Homework 3

Research in Health Economics, Spring 2025

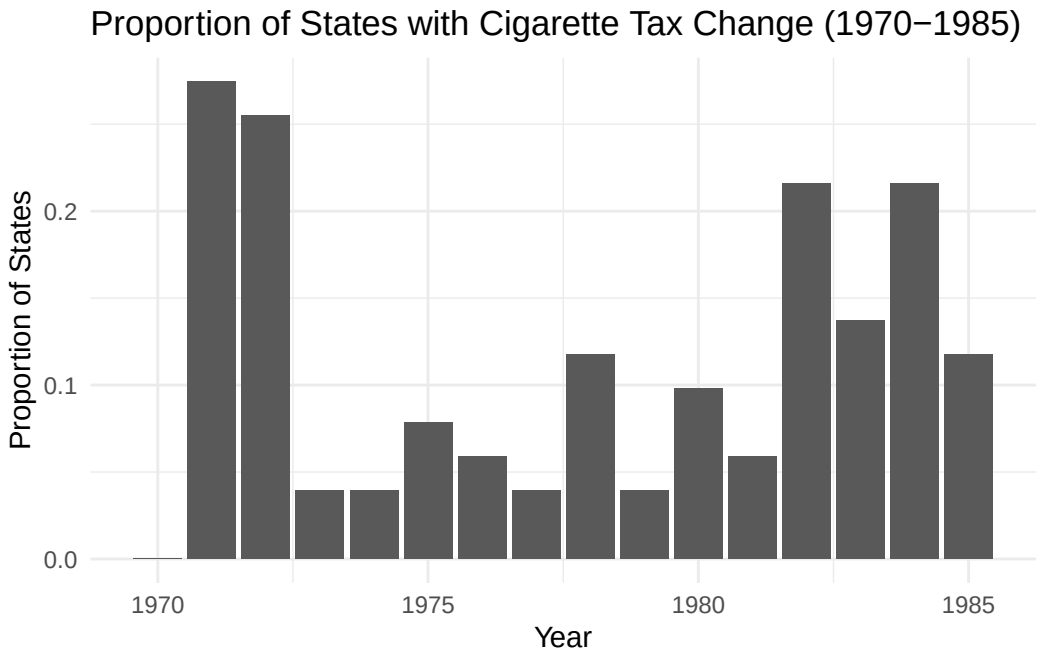
Megan Zheng

My answers in the following file. Check out my repository here. I have suddenly come across many errors all over my code that I am attempting to fix, so sorry if this information looks similar to, or even less than, submission 1.

