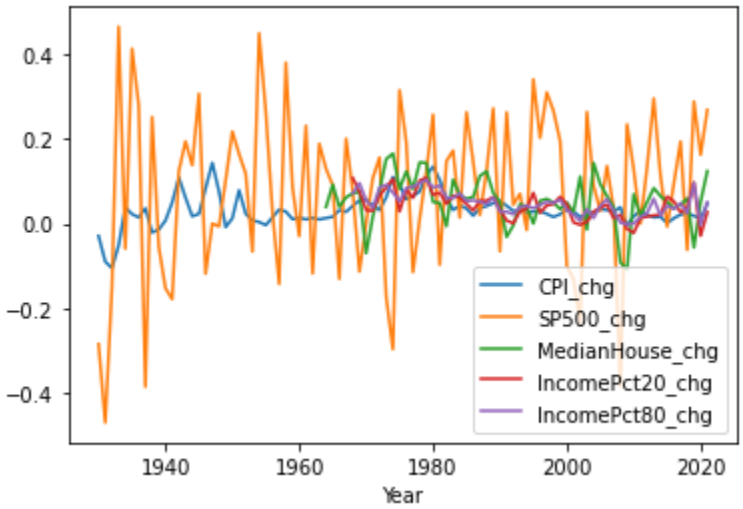
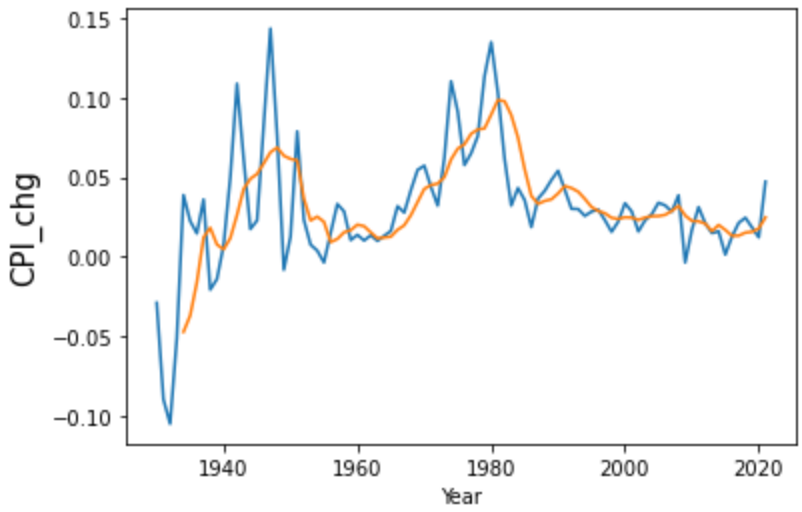
1. Compute descriptive statistics for each rate of change. What observations can you make about the annual rate of change in these columns?

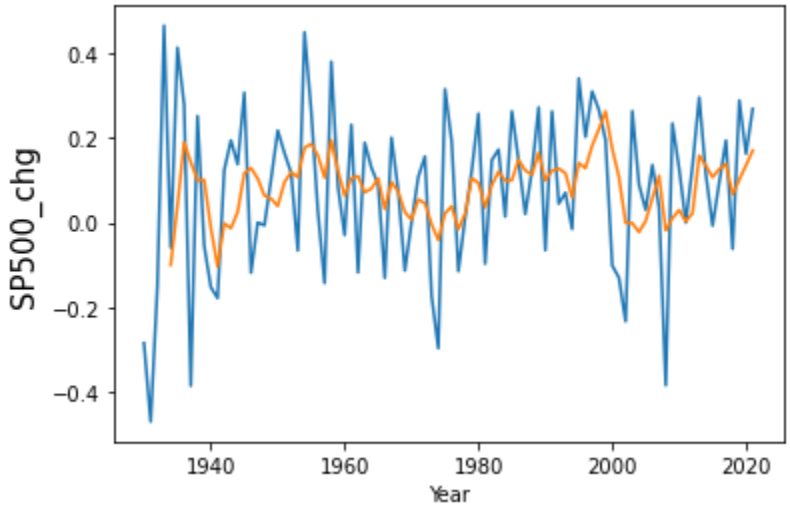


It's hard to label a general trend or line of best fit in all these data sets but all of them do have some loose correlation and varying annual rates of change. The SP500 changes wildly every couple of years, while income experiences a relatively moderate annual change. Positive and negative spikes in the SP500 also coincide with similar ripples in the other variables.

