Homework 3 - Submission 3

ECON 470

Baran Pasa

Homework 3 Summary Statistics and ATE Analysis

Link to Github

Collecting and Cleaning Data

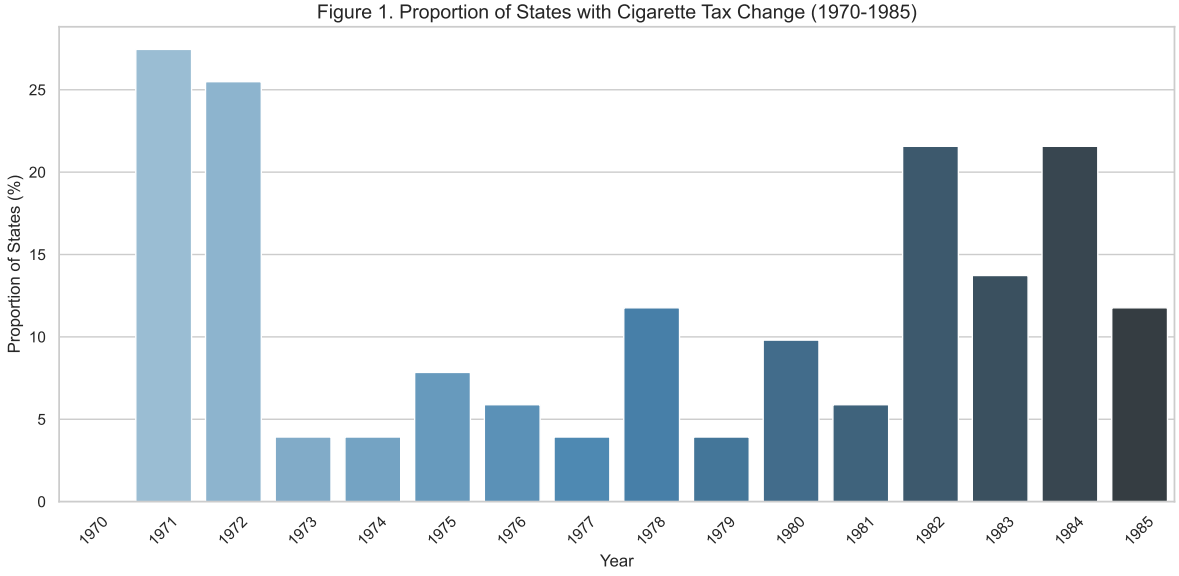
CDC Tax Burden on Tobacco Data was collected from a provided repository and inflation data was collected from the BLI database. Raw data was downloaded and then put into real dollars using 2012 as the base year.

1. Summarizing the Data

1. Proportion of States with a change in their cigarette tax each year from 1970 to 1985

As seen in the figure below, there is a large spike in proportion of states at 1983. It is possible that this conincides with a major policy shift on cigarettes.

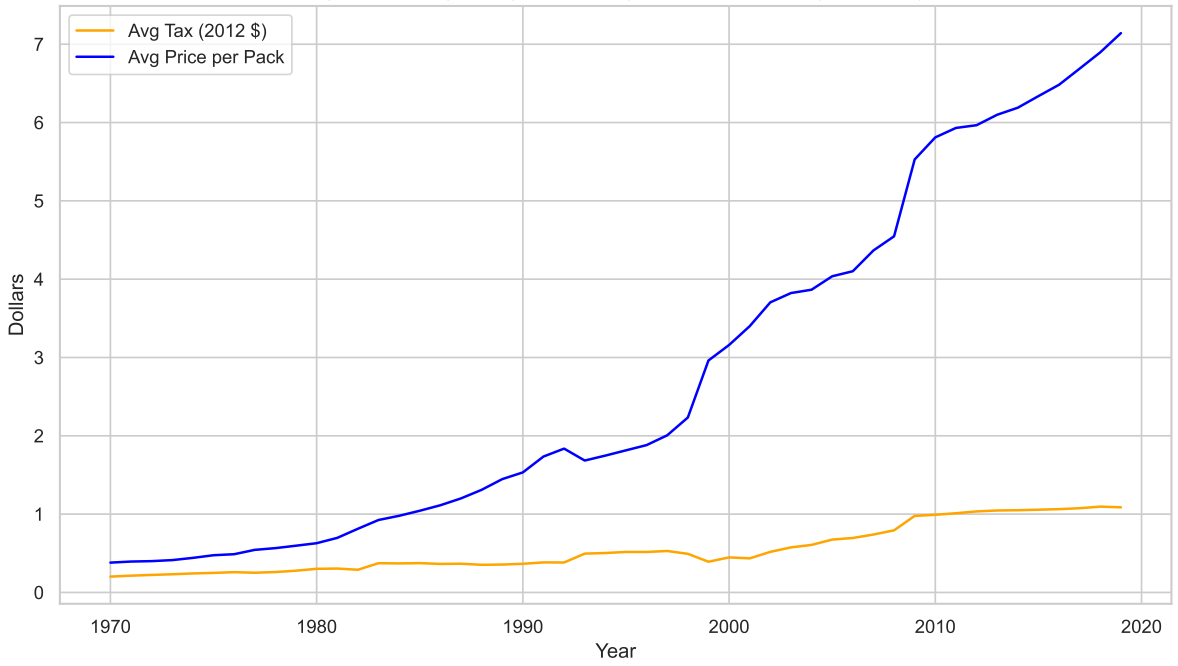
Figure 1. Proportion of States with Cigarette Tax Change (1970-1985)



