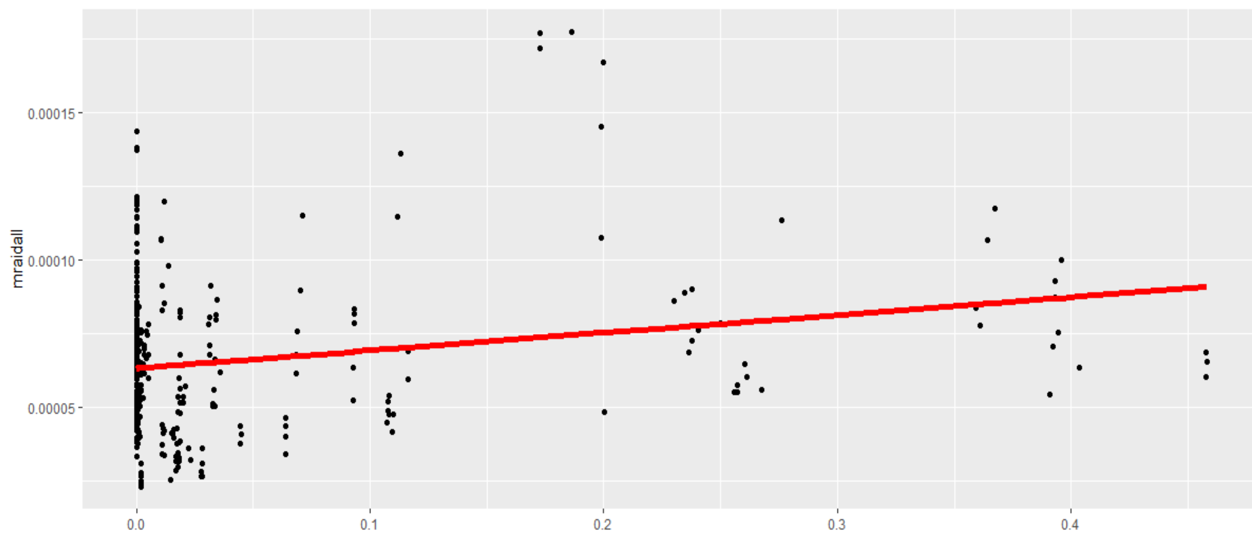
- **perinc** - There is a negative correlation with mraidall which is same as economic theory.

```
> p3<-lm(mraidall~perinc,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ perinc, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max
-5.072e-05 -1.269e-05 -3.180e-06  9.996e-06  8.503e-05

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.324e-04  7.125e-05   13.09   <2e-16 ***
perinc      -9.096e-05  7.478e-06  -12.16   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.165e-05 on 334 degrees of freedom
Multiple R-squared:  0.307,     Adjusted R-squared:  0.3049
F-statistic:   148 on 1 and 334 DF,  p-value: < 2.2e-16
```



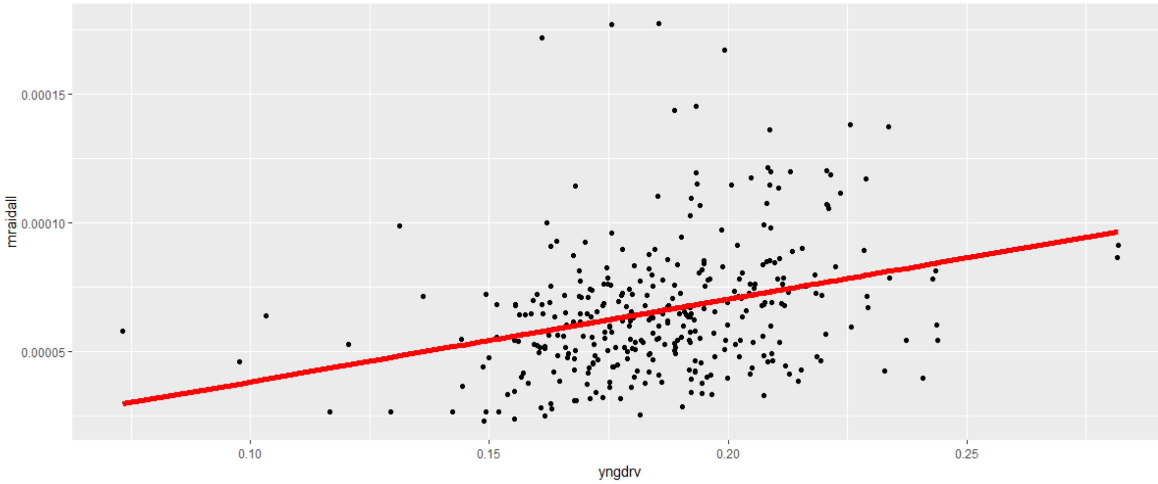
- **dry** - There is positive correlation with mraidall which is against the economic theory.

```
> p3<-lm(mraidall~dry,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ dry, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-4.010e-05 -1.739e-05 -3.332e-06  1.124e-05  1.032e-04 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.337e-05  1.518e-06   41.75  < 2e-16 ***
dry         6.011e-05  1.459e-05    4.12 4.78e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.537e-05 on 334 degrees of freedom
Multiple R-squared:  0.04837,	Adjusted R-squared:  0.04552 
F-statistic: 16.98 on 1 and 334 DF,  p-value: 4.779e-05
```



