Homework 3

Submission 3

Conor Mulligan

Third submission of homework 3.

Link to Github

Summarize The Data

1. Present a bar graph showing the proportion of states with a change in their cigarette tax in each year from 1970 to 1985.

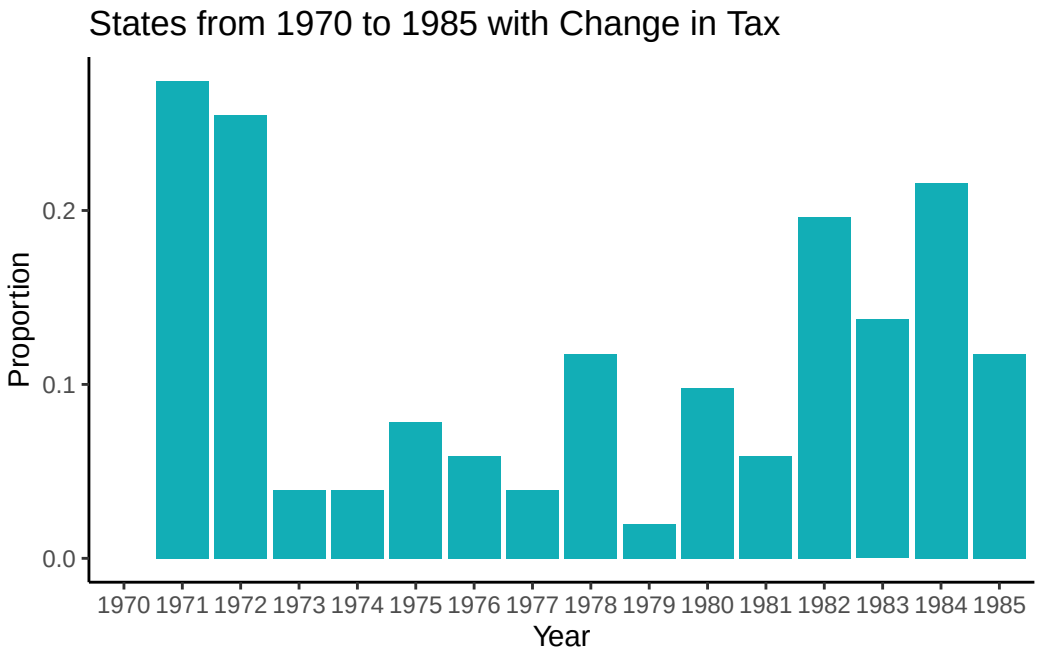

Figure 1: Proportion of States with Change in Their Cigarette Tax from 1970 to 1985

2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.

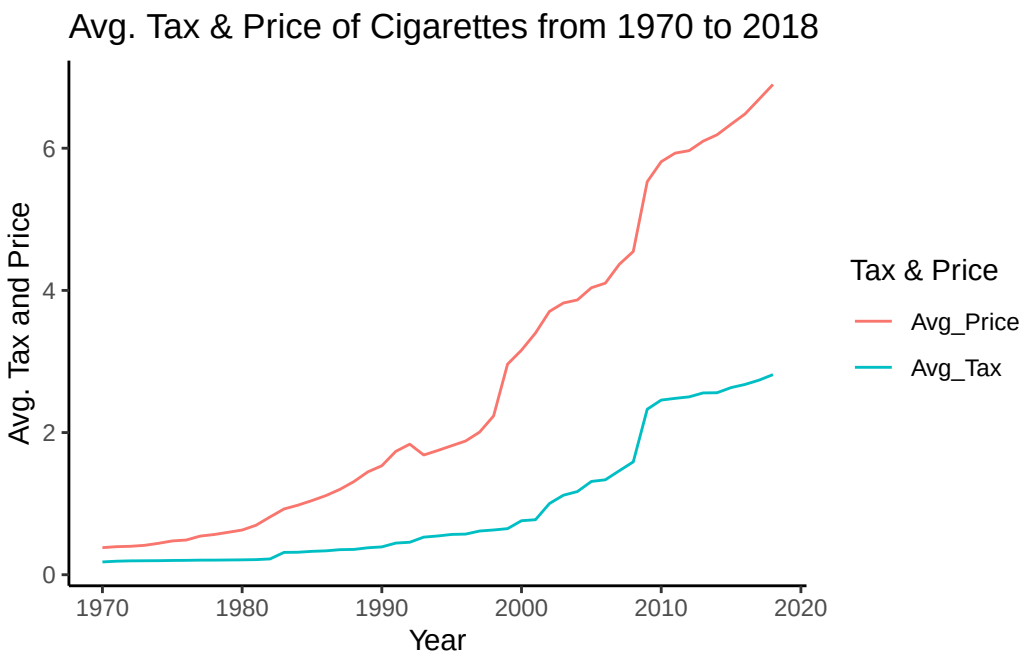

Figure 2: Average Tax in 2012 dollars on Cigarettes & Average Price of a Pack of Cigarettes from 1970 to 2018

