Homework 3: Submission 1

Research Methods, Spring 2024

Leila Mulveny

Homework 3: Repository

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.

@tax-change:

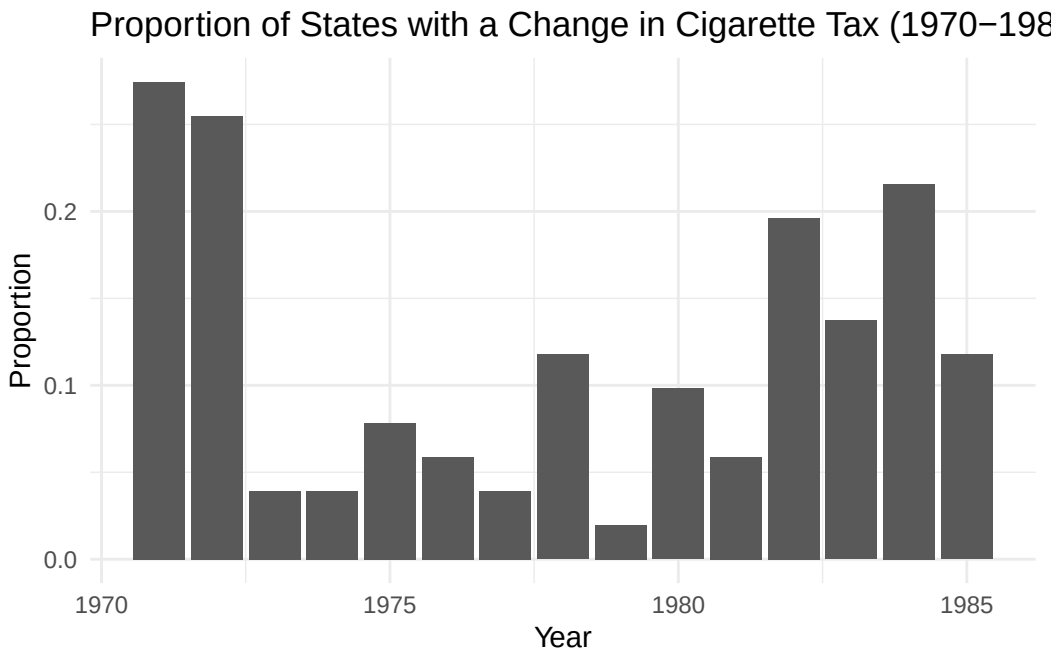

Figure 1: Proportion of States With a Change in Their Cigarette Tax from 1970-1985

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.

@ave-tax:

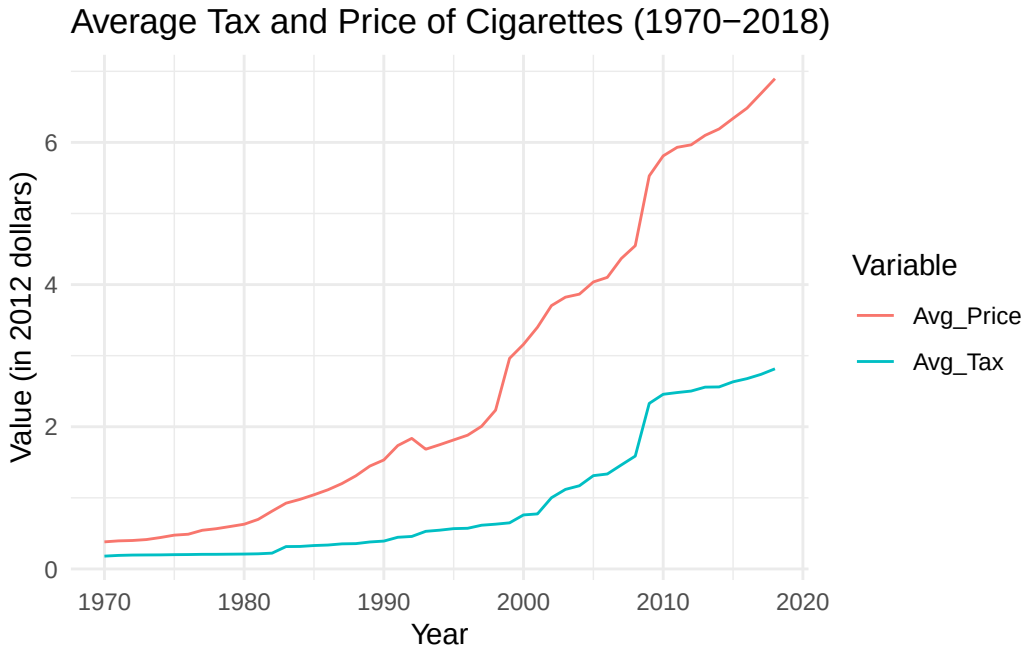

Figure 2: Average Tax (in 2012 dollars) on Cigarettes and the Average Price of a Pack of Cigarettes from 1970-2018

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.

