Homework 3

Submission 2

Conor Mulligan

Second 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.
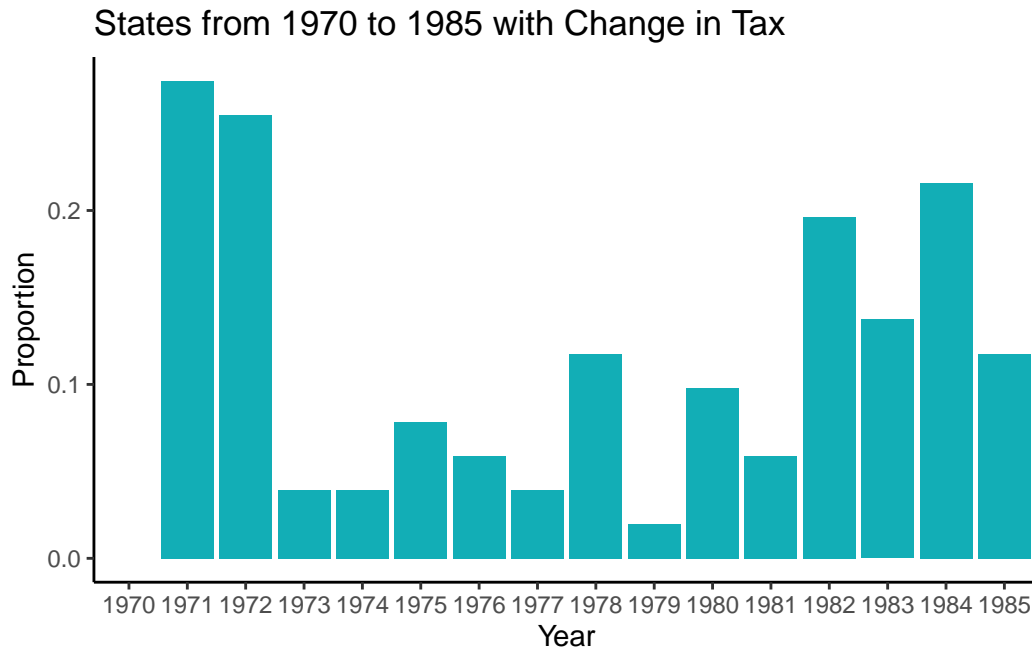
States from 1970 to 1985 with Change in Tax

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.
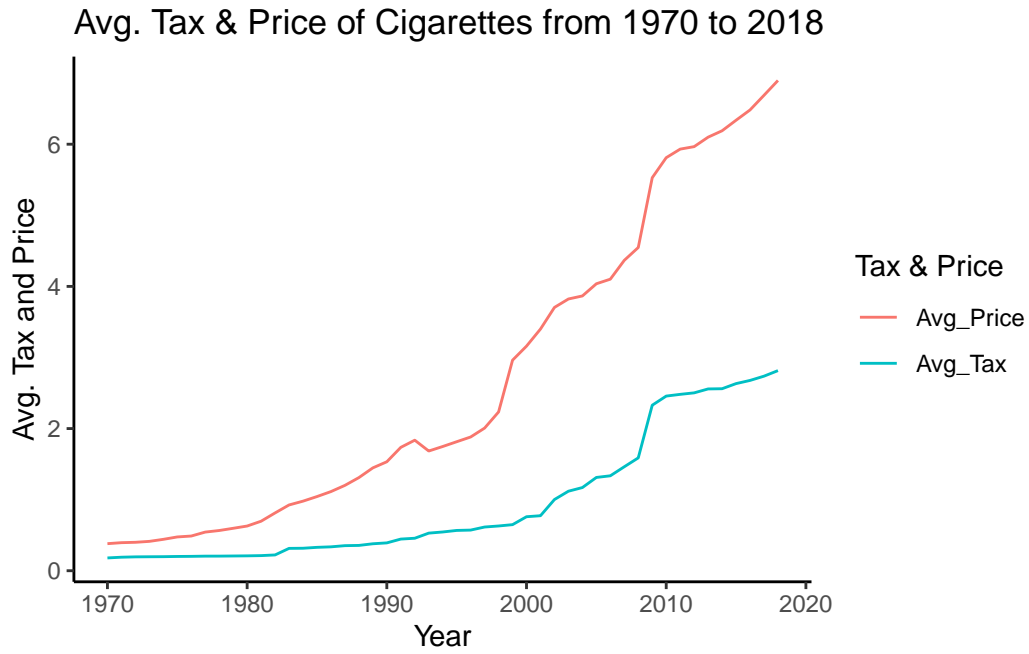
Avg. Tax & Price of Cigarettes from 1970 to 2018

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.