3.Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

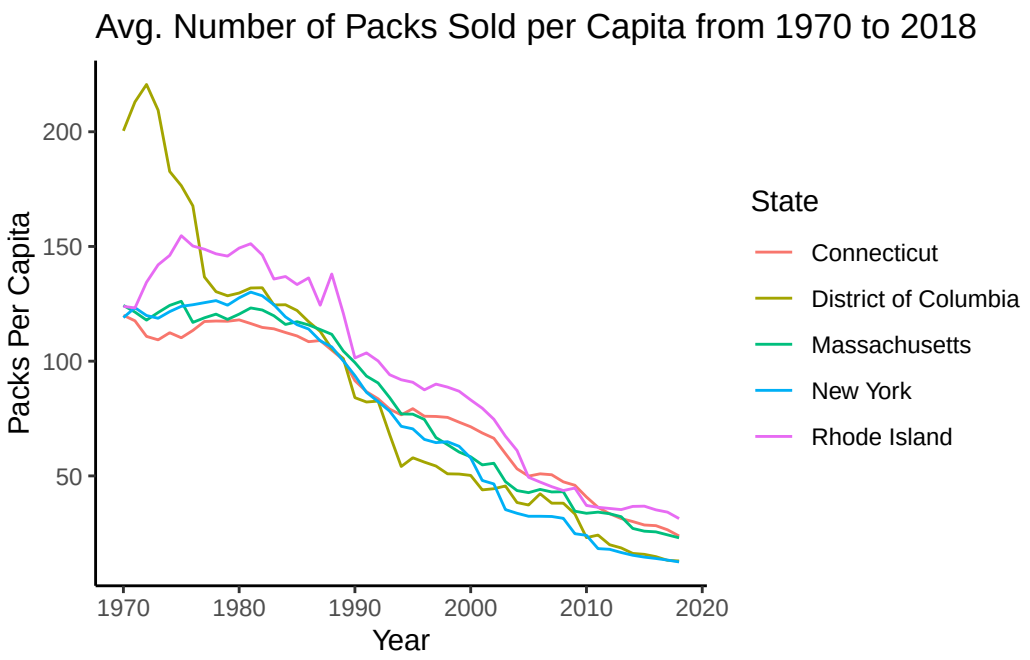

Figure 3: Average Number of Packs Sold per Capita among the Five States with the Highest Increases in Cigarette Prices from 1970 to 2018

4. Identify the 5 states with the lowest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

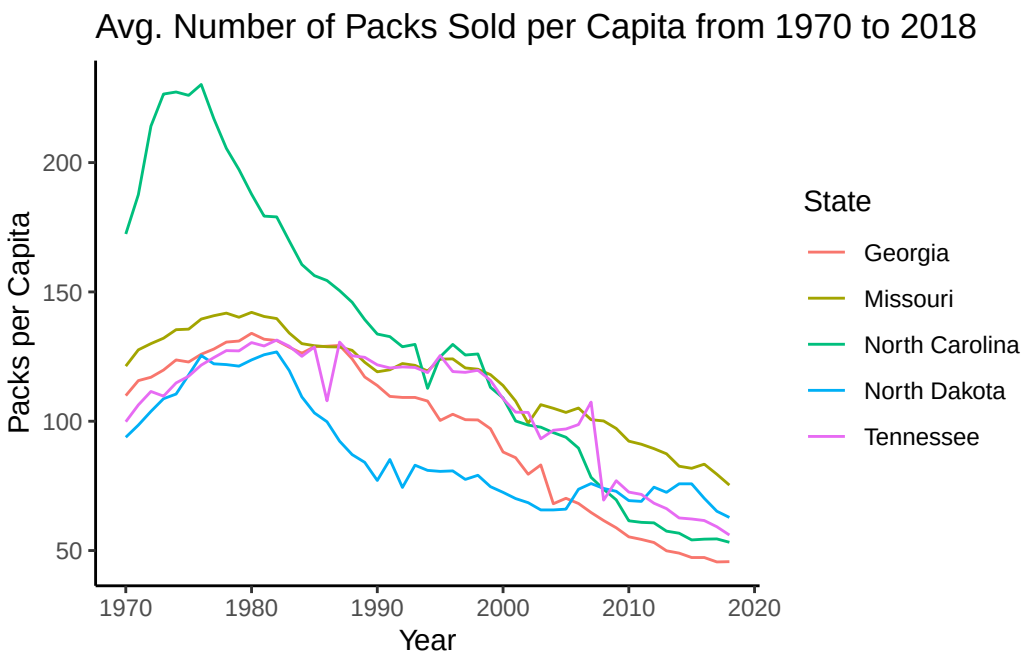

Figure 4: Average Number of Packs Sold per Capita among the Five States with the Lowest Increases in Cigarette Prices from 1970 to 2018

5.Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.

Looking at the graphs, the average packs sold decreased as price increased over time. States with lower price increases (the second of the top 5 graphs) saw less of a decrease in sales per capita. This makes sense given price did not increase as much, so sales stayed a bit higher than the 5 states that saw greatee price increases.