@top-5-states:

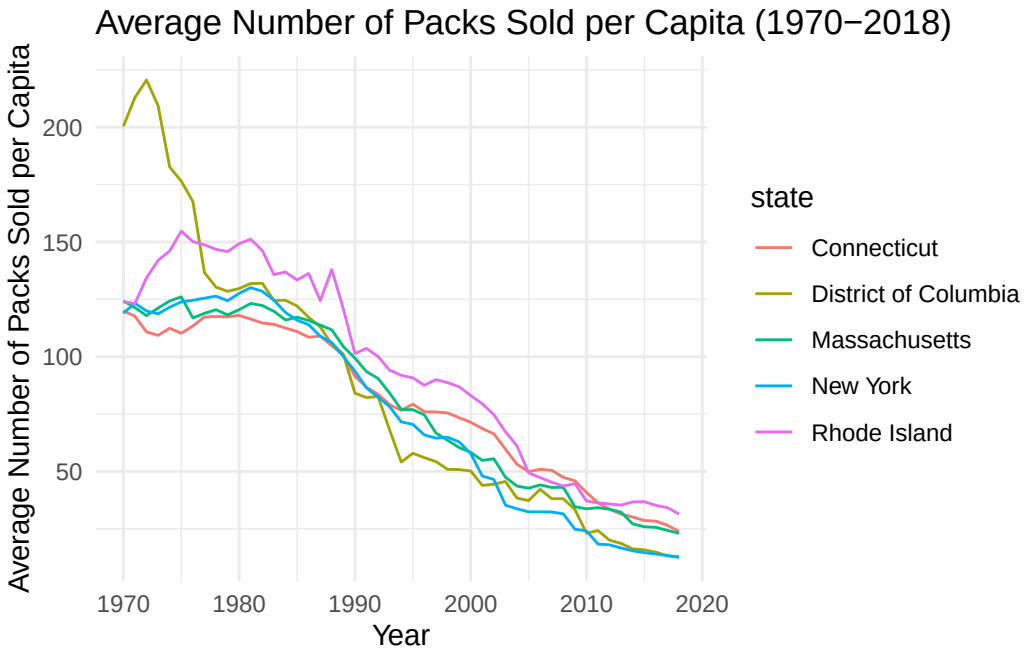

Figure 3: Average Number of Packs Sold per Capita for the Top Five States with the Highest increases in Cigarette Prices (in dollars) from 1970-2018

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.

@bottom-5-states:

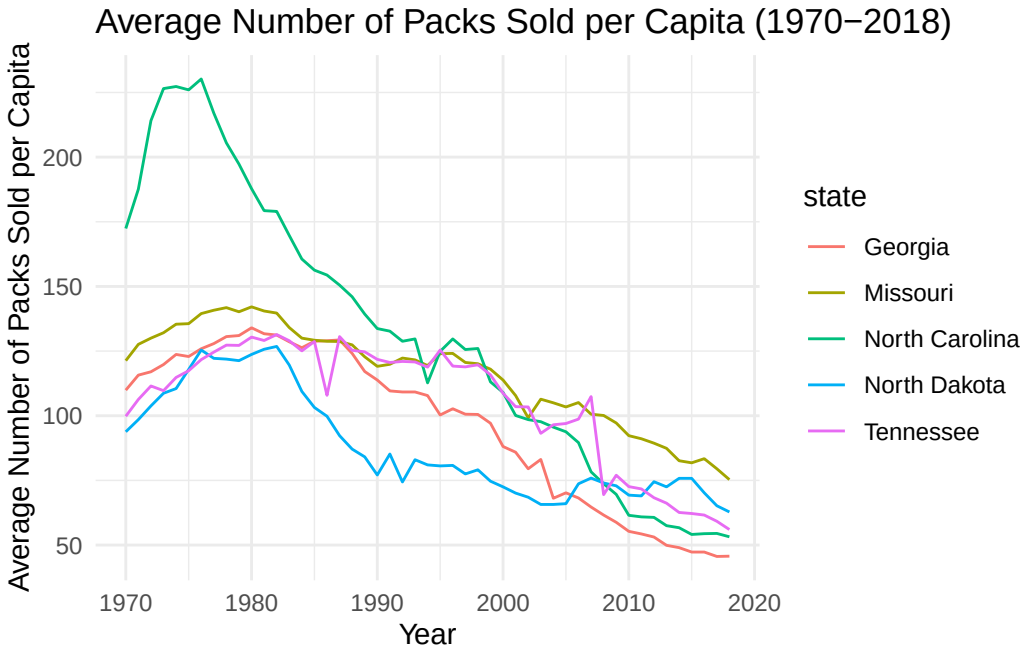

Figure 4: Average Number of Packs Sold per Capita for the Bottom Five States with the Lowest increases in Cigarette Prices (in dollars) from 1970-2018

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

@compare-5-states:

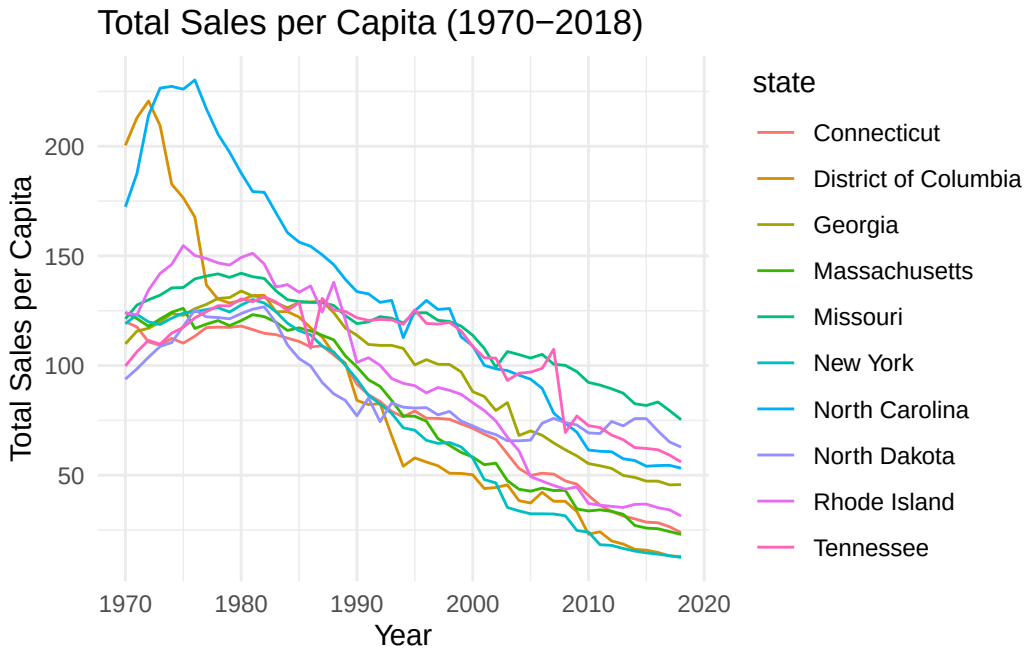

Figure 5: Comparison of the Trends in Sales From the 5 States with the Highest Price Increase to Those with the Lowest Price Increases

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.

@model1:

```
Call:
lm(formula = ln_sales ~ ln_price, data = cig_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 ***
ln_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
```

The results displayed in @model1 show a coefficient of -0.172 which is the estimated price elasticity of demand. It means that a 1% increase in cigarette price is associated with a decrease of about 0.172% in 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?

@model2:

```
Call:
ivreg(formula = ln_sales ~ ln_price | ln_total_tax, data = cig_data_1970_1990)

Residuals:
     Min       1Q   Median       3Q      Max
-1.34832 -0.25750  0.05651  0.22601  1.20077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.99111    0.03411 146.342  < 2e-16 ***
ln_price     0.50237    0.08984   5.592 2.85e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3782 on 1069 degrees of freedom
Multiple R-Squared: -1.816, Adjusted R-squared: -1.819
Wald test: 31.27 on 1 and 1069 DF,  p-value: 2.848e-08
```

The results displayed in @model2 show a coefficient of 0.502 which is the estimated price elasticity of demand. It means that a 1% increase in cigarette price is associated with an increase of about 0.502% in sales. This is not intuitive because an increase in price should result in a decrease in demand, so I likely made an error in my code. It could be that the ln_price variable is endogenous (correlated with the error term) in @model1, but because my result is positive I think there is also something else going on.

