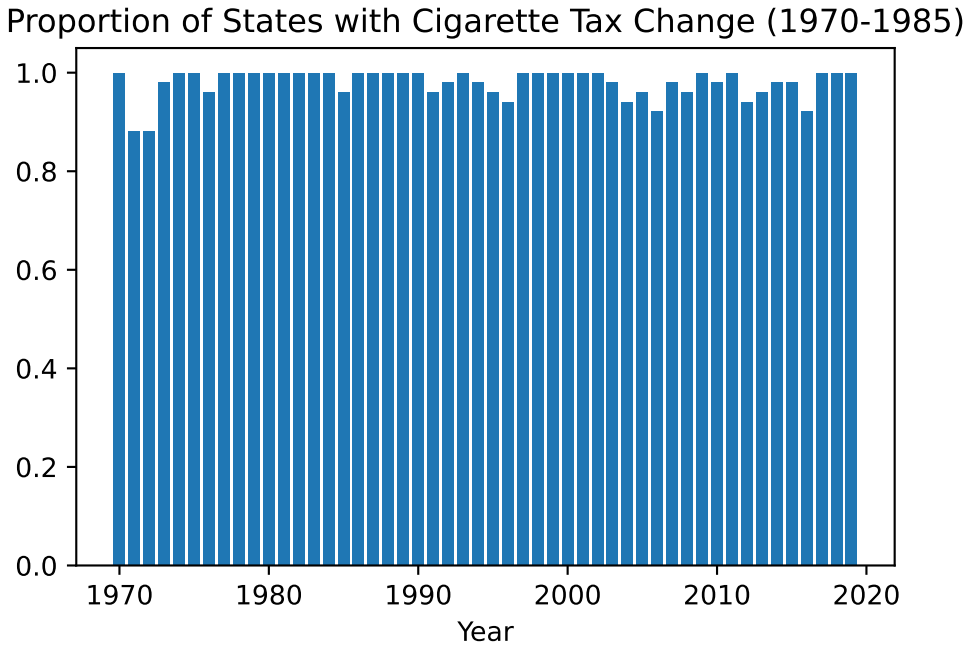
Homework 3

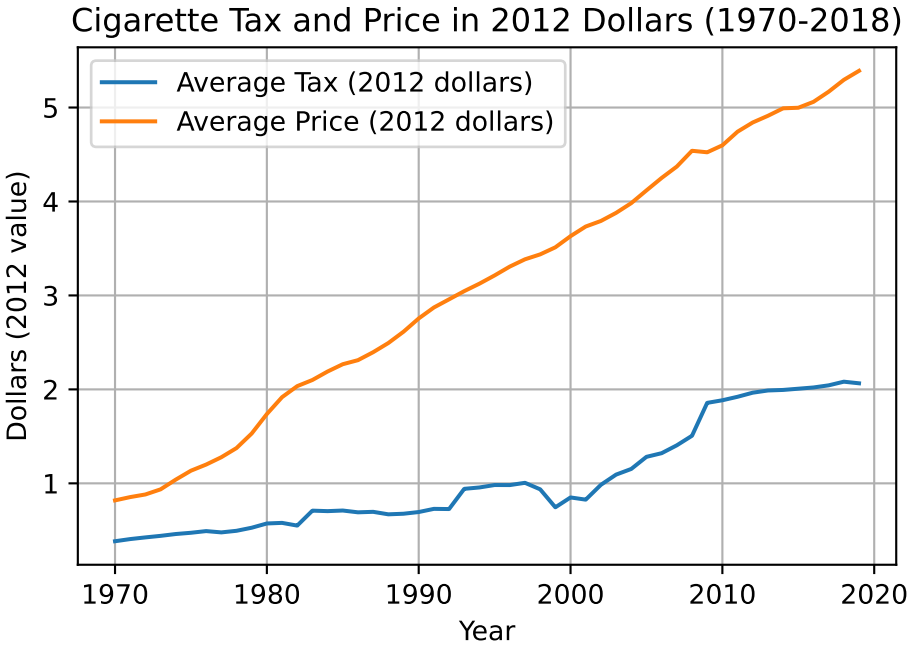
Research Methods, Spring 2025

Ryan Scholte

You can access the Repository # 1 Bar Graph

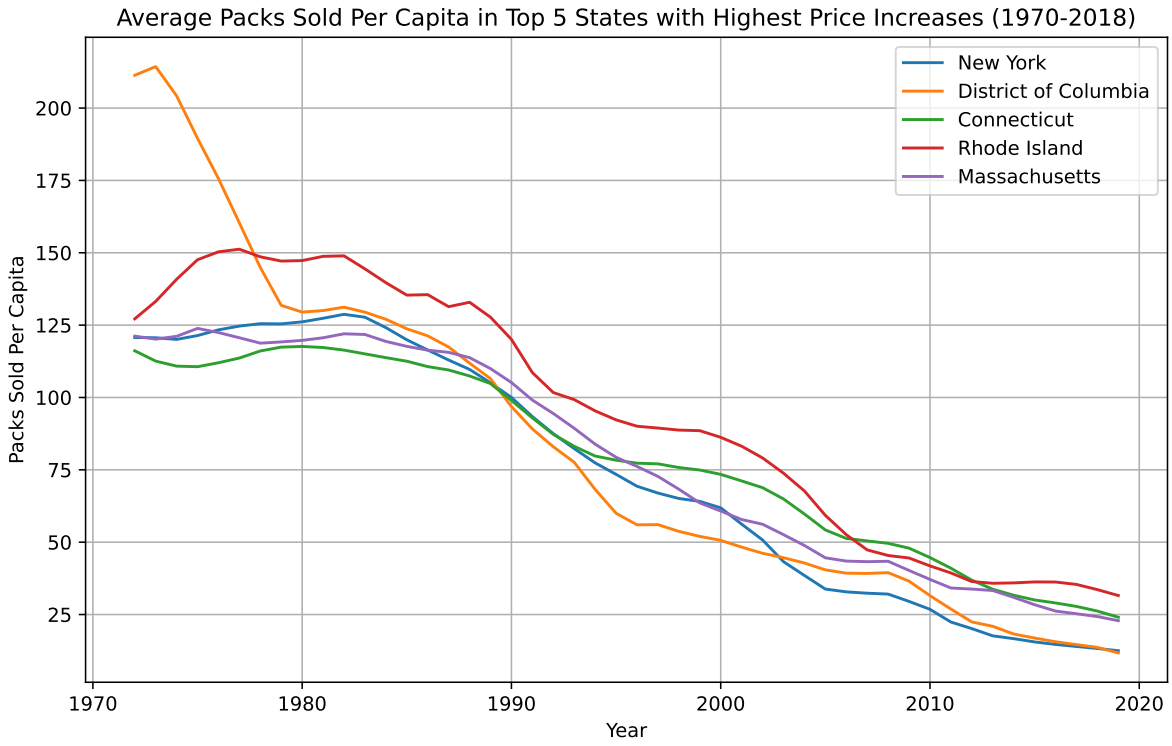


2



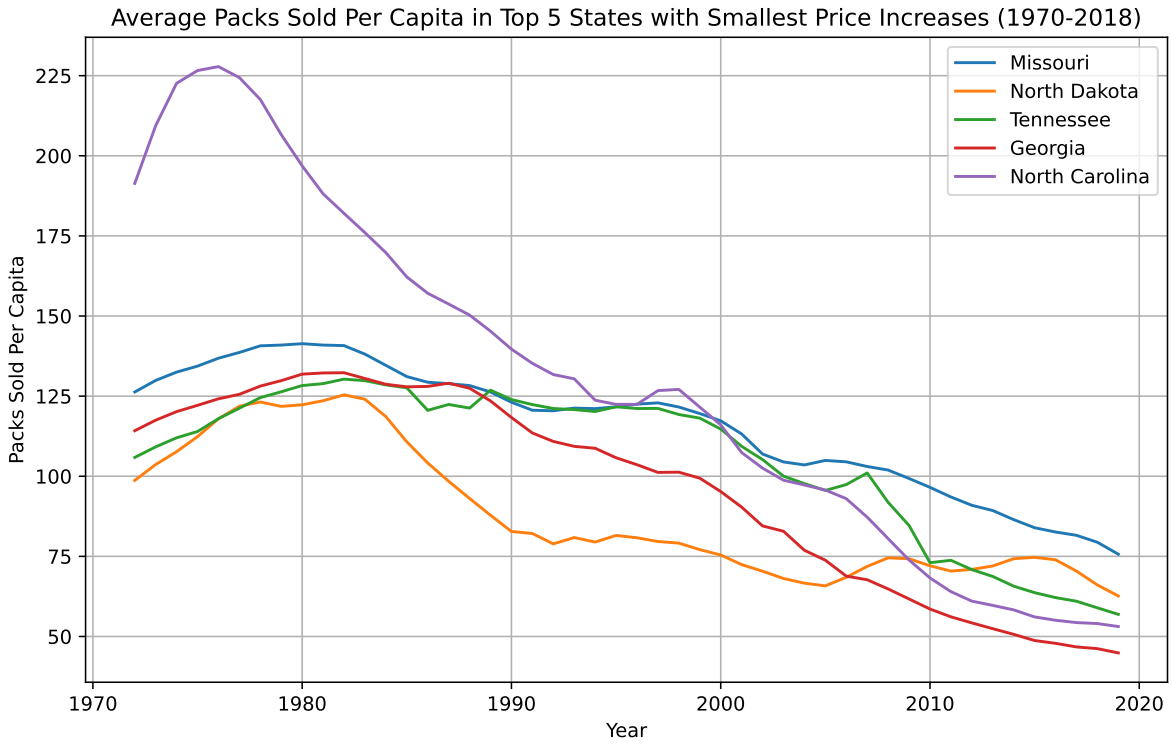