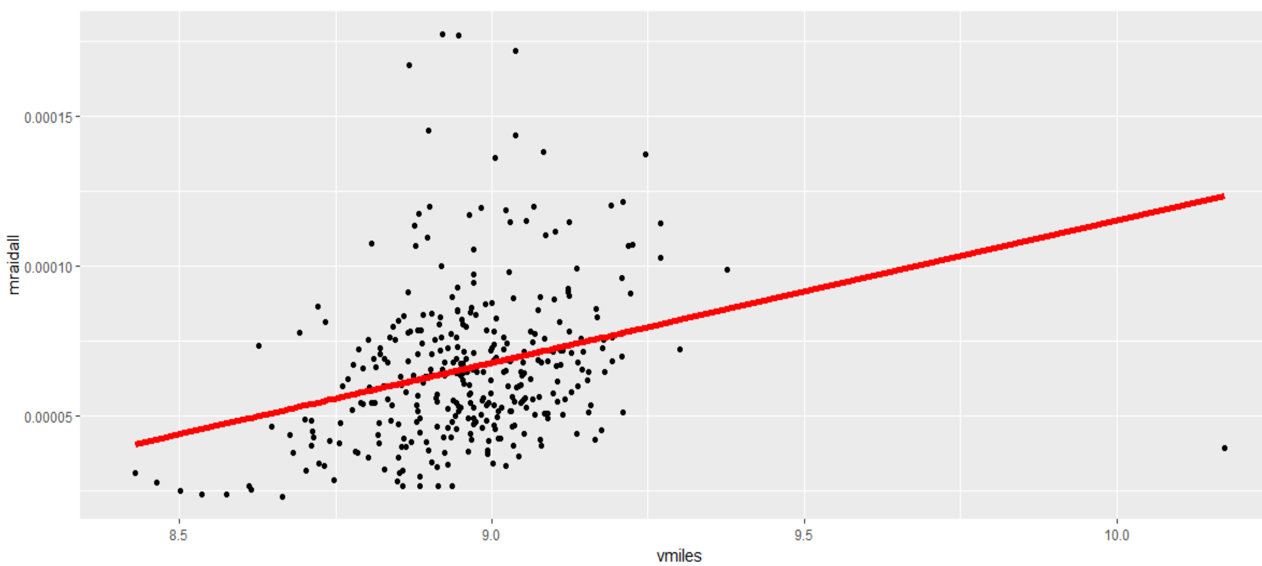
- **yngdrv** - There is a positive correlation with mraidall which is complaint with the economic theory and is significant.

```
> p3<-lm(mraidall~yngdrv,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ yngdrv, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-4.350e-05 -1.640e-05 -3.242e-06  1.101e-05  1.143e-04 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.334e-06  1.020e-05   0.621    0.535    
yngdrv      3.205e-04  5.437e-05   5.896 9.13e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.475e-05 on 334 degrees of freedom
Multiple R-squared:  0.09426,	Adjusted R-squared:  0.09155 
F-statistic: 34.76 on 1 and 334 DF,  p-value: 9.13e-09
```