Summarize the Data

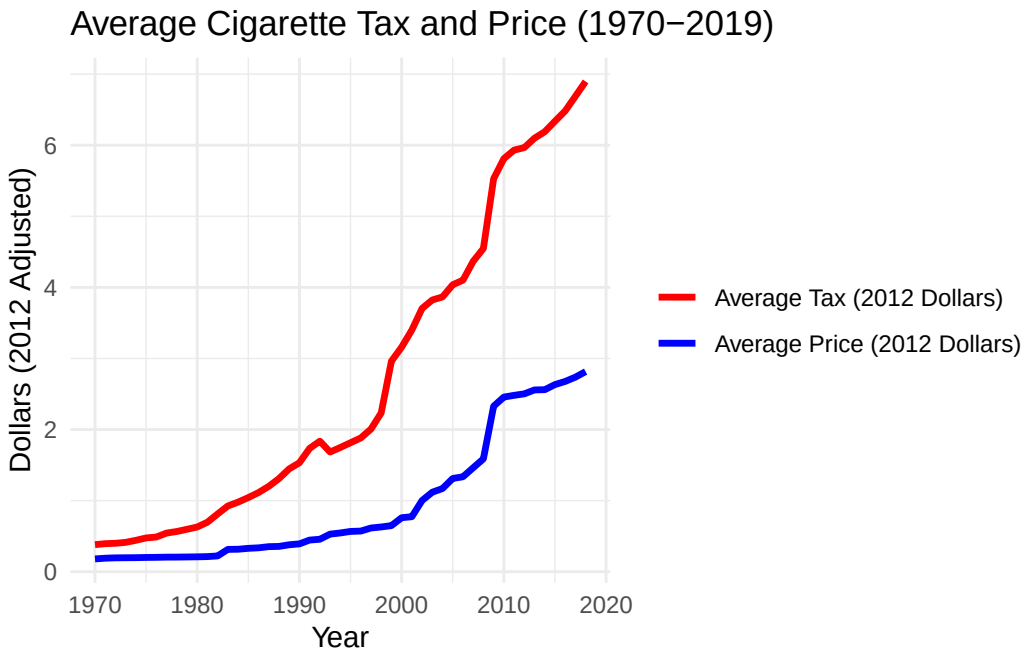
Question 1

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



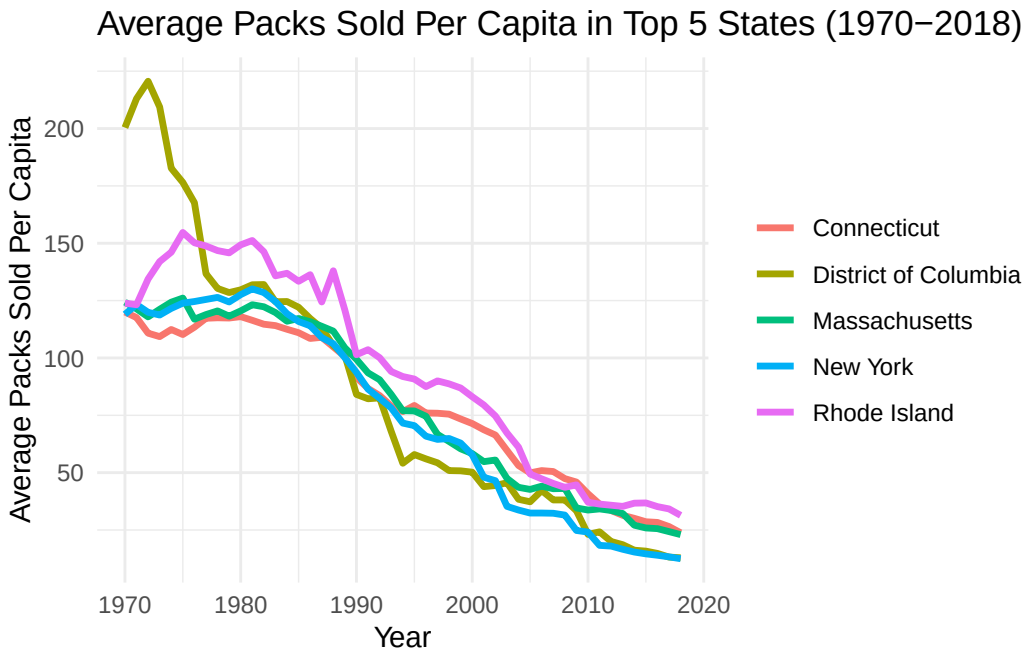
Question 2

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.



