huang-r-hwk3

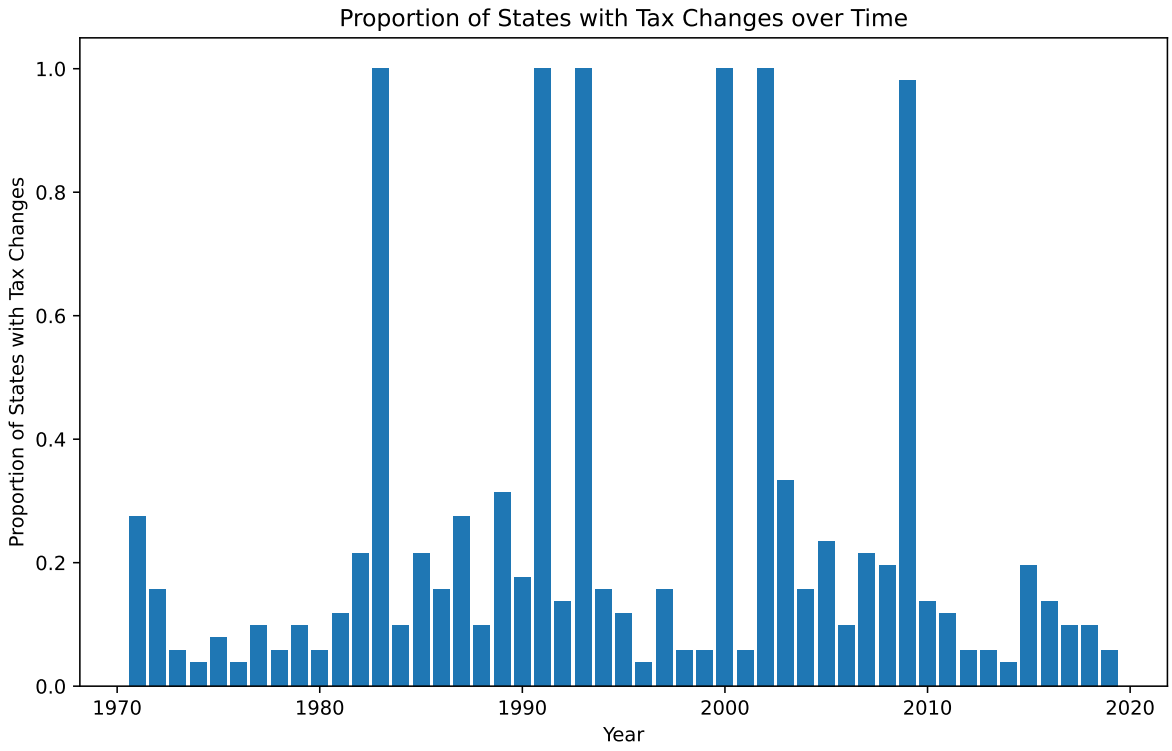
Ron Huang

Table of contents

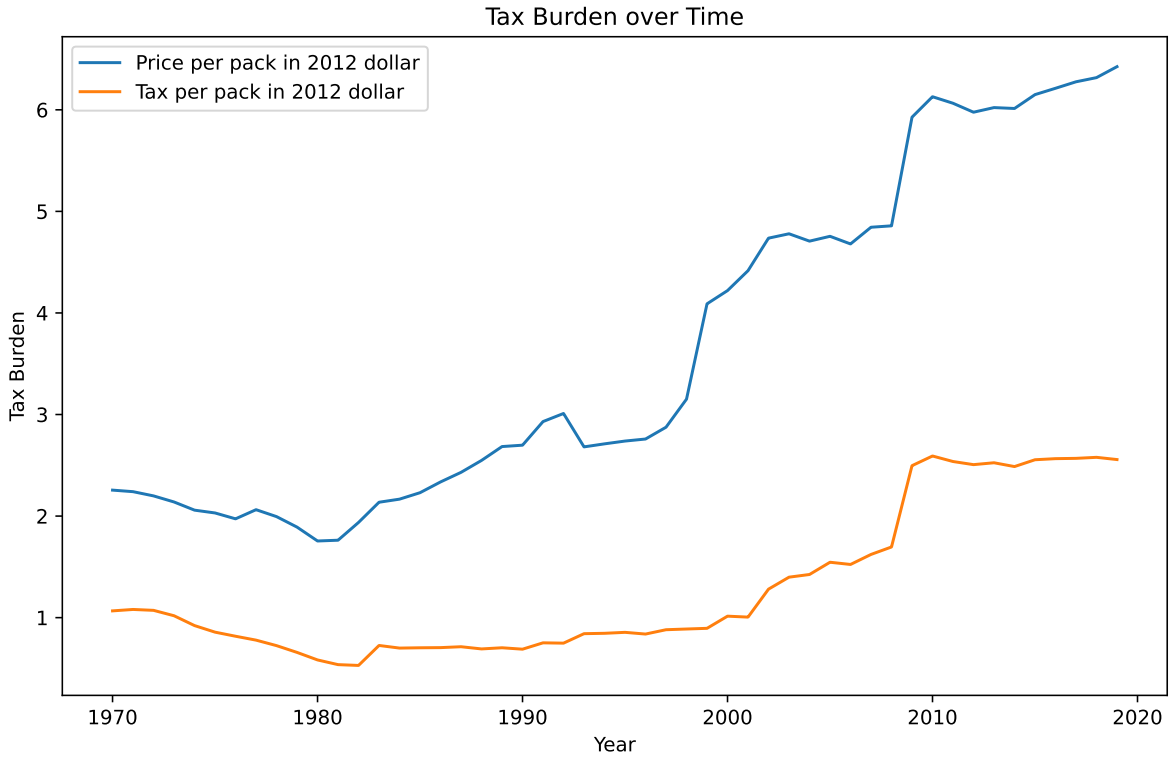
git@github.com:Rorn001/econ470-hwk3.git

1 Summarize the data

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

