# How do drunk driving laws affect traffic deaths?

PANEL DATA ANALYSIS

**BUAN6312.003**

PREPARED BY:

ACHINTYA SEN
HAMEER HANDU
RAMA KRISHNA REDDY
YE LIM SEO

# Table of Content

# Data Description

| 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 |
| 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 |

# Introduction

Car fatality is the major issue faced by transport industry wherein majority of the cause is substance abuse or negligent driving. In 2014, 9,967 people were killed in alcohol-impaired driving crashes, accounting for nearly one-third (31%) of all traffic-related deaths in the United States. Motivated to find the cause and solution, we are analyzing the panel data to gain evidence to prove the major factors of vehicle fatality due to alcohol.

We are also trying to find useful insights regarding the factors causing the increment and decrement of the fatality rate. The data set consist of multiple plausible explanatory variables, but feature selection is the key to find trends and insights. We perform an initial correlation plot on all the variable in the dataset, as we can see in APPENDIX A, there are few correlated variables. Highlighting the ones above 0.5 we find some new insights like:

- *beertax* (Tax on Case of Beer ($)) is related to the *sobapt* (% Southern Baptist), which is related to the *dry* (% Residing in Dry Counties). These relations are as expected, the states having higher Christian population will implement stringent taxes.

- Higher populations area has more demand for allnite drivers hence have higher *allnite* (# of Night-time VF (#NVF)

Based on logic and the correlation table we select *UNRATE* (State Unemployment rate %), *PERINC* (Per Capita Personal Income), *BEERTAX* (Tax on Case of Beer), *VMILES* (Average miles per Driver), *MLDA* (Minimum Legal Drinking Age in years), *JAILD* (Mandatory Jail Sentence), and *COMSERD* (Mandatory Community Service) to estimate our model. We expect to understand the relation how drunk driving laws affect traffic deaths, some given variables are negligible in the scenario. Based on the variables we will perform our regression equation and try to understand the relation between them and find useful insight.

# Analysis

## Correlation Test

The selected features are used to perform a correlation test to check if there are any linear correlation between dependent variable and explanatory variables. *perinc, vmiles*, and *beertax* have relatively higher correlation with dependent variable. *mlda* has very weak correlation with dependent variable. What should draw our attention is there is a highly negative correlation between *unrate* and *perinc*, and a highly positive correlation between *jaild* and *comserd*.

| | mrall | unrate | perinc | beertax | mlda | vmiles | jaild | comserd |
|---|---|---|---|---|---|---|---|---|
| mrall | 1.0000 | | | | | | | |
| unrate | 0.1750 | 1.0000 | | | | | | |
| perinc | -0.5001 | -0.5525 | 1.0000 | | | | | |
| beertax | 0.3053 | 0.0547 | -0.3951 | 1.0000 | | | | |
| mlda | -0.0905 | -0.2585 | 0.2009 | -0.0585 | 1.0000 | | | |
| vmiles | 0.3973 | -0.2774 | -0.0831 | 0.1433 | 0.0585 | 1.0000 | | |
| jaild | 0.2779 | 0.1445 | -0.1507 | -0.0392 | -0.1085 | 0.0716 | 1.0000 | |
| comserd | 0.1736 | -0.0619 | 0.0522 | 0.1075 | 0.0401 | -0.0228 | 0.5236 | 1.0000 |

## Transformation

We have transformed the variables to have an easier understanding of the model due to their units as well as standardize the variables so that we don't have bias in the model.

- The dependent variable *mrall* (Traffic Fatality Rate) has been multiplied by 10,000 to show the number of traffic deaths every 10,000 people. Since the dependent variable was a vehicle fatality rate given per every 10,000 people, now our new dependent variable will represent the expected number of traffic death per 10,000 people instead of the rate.
- We performed a natural log of *perinc* (Per Capita Personal Income ($)) because it is an income variable and its effect falls over time and the distribution is right skewed (Figure 2).
- We also reformed the *vmiles* (Avg. Mile per Driver) in units of thousands to interpret the data more simply.
- We have additionally considered the variable jaild (Mandatory Jail Sentence) or *comserd* (Mandatory Community Service) to see how it affects the model as they had very high correlation.