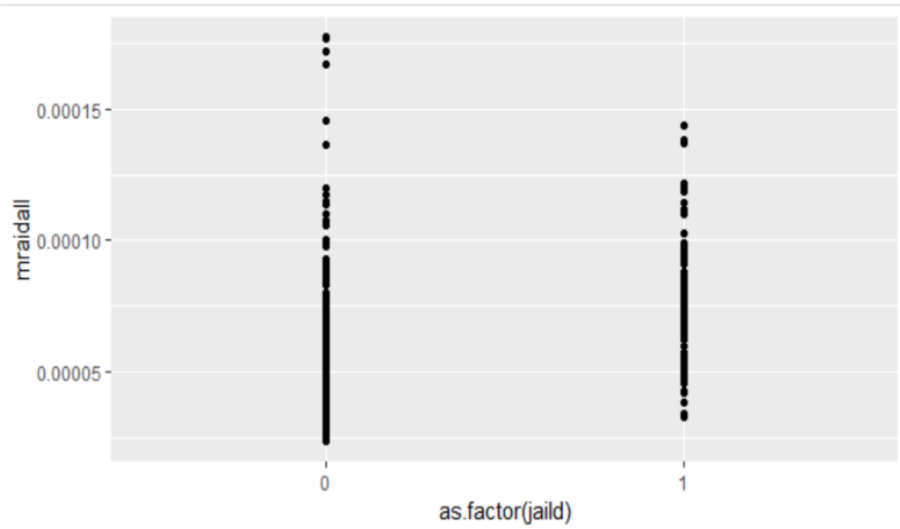
- **vmiles** - There is a positive correlation with mraidall which is same as the economic theory

```
> p3<-lm(mraidall~vmiles,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ vmiles, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-8.396e-05 -1.653e-05 -4.242e-06  1.165e-05  1.131e-04 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.595e-04  7.808e-05  -4.604 5.90e-06 ***
vmiles       4.748e-05  8.713e-06   5.449 9.85e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.492e-05 on 334 degrees of freedom
Multiple R-squared:  0.08164,	Adjusted R-squared:  0.07889 
F-statistic: 29.69 on 1 and 334 DF,  p-value: 9.846e-08
```



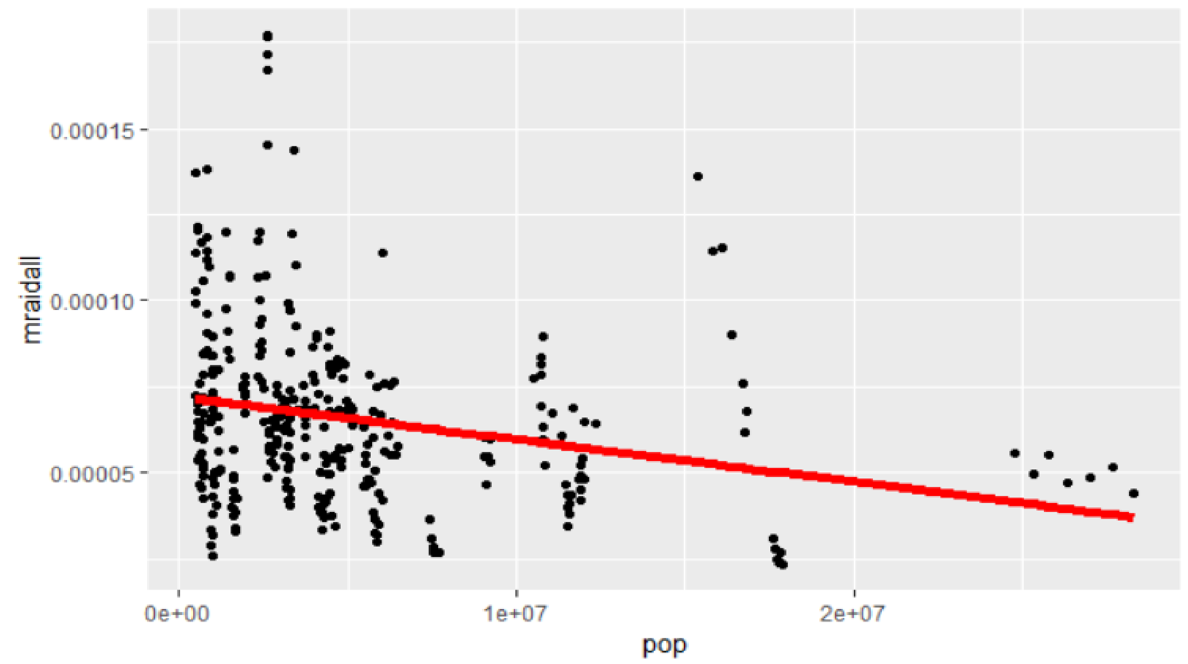
- **jaild** - If mandatory jail sentence is applicable then the alcohol involved VFR should reduce according to economic theory. Below we see that, if jaild is 0, then mraidall increases

```
> p3<-lm(mraidall~as.factor(jaild),data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ as.factor(jaild), data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max
-4.270e-05 -1.649e-05 -3.905e-06  9.732e-06  1.151e-04

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       6.212e-05  1.625e-06   38.24  < 2e-16 ***
as.factor(jaild)1 1.361e-05  3.072e-06    4.43 1.28e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.527e-05 on 334 degrees of freedom
Multiple R-squared:  0.05549,   Adjusted R-squared:  0.05266
F-statistic: 19.62 on 1 and 334 DF,  p-value: 1.281e-05
```