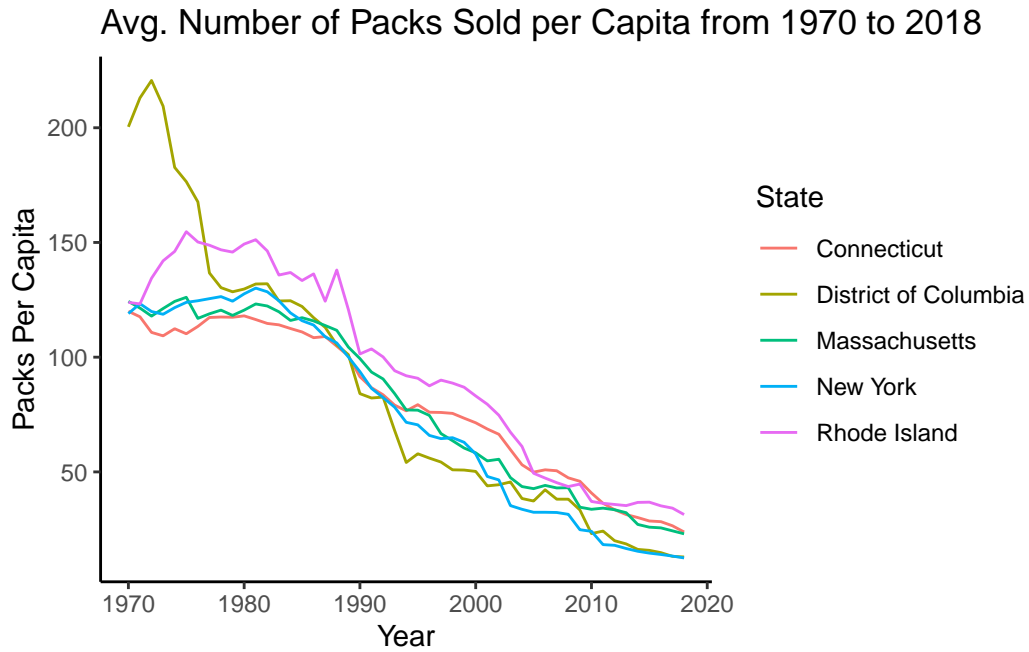Avg. Number of Packs Sold per Capita from 1970 to 2018

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.
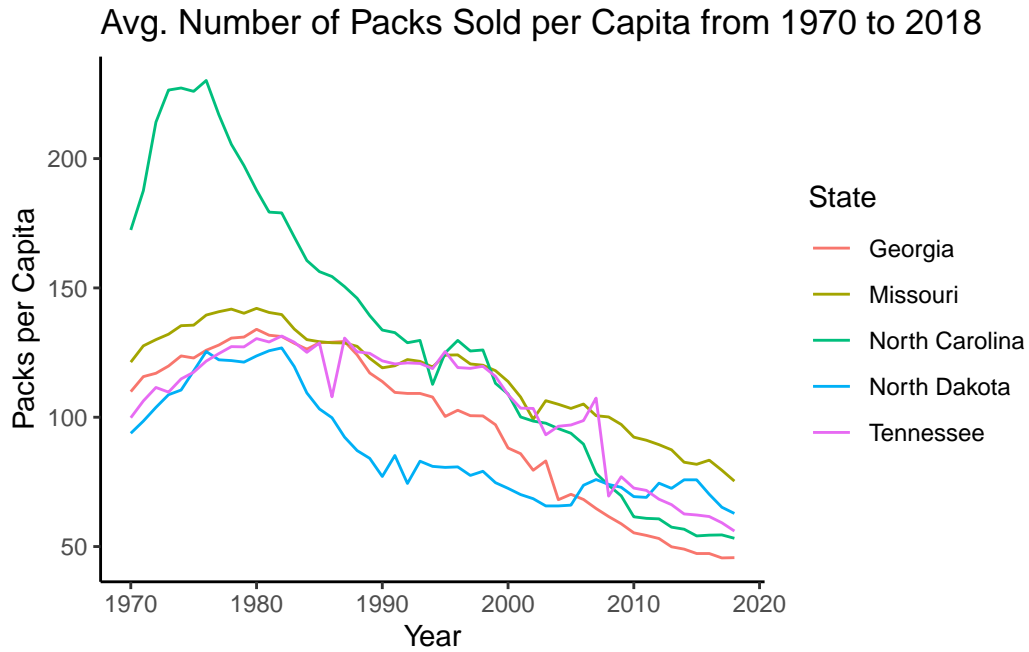

