Replication on "Alcohol Policies and Highway Vehicle Fatalities"

Brandon Tao
STAT 463: Exploratory Data Analysis
April 26, 2019

This project focus on the study of the impact of beer taxes and a variety of alcohol-control policies on motor vehicle fatality rates. Special attention is paid to omitted variables biases resulting from failing to adequately control for grassroots efforts to reduce drunk driving, the enactment of other laws which simultaneously operate to reduce highway fatalities.[1]

Regression using panel data may mitigate omitted variable bias when there is no information on variables that correlate with both the regressors of interest and the independent variable and if these variables are constant in the time dimension.

For applications we make use of the data set Fatalities[2] which is a panel data set reporting annual state level observations on U.S. traffic fatalities for the period 1982 through 1988. The applications analyze if there are effects of alcohol taxes and drunk driving laws on road fatalities and, if present, how strong these effects are.

We find that the data set consists of 336 observations on 34 variables. Notice that the variable state is a factor variable with 48 levels (one for each of the 48 contiguous federal states of the U.S.). The variable year is also a factor variable that has 7 levels identifying the time period when the observation was made. This gives us 7×48=336 observations in total. Since all variables are observed for all entities and over all time periods, the panel is balanced. (Data Description See **Appendix A**)

Simple regressions

I first estimate simple regressions using data for years 1982 to 1988 that model the relationship between beer tax (adjusted for 1988 dollars) and the traffic fatality rate, measured as the number of fatalities per 10000 inhabitants. Afterwards, I plot the data and add the corresponding estimated regression functions.

[1] Ruhm, C. (1996). Alcohol Policies and Highway Vehicle Fatalities. *Journal of Health Economics* 15(4): 435-454.

[2] Fatalities.rda, https://github.com/cran/AER/blob/master/data/Fatalities.rda?raw=true (4/15/2019 Date Accessed)

Figure (1): Traffic fatality rates and Beer Taxes separated by year (1982-1988)

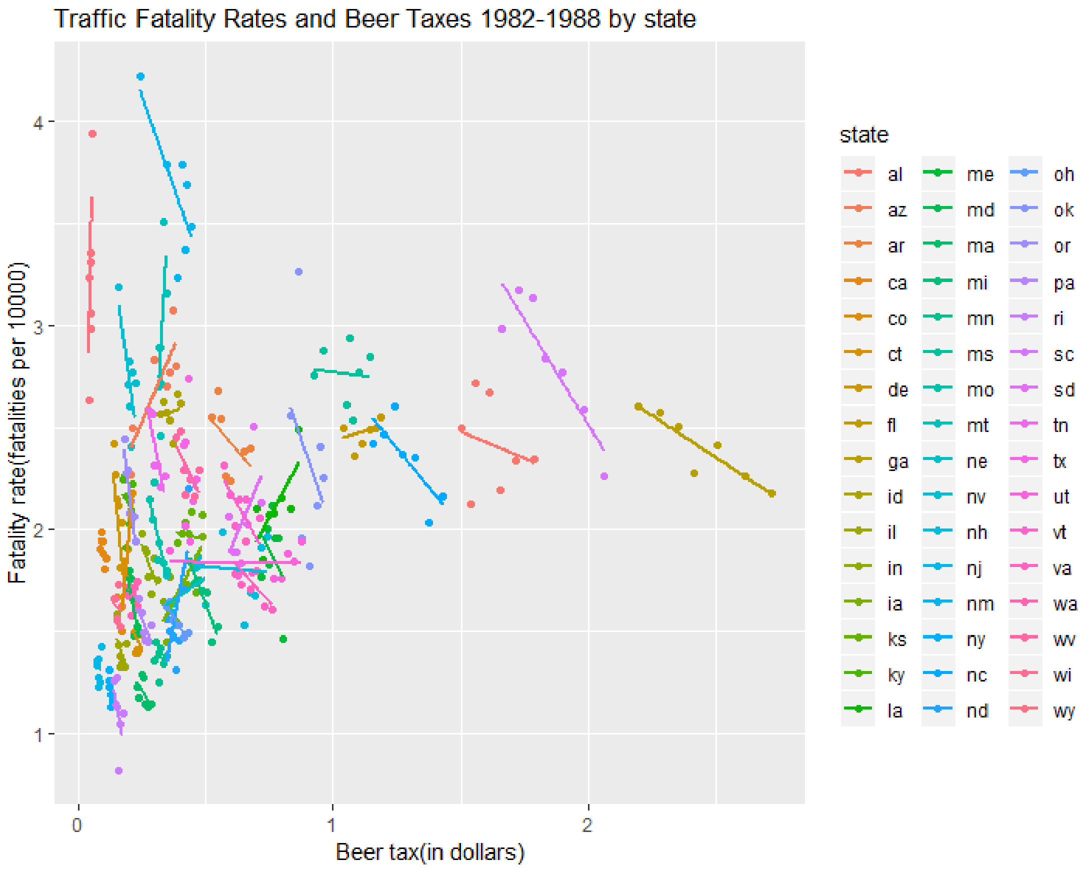
I also selected 1982 and 1988 as specific year in estimated coefficient for convenience.