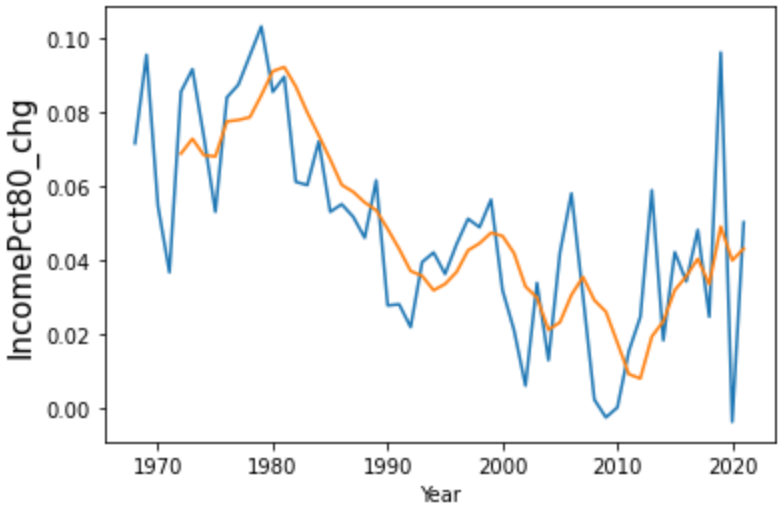
2. Create plots of the annual rate of change, and the 5-year rolling mean, for the `'CPI'`, `'SP500'`, and `'IncomePct80'` data series.
   For each plot, describe in words how well/closely the rolling mean tracks the individual rates of change. Does the rolling mean appear representative representative of these values?



There is a moderate relationship between the annual rates of change for CPI and its rolling mean, so the rolling mean is somewhat representative of CPI_chg.



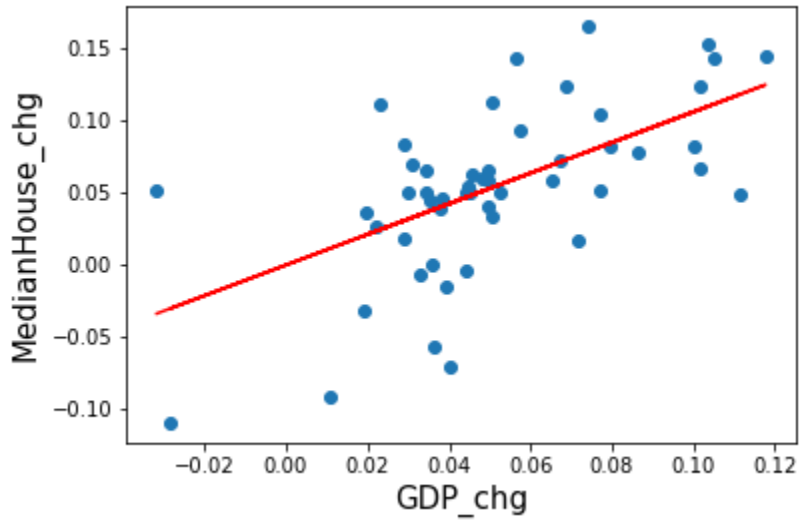
Out of all of these annual rates of change, SP500 seems to be the strongest. The rolling mean nearly echoes all of its falls and rises.



Here is the weakest relationship between the rolling mean and its respective annual rates. The rolling mean follows long term trends in the income over the span of 5-10 year intervals, but doesn't follow shorter instances of change. The only exception is the late 2010s to 2020 where the income spiked and dipped super hard.

3. Create a slice of data from 1970 to 2020. Using regression analysis, try using different independent variables to explain the rate of change in each of the dependent variables that follow.

a. the annual change in the price of a median house



['GDP_chg']

OLS Regression Results

==========

Dep. Variable: MedianHouse_chg R-squared (uncentered): 0.654

Model: OLS Adj. R-squared (uncentered): 0.647

Method: Least Squares F-statistic: 94.36

Date: Tue, 21 Nov 2023 Prob (F-statistic): 4.23e-13

Time: 17:30:17 Log-Likelihood: 84.330

No. Observations: 51 AIC: -166.7

Df Residuals: 50 BIC: -164.7

Df Model: 1

Covariance Type: nonrobust

==========

	coef	std err	t	P>|t|	[0.025	0.975]
GDP_chg	1.0614	0.109	9.714	0.000	0.842	1.281

Omnibus:	0.950	Durbin-Watson:	1.601
Prob(Omnibus):	0.622	Jarque-Bera (JB):	0.644
Skew:	-0.275	Prob(JB):	0.725
Kurtosis:	2.998	Cond. No.	1.00