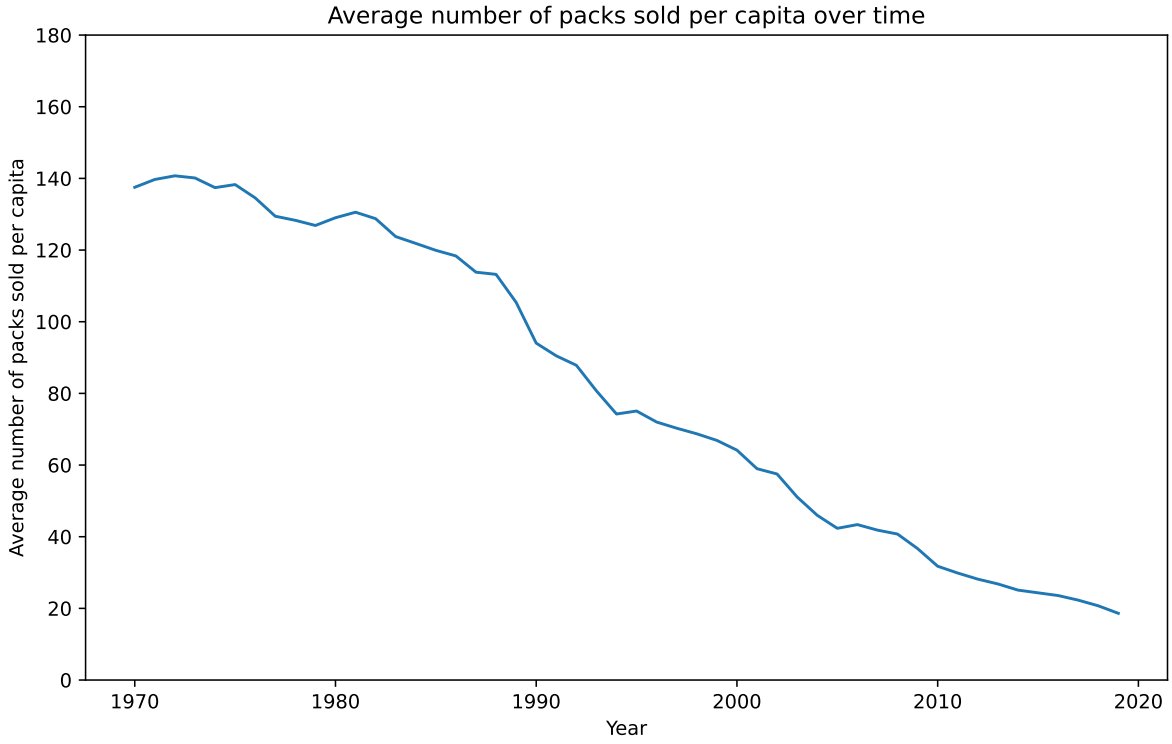


1.2 Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.



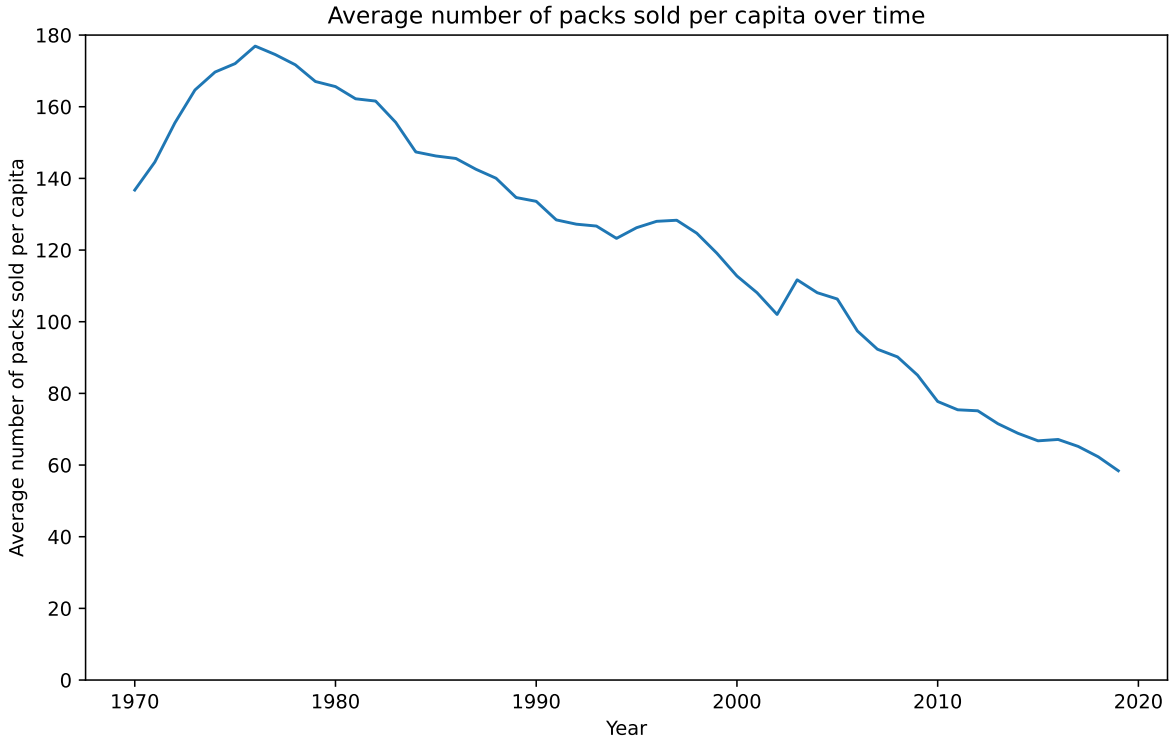
1.3 Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

	state	price_change	price_1970	price_2019
8	District of Columbia	7.441234	1.931230	9.372464
32	New York	7.005608	2.470316	9.475923
39	Rhode Island	6.500704	2.381455	8.882160
21	Massachusetts	6.346255	2.428847	8.775102
6	Connecticut	6.339671	2.695428	9.035099