2. Average Price per Pack, 1970-2018 (in 2012 dolalrs)

While both tax and price per pack have risen in past years, price per pack has grown at an exponential rate. This makes sense as the government has tried to use economic disincentives to sway people away from cigarettes.

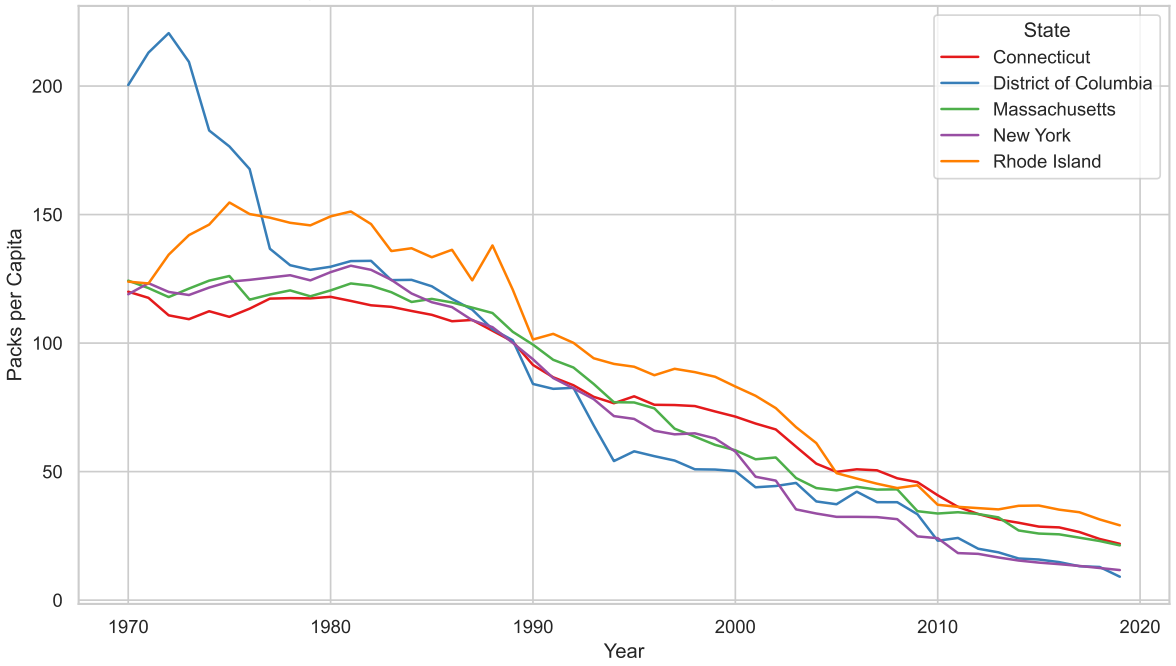
Figure 2. Average Tax (2012 Dollars) and Price per Pack (1970-2018)



3. Top 5 states with highest increases in cigarette prices

In the top 5 states with the highest price increases, there has been a significant decrease in the number of packs per person. This might point to some movement in the right direction for policy makers.