The estimated regression functions are

$$\widehat{FatalityRate} = \mathbf{2.01} + \mathbf{0.15} \times BeerTax(1982\ data),$$

$$\widehat{FatalityRate} = \mathbf{1.86} + \mathbf{0.44} \times BeerTax(1988\ data)$$

In the plot above, each point represents observations of beer tax and fatality rate for a given respective year. The regression results first indicate a positive relationship between the beer tax and the fatality rate. However, the estimated coefficient on beer tax for the 1988 data is almost three times as large as for the 1982 data set. This is contrary to our expectations: alcohol taxes are supposed to lower the rate of traffic fatalities. This is possibly due to omitted variable bias, since models do not include any covariates, e.g., economic conditions, state effect.



As I replot the data separated each observation for given state, I discover many subgroups that's far away from the main scatter even with different trend. This could be corrected for using a multiple regression approach. However, this cannot account for omitted unobservable factors that differ from state to state but can be assumed to be constant over the observation span, e.g., the populations' attitude towards drunk driving. In the next model, panel data allow us to hold such factors constant.

Panel Data with "Before and After"

Suppose there are only T=2 time periods t=1982,1988. This allows us to analyze differences in changes of the the fatality rate from year 1982 to 1988. We start by considering the population regression model

$$\text{FatalityRate}_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{it}$$

where the Z_i are state specific characteristics that differ between states but are constant over time. For t=1982 and t=1988 we have

$$\text{FatalityRate}_{i1982} = \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982}$$
$$\text{FatalityRate}_{i1988} = \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988}$$

We can eliminate the Z_i by regressing the difference in the fatality rate between 1988 and 1982 on the difference in beer tax between those years:

$$\text{FatalityRate}_{i1988} - \text{FatalityRate}_{i1982} = \beta_1(BeerTax_{i1988} - BeerTax_{i1982}) + u_{i1988} - u_{i1982}$$

This regression model yields an estimate for$\beta_1$robust a possible bias due to omission of theZ_i, since these influences are eliminated from the model.

Table(a):

