

# Homework 3

Research Methods, Spring 2025

Ryan Scholte

You can access the [# 1 Bar Graph](https://github.com/rscholt/HW3)

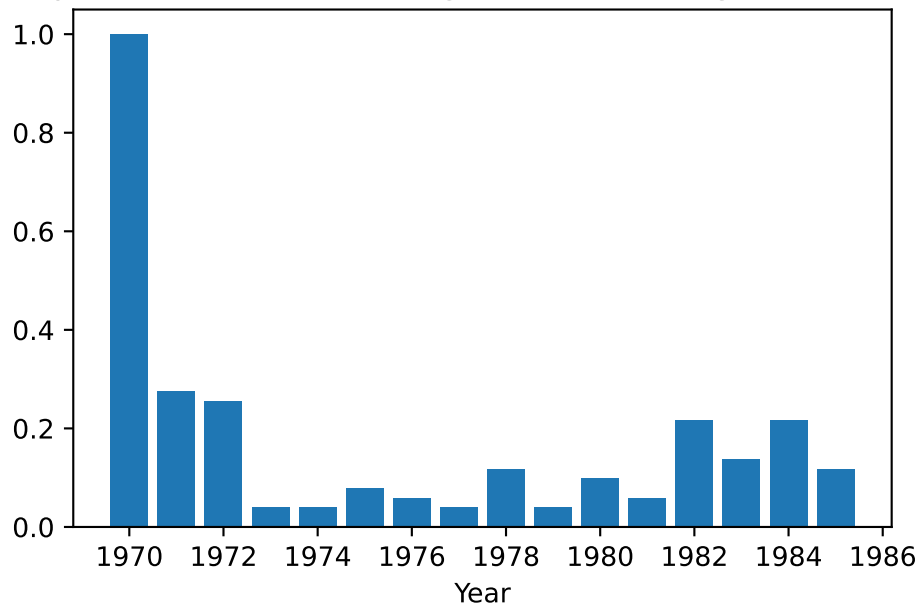
Unable to display output for mime type(s): text/html

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

/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel\_58123/2100850660.py:20: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

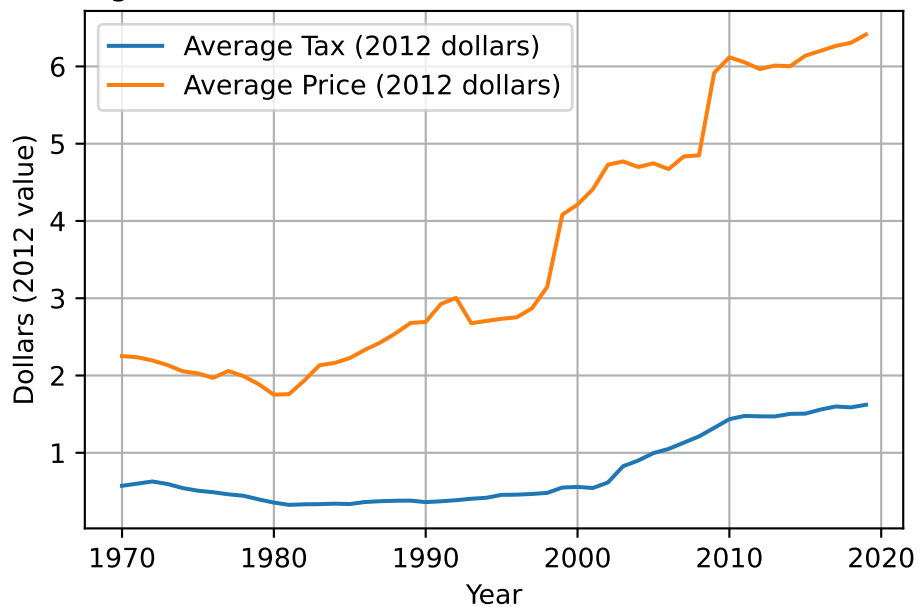
See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide](https://pandas.pydata.org/pandas-docs/stable/user_guide)  
dataq1['tax\_change'] = dataq1.groupby('state')['tax\_state'].diff().ne(0).astype(int)