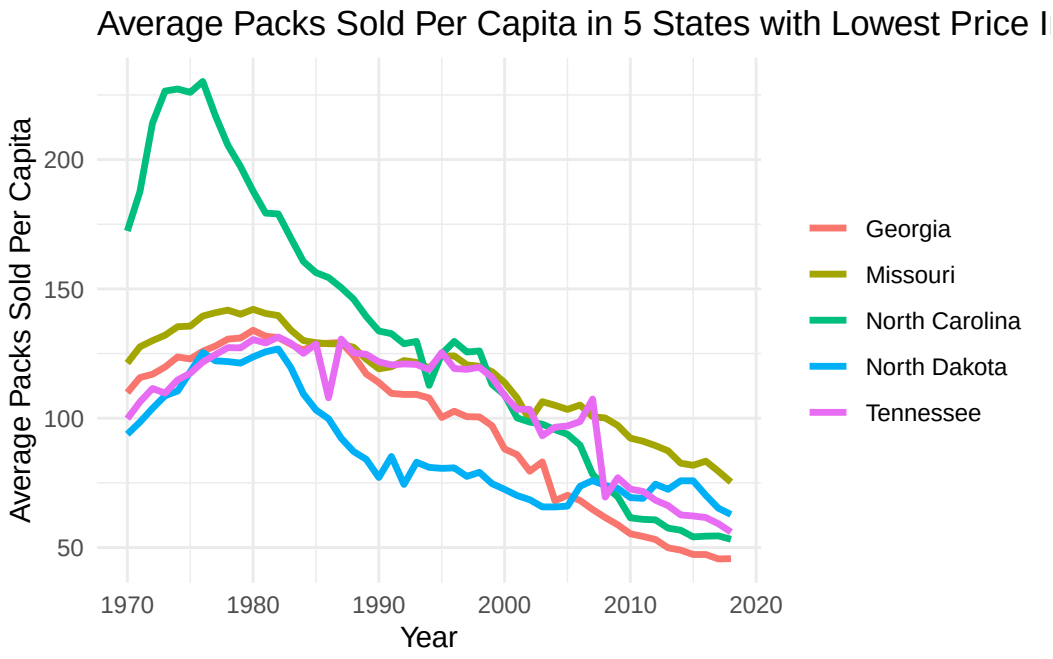
Question 3

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.



Question 4

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



Question 5

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

Here, the biggest trend we see is that the average packs per capita for all states has been generally decreasing over time. For the states with the highest price increases, seen in question 3, the number of packs sold has decreased more significantly than those in the states with the lowest increases in cigarette prices.

Estimate ATEs

Question 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.: ln_sales
Observations: 1,071
Fixed-effects: state: 51,  Year: 21
Standard-errors: Clustered (state)
          Estimate Std. Error  t value   Pr(>|t|)
ln_price -0.74166   0.131926 -5.62181 8.4191e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.062859     Adj. R2: 0.916544
                 Within R2: 0.220887
```

Question 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.: ln_sales
                  Endo.    : ln_price
                  Instr.   : tax_dollar
Second stage: Dep. Var.: ln_sales
Observations: 1,071
Fixed-effects: state: 51,  Year: 21
Standard-errors: Clustered (state)
              Estimate Std. Error  t value   Pr(>|t|)
fit_ln_price -0.832286   0.201542 -4.12959 0.00013787 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.062992     Adj. R2: 0.916191
                 Within R2: 0.217589
F-test (1st stage), ln_price: stat = 502.7    , p < 2.2e-16 , on 1 and 1,049 DoF.
                  Wu-Hausman: stat =   2.02852, p = 0.154684, on 1 and 998 DoF.
```

Question 8

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

```
OLS estimation, Dep. Var.: ln_price
Observations: 1,071
Fixed-effects: state: 51,  Year: 21
Standard-errors: Clustered (state)
           Estimate Std. Error t value   Pr(>|t|)
tax_dollar 0.825616   0.114165  7.2318 2.6049e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.037105     Adj. R2: 0.993199
                 Within R2: 0.323954


OLS estimation, Dep. Var.: ln_sales
Observations: 1,071
Fixed-effects: state: 51,  Year: 21
Standard-errors: Clustered (state)
            Estimate Std. Error  t value  Pr(>|t|)
tax_dollar -0.687148   0.196652 -3.49423 0.0010054 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.06793     Adj. R2: 0.902536
                Within R2: 0.090113
```

Question 9

Repeat questions 1-3 focusing on the period from 1991 to 2015.

```
OLS estimation, Dep. Var.: ln_sales
Observations: 1,275
Fixed-effects: state: 51,  Year: 25
Standard-errors: Clustered (state)
          Estimate Std. Error t value   Pr(>|t|)
ln_price -0.858241    0.13923 -6.1642 1.2155e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.112139     Adj. R2: 0.933018
                 Within R2: 0.23069


TSLS estimation - Dep. Var.: ln_sales
                  Endo.    : ln_price
                  Instr.   : tax_dollar
Second stage: Dep. Var.: ln_sales
Observations: 1,275
Fixed-effects: state: 51,  Year: 25
Standard-errors: Clustered (state)
             Estimate Std. Error t value   Pr(>|t|)
fit_ln_price -1.23796   0.173908 -7.1185 3.9163e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.115383     Adj. R2: 0.929086
                 Within R2: 0.185531
F-test (1st stage), ln_price: stat = 2,588.9, p < 2.2e-16, on 1 and 1,249 DoF.
                  Wu-Hausman: stat =   166.0, p < 2.2e-16, on 1 and 1,198 DoF.


OLS estimation, Dep. Var.: ln_price
Observations: 1,275
Fixed-effects: state: 51,  Year: 25
Standard-errors: Clustered (state)
           Estimate Std. Error t value  Pr(>|t|)
tax_dollar 0.150091   0.008128 18.4648 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.040818     Adj. R2: 0.992619
                 Within R2: 0.674557
```

Question 10

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

I am comparing everything here.