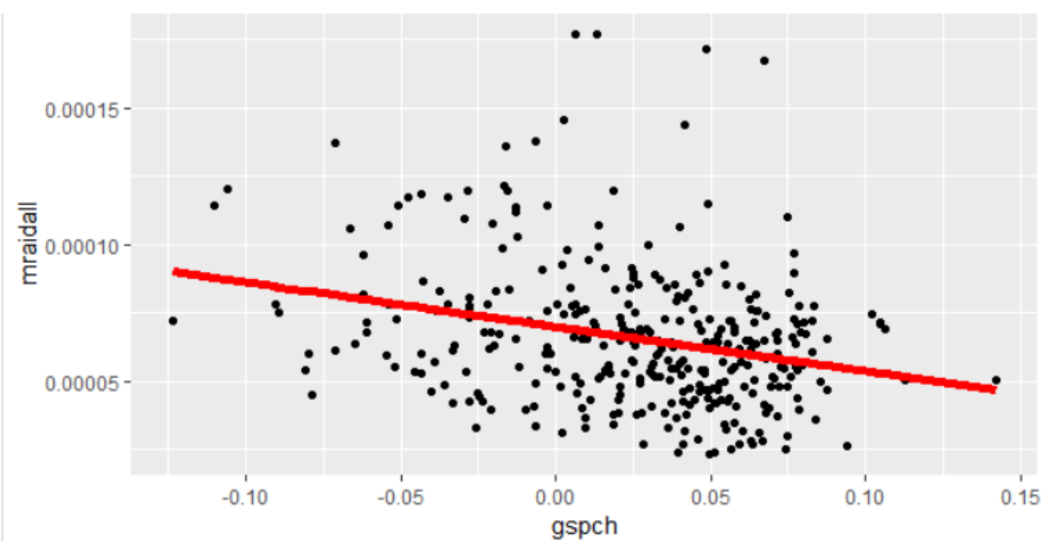
- **pop** - There is a negative correlation with mraidall which is against the economic theory.

```
> p3<-lm(mraidall~pop,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ pop, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max
-4.526e-05 -1.685e-05 -3.404e-06  1.129e-05  1.084e-04

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.202e-05  1.921e-06  37.489  < 2e-16 ***
pop         -1.236e-12  2.718e-13  -4.547 7.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.524e-05 on 334 degrees of freedom
Multiple R-squared:  0.0583,    Adjusted R-squared:  0.05548
F-statistic: 20.68 on 1 and 334 DF,  p-value: 7.612e-06
```