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



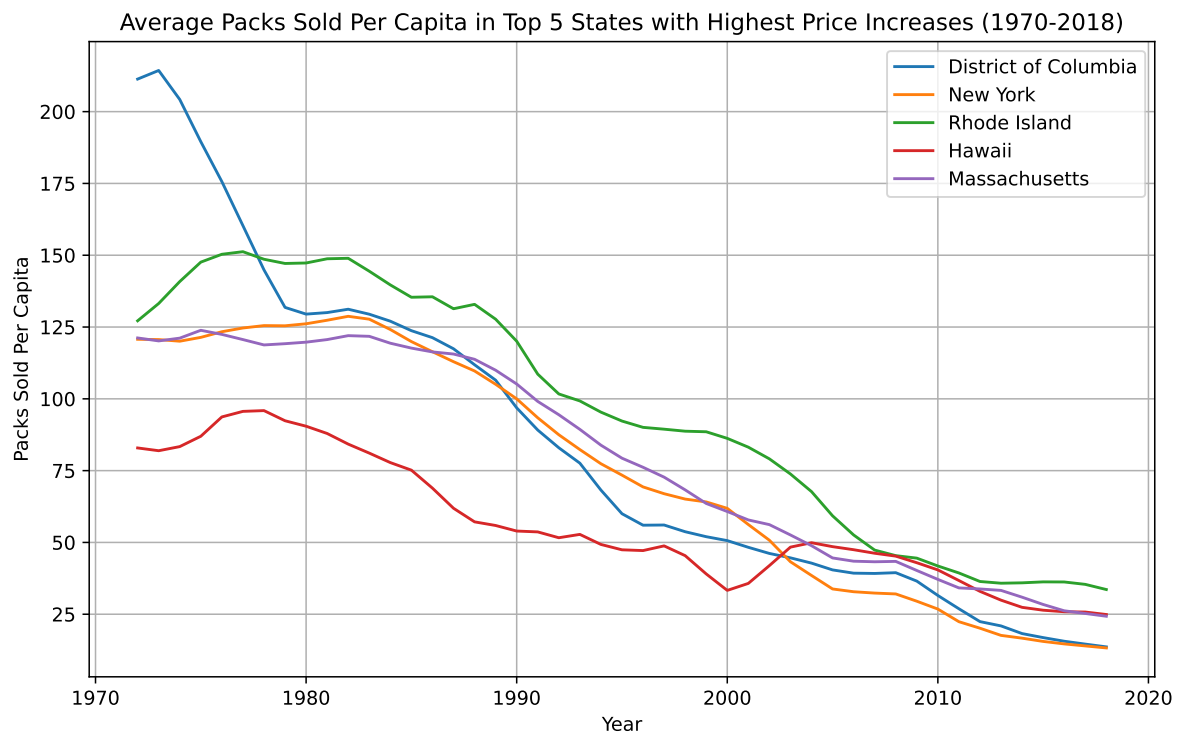
2

Cigarette Tax and Price in 2012 Dollars (1970-2018)