The estimated coefficient:

```
t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.072037   0.065355 -1.1022 0.276091
diff_beertax -1.040973   0.355006 -2.9323 0.005229 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note: the OLS estimate of the parameter of interest β_1 (from R)

Including the intercept allows for a change in the mean fatality rate in the time between 1982 and 1988 in the absence of a change in the beer tax.

We obtain the OLS estimated regression function

$$\widehat{\text{FatalityRate}_{i1988} - \text{FatalityRate}_{i1982}} = \mathbf{-0.072 - 1.04} \times (BeerTax_{i1988} - BeerTax_{i1982})$$

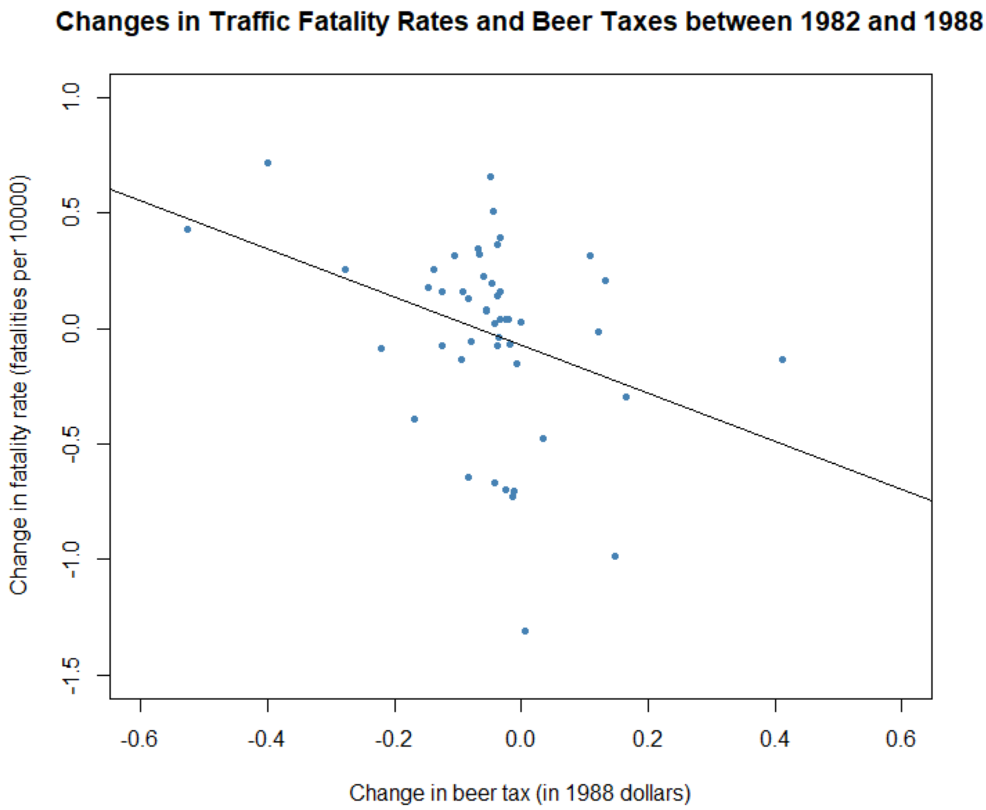


Figure (2): Changes in Traffic fatality Rates and Beer Taxes between 1982-1988

The estimated coefficient on beer tax is now negative and significantly different from zero at 5%. Its interpretation is that raising the beer tax by 1$causes traffic fatalities to decrease by 1.04 per 10000 people. This is rather large as the average fatality rate is approximately 2 persons per 10000 people.
However, this outcome is likely to be a consequence of omitting factors in the single-year regression that influence the fatality rate and are correlated with the beer tax and change over time. We need to be more careful and control for such factors before drawing conclusions about the effect of a raise in beer taxes.

Fixed Effects Regression

A method that allows to use data for more than T=2 time periods and enables us to add control variables is the fixed effects regression approach.
The alcohol regulations controlled could be correlated with related policies, resulting in possible omitted variables bias.
Consider the panel regression model

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}$$

We aim to estimate $\beta_1$the effect on Y_i of a change in X_i holding constant Z_i, where Z_i is the state policy factor that constant over time.

Letting $\alpha_i = \beta_0 + \beta_2 X_{it}$, formula can be rewritten as

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$$

And the averages equation is

$$\bar{Y} = \alpha_i + \beta_1 \overline{X_{it}} + \overline{u_{it}}$$

Subtract averages equation from the last equation, we obtain change in fatality rate

$$Y_{it} - \bar{Y} = \beta_1 (X_{it} - \overline{X_{it}}) + (u_{it} - \overline{u_{it}})$$

$$\boldsymbol{\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}}$$

The simple fixed effects model is a regression of the traffic fatality rate on beer tax and 48 binary regressors — one for each federal state. We use it to obtain state specific averages of the fatality rate and the beer tax.

The fixed effects model can be generalized to contain more than just one determinant of Y that is correlated with X and changes over time. Note that the variation in the α_i come from Z_i.

The simple fixed effects model for estimation of the relation between traffic fatality rates and the beer taxes is

$$\text{FatalityRate}_{it} = \beta_1 BeerTax_{it} + StateFixedEffects + u_{it}$$

Table(b):

the OLS estimate of the parameter of interest β_1:

```
Call:
lm(formula = fatal_rate ~ beertax + state - 1, data = Fatalities)