- **gspch** - There is a negative correlation with mraidall which is against the economic theory.

```
> p3<-lm(mraidall~gspch,data=carpp)
> summary(p3)

Call:
lm(formula = mraidall ~ gspch, data = carpp)

Residuals:
       Min         1Q     Median         3Q        Max 
-4.088e-05 -1.715e-05 -2.967e-06  1.137e-05  1.096e-04 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.004e-05  1.584e-06  44.210  < 2e-16 ***
gspch       -1.623e-04  3.169e-05  -5.122 5.11e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.504e-05 on 334 degrees of freedom
Multiple R-squared:  0.07284,	Adjusted R-squared:  0.07006 
F-statistic: 26.24 on 1 and 334 DF,  p-value: 5.112e-07
```



Model selection

Based on the insights above and analysis performed earlier, the following variables are considered for the model selection :

mraidall = f(spircons, beertax, unrate, perinc, dry, yngdrv, vmiles, jaild, pop, gspch)

- **Pooled OLS Model**

```
> summary(model1)
Pooling Model

Call:
plm(formula = mraidall ~ spircons + beertax + unrate + perinc +
    dry + yngdrv + vmiles + I(jaild) + pop + gspch, data = carpp,
    model = "pooling")

Balanced Panel: n = 48, T = 7, N = 336

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-6.8678e-05 -1.0707e-05 -1.8849e-06  8.9473e-06  9.1062e-05

Coefficients:
               Estimate  Std. Error t-value  Pr(>|t|)
(Intercept)  4.4263e-04  1.5882e-04  2.7870  0.005632 **
spircons     8.6402e-06  1.9667e-06  4.3932 1.514e-05 ***
beertax      2.2249e-04  2.7792e-04  0.8006  0.423974
unrate      -3.1623e-05  6.6961e-05 -0.4723  0.637059
perinc      -8.5002e-05  1.3601e-05 -6.2495 1.292e-09 ***
dry          4.0007e-05  1.3361e-05  2.9943  0.002962 **
yngdrv       6.8215e-05  5.3291e-05  1.2800  0.201441
vmiles       4.4652e-05  8.0211e-06  5.5667 5.439e-08 ***
I(jaild)1    1.1482e-05  2.6671e-06  4.3049 2.214e-05 ***
pop          6.6816e-13  2.8255e-13  2.3647  0.018630 *
gspch       -7.7844e-05  3.1171e-05 -2.4973  0.013007 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    2.259e-07
Residual Sum of Squares: 1.2589e-07
R-Squared:      0.4427
Adj. R-Squared: 0.42555
F-statistic: 25.8167 on 10 and 325 DF, p-value: < 2.22e-16
```

In the pooled OLS model, alcohol involved VFR is regressed against various explanatory variables as shown above. There are many variables which are significant, however, not complaint with the economic theory. The beertax variable is insignificant which states that this is not good model to be considered for omitted variable bias. The fixed effect model would be a better model to be considered for capturing the unobserved heterogeneity like cultural attitude of people towards drinking.

