Homework 3-3

Yaseen Bhat

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.

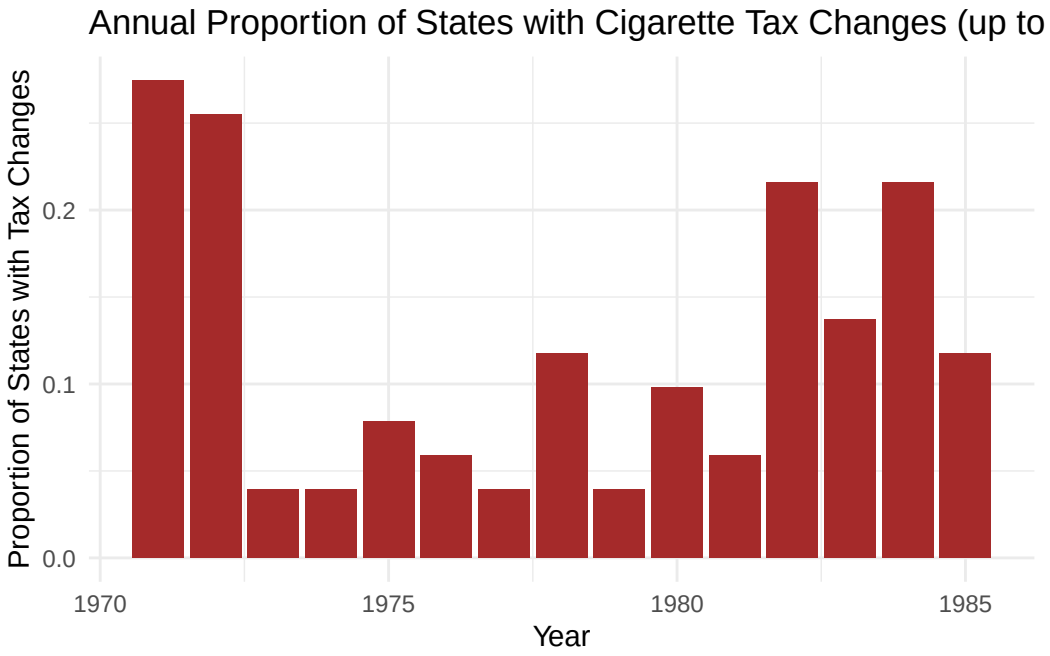

Figure 1: The Proportion of States with A Change in Their Cigarette Tax in Each Year from 1970 to 1985

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.

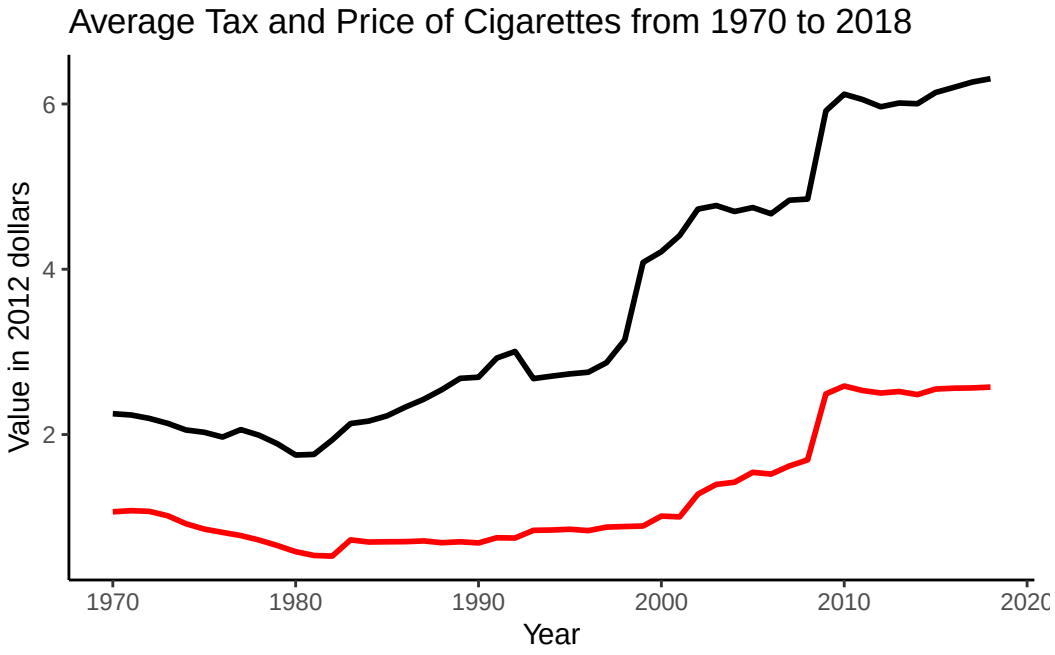

Figure 2: The Average Tax (in 2012 dollars) On Cigarettes And The Average Price of A Pack of Cigarettes From 1970 to 2018

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.