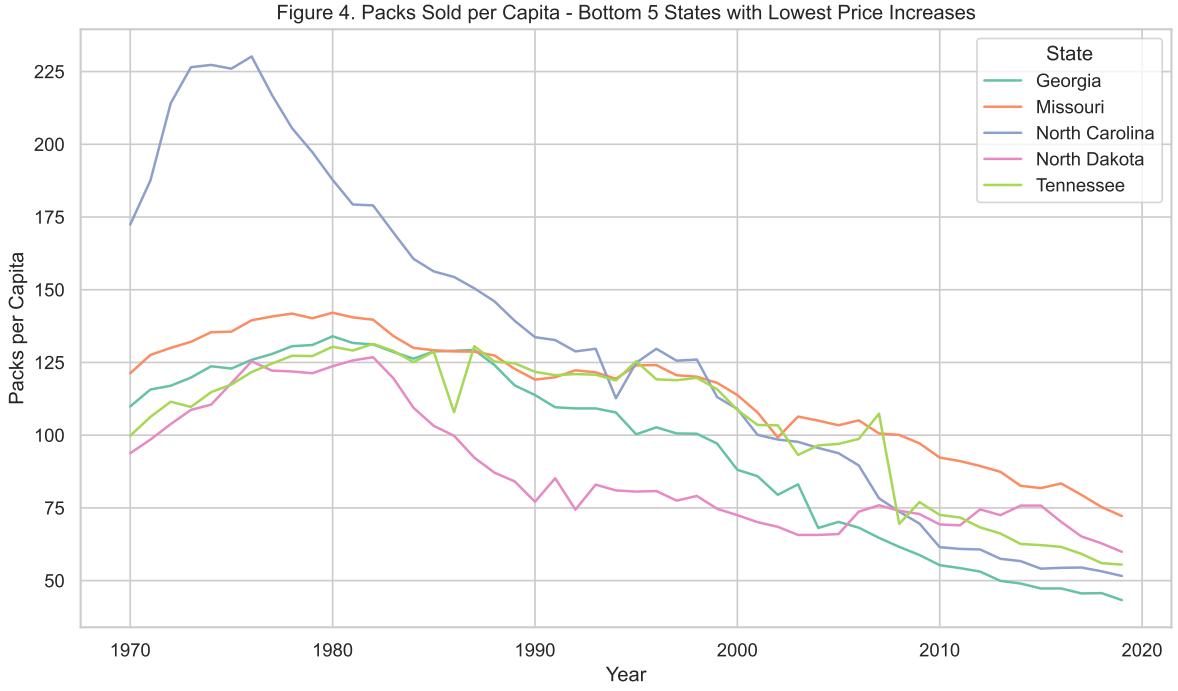
Figure 3. Packs Sold per Capita - Top 5 States with Highest Price Increases



4. Top 5 states with lowest increases in cigarette prices

Decline in the top 5 states with the lowest increases mimic the previous graph, just a higher number of packs per person. While they directionally are comaprable, their scales are different.