Coefficients:
beertax  stateal  stateaz  statear  stateca  stateco  statect  statede  statefl
-0.6559   3.4776   2.9099   2.8227   1.9682   1.9933   1.6154   2.1700   3.2095
statega  stateid  stateil  statein  stateia  stateks  stateky  statela  stateme
 4.0022   2.8086   1.5160   2.0161   1.9337   2.2544   2.2601   2.6305   2.3697
statemd  statema  statemi  statemn  statems  statemo  statemt  statene  statenv
 1.7712   1.3679   1.9931   1.5804   3.4486   2.1814   3.1172   1.9555   2.8769
statenh  statenj  statenm  stateny  statenc  statend  stateoh  stateok  stateor
 2.2232   1.3719   3.9040   1.2910   3.1872   1.8542   1.8032   2.9326   2.3096
statepa  stateri  statesc  statesd  statetn  statetx  stateut  statevt  stateva
 1.7102   1.2126   4.0348   2.4739   2.6020   2.5602   2.3137   2.5116   2.1874
statewa  statewv  statewi  statewy
 1.8181   2.5809   1.7184   3.2491
```

Note: a regression of the traffic fatality rate on beer tax and 48 binary regressors — one for each federal state.

Table(c):
Estimate the regression:

```
Call:
lm(formula = fatal_rate ~ beertax - 1, data = Fatalities_demeaned)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.58696 -0.08284 -0.00127  0.07955  0.89780 