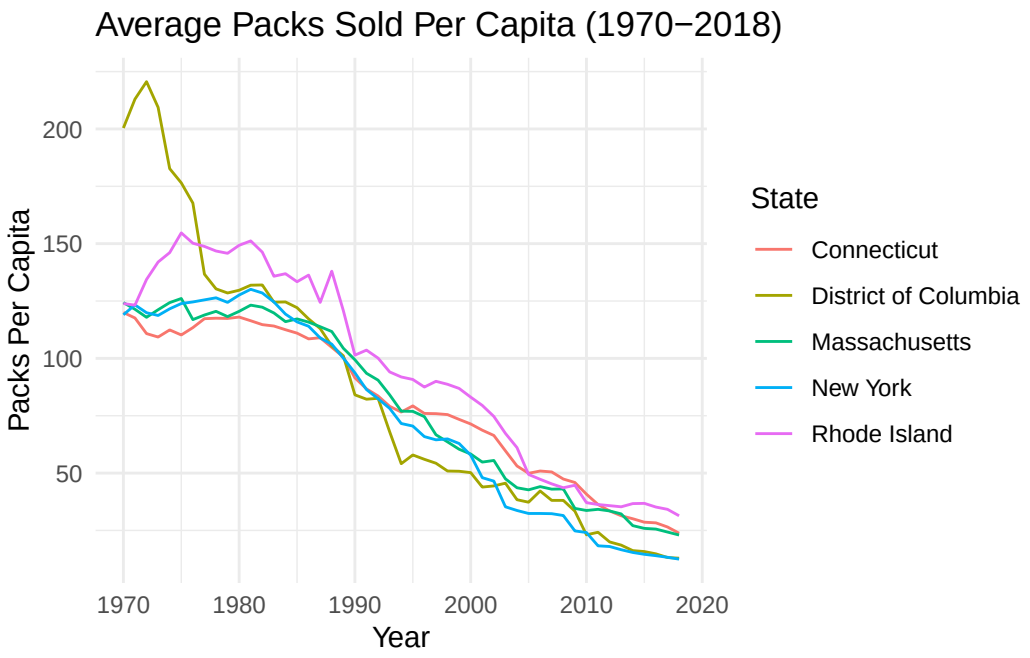

Figure 3: Average Number of Packs sold per capita for Top 5 States from 1970 to 2018

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.