Figure 4. Packs Sold per Capita - Bottom 5 States with Lowest Price Increases

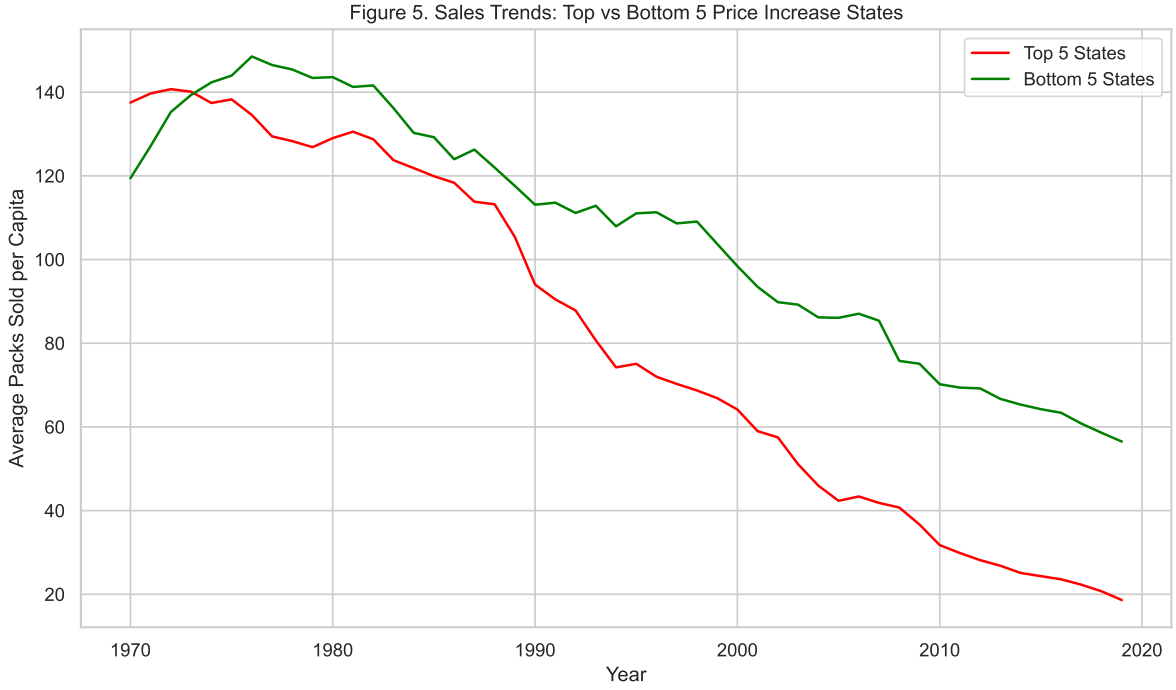


5. Comparison of trends between these 10 states:

Contrasting the 10 states, the differing scales mentioned above are clearer. These graphs do show some indication that increasing the tax burden might decrease the number of packs per person, but there are many outside variables that have not been taken into account.

From this graph, we can extrapolate that across the country, cigarette sales have been decreasing at a significant rate.

Figure 5. Sales Trends: Top vs Bottom 5 Price Increase States



ATE

For the following regressions I used the pyfixest model.

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

6. Regressing Log Sales on Log Prices 1970 - 1990

	log_sales (1)
coef	
log_price	-0.809***
	(0.038)
Intercept	5.427***
	(0.030)
stats	
Observations	1071
S.E. type	iid
R^2	0.294

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

With a price coefficient of -0.809, the regression indicates that for every 1 unit increase of price, sales decrease by 0.809 units. This does make sense, as an increase in price should decrease the demand.

It is worth noting that the p-value is very small, $p < 0.001$. This signifies that, within our sample data, evidence supporting the null hypothesis is very weak.

7. IV: Regressing Log Sales on Log Prices using Log Total Tax 1970 - 1990

	log_sales (1)
coef	
log_price	-0.867*** (0.116)
Intercept	5.471*** (0.089)
stats	
Observations	1071
S.E. type	iid
R^2	-

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

After switching from an OLS regression to using the total of Federal and State cigarette tax as an instrumental variable, the regression coefficient of price is -0.867. This is very similiar to what we got when running the OLS regression. Taking into account the standard error, we can assume that they are telling us the same result. As discussed later, however, neither regression should be trusted as accurate.

Using IV yielded a different regression result as there is some model-accuracy lost when using an insturmental variable rather than the target variable. A perfect insturmental variable would yield the same result as the original model, but there are many outside variables that could influence the relationship of total tax on price.

Again, it is worth noting that the p-value is very small, $p < 0.001$. This signifies that, within our sample data, evidence supporting the null hypothesis is very weak.

8. First Stage and Reduced-Form Instrument Results

First Stage Regression (1) & Reduced-Form Regression (2)

	log_price (1)	log_sales (2)
coef		
log_total_tax	0.044***	-0.038***
	(0.004)	(0.006)
Intercept	0.885***	4.703***
	(0.012)	(0.018)
stats		
Observations	1071	1071
S.E. type	iid	iid
R^2	0.109	0.037

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Format of coefficient cell: Coefficient (Std. Error)

9. Repeating the regression above for 1991-2015

The regressions were reran, this time with data from 1991 to 2015. Despite my best efforts to create a cohesive table, this is the most presentable I could get.