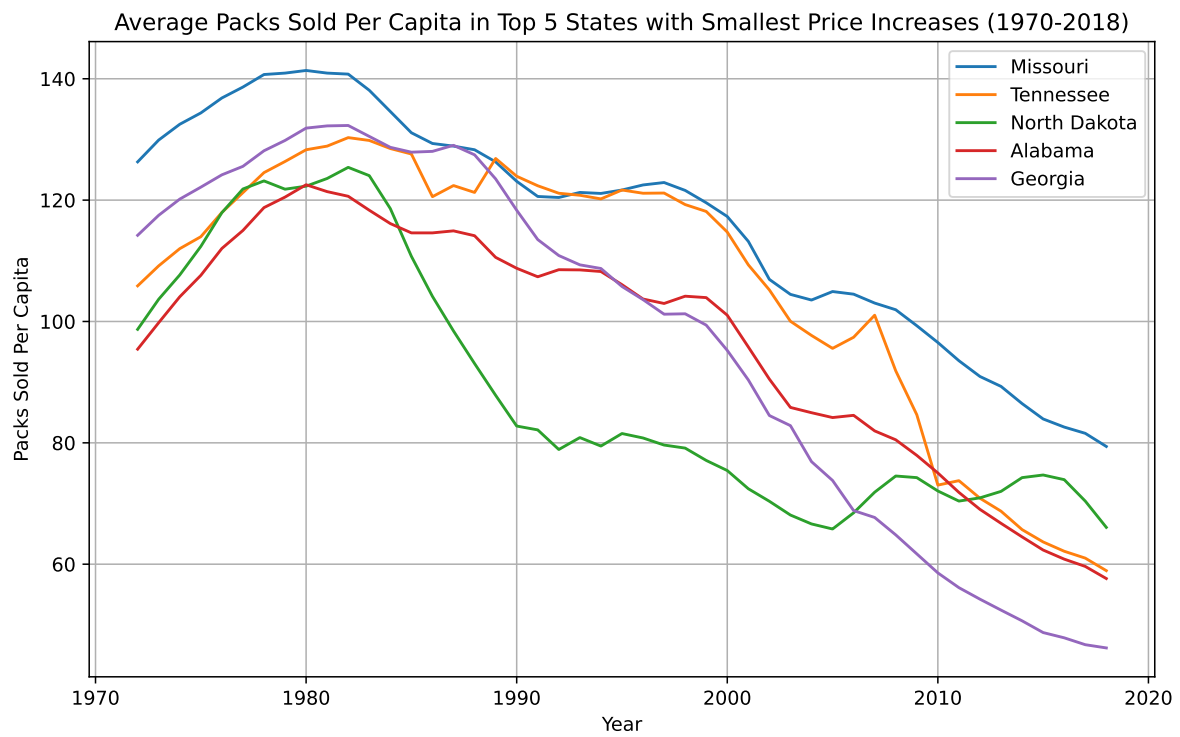
### 3

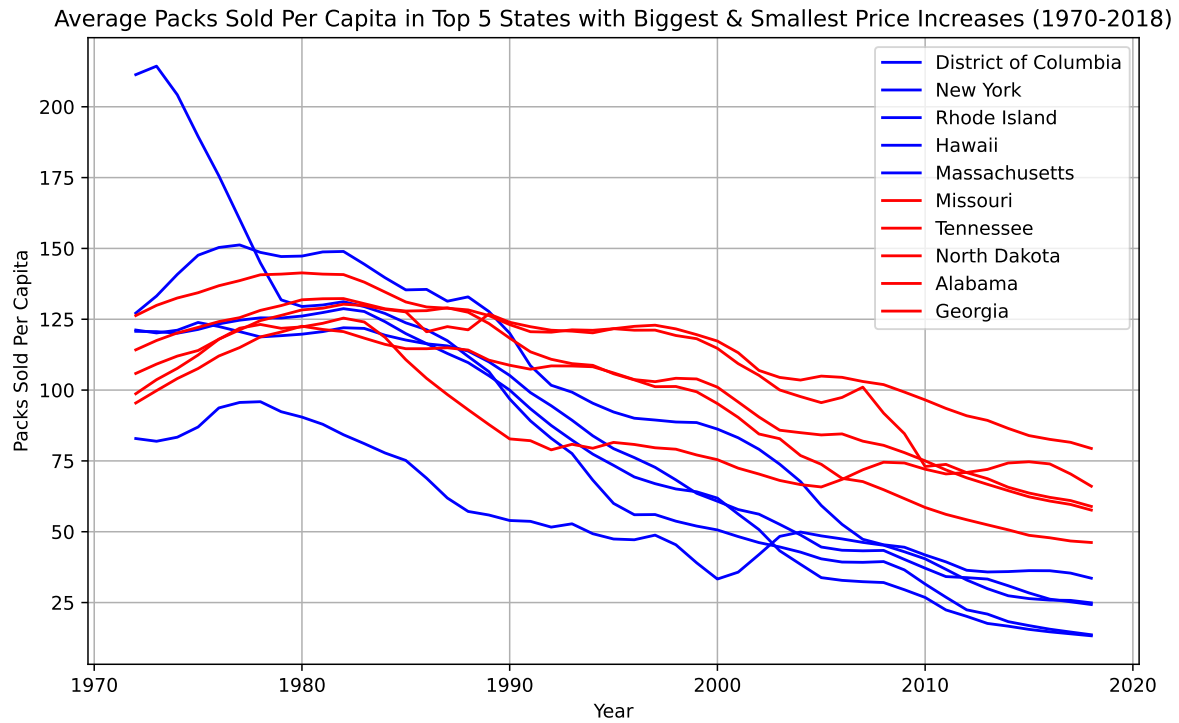
5 highest states in legend



4

5 lowest states in legend





Both start with similar sales per capita, but the states with the highest price increases have a steeper decline in sales per capita compared to the states with the smallest price increases. You can see this as they all end with a lower number of packs sold per capita than all the red (low price increase) states. This suggests that significant price increases do decrease cigarette sales per capita further. I like this graph more than the mean to show it is not just an average effect due to outlier but a comprehensive trend and the graph is still very clear with the colors.

