Question 1

```
ValueError: numpy.dtype size changed, may indicate binary incompatibility. Expected 96 from C
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Input In [2], in <cell line: 3>()
      1 #| echo: false
      2 #calling packages
----> 3 import pandas as pd
      4 import numpy as np
      5 import matplotlib.pyplot as plt
File ~/anaconda/lib/python3.9/site-packages/pandas/__init__.py:22, in <module>
     19 del hard_dependencies, dependency, missing_dependencies
     21 # numpy compat
---> 22 from pandas.compat import is_numpy_dev as _is_numpy_dev
     24 try:
     25     from pandas._libs import hashtable as _hashtable, lib as _lib, tslib as _tslib
File ~/anaconda/lib/python3.9/site-packages/pandas/compat/__init__.py:15, in <module>
     12 import sys
     14 from pandas._typing import F
---> 15 from pandas.compat.numpy import (
     16     is_numpy_dev,
     17     np_version_under1p19,
     18     np_version_under1p20,
     19 )
     20 from pandas.compat.pyarrow import (
     21     pa_version_under1p01,
     22     pa_version_under2p0,
     23     pa_version_under3p0,
     24     pa_version_under4p0,
     25 )
     27 PY39 = sys.version_info >= (3, 9)
File ~/anaconda/lib/python3.9/site-packages/pandas/compat/numpy/__init__.py:4, in <module>
      1 """ support numpy compatibility across versions """
      2 import numpy as np
```

```
----> 4 from pandas.util.version import Version
      6 # numpy versioning
      7 _np_version = np.__version__
File ~/anaconda/lib/python3.9/site-packages/pandas/util/__init__.py:1, in <module>
----> 1 from pandas.util._decorators import (  # noqa:F401
      2     Appender,
      3     Substitution,
      4     cache_readonly,
      5 )
      7 from pandas.core.util.hashing import (  # noqa:F401
      8     hash_array,
      9     hash_pandas_object,
     10 )
     13 def __getattr__(name):
File ~/anaconda/lib/python3.9/site-packages/pandas/util/_decorators.py:14, in <module>
      6 from typing import (
      7     Any,
      8     Callable,
      9     Mapping,
     10     cast,
     11 )
     12 import warnings
---> 14 from pandas._libs.properties import cache_readonly  # noqa:F401
     15 from pandas._typing import F
     18 def deprecate(
     19     name: str,
     20     alternative: Callable[..., Any],
   (...)
     25     msg: str | None = None,
     26 ) -> Callable[[F], F]:
File ~/anaconda/lib/python3.9/site-packages/pandas/_libs/__init__.py:13, in <module>
      1 __all__ = [
      2     "NaT",
      3     "NaTType",
   (...)
      9     "Interval",
     10 ]
---> 13 from pandas._libs.interval import Interval
     14 from pandas._libs.tslibs import (
     15     NaT,
     16     NaTType,
   (...)
     21     iNaT,
```

```
    22 )
File ~/anaconda/lib/python3.9/site-packages/pandas/_libs/interval.pyx:1, in init pandas._libs
ValueError: numpy.dtype size changed, may indicate binary incompatibility. Expected 96 from C
```

```
#summarize the data
#question 1

#finding prop of states that had change in tax from 1970-1986
tax_data = tax_data.sort_values(by=['state', 'Year'])
tax_data_1 = tax_data[(tax_data['Year']>=1970) & (tax_data['Year']<=1985)]
tax_data_1['tax_change'] = tax_data_1.groupby('state')['tax_state'].diff().ne(0).astype(int)
tax_change = tax_data_1.groupby('Year')['tax_change'].mean()


plt.bar(tax_change.index, tax_change.values)
plt.title('Proportion of States with Change in Cigarette Tax (1970-1985)')
plt.xlabel('Year')
plt.show()
```