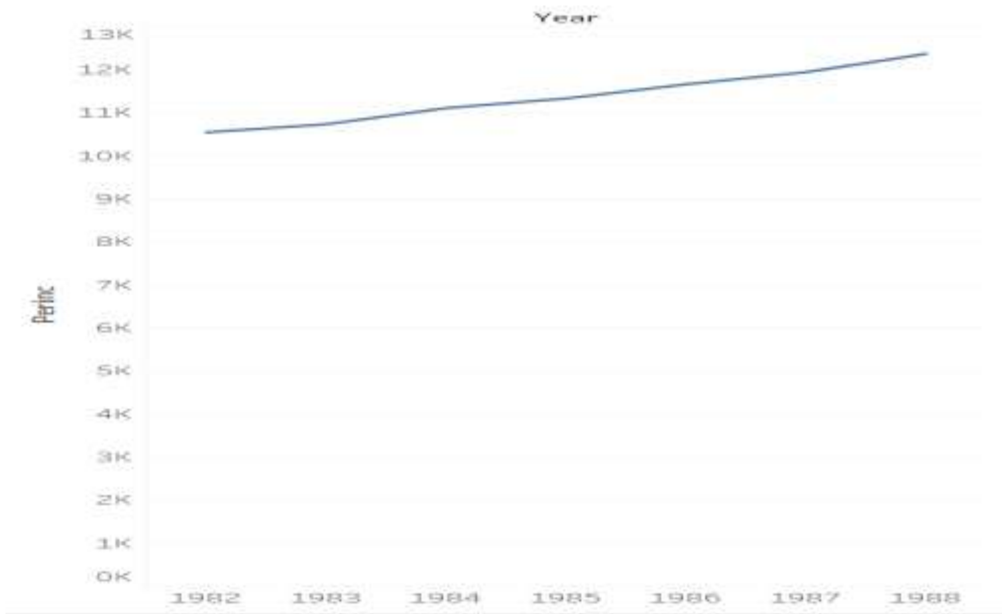


Figure 2

## Regression Analysis

Initiating the analysis with a dry regression model we found many variables to be significant but the overall model has a low poor R-square value. Based on the dataset which is a panel data we should run a panel data model because there are exactly same number of years with each state's data i.e. it is a balanced panel data. Based on Hausman test (Figure 3) we decide whether we should use fixed effect or random effect model.

```
. *HAUSMAN TEST

.
. hausman fixed random

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
      unrate |   -.0197098    -.0440225        .0243127        .0034582
    lnperinc |    .6784048    -.6241116        1.302516        .2030093
     beertax |   -.3647406     .1094064       -.4741471        .1610578
        mlda |   -.0378514    -.0115529       -.0262986               .
      mmiles |   -.0035239     .0160744       -.0195983               .
     1.jaild |   -.0192456     .1211803       -.1404259        .0930877
   1.comserd |   -.0198587     -.064482        .0446233         .106008
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       97.91
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

Figure 3

After comparing whether fixed effect or random effect model (Model 2) is appropriate, we test if the time effect matters in this model. With a result of the time effects being jointly statistically significant, we decided to use state and time fixed effect from model 3 to 5.

## Model Comparison

| Explanatory Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| State Unemployment Rate (%) | 0.0046 (0.0127) | -0.0197 (0.0116) | -0.0625** (0.3632) | | -0.0625** (0.0111) | -0.0630** (0.0110) |
| Per Capita Personal Income (Logarithmic) | -1.5492** (0.2130) | 0.6784 (0.3760) | 1.7866** (0.3632) | | 1.7883** (0.3623) | 1.8103** (0.3581) |
| Tax on Case of Beer ($) | 0.0806 (0.057) | -0.3647 (0.1952) | -0.4597** (0.1668) | -0.6512** (0.2003) | -0.4599** (0.1665) | -0.4604** (0.1633) |
| Minimum Legal Drinking Age (years) | -0.0079 (0.0274) | -0.0379 (0.0199) | -0.0019 (0.0178) | 0.0199 (0.0214) | -0.0018 (0.0178) | |
| Avg. Mile per Driver (1000 miles) | 0 .1391** (0.0176) | -0.0035 (0.0101) | 0.0090 (0.0088) | 0.0153 (0.0106) | 0.0090 (0.0087) | |
| Mandatory Jail Sentence (Indicator) | 0 .1405* (0.0661) | -0.0192 (0.1403) | 0.0140 (0.1203) | -0.0170 (0.1452) | | |
| Mandatory Community Service (Indicator) | 0.2056** (0.0754) | -0.0199 (0.1620) | 0.0330 (0.1380) | 0.1315 (0.1664) | | |
| Mandatory Jail Sentence or Mandatory Community Service | | | | | 0.0469 (0.0685) | |
| _Intercept | 15.7079 (2.1942) | -3.2771 (3.5598) | -14.02608 (3.4891) | 1.9149 (0.4488) | -14.0412 (3.4800) | -14.2148 (3.4580) |
| State Effect | No | Yes | Yes | Yes | Yes | Yes |
| Time Effect | No | No | Yes | Yes | Yes | Yes |
| Adjusted R-Square | 0.4268 | 0.8969 | 0.9259 | 08917 | 0.9262 | 0.9266 |
| R-Square | 0.4388 | 0.9136 | 0.9392 | 0.9105 | 0.9392 | 0.9389 |

* indicates the coeff. are significant at 99% ignificance/ ** indicates the coeff. are at 95% significance

## Regression Model 3

Based on this regression model, we discovered that *unrate*, *lnperinc* (natural log of personal income per capita), and *beertax* are significant at 1% while other variables are not that significant.

**Inference**

- When unemployment rate goes up by 1%, the number of traffic death will likely decrease by 0.0625 per 10000.