(1) -> OLS 1991 - 2015
(2) -> IV 1991 - 2015
(3) -> First Stage 1991 - 2015
(4) -> Reduced Form 1991 - 2015 etc.

	log_sales		log_price	log_sales
	(1)	(2)	(3)	(4)
coef				
log_price	-0.997***	-1.169***		
	(0.025)	(0.042)		
log_total_tax			0.169***	-0.197***
			(0.006)	(0.009)
Intercept	5.660***	5.908***	1.683***	3.940***
	(0.036)	(0.061)	(0.012)	(0.017)
stats				
Observations	1275	1275	1275	1275
S.E. type	iid	iid	iid	iid
R^2	0.561	-	0.354	0.273

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Format of coefficient cell: Coefficient (Std. Error)

10. Further Interpretations

10.1 Comparing OLS Regressions from 1970 - 1990 and 1991 - 2015

	log_sales	
	(1)	(2)
coef		
log_price	-0.809***	-0.997***
	(0.038)	(0.025)
Intercept	5.427***	5.660***
	(0.030)	(0.036)
stats		
Observations	1071	1275
S.E. type	iid	iid
R^2	0.294	0.561

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

When compared side-by-side we see that in both time periods, an increase in price by 1-unit results in a decrease in sales of 1-unit (-0.809 and -0.997 respectively).

The 1970-1990 OLS (1), however, has an R^2 of 0.294. This indicates that for this time period, the price is not effectivley explaining sales. The 1991-2015 OLS (2), however, has an R^2 of 0.561. This indicates that price is moderately explaining change in sales. Both time periods show a moderate to weak fit for price on sales.

10.2 Comparing IV Regressions from 1970 - 1990 and 1991 - 2015

	log_sales	
	(1)	(2)
coef		
log_price	-0.867***	-1.169***
	(0.116)	(0.042)
Intercept	5.471***	5.908***
	(0.089)	(0.061)
stats		
Observations	1071	1275
S.E. type	iid	iid
R^2	-	-

	log_sales	
	(1)	(2)

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

Using total tax on cigarettes as an instrumental varibale, both regressions yielded a negative coefficient. The coefficient for 1991-2015 (2) is larger at -1.169 than 1970-1990's of -0.867.

Looking at Figure 2 from earlier, there is a clear exponential increase in the slope of the average price of cigarettes. The increases in average price per pack from 1991-2015 are much higher than the increases between 1970-1990.

In that same graph, the increase in average tax does not increase significantly until 2000. This leads me to believe that total tax, as collected and calculated here, is not an accurate instrumental variable for the increasing price per pack.

10.3 Comparing First Stage and Reduced-Form Regressions from 1970 - 1990 and 1991 - 2015

First Stage Regressions:

	log_price	
	(1)	(2)
coef		
log_total_tax	0.044***	0.169***
	(0.004)	(0.006)
Intercept	0.885***	1.683***
	(0.012)	(0.012)
stats		
Observations	1071	1275
S.E. type	iid	iid
R^2	0.109	0.354

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

Running the 1970-1990 first stage regression, price (1) yielded a total tax coefficient of 0.044. This indicates that a 1-unit increase in total tax would lead to a 0.044 increase in cigarette price. If true, this increase would have little impact out in the real world.

Looking at the reduced form regression (2), the total tax coefficient on sales is -0.038. Accordning to the model, a 1-uint increase in total tax would lead to a 0.169 unit increase in price. While this regression coefficient is more significant than the 1970-1990 coefficient, it still does not explain the large difference in increases seen in the Figure 2 graph.

Assessing the R^2 values of both regressions, we can assume that there is a low measure of fit between our variables.

Reduced Form Regressions:

	log_sales	
	(1)	(2)
coef		
log_total_tax	-0.038***	-0.197***
	(0.006)	(0.009)
Intercept	4.703***	3.940***
	(0.018)	(0.017)
stats		
Observations	1071	1275
S.E. type	iid	iid
R^2	0.037	0.273

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

Comparing the reduced form regressions, the 1970-1990 and 1991-2015 rgression coefficients of total tax on sales is -0.038 and -0.197 respectively. These coefficients indicate that in both time periods, increasing total tax led to a small decrease in sales of cigaretttes.

Assessing the R^2 values of both regressions, we can assume that there is a low measure of fit between our variables.

Conclusion

Using the regression coefficients as an indicator for price changes and considering the sharp decreases in cigarette packs sold as shown in Figure 5, it can be inferred that price is not what is driving cigarette sales down. Price changes are not what driving sales down; rather, I believe that the anti-smoking social movement was the main driver.

After conducting these regressions, I put further weight into the impact of the anti-smoking social movement on cigarette sales. An interesting next experiment would be to try and quantify the impact that the movement had on cigarette sales, and compare it to the regressions above.