3

5 highest states in legend

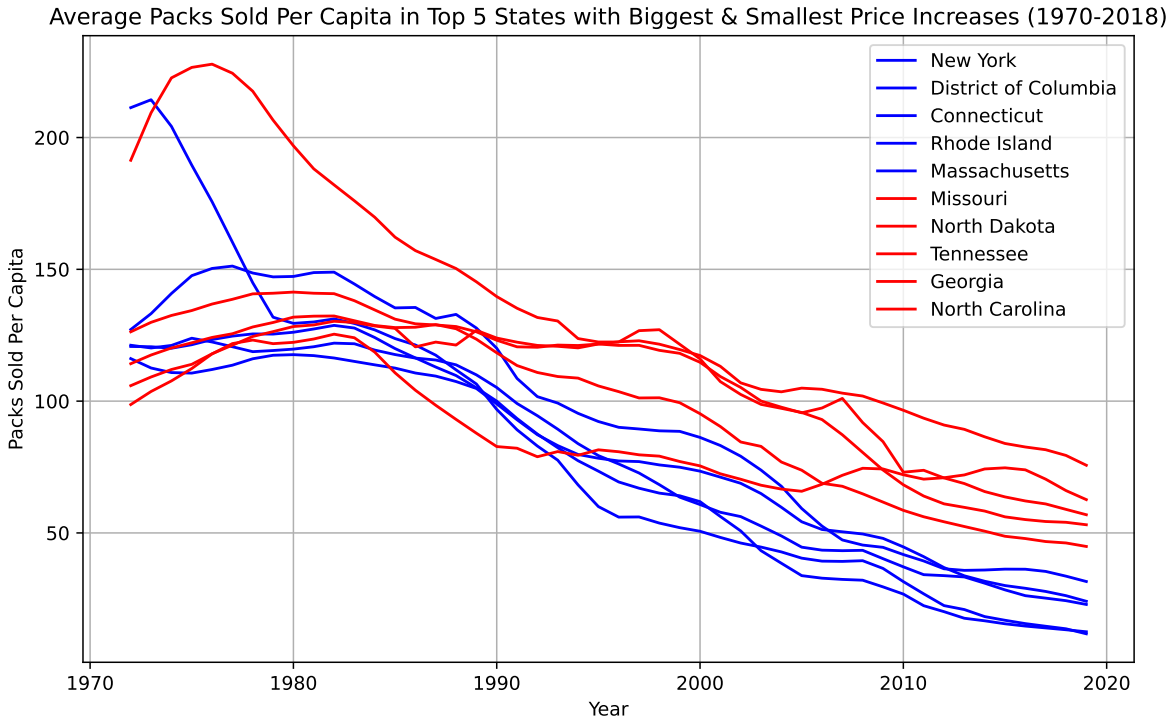


4

5 lowest states in legend


Average Packs Sold Per Capita in Top 5 States with Smallest Price Increases (1970-2018)
Missouri
North Dakota
Tennessee
Georgia
North Carolina
Packs Sold Per Capita
225
200
175
150
125
100
75
50
1970
1980
1990
2000
2010
2020
Year

5



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. This suggests that significant price increases do decrease cigarette sales per capita further.