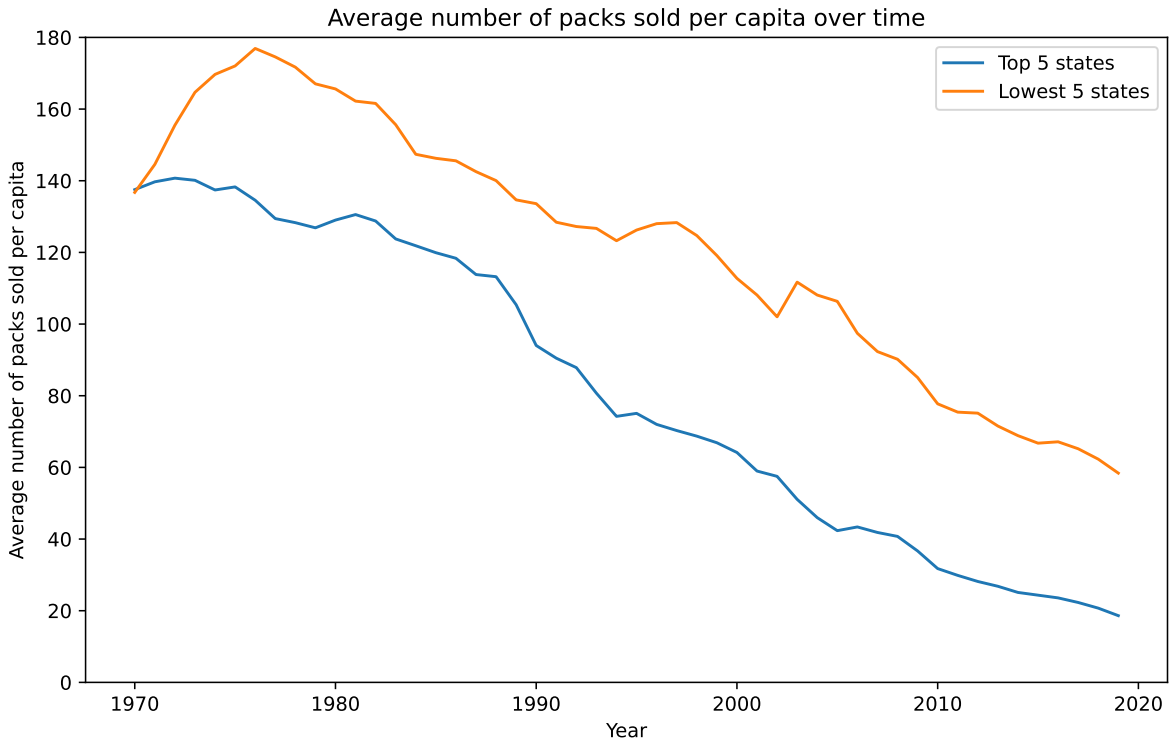


1.4 Identify the 5 states with the lowest increases in cigarette prices over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

	state	price_change	price_1970	price_2019
25	Missouri	2.505294	2.180039	4.685332
42	Tennessee	2.524120	2.464392	4.988512
34	North Dakota	2.579710	2.304443	4.884153
0	Alabama	2.673970	2.529556	5.203526
10	Georgia	2.695882	2.120798	4.816680



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



The lowest 5 and highest 5 states have mostly same sales per capita in 1970. The lowest 5 states increased afterward but started to decrease at the end of 1970s. However, lowest 5 states consistently have higher sales per capita than the highest 5 states, while highest 5 states have a consistent decrease in sales per capita from 1970 to 2018.

2 Estimate ATEs

2.1 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.

