

Homework 3

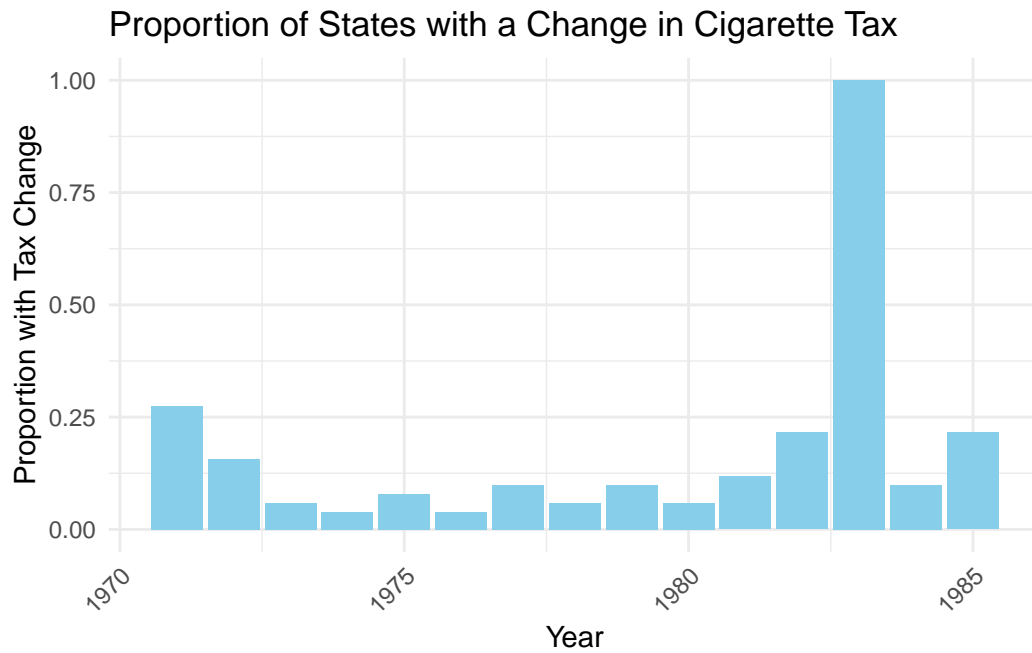
Research in Health Economics, Spring 2025

Lisbeth Vargas

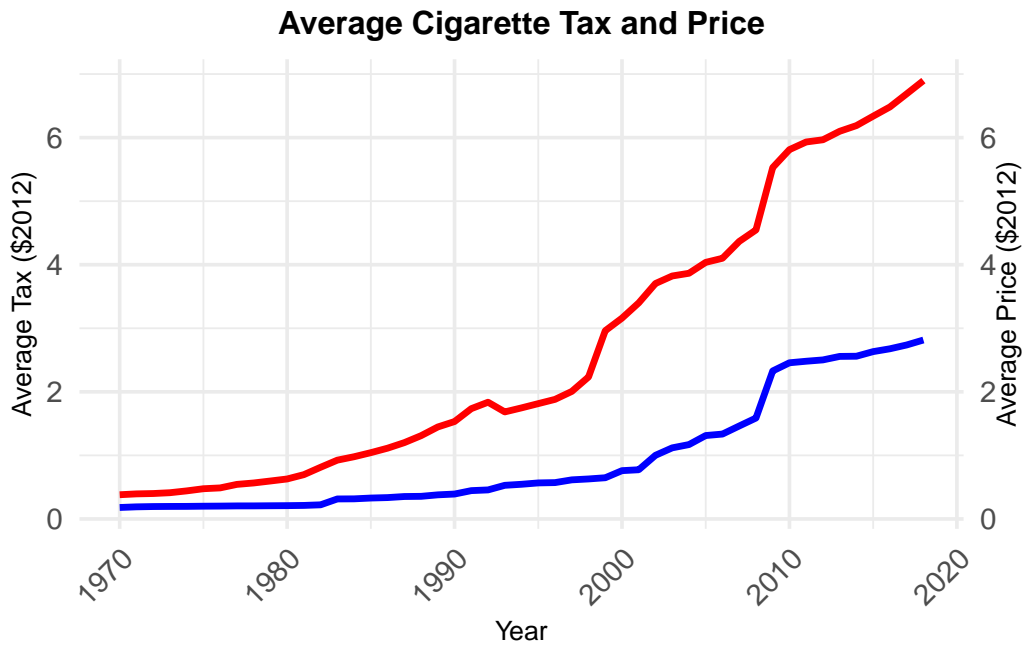
The following is my submission for Homework 3. Note that the setup and analysis for these responses are in a separate R script. The GitHub repository for this work is available [here](#).

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.