Figure 4: Average Number of Packs Sold per Capita among the Five States with the Lowest Increases in Cigarette Prices from 1970 to 2018

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)  4.750402   0.008116 585.3207 < 2.2e-16 ***
log_price   -0.171540   0.013829 -12.4039 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.210546   Adj. R2: 0.125
```

We see a coefficient of -0.171 which is the estimated price elasticity of demand. This should mean a 1% increase in cigarette price is associated with a decrease of about 0.171% 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)   4.991108   0.034106 146.34225  < 2.2e-16 ***
fit_log_price 0.502373   0.089837   5.59207 2.8482e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.377891   Adj. R2: -1.81869
F-test (1st stage), log_price: stat =  88.4, p < 2.2e-16, on 1 and 1,069 DoF.
                    Wu-Hausman: stat = 240.2, p < 2.2e-16, on 1 and 1,068 DoF.
```

The estimated coefficient is about 0.502, which means that there is an increase in sales of about .502% per 1% increase in price. This does not make much sense intuitively, so I may have done something wrong when estimating the value.

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)   -0.503589   0.020736 -24.28593 < 2.2e-16 ***
log_total_tax -0.411813   0.043812  -9.39948 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.447517   Adj. R2: 0.075474
```

The first stage results show a coefficient of -0.412, which means that a decrease of .412 percent occurs per 1% increase in cigarette price.

```
OLS estimation, Dep. Var.: log_sales
Observations: 1,071
Standard-errors: IID
           Estimate Std. Error   t value  Pr(>|t|)
(Intercept) 4.991108   0.019469 256.36475 < 2.2e-16 ***
           0.502373   0.051282   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 .502 which is a different effect than the first stage result, but the same as my answer for question 7. This is unexpected and may be a result of error as one might expect it to be negative.

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)  5.039485   0.022914 219.9340 < 2.2e-16 ***
log_price   -0.665626   0.017473 -38.0943 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.305314   Adj. R2: 0.532335
```

The coefficient is -0.665 which suggests a 1% increase in price leads to a 0.665% 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)   5.218896   0.026498 196.9549 < 2.2e-16 ***
fit_log_price -0.813109   0.020548 -39.5712 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.313741   Adj. R2: 0.506163
F-test (1st stage), log_price: stat = 4,111.4, p < 2.2e-16, on 1 and 1,273 DoF.
                  Wu-Hausman: stat =   280.6, p < 2.2e-16, on 1 and 1,272 DoF.
```

The coefficient of -0.813 suggests that there is a decrease of 0.813 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)   1.084827   0.006983 155.3561 < 2.2e-16 ***
log_total_tax 0.726380   0.011328  64.1199 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.238127   Adj. R2: 0.763389


OLS estimation, Dep. Var.: log_sales
Observations: 1,275
Standard-errors: IID
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   5.218896   0.023648 220.6885 < 2.2e-16 ***
             -0.813109   0.018338 -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 first stage regression shows a coefficient of 0.726, or a 0.726% increase in price for each 1% increase in tax.

The second stage coefficient of -0.813 means a 0.813 percent sales decrease for every 1 percent increase in price. This is different than the first reduced form output which was positive, meaning the difference is substantial. It is also the same as the second calculation from part 9, which is interesting (could be wrong).

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 positive in the first and negative in the second. The negative elasticity in the second model suggests that a decrease in price means an increase in sales. I likely did something wrong to have the first elasticity be positive, as it does not make sense intuitively this way. Some of the coefficients in the second stage regressions matched earlier calculations, so unsure if thst is also wrong.