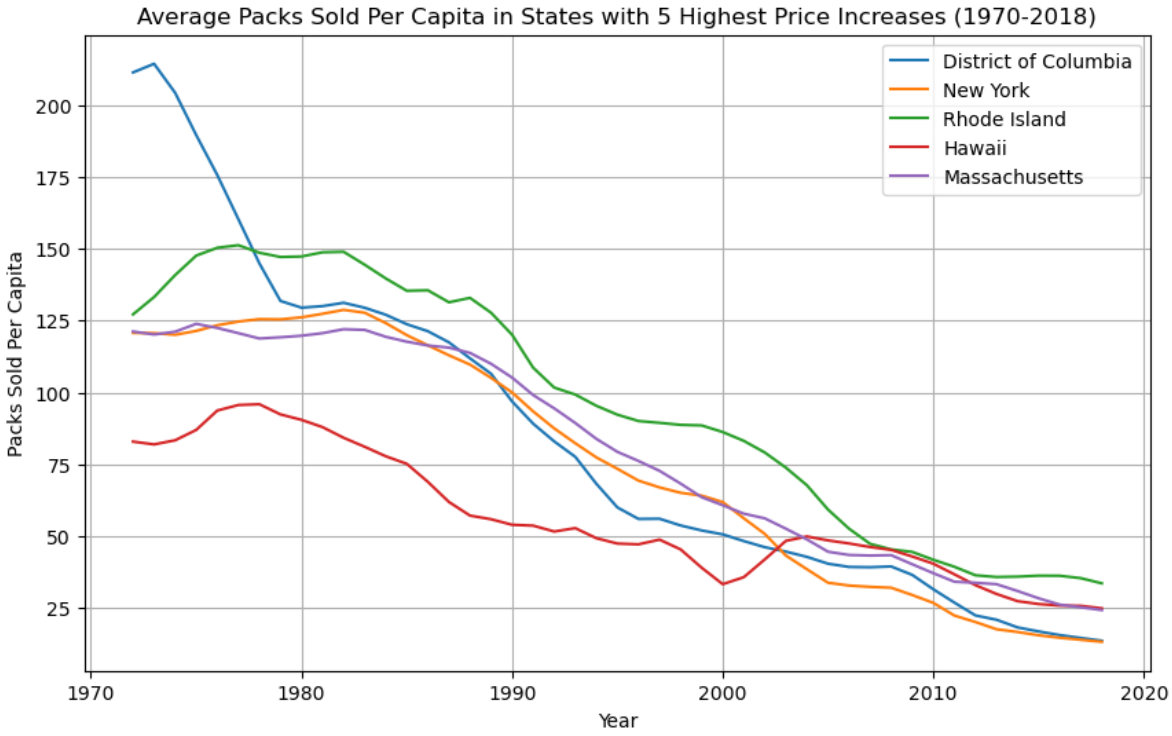
```
NameError: name 'tax_data' is not defined
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [3], in <cell line: 6>()
      1 #| echo: false
      2 #summarize the data
      3 #question 1
      4
      5 #finding prop of states that had change in tax from 1970-1986
----> 6 tax_data = tax_data.sort_values(by=['state', 'Year'])
      7 tax_data_1 = tax_data[(tax_data['Year']>=1970) & (tax_data['Year']<=1985)]
      8 tax_data_1['tax_change'] = tax_data_1.groupby('state')['tax_state'].diff().ne(0).asty
NameError: name 'tax_data' is not defined
```