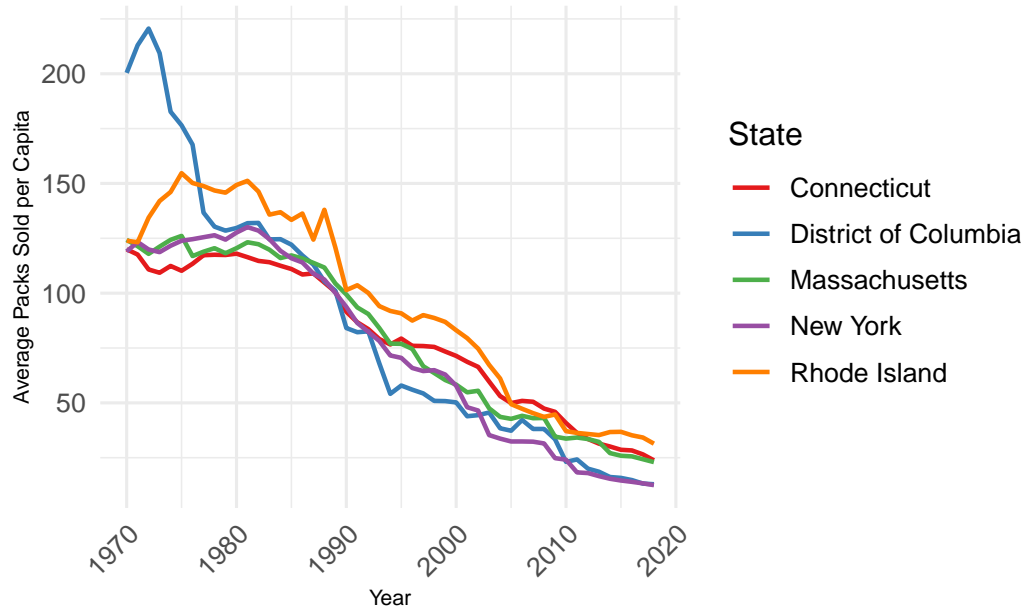


Question 2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.



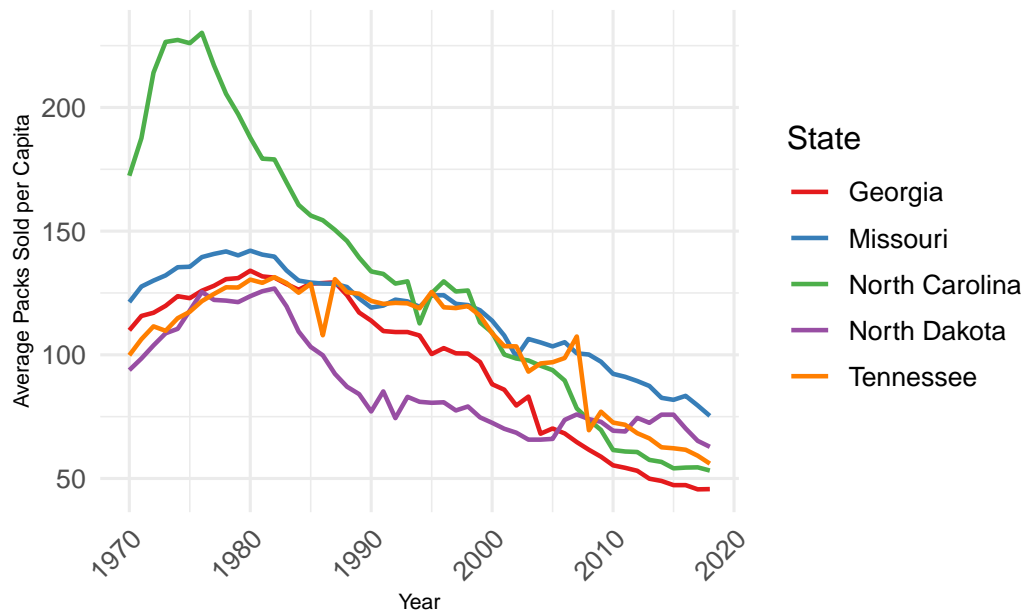
Question 3. Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Average Cigarette Packs Sold per Capita for States with Highest Price Increases



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.

Average Number of Packs Sold per Capita for States with Lowest Price Increases



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

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

I am interpreting my results.

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?

Loading required package: car

Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:dplyr':

recode

The following object is masked from 'package:purrr':

some

Loading required package: lmtest

Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

Loading required package: sandwich

Table 1: Instrumental Variable Regression Results

term	estimate	std.error	statistic	p.value
(Intercept)	4.710	0.009	520.12	0
log_price	-0.284	0.017	-16.48	0

I am interpreting results.

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

Table 2: First Stage Regression: $\log_price \sim \log_tax$

term	estimate	std.error	statistic	p.value
(Intercept)	1.179	0.033	35.71	0
\log_tax	1.080	0.023	47.97	0

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

Table 3: Price Elasticity for Demand: $\log_sales \sim \log_price$

term	estimate	std.error	statistic	p.value
(Intercept)	5.039	0.023	219.93	0
\log_price	-0.666	0.017	-38.09	0

Table 4: IV Regression: $\log_sales \sim \log_price \mid \log_tax$

term	estimate	std.error	statistic	p.value
(Intercept)	5.158	0.025	209.25	0
\log_price	-0.763	0.019	-40.20	0

Table 5: First Stage Regression: $\log_price \sim \log_tax$

term	estimate	std.error	statistic	p.value
(Intercept)	1.207	0.005	242.91	0
\log_tax	0.630	0.007	91.88	0

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