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

@stages_1:

```
Call:
lm(formula = ln_price ~ ln_total_tax, data = cig_data_1970_1990)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0060 -0.3359 -0.1095  0.3779  1.0872

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.51899    0.02199 -23.598   <2e-16 ***
ln_total_tax -0.41181    0.04381  -9.399   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4479 on 1069 degrees of freedom
Multiple R-squared:  0.07634,   Adjusted R-squared:  0.07547
F-statistic: 88.35 on 1 and 1069 DF,  p-value: < 2.2e-16
```

@reduced_1:

```
Call:
lm(formula = ln_sales ~ ln_total_tax, data = cig_data_1970_1990)

Residuals:
     Min       1Q   Median       3Q      Max
-0.86239 -0.09798  0.00549  0.09359  0.95094

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.73038    0.01060 446.216   <2e-16 ***
ln_total_tax -0.20688    0.02112  -9.796   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2159 on 1069 degrees of freedom
Multiple R-squared:  0.08238,   Adjusted R-squared:  0.08152
F-statistic: 95.97 on 1 and 1069 DF,  p-value: < 2.2e-16
```

Estimate ATEs

Question 9 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.

@model_1:

```
Call:
lm(formula = ln_sales ~ ln_price, data = cig_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 ***
ln_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
```

The results in @model_1 display a coefficient for ln_price of -0.66563. This is the estimated price elasticity of demand. It means that a 1% increase in price is associated with a decrease of about 0.666% in sales. This suggests that cigarettes are a normal good (a good that people consume less of as the price rises).

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?

@model_2:

```
Call:
ivreg(formula = ln_sales ~ ln_price | ln_total_tax, data = cig_data_1991_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-0.81957 -0.19431 -0.01266  0.21094  1.17694

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.21890    0.02650  196.96   <2e-16 ***
ln_price    -0.81311    0.02055  -39.57   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.314 on 1273 degrees of freedom
Multiple R-Squared: 0.5066, Adjusted R-squared: 0.5062
Wald test:  1566 on 1 and 1273 DF,  p-value: < 2.2e-16
```

The results displayed in @model_2 show a coefficient of -0.813 which is the estimated price elasticity of demand. Both models estimate a negative price elasticity of demand, however, the estimated elasticity is larger (in absolute value) in @model_2 (-0.81311) than in @model_1 (-0.66563). This suggests that when accounting for the potential endogeneity of price (using ln_total_tax as an instrument), a 1% increase in price is associated with a larger decrease in sales (about 0.81% instead of 0.67%). The differences could be due to the endogeneity of ln_price, which model_2 tries to correct for by using ln_total_tax as an instrument.

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

@stages_2:

```
Call:
lm(formula = ln_price ~ ln_total_tax, data = cig_data_1991_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-0.57946 -0.17145  0.02125  0.16675  0.66985

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.11198    0.00687  161.85   <2e-16 ***
ln_total_tax  0.72638    0.01133   64.12   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2383 on 1273 degrees of freedom
Multiple R-squared:  0.7636,    Adjusted R-squared:  0.7634
F-statistic:  4111 on 1 and 1273 DF,  p-value: < 2.2e-16
```

@reduced_2:

```
Call:
lm(formula = ln_sales ~ ln_total_tax, data = cig_data_1991_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-0.90878 -0.15465  0.01119  0.15334  1.16925

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.314731   0.008078  534.11   <2e-16 ***
ln_total_tax -0.590626   0.013320  -44.34   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2802 on 1273 degrees of freedom
Multiple R-squared:  0.607, Adjusted R-squared:  0.6067
F-statistic:  1966 on 1 and 1273 DF,  p-value: < 2.2e-16
```

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 @model2 (0.50237) and negative in @model_2 (-0.81311). This is a substantial difference. The positive elasticity in model2 is unusual and suggests that an increase in price is associated with an increase in sales, which is not typically expected for normal goods (so I likely have an error). The negative elasticity in @model_2 is more typical and suggests that an increase in price is associated with a decrease in sales. Theoretically, I would expect a greater price elasticity estimate for the years 1991-2015 because people are more aware of the health consequences of cigarettes, and therefor more responsive to price changes versus earlier when everyone smoked with little concern (more people addicted would seem to drive a lower elasticity).