- When the personal income increase by 1%, the number of traffic death will likely increase by 0.0179 per 10000.
- When the tax on case of beer increase by 1 dollar, the number of traffic death will likely decrease by 0.46 per 10000.
- When the minimum drinking age increase by 1 year, the number of traffic death will likely decrease by 0.0079 per 10000.
- When average miles per driver increase by 1000 miles in a year, the number of traffic death will likely increase by 0.009 per 10000.
- When the state has jail sentence mandatory, the number of traffic death will likely increase by 0.14 per 10000.

- When the state has community service mandatory, the number of traffic death will likely increase by 0.03 per 10000. We decided to change the input variables to get a better estimate model.

**Regression Model 4**

Based on this regression model, we deleted *unrate* and *lnperinc* variables which represent economic conditions in the model. The output shows that the coefficient of beer tax changes from 0.46 to 0.65. However, there is no significant change in other variables.

**Inference**

- We can infer that the beer tax will make a bigger effect on fatalities because the effect of economic condition is added into the beer tax.
- When compared with the model 3, we conclude that the two variables unemployment rate and real income per capita should persist in the regression model.

**Regression Model 5**

Based on this regression model, we take the joint effect of mandatory jail sentence or mandatory community service into consideration, based on the result found from correlation. We define a new variable *jailcoms* (not shown on the table) to describe the two binary punishment variables. The new variable measures whether there are any punishment laws. Either one of these two punishment variables are included in the laws, the new variable will tell us there is punishment when the people are drunk driving. When compared with model D, model E does not change much in the coefficients of the other five variables. Hence, model E is not sensitive to the joint effect of two punishments.

**Regression Model 6**

In model F, we keep these variables which are statistically significant in model D. We found the coefficients of these three variables increase a little because they include the effect the other four variables in model D. Three variables in model F stayed significant and model F has the highest R-square value.

# Conclusion

We conclude that drunk driving laws affect traffic deaths. Our models show that unemployment rate, personal income, and beer tax variables are significant variables that affect the traffic death. Unemployment rate and beer tax decrease the number of traffic deaths, and personal income increase the number of traffic deaths. When tax on case of beer increased by $1, it decreases the number of traffic death by 0.45 per 10,000 people. Based on this result, we clearly observe that it is likely to decrease the number of traffic deaths by increasing the beer tax. In addition, mandatory jail sentence and mandatory community service were not significant. To reduce the traffic deaths by altering drunk driving laws, any laws that affects the increase in beer tax would reduce the deaths caused by traffic.