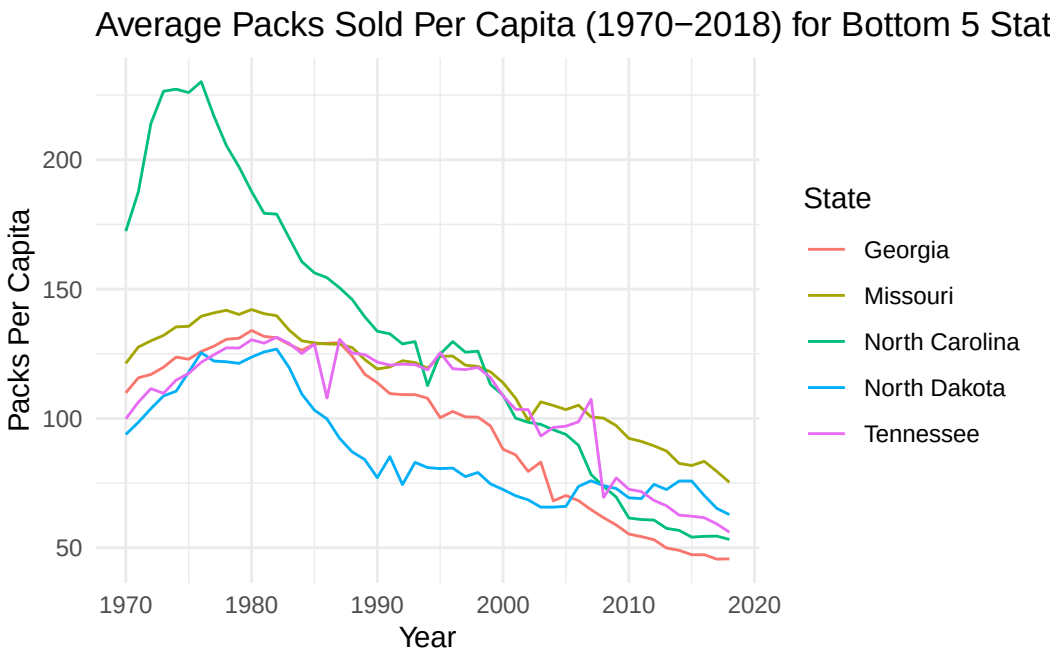

Figure 4: Average Number of Packs sold per capita for Bottom 5 States from 1970 to 2018

5

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

Both sets of groups showed a decreasing trend of cig sales over time. But the states with the higher price increases led to lesser sales than the ones with the lower increases. i.e. the lower decreases didn't have as steep of a decline.

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.

```
Call:
lm(formula = ln_sales ~ ln_price_2012, data = filter(cig.data,
    Year >= 1970 & Year <= 1990))

Residuals:
     Min       1Q   Median       3Q      Max
-0.68335 -0.08598 -0.00284  0.08778  0.83516

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.42738    0.02975   182.4   <2e-16 ***
ln_price_2012 -0.80944    0.03837   -21.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1894 on 1069 degrees of freedom
Multiple R-squared:  0.294, Adjusted R-squared:  0.2933
F-statistic: 445.1 on 1 and 1069 DF,  p-value: < 2.2e-16
```

The coefficient is -0.80944, which means as price goes up, sales go down and vice-versa.

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_2012, Instr.: ln_tax_2012
Second stage: Dep. Var.: ln_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_ln_price_2012 -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), ln_price_2012: 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 significant negative coefficient for again confirms that higher prices lead to lower sales. It doesn't seem like there's a significant difference with the estimates without the instrument, however it is slightly lower.

8

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

```
#q8
#| echo: false
#| warning: false
#| fig-cap: "Number 8"
print(first_stage_model)
```

```
Call:
lm(formula = ln_price_2012 ~ ln_tax_2012, data = price_elasticity_model8)

Coefficients:
(Intercept)  ln_tax_2012
     0.8396       0.2601
```

```
print(second_stage_model)
```

```
Call:
lm(formula = ln_sales ~ predicted_price, data = price_elasticity_model8)

Coefficients:
    (Intercept)  predicted_price
         5.4168          -0.7955
```

Clearly it has a significant effect on the price, as shown by the first stage, but the coefficients are quite different with/without the fixed effects.

9

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

```
#q9
#| echo: false
#| warning: false
#| fig-cap: "Number 9"
print(pem_2)
```

```
Call:
lm(formula = ln_sales ~ ln_price_2012, data = regdata2)

Coefficients:
  (Intercept)  ln_price_2012
       5.6600        -0.9968
```

```
print(feols(ln_sales ~ 1 | ln_tax_2012,
             data= regdata2))
```

```
OLS estimation, Dep. Var.: ln_sales
Observations: 1,275
Fixed-effects: ln_tax_2012: 1,024
RMSE: 0.112223   Adj. R2: 0.679548
```

```
print(first_stage2)
```

```
Call:
lm(formula = ln_price_2012 ~ ln_tax_2012, data = regdata2)

Coefficients:
(Intercept)  ln_tax_2012
     1.3151       0.5136
```

```
print(second_stage2)
```

```
Call:
lm(formula = ln_sales ~ predicted_price2, data = regdata2)

Coefficients:
     (Intercept)  predicted_price2
            5.88             -1.15
```

10

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

For the years 1991-2015, the elasticity estimate of -0.99681 mean there is a stronger reaction of cigarette sales to price changes than in the earlier period of 1970-1990, which had an elasticity estimate of -0.795524. It seems that as time went on, people became more sensitive to the price of cigarettes. This shift could be due to several factors—perhaps there's been a change in how society views smoking, or maybe there are now more alternatives available to smokers.