Estimate ATEs

Now let's work on estimating a demand curve for cigarettes. Specifically, we're going to estimate the price elasticity of demand for cigarettes. When explaining your findings, try to limit your discussion just to a couple of sentences.

6. Focusing only on the time period from 1970 to 1990, regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

```
OLS estimation, Dep. Var.: log_sales
Observations: 1,071
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  5.427381   0.029752 182.4238 < 2.2e-16 ***
log_price   -0.809438   0.038366 -21.0980 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.189214   Adj. R2: 0.293322
```

We see a coefficient of -0.809 which is the estimated price elasticity of demand. This should mean a 1% increase in cigarette price is associated with a decrease of about 0.809% in cigarette sales.

7. Again limiting to 1970 to 1990, regress log sales on log prices using the total (federal and state) cigarette tax (in dollars) as an instrument for log prices. Interpret your results and compare your estimates to those without an instrument. Are they different? If so, why?

```
TSLS estimation, Dep. Var.: log_sales, Endo.: log_price, Instr.: log_total_tax
Second stage: Dep. Var.: log_sales
Observations: 1,071
Standard-errors: IID
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.416797   0.054494  99.4017 < 2.2e-16 ***
fit_log_price -0.795524   0.071235 -11.1676 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.189226   Adj. R2: 0.293235
F-test (1st stage), log_price: stat = 436.8     , p < 2.2e-16 , on 1 and 1,069 DoF.
                  Wu-Hausman: stat =   0.053709, p = 0.816775, on 1 and 1,068 DoF.
```

The estimated coefficient is about -0.795, which means that there is an decrease in sales of about 0.795% per 1% increase in price.

8. Show the first stage and reduced-form results from the instrument.

```
OLS estimation, Dep. Var.: log_price
Observations: 1,071
Standard-errors: IID
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   0.839646   0.005421 154.8788 < 2.2e-16 ***
log_total_tax 0.260060   0.012443  20.9009 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.127093   Adj. R2: 0.289437
```

The first stage results show a coefficient of 0.260.

```
OLS estimation, Dep. Var.: log_sales
Observations: 1,071
Standard-errors: IID
               Estimate Std. Error   t value  Pr(>|t|)
(Intercept)    4.748839   0.009202 516.09178 < 2.2e-16 ***
log_total_tax -0.206884   0.021119  -9.79629 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.215714   Adj. R2: 0.081519
```

The coefficient for the second stage is -0.206 which is a different effect than the first stage result.This is expected as there is some difference when accounting for endogeneity.

9. Repeat questions 6-8 focusing on the period from 1991 to 2015.

9.1 Focusing only on the time period from 1991 to 2015, regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

```
OLS estimation, Dep. Var.: log_sales
Observations: 1,275
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  5.659955   0.036384 155.5601 < 2.2e-16 ***
log_price   -0.996814   0.024692 -40.3697 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.295775   Adj. R2: 0.561101
```

The coefficient is -0.997 which suggests a 1% increase in price leads to a 0.997% decrease in sales.

9.2 Again limiting to 1991 to 2015, regress log sales on log prices using the total (federal and state) cigarette tax (in dollars) as an instrument for log prices. Interpret your results and compare your estimates to those without an instrument. Are they different? If so, why?

```
TSLS estimation, Dep. Var.: log_sales, Endo.: log_price, Instr.: log_total_tax
Second stage: Dep. Var.: log_sales
Observations: 1,275
Standard-errors: IID
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.87986   0.040780 144.1859 < 2.2e-16 ***
fit_log_price -1.15008   0.027811 -41.3536 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.300218   Adj. R2: 0.547816
F-test (1st stage), log_price: stat = 5,503.6, p < 2.2e-16, on 1 and 1,273 DoF.
                   Wu-Hausman: stat =   191.5, p < 2.2e-16, on 1 and 1,272 DoF.
```

The coefficient of -1.15 suggests that there is a decrease of 1.15 percent per 1% increase in price.

9.3 Show the first stage and reduced-form results from the instrument.

```
OLS estimation, Dep. Var.: log_price
Observations: 1,275
Standard-errors: IID
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   1.315073   0.004386 299.8415 < 2.2e-16 ***
log_total_tax 0.513550   0.006922  74.1860 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.145512   Adj. R2: 0.811999

OLS estimation, Dep. Var.: log_sales
Observations: 1,275
Standard-errors: IID
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    4.367418    0.00844 517.4962 < 2.2e-16 ***
log_total_tax -0.590626    0.01332 -44.3396 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.28   Adj. R2: 0.606669
```

The coefficient of -0.591 means a 0.591 sales decrease for every 1 percent increase in price.

10. Compare your elasticity estimates from 1970-1990 versus those from 1991-2015. Are they different? If so, why?.

The estimated price elasticity of demand is negative in both instances. The second time-frame estimates less of a decrease in sales per price increase. This makes sense intuitively as regulations in later years likely meant that the demand elasticity of cigarettes decreased.