However, we are not able to conclude based on our model because there are high possibilities that biased omitted variables may exist. Such limitation would give us the imperfect model, and it is impossible to find a perfect real-life model. There are many other factors that could cause the traffic deaths that were not present in the dataset. Signs of the coefficients were observed as expected. As mentioned above, increasing the beer tax will reduce the traffic deaths. We could also conclude that other two significant variables, unemployment rate and personal income, affect the result of traffic deaths. However, they are not drunk driving laws, which implies that they are not desired explanatory variables that we were looking for.

# Appendix A

| | spircons | unrate | perinc | beertax | sobapt | mormon | mlda | dry | yngdrv | vmiles | jaild | comserd | allmort | mrall | allnite | aidall | mraidall | pop | pop1517 | pop1820 | pop2124 | miles | gspch |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| spircons | 1.00 | | | | | | | | | | | | | | | | | | | | | | |
| unrate | -0.24 | 1.00 | | | | | | | | | | | | | | | | | | | | | |
| perinc | 0.45 | -0.55 | 1.00 | | | | | | | | | | | | | | | | | | | | |
| beertax | -0.09 | 0.05 | -0.40 | 1.00 | | | | | | | | | | | | | | | | | | | |
| sobapt | -0.29 | 0.26 | -0.47 | 0.63 | 1.00 | | | | | | | | | | | | | | | | | | |
| mormon | -0.18 | -0.01 | -0.22 | 0.00 | -0.15 | 1.00 | | | | | | | | | | | | | | | | | |
| mlda | -0.08 | -0.26 | 0.20 | -0.06 | 0.06 | 0.01 | 1.00 | | | | | | | | | | | | | | | | |
| dry | -0.27 | 0.26 | -0.34 | 0.18 | 0.57 | -0.09 | 0.14 | 1.00 | | | | | | | | | | | | | | | |
| yngdrv | -0.06 | 0.38 | -0.47 | 0.25 | 0.17 | 0.21 | -0.28 | 0.06 | 1.00 | | | | | | | | | | | | | | |
| vmiles | -0.06 | -0.28 | -0.08 | 0.14 | 0.14 | 0.00 | 0.06 | -0.08 | -0.05 | 1.00 | | | | | | | | | | | | | |
| jaild | -0.05 | 0.14 | -0.15 | -0.04 | -0.08 | 0.22 | -0.11 | -0.21 | 0.01 | 0.07 | 1.00 | | | | | | | | | | | | |
| comserd | 0.14 | -0.06 | 0.05 | 0.11 | -0.04 | 0.25 | 0.04 | -0.19 | 0.04 | -0.02 | 0.52 | 1.00 | | | | | | | | | | | |
| allmort | -0.10 | 0.10 | 0.22 | 0.07 | 0.17 | -0.15 | 0.04 | 0.08 | -0.15 | -0.13 | -0.21 | -0.02 | 1.00 | | | | | | | | | | |
| mrall | -0.06 | 0.18 | -0.50 | 0.31 | 0.44 | 0.08 | -0.09 | 0.13 | 0.23 | 0.40 | 0.28 | 0.17 | -0.05 | 1.00 | | | | | | | | | |
| allnite | -0.09 | 0.14 | 0.23 | 0.01 | 0.13 | -0.17 | 0.00 | 0.05 | -0.12 | -0.16 | -0.22 | -0.05 | 0.98 | -0.10 | 1.00 | | | | | | | | |
| aidall | -0.14 | 0.17 | 0.11 | 0.09 | 0.24 | -0.17 | -0.02 | 0.13 | -0.06 | -0.12 | -0.16 | -0.03 | 0.94 | -0.01 | 0.96 | 1.00 | | | | | | | |
| mraidall | -0.12 | 0.28 | -0.54 | 0.29 | 0.46 | -0.06 | -0.16 | 0.22 | 0.30 | 0.21 | 0.23 | 0.06 | -0.05 | 0.75 | -0.05 | 0.14 | 1.00 | | | | | | |
| pop | -0.07 | 0.09 | 0.35 | -0.08 | 0.01 | -0.16 | 0.06 | 0.03 | -0.21 | -0.26 | -0.25 | -0.10 | 0.95 | -0.27 | 0.94 | 0.85 | -0.24 | 1.00 | | | | | |
| pop1517 | -0.07 | 0.13 | 0.33 | -0.07 | 0.03 | -0.17 | 0.04 | 0.05 | -0.18 | -0.27 | -0.26 | -0.13 | 0.93 | -0.28 | 0.93 | 0.85 | -0.23 | 1.00 | 1.00 | | | | |
| pop1820 | -0.07 | 0.13 | 0.32 | -0.06 | 0.03 | -0.17 | 0.04 | 0.05 | -0.16 | -0.27 | -0.26 | -0.12 | 0.94 | -0.28 | 0.94 | 0.86 | -0.22 | 1.00 | 1.00 | 1.00 | | | |
| pop2124 | -0.06 | 0.13 | 0.33 | -0.07 | 0.02 | -0.16 | 0.04 | 0.04 | -0.16 | -0.26 | -0.25 | -0.11 | 0.94 | -0.27 | 0.94 | 0.86 | -0.22 | 0.99 | 0.99 | 1.00 | 1.00 | | |
| miles | -0.09 | 0.05 | 0.33 | -0.02 | 0.08 | -0.16 | 0.08 | 0.04 | -0.21 | -0.12 | -0.25 | -0.09 | 0.98 | -0.20 | 0.96 | 0.89 | -0.18 | 0.97 | 0.96 | 0.96 | 0.97 | 1.00 | |
| gspch | 0.27 | -0.43 | 0.39 | 0.11 | -0.02 | -0.05 | 0.20 | 0.01 | -0.33 | 0.00 | -0.06 | 0.09 | 0.12 | -0.18 | 0.08 | 0.05 | -0.27 | 0.13 | 0.11 | 0.12 | 0.12 | 0.15 | 1.00 |

# Code File

