Homework 3

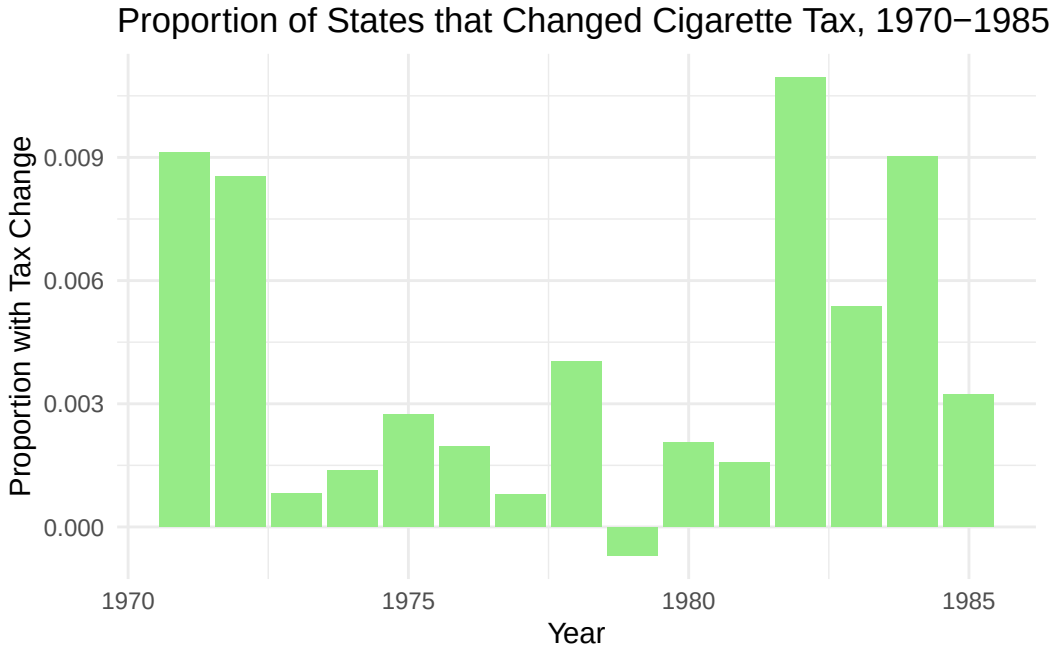
Research in Health Economics, Spring 2025

Sammy Ramacher

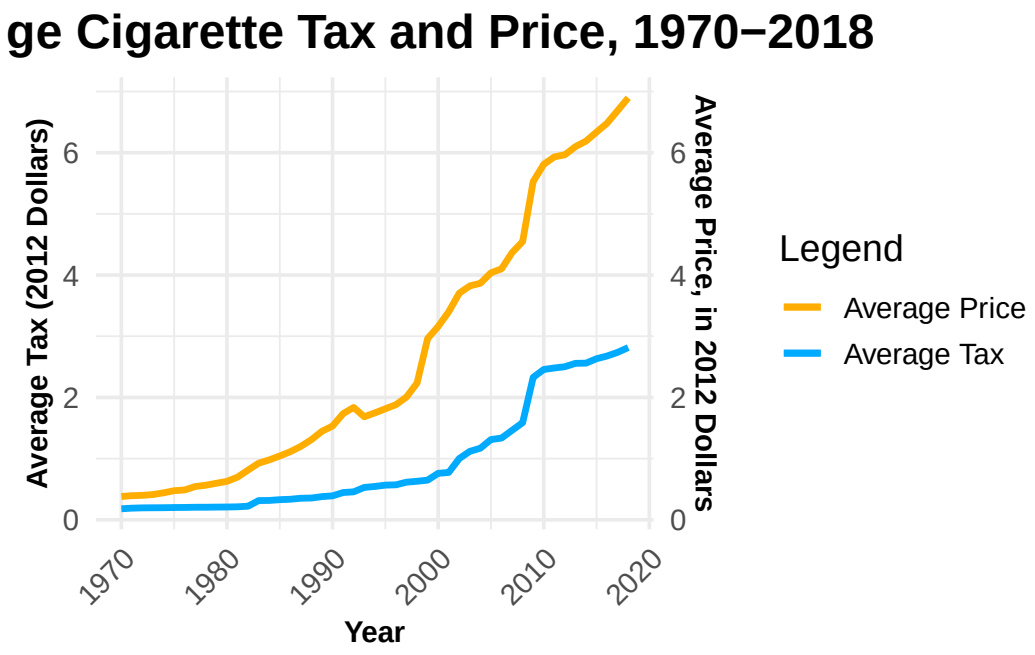
The GitHub repository for this work is available here.