Coefficients:
        Estimate Std. Error t value Pr(>|t|)    
beertax  -0.6559     0.1739  -3.772 0.000191 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1757 on 335 degrees of freedom
Multiple R-squared:  0.04074,	Adjusted R-squared:  0.03788 
F-statistic: 14.23 on 1 and 335 DF,  p-value: 0.0001913
```

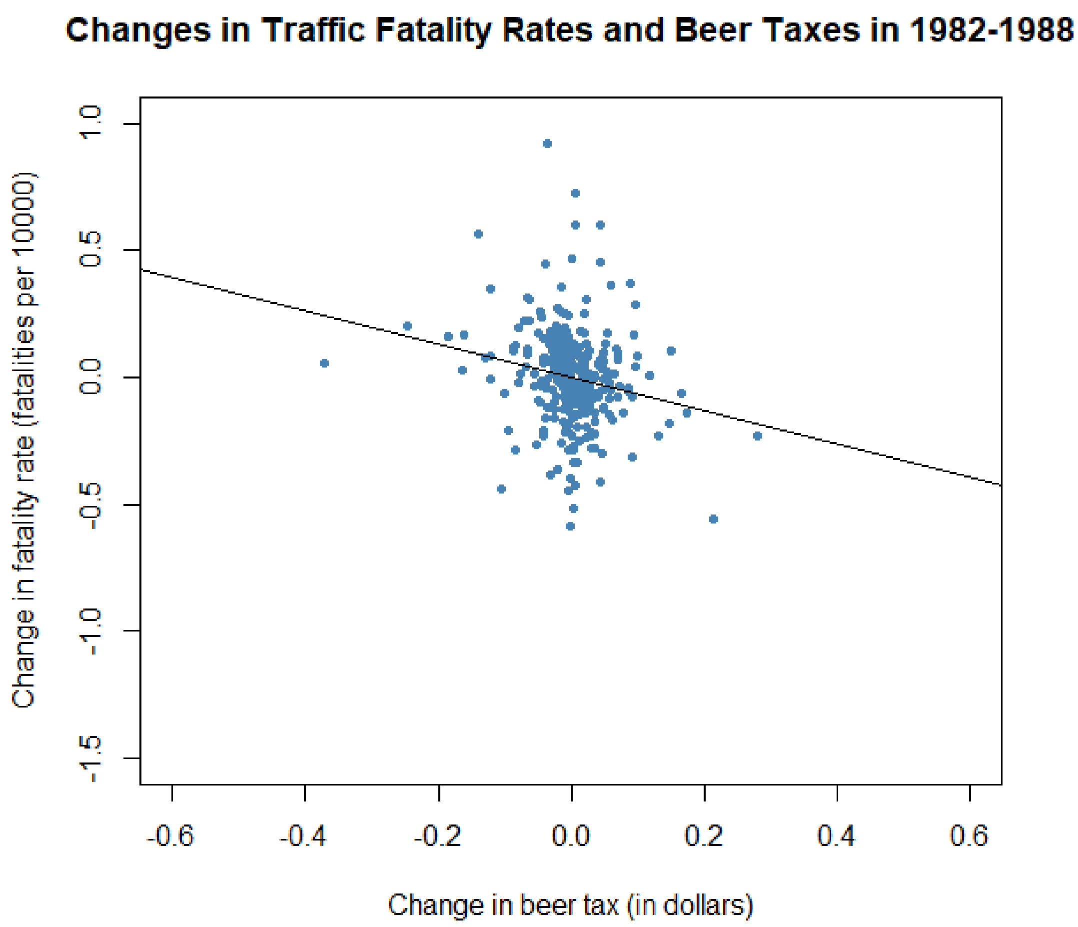
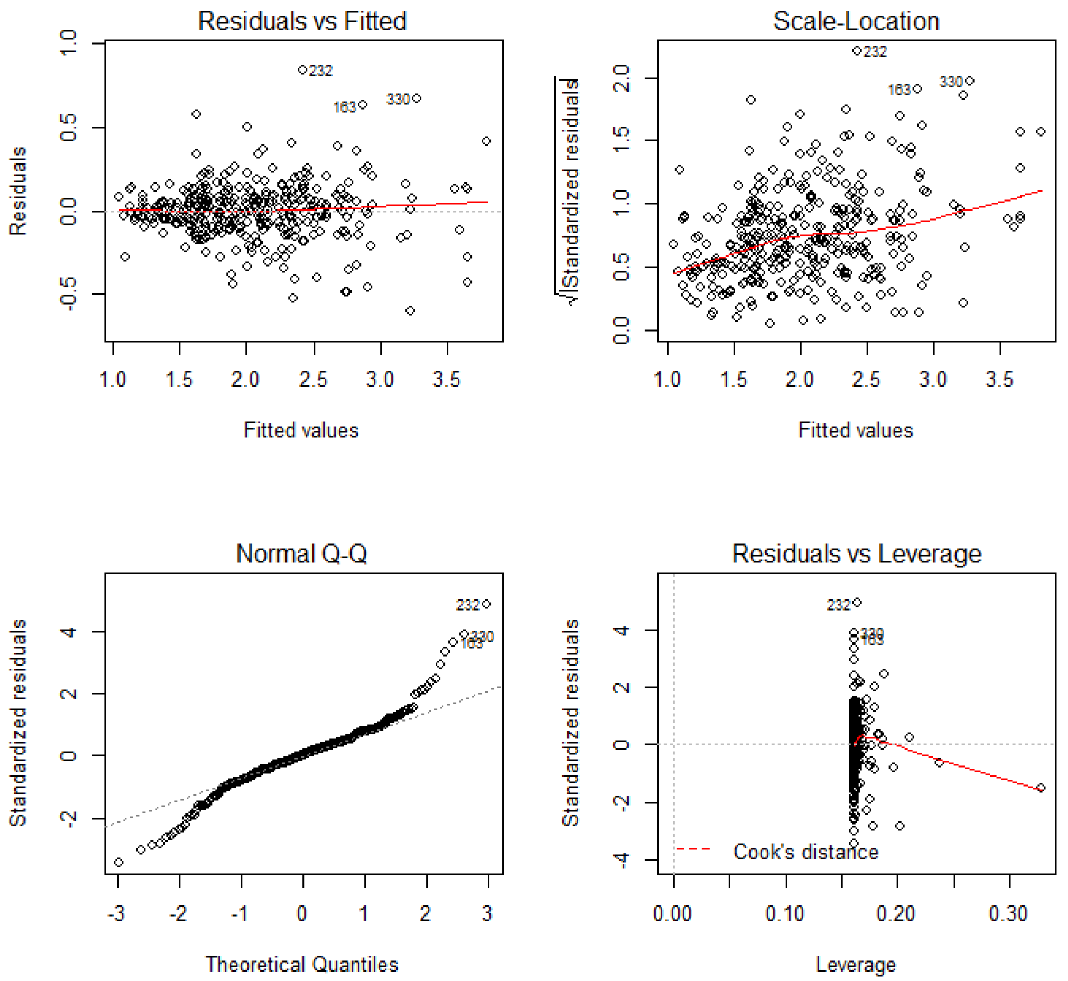


Figure (3): Changes in Traffic fatality Rates and Beer Taxes in 1982-1988 with state effects.

The estimated regression function is

$$FatalityRate = -\mathbf{0.66} \times BeerTax + StateFixedEffects$$

The coefficient on $BeerTax$ is negative and significant (small p). The interpretation is that the estimated reduction in traffic fatalities due to an increase in the real beer tax by 1$ is 0.66 per 10,000 people, which is still rather large.



A quick look at the diagnostic plots for the fixed effect model, outliers still existing and effect on the long tail QQ plot. Although including state fixed effects eliminates the risk of a bias due to omitted factors that vary across states but not over time, there's reason to suspect that there are other omitted variables that vary over time and thus cause a bias.

Regression continues with Time Fixed Effects

Controlling for variables that are constant across entities but vary over time can be done by including time fixed effects. The combined model allows to eliminate bias from unobservable that change over time but are constant over entities, and it controls for factors that differ across entities but are constant over time.

$$\text{FatalityRate} = \beta_1 BeerTax_{it} + StateEffects + TimeFixedEffects + u_{it}$$

using lm() function,

Table(d):

Estimated using the OLS algorithm that is implemented in R:

```
Call:
lm(formula = fatal_rate ~ beertax + state + year - 1, data = Fatalities)

Coefficients:
 beertax   stateal   stateaz   statear   stateca   stateco   statect   statede
-0.63998   3.51137   2.96451   2.87284   2.02618   2.04984   1.67125   2.22711
 statefl   statega   stateid   stateil   statein   stateia   stateks   stateky
 3.25132   4.02300   2.86242   1.57287   2.07123   1.98709   2.30707   2.31659
 statela   stateme   statemd   statema   statemi   statemn   statems   statemo
 2.67772   2.41713   1.82731   1.42335   2.04488   1.63488   3.49146   2.23598
 statemt   statene   statenv   statenh   statenj   statenm   stateny   statenc
 3.17160   2.00846   2.93322   2.27245   1.43016   3.95748   1.34849   3.22630
 statend   stateoh   stateok   stateor   statepa   stateri   statesc   statesd
 1.90762   1.85664   2.97776   2.36597   1.76563   1.26964   4.06496   2.52317
 statetn   statetx   stateut   statevt   stateva   statewa   statewv   statewi
 2.65670   2.61282   2.36165   2.56100   2.23618   1.87424   2.63364   1.77545
 statewy  year1983  year1984  year1985  year1986  year1987  year1988
 3.30791  -0.07990  -0.07242  -0.12398  -0.03786  -0.05090  -0.05180
```

The lm() functions converts factors into dummies automatically. Since we exclude the intercept by adding -1 to the right-hand side of the regression formula, lm() estimates coefficients for $n + (T - 1) = 48 + 6 = 54$ binary variables (6 year dummies and 48 state dummies).

Table(e):

The estimated coefficient on $BeerTax$.

```
t test of coefficients:

        Estimate Std. Error t value Pr(>|t|)  
beertax -0.63998    0.35015 -1.8277  0.06865 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimated regression function is

$$FatalityRate = \mathbf{-0.64} \times BeerTax + StateEffects + TimeFixedEffects$$

Base on results of both Fixed Effects and Time Fixed Effects Regression, comparatively the simple fixed effects model has smaller variance of estimated coefficient on β_1 and yield a higher significant (smaller p-value of 0.0002 compared to 0.069).

The result -0.66 is also close to the estimated coefficient for the regression model including only entity fixed effects. I conclude that the estimated relationship between traffic fatalities and the real beer tax is not affected by omitted variable bias due to factors that are constant over time.

Appendix A (variables used only)

state factor indicating state. (with 48 levels)
year factor indicating year. (with 7 levels)
beertax numeric. Tax on case of beer.
fatal numeric. Number of vehicle fatalities.
pop numeric. Population.

Note. A data frame containing 336 observations.
Traffic fatalities are from the US Department of Transportation Fatal Accident Reporting System.
The beer tax is the tax on a case of beer, which is an available measure of state alcohol taxes more generally.
Sourse: Ruhm, C. (1996). Alcohol Policies and Highway Vehicle Fatalities. Journal of Health Economics 15(4): 435-454.