Dep. Variable:	log_sales	**R-squared:**	0.922
Model:	OLS	**Adj. R-squared:**	0.917
Method:	Least Squares	**F-statistic:**	166.5
Date:	Sat, 16 Mar 2024	**Prob (F-statistic):**	0.00
Time:	12:11:10	**Log-Likelihood:**	1443.6
No. Observations:	1071	**AIC:**	-2743.
Df Residuals:	999	**BIC:**	-2385.
Df Model:	71		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
log_price	-0.7417	0.044	-16.829	0.000	-0.828	-0.655
const	5.2905	0.040	131.505	0.000	5.212	5.369
individaul dummies	Yes					
time dummies	Yes					

Omnibus:	97.631	**Durbin-Watson:**	0.693
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	577.353
Skew:	0.119	**Prob(JB):**	4.26e-126
Kurtosis:	6.589	**Cond. No.**	67.0

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Applying OLS and states and time fixed effects, the price elasticity of demand is -0.7417. This means that a 10% increase in price will lead to a 7.417% decrease in sales, which is inelastic. However, there might be endogeneity in the price variable as we did not control for any other factors that might affect both sales and prices, so this estimate might be causal.

2.2 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?

Dep. Variable:	log_sales	**R-squared:**	0.9204
Estimator:	IV-2SLS	**Adj. R-squared:**	0.9147
No. Observations:	1071	**F-statistic:**	1.343e+04
Date:	Sat, Mar 16 2024	**P-value (F-stat)**	0.0000
Time:	12:23:08	**Distribution:**	chi2(71)
Cov. Estimator:	robust		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	5.4621	0.0643	84.952	0.0000	5.3361	5.5882
log_price	-0.9484	0.0744	-12.755	0.0000	-1.0942	-0.8027
individual dummies	Yes					
time dummies	Yes					

Endogenous: log_price
Instruments: log_tax
Robust Covariance (Heteroskedastic)
Debiased: False

The IV regression estimates the second stage coefficient to be -0.9484, which is more elastic than the OLS estimate of -0.7417. This tends to be more causal if tax is exogenous and satisfies the exclusion restriction because it could recover the local treatment effect of log price on log sales of those whose price is positively affected by tax. If tax is a valid instrument, this comparison between the results of OLS and IV suggests that there do exists some omitted variables that correlates with price and sales. In particular, it may increase with price while decrease with sales.

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

Table 1: Table 1 - OLS, Reduced Form, and First Form Results

	OLS	IV	First Form	Reduced Form
const	5.29***	5.462***	0.81***	4.70***
	(0.04)	(0.06)	(0.01)	(0.02)
log_price	-0.74***	-0.95***		
	(0.04)	(0.07)		
log_tax			0.36***	-0.34***
			(0.01)	(0.02)
Individual dummies	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
R-squared	0.92	0.92	0.96	0.92
R-squared Adj.	0.92	0.92	0.96	0.92
No. observations	1071	1071	1071	1071

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

We can see from the first stage that the instrument is at least strong and relevant, as it significantly increases the price.

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

Table 2: Table 1 - OLS, Reduced Form, and First Form Results

	OLS	IV	First Form	Reduced Form
const	5.62***	5.63***	1.18***	4.61***
	(0.05)	(0.06)	(0.01)	(0.03)
log_price	-0.86***	-0.87***		
	(0.05)	(0.05)		
log_tax			0.33***	-0.29***
			(0.01)	(0.02)
individual dummies	Yes	Yes	Yes	Yes
time dummies	Yes	Yes	Yes	Yes
R-squared	0.94	0.93	0.98	0.93
R-squared Adj.	0.93	0.93	0.98	0.93
No. observations	1275	1275	1275	1275

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

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

Compared to the 1970-1990 period, the price elasticity of demand is more elastic in the 1991-2015 period, but they are both inelastic. The estimate for the 1991-2015 period is -0.87, while the estimate for the 1970-1990 period is -0.95, which is slightly more elastic. There are might be avfew reasons for this difference. First, there might be some omitted variables that are correlated with price and sales, and they might have changed over time. Second, we can see previously that the tax burden is increasing over time, so the expectation of increasing in the real price of cigs might exclude consumers who are more sensitive to its price, left those who are the most addicted to cigarettes and least responsive to price changes.