- **Fixed Effect Model**

```
> model2<-plm(mraidall~spircons+beertax+unrate+perinc+ mlda+jaild+comserd+ dry+
yngdrv+ vmiles+jaild*comserd*mlda+pop,data=carpp,model="within")
> coeftest(model2,method=vcovHC)

t test of coefficients:

                          Estimate  Std. Error t value  Pr(>|t|)    
spircons                3.1807e-05  8.2745e-06  3.8439 0.0001514 ***
beertax                -2.7817e-03  1.4361e-03 -1.9370 0.0537963 .  
unrate                 -1.4932e-04  8.6594e-05 -1.7244 0.0857978 .  
perinc                  2.1500e-05  3.1545e-05  0.6816 0.4961040    
mlda19                  5.8346e-06  6.6493e-06  0.8775 0.3810226    
mlda20                  3.9276e-06  7.3583e-06  0.5338 0.5939436    
mlda21                  4.3925e-06  6.9283e-06  0.6340 0.5266260    
jaild1                  3.0630e-05  1.6631e-05  1.8418 0.0666206 .  
comserd1               -2.9744e-05  2.0380e-05 -1.4595 0.1456145    
dry                     1.2483e-05  1.0664e-04  0.1171 0.9069002    
yngdrv                  1.2251e-04  6.3563e-05  1.9274 0.0549853 .  
vmiles                 -3.7492e-06  1.0049e-05 -0.3731 0.7093631    
pop                    -4.8776e-12  3.9110e-12 -1.2471 0.2134376    
mlda19:jaild1          -4.7491e-06  1.3015e-05 -0.3649 0.7154716    
mlda20:jaild1          -9.1320e-08  1.5771e-05 -0.0058 0.9953843    
mlda21:jaild1          -9.8348e-06  1.3450e-05 -0.7312 0.4653029    
mlda19:comserd1         1.0282e-05  2.0580e-05  0.4996 0.6177613    
mlda20:comserd1         5.3157e-07  2.2843e-05  0.0233 0.9814516    
mlda21:comserd1         9.3377e-06  1.6954e-05  0.5508 0.5822449    
mlda19:jaild1:comserd1 -2.7176e-05  1.6525e-05 -1.6446 0.1012373    
mlda20:jaild1:comserd1 -5.7110e-06  2.0750e-05 -0.2752 0.7833558    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

On running the fixed effect model, many variables become insignificant. In the above model, we have considered the interaction between jaild, comserd and mlda considering the fear of mandatory jail and community service will be high among young individuals. Hence this can reduce the alcohol involved VFR.

On close observation, the interaction terms are insignificant and hence can be omitted for further analysis although it did improve the R^2 of the model. The beertax variable is negatively correlated with mraidall and is significant with economic theory. However, there can be an omitted variable bias for which I considered to run the entity fixed and time fixed effect model for further analysis.

- **Fixed Effect Model – Fixed Entity and Fixed Time**

Fixed Time model -

```
> model3 <- plm(mraidall~as.factor(year)+spircons+beertax+unrate+perinc+mlda+jaild+comserd+dry+yngdrv+vmiles+pop,
+               data = carpp, model = "within")
> summary(model3)
Oneway (individual) effect Within Model

Call:
plm(formula = mraidall ~ as.factor(year) + spircons + beertax +
    unrate + perinc + mlda + jaild + comserd + dry + yngdrv +
    vmiles + pop, data = carpp, model = "within")

Unbalanced Panel: n = 48, T = 6-7, N = 335

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-9.4050e-05 -5.1318e-06 -2.6069e-07  4.6017e-06  4.5217e-05

Coefficients:
                       Estimate  Std. Error t-value  Pr(>|t|)
as.factor(year)1983 -6.5256e-06  2.6484e-06 -2.4639 0.0143701 *
as.factor(year)1984 -1.1769e-05  3.4792e-06 -3.3828 0.0008247 ***
as.factor(year)1985 -1.5250e-05  4.0510e-06 -3.7646 0.0002050 ***
as.factor(year)1986 -9.9282e-06  5.2038e-06 -1.9079 0.0574754 .
as.factor(year)1987 -1.4292e-05  6.1240e-06 -2.3337 0.0203512 *
as.factor(year)1988 -1.5717e-05  7.2120e-06 -2.1793 0.0301786 *
spircons             2.6725e-05  9.6903e-06  2.7579 0.0062177 **
beertax             -1.9912e-03  1.3523e-03 -1.4724 0.1420781
unrate              -2.4939e-04  9.2451e-05 -2.6975 0.0074286 **
perinc               3.3033e-05  3.1185e-05  1.0593 0.2904322
mlda19               3.8835e-06  5.1408e-06  0.7554 0.4506598
mlda20               3.3493e-06  5.6394e-06  0.5939 0.5530686
mlda21               2.5289e-06  5.3013e-06  0.4770 0.6337255
jaild1               2.3424e-05  9.5779e-06  2.4456 0.0151052 *
comserd1            -2.0294e-05  1.0999e-05 -1.8451 0.0661281 .
dry                 -1.1563e-05  1.0367e-04 -0.1115 0.9112733
yngdrv               3.6886e-05  7.0950e-05  0.5199 0.6035668
vmiles              -1.3526e-09  9.5761e-06 -0.0001 0.9998874
pop                 -3.9764e-12  4.1821e-12 -0.9508 0.3425602
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    5.3855e-08
Residual Sum of Squares: 4.0738e-08
R-Squared:      0.24355
Adj. R-Squared: 0.057264
F-statistic: 4.54147 on 19 and 268 DF, p-value: 5.8305e-09
```

Here we see that there is a significant impact of the year-panel on our model. This goes to show that there is a huge impact of the year of occurrence on Alcohol involved VFR. This is a critical understanding for our model.

Next we will look at the "between" model.