Appendix B (R.code)

```
library(ggplot2)

load("C:/Users/enemy/OneDrive - George Mason University/Desktop/GMU/STAT 463/Final/Fatalities.rda")
data(Fatalities)

# obtain the dimension and inspect the structure
dim(Fatalities)
str(Fatalities)

# Simple regression

# define the fatality rate
Fatalities$fatal_rate <- Fatalities$fatal / Fatalities$pop * 10000

# subset the data
Fatalities1982 <- subset(Fatalities, year == "1982")
Fatalities1988 <- subset(Fatalities, year == "1988")
# estimate simple regression models using 1982 and 1988 data
fatal1982_mod <- lm(fatal_rate ~ beertax, data = Fatalities1982)
fatal1988_mod <- lm(fatal_rate ~ beertax, data = Fatalities1988)
library(stargazer)
coeftest(fatal1982_mod, vcov. = vcovHC, type = "HC1")

coeftest(fatal1988_mod, vcov. = vcovHC, type = "HC1")
```

```
# plot observations and add estimated regression line for 1982-1988 data

ggplot(Fatalities,aes(beertax,fatal_rate,color=year)) + geom_point() +
        geom_smooth(method="lm",se=F) +
        labs(x = 'Beer tax(in dollars)',y = 'Fatality rate(fatalities per 10000)',
                  title = 'Traffic Fatality Rates and Beer Taxes 1982-1988')
ggplot(Fatalities,aes(beertax,fatal_rate,color=state)) + geom_point() +
    geom_smooth(method="lm",se=F) +
    labs(x = 'Beer tax(in dollars)',y = 'Fatality rate(fatalities per 10000)',
              title = 'Traffic Fatality Rates and Beer Taxes 1982-1988 by state')

# Panel Data with ??Before and After??

# compute the differences
diff_fatal_rate <- Fatalities1988$fatal_rate - Fatalities1982$fatal_rate
diff_beertax <- Fatalities1988$beertax - Fatalities1982$beertax

# estimate a regression using differenced data
fatal_diff_mod <- lm(diff_fatal_rate ~ diff_beertax)

coeftest(fatal_diff_mod, vcov = vcovHC, type = "HC1")
# plot the differenced data
plot(x = diff_beertax,
          y = diff_fatal_rate,
          xlab = "Change in beer tax (in 1988 dollars)",
          ylab = "Change in fatality rate (fatalities per 10000)",
          main = "Changes in Traffic Fatality Rates and Beer Taxes between 1982 and 1988",
          xlim = c(-0.6, 0.6),
          ylim = c(-1.5, 1),
          pch = 20,
          col = "steelblue")

# add the regression line to plot
abline(fatal_diff_mod, lwd = 1.5)



# Fixed Effects Regression

fatal_fe_lm_mod <- lm(fatal_rate ~ beertax + state - 1, data = Fatalities)
fatal_fe_lm_mod

# obtain demeaned data
```

```
Fatalities_demeaned <- with(Fatalities,
                            data.frame(fatal_rate = fatal_rate - ave(fatal_rate, state),
                                       beertax = beertax - ave(beertax, state)))

# estimate the regression
summary(lm(fatal_rate ~ beertax - 1, data = Fatalities_demeaned))

# plot the demeaned data
plot(x = Fatalities_demeaned$beertax,
     y = Fatalities_demeaned$fatal_rate,
     xlab = "Change in beer tax (in dollars)",
     ylab = "Change in fatality rate (fatalities per 10000)",
     main = "Changes in Traffic Fatality Rates and Beer Taxes in 1982-1988",
     xlim = c(-0.6, 0.6),
     ylim = c(-1.5, 1),
     pch = 20,
     col = "steelblue")

abline(fatal_fe_lm_mod, lwd = 1.5)


# Regression with Time Fixed Effects

# estimate a combined time and entity fixed effects regression model

fatal_tefe_lm_mod <- lm(fatal_rate ~ beertax + state + year - 1, data = Fatalities)
fatal_tefe_lm_mod
lm(formula = fatal_rate ~ beertax + state + year - 1, data = Fatalities)
library(plm)
# set effect = "twoways" for inclusion of entity and time dummies.
fatal_tefe_mod <- plm(fatal_rate ~ beertax,
                      data = Fatalities,
                      index = c("state", "year"),
                      model = "within",
                      effect = "twoways")

coeftest(fatal_tefe_mod,vcov = vcovHC, type = "HC1")

#for presentation
presentation <- subset(Fatalities,select=c(year,state,beertax,fatal_rate))
par(mfcol=c(2,2))
plot(fatal_tefe_lm_mod)
```

Bibliography

Ruhm, C. (1996). Alcohol Policies and Highway Vehicle Fatalities. Journal of Health Economics 15(4): 435-454.

Fatalities.rda (Data)https://github.com/cran/AER/blob/master/data/Fatalities.rda?raw=true (4/15/2019 Date Accessed)

Kate L. Turabian, "The Chicago Manual of Style", *A Manual for Writers, Eighth Edition,* Date Accessed 4/20/2019. https://www.chicagomanualofstyle.org/turabian/citation-guide.html