Summarize the Data

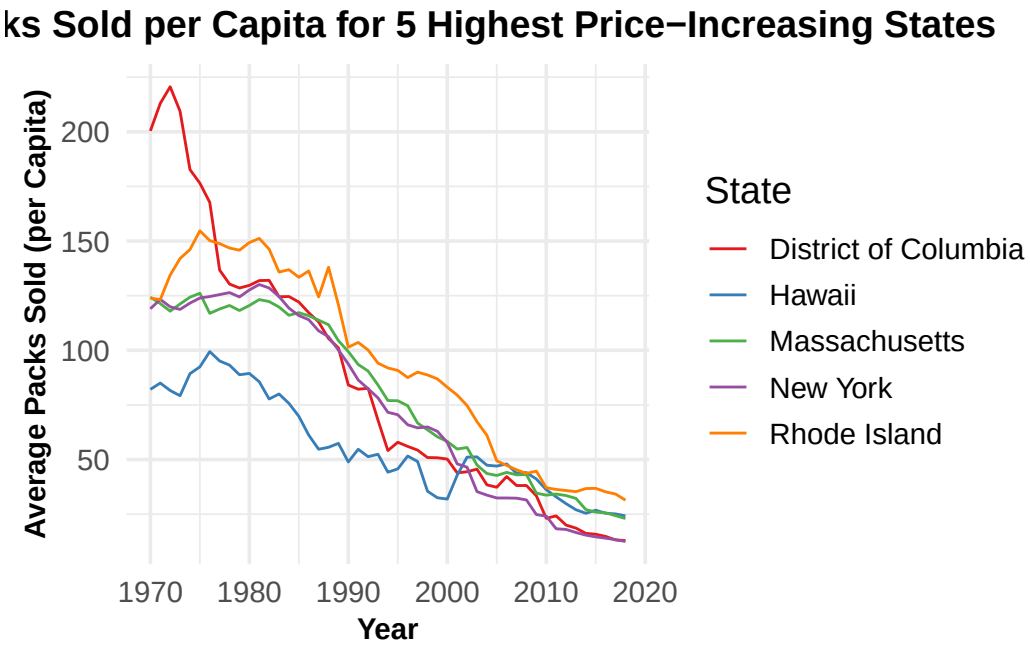
Question 1. Present a bar graph showing the proportion of states with a change in their cigarette tax in each year from 1970 to 1985.



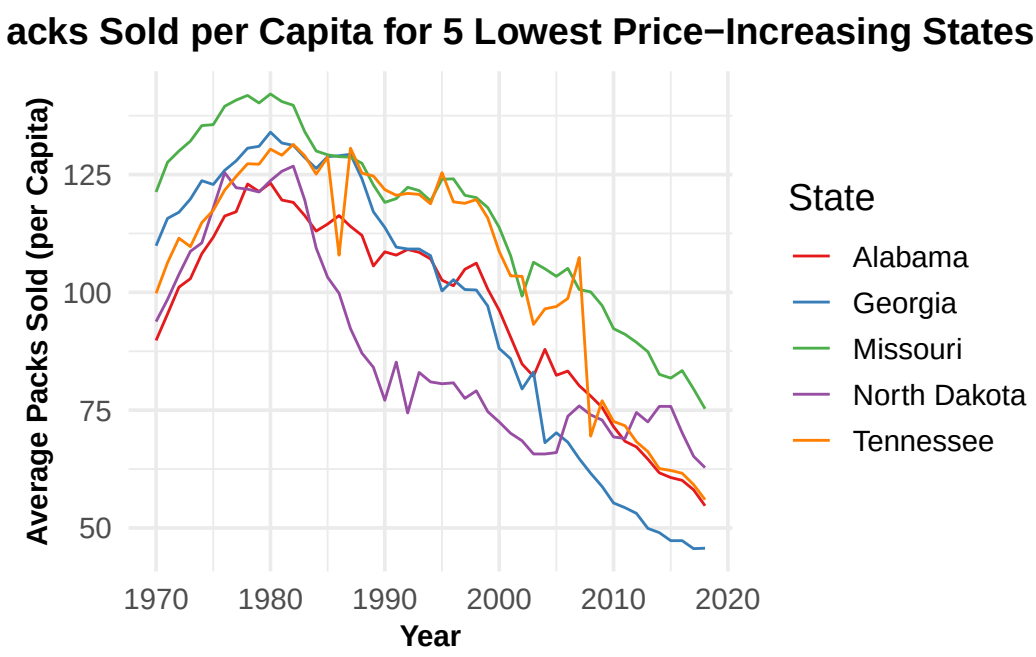
Question 2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.