Question 2

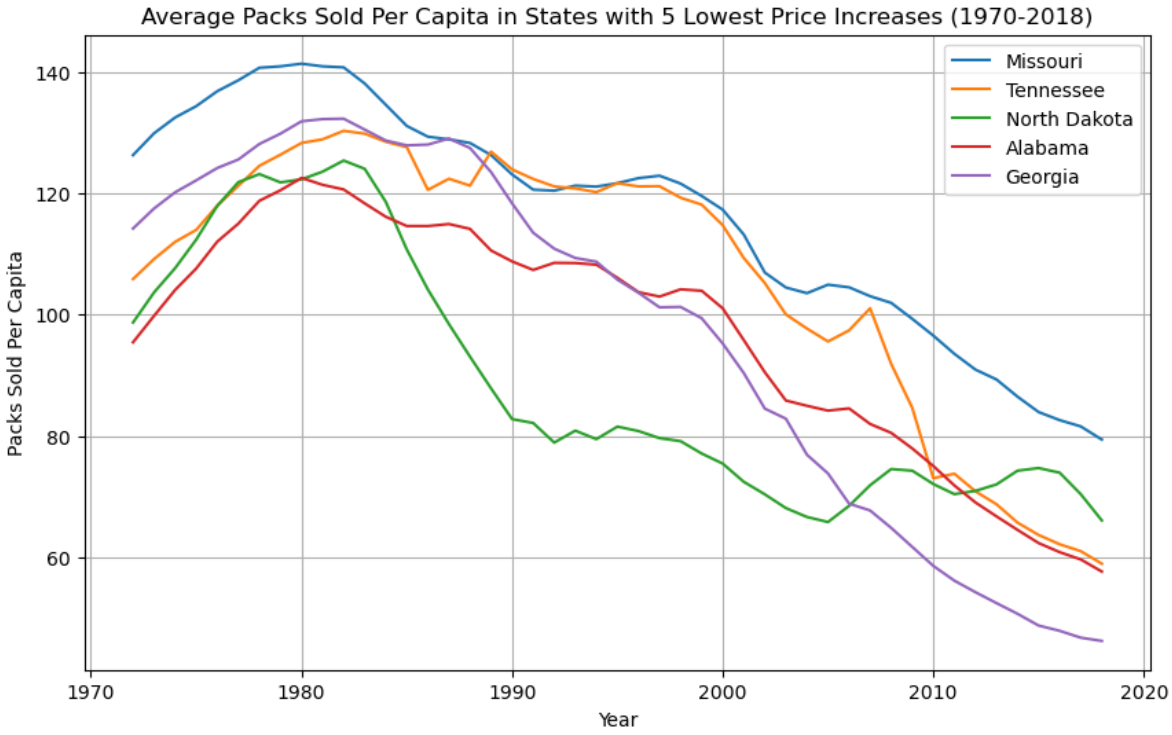
```
NameError: name 'tax_data' is not defined
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [5], in <cell line: 8>()
      1 #| echo: false
      2 #question 2
      3 #in 2012 $$$
```

```
      4 #cpi_2012 = tax_data.loc[tax_data['Year'] == 2012, 'price_cpi'].iloc[0]
      5 #tax_data['tax_dollar'] = tax_data['tax_dollar'] * (cpi_2012 / tax_data['price_cpi'])
      6 #tax_data['price_per_pack_2012'] = tax_data['cost_per_pack'] * (cpi_2012 / tax_data[
----> 8 avg_values = tax_data.groupby('Year')[['tax_dollar', 'price_cpi']].mean()
     10 #plot graph
     11 plt.plot(avg_values.index, avg_values['tax_dollar'], label='Average Tax (2012 dollars
NameError: name 'tax_data' is not defined
```