```
*Summarize the variables to get the descriptive statistics
Sum

*variable transformation
gen vfrall=mrall*10000
gen mmiles=vmiles/1000
gen lnperinc=ln(perinc)

*Correlation Test
corr mrall unrate perinc beertax mlda vmiles jaild comserd

*dry Regression model 1
reg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd

*PANEL DATA (Fixed effect State fixed) model 2
xtset state year
xtreg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd, fe
estimates store fixed
areg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd, absorb(state)

*PANEL DATA (Random effect)
xtreg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd, re
estimates store random

*HAUSMAN TEST
hausman fixed random

*PANEL DATA (Fixed effectState and Time fixed) model 3
xtreg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd i.year ,fe
testparm i.year
areg vfrall unrate lnperinc beertax mlda mmiles i.jaild i.comserd i.year, absorb(state)

*NO ECONOMIC CONDITION model 4
xtreg vfrall beertax mlda mmiles i.jaild i.comserd i.year, fe
areg vfrall beertax mlda mmiles i.jaild i.comserd i.year, absorb(state)

*JAIL OR COMMUNITY model 5
gen jailcoms=1jaild*
comserd
xtreg vfrall unrate lnperinc beertax mlda mmiles jailcoms i.year, fe
areg vfrall unrate lnperinc beertax mlda mmiles jailcoms i.year, absorb (state)

*SIGNIFICANT VARIABLES model 6
xtreg vfrall unrate lnperinc beertax i.year, fe
areg vfrall unrate lnperinc beertax i.year, absorb (state)
```