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.

The 5 states with the highest increase in cigarette prices have a sharp decline over time in the average number of packs sold. Meanwhile, the 5 states with the lowest price increase for cigarettes have a more stable average number of packs sold over time, although there is still a steady declune. Both datasets have a maximum average number of packs sold around 1980, and then the averages drop over time. All 5 of the states with the highest price increase sell less packs per capita than all 5 of the low price increase states in later years, starting around 2015.

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.

```
Call:
lm(formula = log_sales ~ log_price, data = data_1970_1990)

Residuals:
     Min       1Q   Median       3Q      Max
-0.77629 -0.09967 -0.00787  0.09969  0.78423

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.750402   0.008116   585.3   <2e-16 ***
log_price   -0.171540   0.013829   -12.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2107 on 1069 degrees of freedom
Multiple R-squared:  0.1258,    Adjusted R-squared:  0.125
F-statistic: 153.9 on 1 and 1069 DF,  p-value: < 2.2e-16
```

xxx

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?

```
             Length Class      Mode
coefficients     2   -none-     numeric
residuals     1071   -none-     numeric
fitted.values 1071   -none-     numeric
weights          0   -none-     NULL
offset           0   -none-     NULL
n                1   -none-     numeric
nobs             1   -none-     numeric
rank             1   -none-     numeric
```

```
df.residual      1   -none-     numeric
cov.unscaled     4   -none-     numeric
sigma            1   -none-     numeric
call             3   -none-     call
formula          3   formula    call
terms            3   -none-     list
levels           0   -none-     list
contrasts        2   -none-     list
model            3   data.frame list
y             1071   -none-     numeric
```

xxx

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

```
Call:
lm(formula = log_price ~ log_tax, data = data_1970_1990)

Coefficients:
(Intercept)      log_tax
      1.179        1.080
```

```
Call:
lm(formula = log_sales ~ log_tax, data = data_1970_1990)

Residuals:
     Min       1Q   Median       3Q      Max
-0.75589 -0.08447  0.00043  0.09596  0.80589

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.37498    0.02477  176.63   <2e-16 ***
log_tax     -0.30719    0.01690  -18.18   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.197 on 1069 degrees of freedom
Multiple R-squared:  0.2361,    Adjusted R-squared:  0.2353
F-statistic: 330.3 on 1 and 1069 DF,  p-value: < 2.2e-16
```

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

```
Call:
lm(formula = log_sales ~ log_price, data = data_1991_2015)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9375 -0.1781  0.0013  0.1860  1.1433

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.03949    0.02291  219.93   <2e-16 ***
log_price   -0.66563    0.01747  -38.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3056 on 1273 degrees of freedom
Multiple R-squared:  0.5327,    Adjusted R-squared:  0.5323
F-statistic:  1451 on 1 and 1273 DF,  p-value: < 2.2e-16


             Length Class      Mode
coefficients    2   -none-     numeric
residuals    1275   -none-     numeric
fitted.values 1275  -none-     numeric
weights         0   -none-     NULL
offset          0   -none-     NULL
n               1   -none-     numeric
nobs            1   -none-     numeric
rank            1   -none-     numeric
df.residual     1   -none-     numeric
cov.unscaled    4   -none-     numeric
sigma           1   -none-     numeric
call            3   -none-     call
formula         3   formula    call
terms           3   -none-     list
levels          0   -none-     list
contrasts       2   -none-     list
model           3   data.frame list
y            1275   -none-     numeric
```

```
Call:
lm(formula = log_price ~ log_tax, data = data_1991_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-0.44006 -0.10954 -0.00173  0.10152  0.55693

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.207168   0.004970  242.91   <2e-16 ***
log_tax     0.630010   0.006857   91.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1774 on 1273 degrees of freedom
Multiple R-squared:  0.869, Adjusted R-squared:  0.8689
F-statistic:  8442 on 1 and 1273 DF,  p-value: < 2.2e-16



Call:
lm(formula = log_sales ~ log_tax, data = data_1991_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-0.82897 -0.14423  0.00604  0.14668  1.19203

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.236866   0.007842  540.26   <2e-16 ***
log_tax     -0.480477   0.010820  -44.41   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.28 on 1273 degrees of freedom
Multiple R-squared:  0.6077,    Adjusted R-squared:  0.6074
F-statistic:  1972 on 1 and 1273 DF,  p-value: < 2.2e-16
```

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

xxx