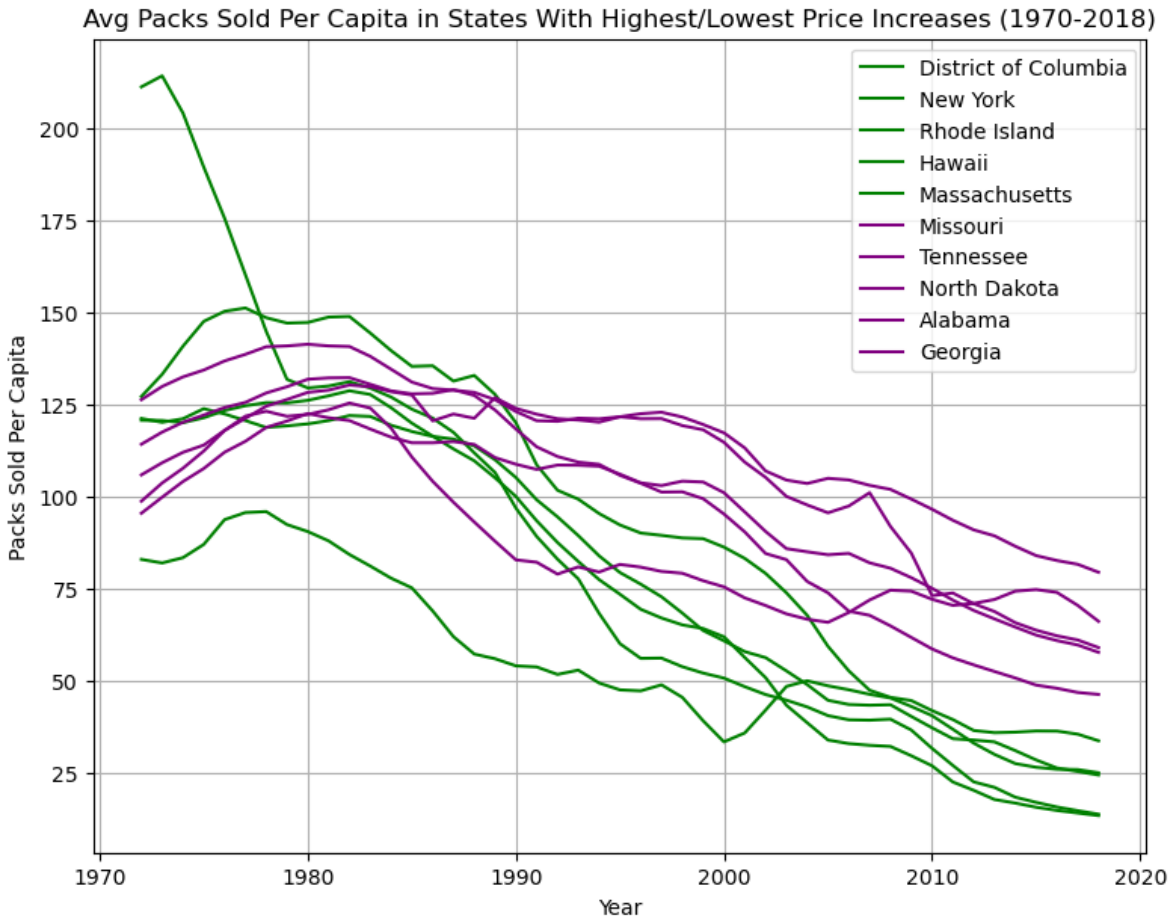
Question 3



Question 4


Average Packs Sold Per Capita in States with 5 Lowest Price Increases (1970-2018)
Missouri
Tennessee
North Dakota
Alabama
Georgia
Packs Sold Per Capita
140
120
100
80
60
1970
1980
1990
2000
2010
2020
Year

Question 5



Add some analysis for question 5

Question 6

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:	14:23:41	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          5.3854      0.028    193.692      0.000       5.331       5.440
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.                         10.0
==============================================================================

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


/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/475737347.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_guid
  cig_data['ln_sales'] = np.log(cig_data['sales_per_capita'])
/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/475737347.py:8: 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_guid
  cig_data['ln_total_tax'] = np.log(cig_data['tax_dollar'])
/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/475737347.py:9: 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_guid
  cig_data['ln_price'] = np.log(cig_data['price_cpi'])
```

Question 7 and 8

```
First stage (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:	14:23:49	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	1.1819	0.012	100.663	0.000	1.159	1.205
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.

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:	14:23:57	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	5.4660	0.037	149.749	0.000	5.394	5.538

```
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.                         12.7
==============================================================================

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

Question 9

```
                            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:                        14:24:03   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          5.6083      0.035    159.600      0.000       5.539       5.677
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.                         8.90
==============================================================================

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

/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/2081332474.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_2['ln_price'] = np.log(cig_data_2['price_cpi'])
/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/2081332474.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_2['ln_sales'] = np.log(cig_data_2['sales_per_capita'])
/var/folders/2q/wzjp_2kd355b8clhzqwmytb40000gn/T/ipykernel_4432/2081332474.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_2['ln_total_tax'] = np.log(cig_data_2['tax_dollar'])

First-stage  (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:                       14:24:11   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          1.3766      0.003    402.973      0.000       1.370       1.383
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.
```

```
Reduced Form 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:                       14:24:23   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          5.7689      0.036    162.335      0.000       5.699       5.839
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.                         9.52
==============================================================================

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

Question 10

```
#QUESTION 10
# summarize results into one table
summary_table = pd.DataFrame({
    'Time Period': ['1970-1990', '1970-1990', '1991-2015', '1991-2015'],
    'Model': ['OLS', '2SLS', 'OLS', '2SLS'],
    'ATE (Price Elasticity)': [
        model.params['ln_price'],
        Reduced_stage.params[0],
        reg2.params['ln_price'],
        Reduced_stage_2.params[0]
```

```
    ]
})


print(summary_table.to_string(index=False, line_width=80))
```

```
Time Period Model  ATE (Price Elasticity)
  1970-1990   OLS               -0.809438
  1970-1990  2SLS               -0.923078
  1991-2015   OLS               -0.996814
  1991-2015  2SLS               -1.112943
```