```
> model4 <- plm(mraidall~as.factor(year)+spircons+beertax+unrate+perinc+mlda+jaild+comserd+dry+yngdrv+vmiles+pop,
+               data = carpp, model = "between")
> summary(model4)
Oneway (individual) effect Between Model

Call:
plm(formula = mraidall ~ as.factor(year) + spircons + beertax +
    unrate + perinc + mlda + jaild + comserd + dry + yngdrv +
    vmiles + pop, data = carpp, model = "between")

Unbalanced Panel: n = 48, T = 6-7, N = 335
Observations used in estimation: 48

Residuals:
      Min.     1st Qu.      Median     3rd Qu.        Max.
-3.2975e-05 -7.3810e-06 -1.5398e-06  9.1812e-06  4.9315e-05

Coefficients: (5 dropped because of singularities)
                      Estimate  Std. Error t-value Pr(>|t|)
(Intercept)          3.0190e-04  4.4111e-04  0.6844  0.49849
as.factor(year)1983 -2.7809e-04  1.0424e-03 -0.2668  0.79130
spircons             8.6610e-06  5.1369e-06  1.6861  0.10122
beertax             -9.9334e-05  6.5705e-04 -0.1512  0.88075
unrate               2.3185e-04  1.9494e-04  1.1893  0.24280
perinc              -8.8903e-05  3.4106e-05 -2.6067  0.01362 *
mlda19               2.0513e-05  2.0035e-05  1.0238  0.31335
mlda20               1.2010e-05  2.4891e-05  0.4825  0.63262
mlda21               9.0119e-06  1.7505e-05  0.5148  0.61011
jaild1               4.9494e-06  7.8763e-06  0.6284  0.53407
comserd1             6.6011e-06  8.7739e-06  0.7524  0.45717
dry                  2.8429e-05  3.1256e-05  0.9095  0.36966
yngdrv              -1.4305e-05  1.7601e-04 -0.0813  0.93572
vmiles               6.7404e-05  2.5918e-05  2.6007  0.01381 *
pop                  6.0810e-13  9.6156e-13  0.6324  0.53148
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    2.454e-08
Residual Sum of Squares: 8.9201e-09
R-Squared:      0.6365
Adj. R-Squared: 0.48229
F-statistic: 4.12751 on 14 and 33 DF, p-value: 0.00039828
```

This tells us that discarding the assumption on intra-group variability is not a great idea and hence I will stick with the within group variability.

4. CONCLUSION

1. Running multiple models and trying out various input variables, I have decided to go with the Fixed Effects model.
2. According to the results of our fixed effects model, we can conclude that implementing beer tax has a significant negative effect on alcohol involved vehicle fatality rate. At 10% significance level, 1$ increase in beer tax decreases the alcohol involved traffic deaths by an average 0.2%. This is in line with economic theory. It is advisable for states with higher alcohol involved fatality rates to increase the beer tax to mitigate the issue.
3. We also notice that the state's unemployment rate has a negative effect on the fatality rate and the effect is significant at 10% significance level. Even though this is against our initial assumption, one could argue that the states with higher unemployment rates would have fewer people owning cars and investing in gas or buying alcohol.
4. The model also indicates that the % of young drivers also have a significant positive effect on alcohol involved fatality rate. This is in line with the economic theory. Introducing stricter age restrictions to

buy or consume alcohol would be an effective measure for states to mitigate drink and drive death rates.

5. Even though our model indicates that making jail sentences mandatory has a positive effect on alcohol involved fatality rates, the initial regression model indicates otherwise. Through our analysis we conclude that making jail sentences mandatory does infact reduce the alcohol involved deaths, which is in line with economic theory.
6. We also believe that fine tuning the model by trying to reduce the effect of omitted variable bias or by increasing the number of data points could help arrive at much clearer and significant results.
7. Overall, our analysis supports the implementation of policies and regulations put in place by states to mitigate alcohol involved traffic fatality rates.