## 7

OLS 1970-1990

OLS Regression Results						
=====						
Dep. Variable:	ln_sales	R-squared:	0.294			
Model:	OLS	Adj. R-squared:	0.293			
Method:	Least Squares	F-statistic:	445.1			
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	6.98e-83			
Time:	15:31:30	Log-Likelihood:	263.40			
No. Observations:	1071	AIC:	-522.8			
Df Residuals:	1069	BIC:	-512.8			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	5.4274	0.030	182.424	0.000	5.369	5.486
ln_price	-0.8094	0.038	-21.098	0.000	-0.885	-0.734
=====						
Omnibus:	89.160	Durbin-Watson:	0.183			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	466.536			
Skew:	0.128	Prob(JB):	4.93e-102			
Kurtosis:	6.223	Cond. No.	10.5			
=====						

Notes:

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

```
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/1708263625.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html
cig_data['ln_sales'] = np.log(cig_data['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/1708263625.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html
cig_data['ln_price'] = np.log(cig_data['price_cpi'])
```

```
/var/folders/mn/l2nrwsxn24g6yz6ygh2fxp40000gn/T/ipykernel_58123/1708263625.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html
cig_data['ln_total_tax'] = np.log(cig_data['tax_dollar'])
```

## 8a

first stage 1970-1990

First-stage Regression (ln\_price ~ ln\_total\_tax):

```

                                OLS Regression Results
=====
Dep. Variable:                  ln_price    R-squared:                  0.617
Model:                          OLS        Adj. R-squared:          0.617
Method:                        Least Squares  F-statistic:              1725.
Date:                          Tue, 18 Mar 2025  Prob (F-statistic):      2.80e-225
Time:                          15:31:30      Log-Likelihood:           1020.7
No. Observations:              1071         AIC:                     -2037.
Df Residuals:                  1069         BIC:                     -2027.
Df Model:                      1
Covariance Type:               nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          1.2337      0.012    105.076      0.000      1.211      1.257
ln_total_tax    0.3328      0.008     41.537      0.000      0.317      0.349
=====
Omnibus:                6.850    Durbin-Watson:                0.303
Prob(Omnibus):          0.033    Jarque-Bera (JB):            5.505
Skew:                   0.081    Prob(JB):                    0.0638
Kurtosis:               2.689    Cond. No.                     8.72
=====
```

Notes:

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



## 8b

second stage 1970-1990

Second-stage Regression (ln\_sales ~ pricehat):

OLS Regression Results						
Dep. Variable:	ln_sales	R-squared:		0.236		
Model:	OLS	Adj. R-squared:		0.235		
Method:	Least Squares	F-statistic:		330.3		
Date:	Tue, 18 Mar 2025	Prob (F-statistic):		1.56e-64		
Time:	15:31:30	Log-Likelihood:		221.17		
No. Observations:	1071	AIC:		-438.3		
Df Residuals:	1069	BIC:		-428.4		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	5.5138	0.039	141.021	0.000	5.437	5.591
0	-0.9231	0.051	-18.175	0.000	-1.023	-0.823
Omnibus:	83.338	Durbin-Watson:		0.157		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		430.014		
Skew:	0.023	Prob(JB):		4.20e-94		
Kurtosis:	6.104	Cond. No.		13.4		

Notes:

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

the value of OLS without the instrument is -0.809 and with the instrument is -0.923. This means that a 1% increase in price will decrease sales per capita by 0.81% or 0.92%. They are different and this is due to the endogeneity in the naive estimate. For example a state could increase the tax rate because it has a high smoking rate, and this would bias the estimate.

## 9a

OLS 1991-2015

OLS Regression Results						
=====						
Dep. Variable:	ln_sales	R-squared:	0.561			
Model:	OLS	Adj. R-squared:	0.561			
Method:	Least Squares	F-statistic:	1630.			
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	4.20e-230			
Time:	15:31:30	Log-Likelihood:	-256.00			
No. Observations:	1275	AIC:	516.0			
Df Residuals:	1273	BIC:	526.3			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	5.6600	0.036	155.560	0.000	5.589	5.731
ln_price	-0.9968	0.025	-40.370	0.000	-1.045	-0.948
=====						
Omnibus:	23.003	Durbin-Watson:	0.208			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	43.688			
Skew:	0.011	Prob(JB):	3.26e-10			
Kurtosis:	3.907	Cond. No.	9.34			
=====						

Notes:

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

```
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/3558264353.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html
cig_data2['ln_sales'] = np.log(cig_data2['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/3558264353.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html
cig_data2['ln_price'] = np.log(cig_data2['price_cpi'])
```

```

/var/folders/mn/l2nrwsxn24g6ywwz6ygh2fxp40000gn/T/ipykernel_58123/3558264353.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide](https://pandas.pydata.org/pandas-docs/stable/user_guide)

```

cig_data2['ln_total_tax'] = np.log(cig_data2['tax_dollar'])

```

## 9b

first stage 1991-2015

First-stage Regression (ln\_price ~ ln\_total\_tax):

OLS Regression Results						
=====						
Dep. Variable:	ln_price	R-squared:	0.868			
Model:	OLS	Adj. R-squared:	0.868			
Method:	Least Squares	F-statistic:	8390.			
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	0.00			
Time:	15:31:30	Log-Likelihood:	874.63			
No. Observations:	1275	AIC:	-1745.			
Df Residuals:	1273	BIC:	-1735.			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	1.4284	0.003	418.142	0.000	1.422	1.435
ln_total_tax	0.4317	0.005	91.598	0.000	0.422	0.441
=====						
Omnibus:	29.255	Durbin-Watson:	0.406			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	30.684			
Skew:	0.371	Prob(JB):	2.17e-07			
Kurtosis:	2.835	Cond. No.	1.38			
=====						

Notes:

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

## 9c

second stage 1991-2015

Second-stage Regression (ln\_sales ~ pricehat):

OLS Regression Results						
=====						
Dep. Variable:	ln_sales	R-squared:	0.608			
Model:	OLS	Adj. R-squared:	0.607			
Method:	Least Squares	F-statistic:	1972.			
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	6.43e-261			
Time:	15:31:30	Log-Likelihood:	-184.97			
No. Observations:	1275	AIC:	373.9			
Df Residuals:	1273	BIC:	384.2			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	5.8266	0.037	158.310	0.000	5.754	5.899
0	-1.1129	0.025	-44.405	0.000	-1.162	-1.064
=====						
Omnibus:	44.690	Durbin-Watson:	0.217			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	107.551			
Skew:	0.134	Prob(JB):	4.42e-24			
Kurtosis:	4.397	Cond. No.	9.99			
=====						

Notes:

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

## 10

Time Period	Model	ATE (Price Elasticity)	P-Value	Intercept	Standard Error	R <sup>2</sup>
1970-1990	OLS1	-0.809438	6.981617e-83	5.427381	0.038366	0.293983
1970-1990	IV1	-0.923078	1.564989e-64	5.513822	0.050789	0.236061
1991-2015	OLS2	-0.996814	4.195667e-230	5.659955	0.024692	0.561445
1991-2015	IV2	-1.112943	6.426875e-261	5.826574	0.025063	0.607682

The trend of the increase in effect shown by the IV estimate is consistent in both time periods. This is due to the same issues of endogeneity in both time periods. Comparing the two time periods, the effect of the price increase on sales per capita is larger in the second time period. This could be due to the fact that the taxes increased more steeply in the second time period. Another explanation for a higher elasticity in the second time period (less addictive/more price sensitive) could be that cultural values have shifted due to more education on the health risks of smoking or preferences. Another explanation could be that increases access to alternative like E-cigarettes or other smoking cessation products.

## attempt with pyfixest package

ATE results not as close still trying to see whats wrong. also cant get the table to display nicely working on all that., but trends are similar

```
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data1['ln_sales'] = np.log(data1['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data1['ln_price_2012'] = np.log(data1['price_cpi'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data1['ln_tax_2012'] = np.log(data1['tax_2012'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data2['ln_sales'] = np.log(data2['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data2['ln_price_2012'] = np.log(data2['price_cpi'])
/var/folders/mn/l2nrwsxn24g6yww6ygh2fxp40000gn/T/ipykernel_58123/2620428547.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/1dindexing.html#copy-on-write
data2['ln_tax_2012'] = np.log(data2['tax_2012'])
```

	est1	est2	est3	est4
depvar	ln_sales	ln_sales	ln_sales	ln_sales
ln_price_2012	-0.809*** \n (0.038)	-1.063*** \n (0.084)	-0.997*** \n (0.025)	-1.286*** \n (0.034)
Intercept	5.427*** \n (0.030)	5.620*** \n (0.064)	5.660*** \n (0.036)	6.075*** \n (0.050)
Observations	1071	1071	1275	1275
S.E. type	iid	iid	iid	iid
R2	0.294	-	0.561	-