7

OLS 1970-1990

OLS Regression Results

Dep. Variable:	ln_sales	R-squared:	0.126
Model:	OLS	Adj. R-squared:	0.125
Method:	Least Squares	F-statistic:	153.9
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	4.18e-33
Time:	14:52:19	Log-Likelihood:	148.99
No. Observations:	1071	AIC:	-294.0
Df Residuals:	1069	BIC:	-284.0
Df Model:	1		

```
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7504      0.008    585.321      0.000       4.734       4.766
ln_price      -0.1715      0.014    -12.404      0.000      -0.199      -0.144
==============================================================================
Omnibus:                       64.611   Durbin-Watson:                   0.139
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              224.414
Skew:                           0.173   Prob(JB):                     1.86e-49
Kurtosis:                       5.216   Cond. No.                         2.48
==============================================================================

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


/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/416662071.py:5: SettingWithC
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
  cig_data['ln_sales'] = np.log(cig_data['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/416662071.py:6: SettingWithC
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
  cig_data['ln_price'] = np.log(cig_data['cost_per_pack'])
/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/416662071.py:7: SettingWithC
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
  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:                       0.683
Model:                           OLS   Adj. R-squared:                  0.683
Method:                Least Squares   F-statistic:                     2301.
Date:               Tue, 18 Mar 2025   Prob (F-statistic):          8.21e-269
Time:                       14:52:19   Log-Likelihood:                -86.164
No. Observations:               1071   AIC:                             176.3
Df Residuals:                   1069   BIC:                             186.3
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1786      0.033     35.712      0.000       1.114       1.243
ln_total_tax   1.0803      0.023     47.973      0.000       1.036       1.125
==============================================================================
Omnibus:                       30.760   Durbin-Watson:                   0.408
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               32.668
Skew:                           0.421   Prob(JB):                     8.06e-08
Kurtosis:                       3.156   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:                       14:52:19   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          4.7101      0.008    573.443      0.000       4.694       4.726
0             -0.2843      0.016    -18.175      0.000      -0.315      -0.254
==============================================================================
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.                         2.98
==============================================================================

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

the value of OLS without the instrument is -0.17 and with the instrument is -0.28. This means that a 1% increase in price will decrease sales per capita by 0.17% or 0.28%. They are different and this is due the 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:                       0.533
Model:                           OLS   Adj. R-squared:                  0.532
Method:                Least Squares   F-statistic:                     1451.
Date:               Tue, 18 Mar 2025   Prob (F-statistic):          1.52e-212
Time:                       14:52:19   Log-Likelihood:                -296.47
No. Observations:               1275   AIC:                             596.9
Df Residuals:                   1273   BIC:                             607.2
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.0395      0.023    219.934      0.000       4.995       5.084
ln_price      -0.6656      0.017    -38.094      0.000      -0.700      -0.631
==============================================================================
Omnibus:                       19.351   Durbin-Watson:                   0.158
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               33.046
Skew:                           0.064   Prob(JB):                     6.67e-08
Kurtosis:                       3.778   Cond. No.                         5.37
==============================================================================

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


/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/1062089729.py:3: SettingWitl
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
  cig_data2['ln_sales'] = np.log(cig_data2['sales_per_capita'])
/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/1062089729.py:4: SettingWitl
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
  cig_data2['ln_price'] = np.log(cig_data2['cost_per_pack'])
```

```
/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_54710/1062089729.py:5: SettingWitl
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_guid
  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:                       0.869
Model:                           OLS   Adj. R-squared:                  0.869
Method:                Least Squares   F-statistic:                     8442.
Date:               Tue, 18 Mar 2025   Prob (F-statistic):               0.00
Time:                       14:52:19   Log-Likelihood:                396.65
No. Observations:               1275   AIC:                           -789.3
Df Residuals:                   1273   BIC:                           -779.0
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.2072      0.005    242.906      0.000       1.197       1.217
ln_total_tax   0.6300      0.007     91.881      0.000       0.617       0.643
==============================================================================
Omnibus:                       10.474   Durbin-Watson:                   0.330
Prob(Omnibus):                  0.005   Jarque-Bera (JB):               10.642
Skew:                           0.223   Prob(JB):                      0.00489
Kurtosis:                       2.965   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:	14:52:19	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	5.1575	0.022	231.116	0.000	5.114	5.201
0	-0.7626	0.017	-44.405	0.000	-0.796	-0.729

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.	5.71

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

10

Time Period	Model	ATE (Price Elasticity)
1970-1990	OLS	-0.171540
1970-1990	2SLS	-0.284348
1991-2015	OLS	-0.665626
1991-2015	2SLS	-0.762650

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 steaply in the second time period. Another explaination 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.