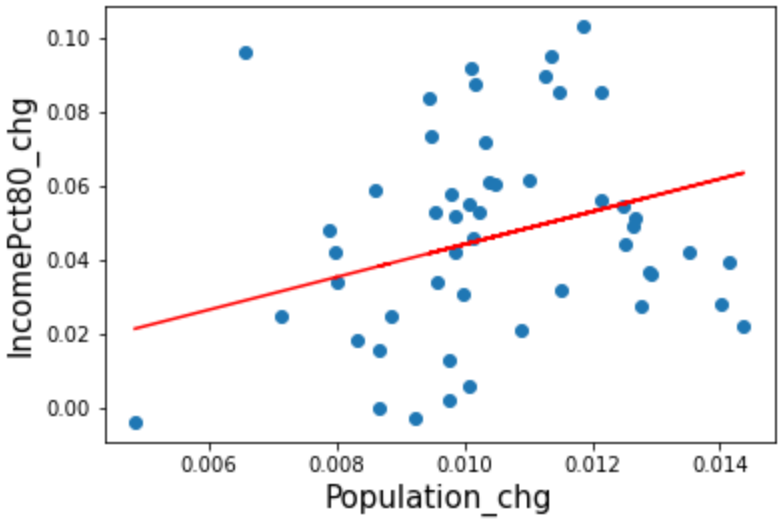
Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

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

The R2 value for annual rates of change of GDP vs. Median House price is 0.654, meaning that GDP has a 65.4% chance of predicting Median House price changes.

b. the annual change in the income for the 80th percentile of the income distribution



['IncomePct60_chg']

OLS Regression Results

==========

Dep. Variable: IncomePct80_chg R-squared (uncentered): 0.977

Model: OLS Adj. R-squared (uncentered): 0.976

Method: Least Squares F-statistic: 2116.

Date: Tue, 21 Nov 2023 Prob (F-statistic): 1.37e-42

Time: 17:37:29 Log-Likelihood: 172.25

No. Observations: 51 AIC: -342.5

Df Residuals: 50 BIC: -340.6

Df Model: 1

Covariance Type: nonrobust

==========

coef std err t P>|t| [0.025 0.975]

IncomePct60_chg 1.0521 0.023 46.003 0.000 1.006 1.098

==========

Omnibus: 0.835 Durbin-Watson: 2.141

Prob(Omnibus): 0.659 Jarque-Bera (JB): 0.804

Skew: -0.035 Prob(JB): 0.669

Kurtosis: 2.389 Cond. No. 1.00

==========

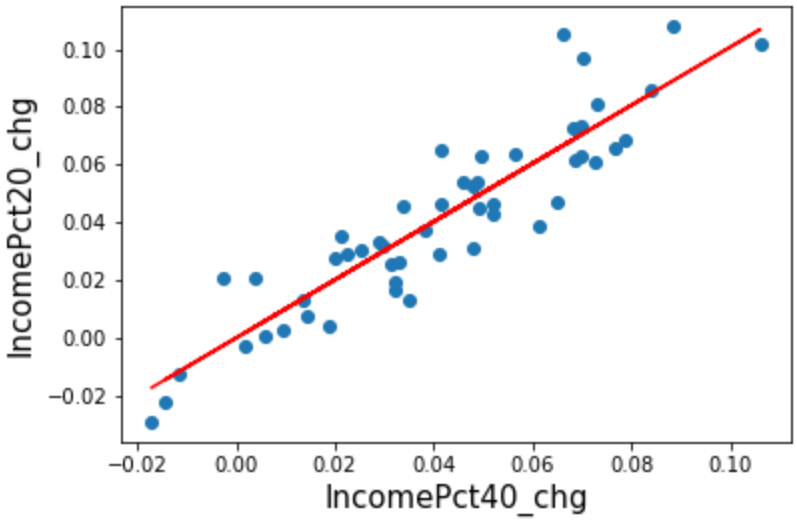
Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

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

The R2 value for the annual rates of change for the 60th percentile vs. 80th percentile of income is 0.977, making the 60th percentile of income a near perfect predictor of the 80th percentile.

c. the annual change in the income for the 20th percentile of the income distribution



['IncomePct40_chg']

OLS Regression Results

==========

Dep. Variable: IncomePct20_chg R-squared (uncentered): 0.938

Model: OLS Adj. R-squared (uncentered): 0.937

Method: Least Squares F-statistic: 756.5

Date: Tue, 21 Nov 2023 Prob (F-statistic): 7.47e-32

Time: 17:43:06 Log-Likelihood: 149.64

No. Observations: 51 AIC: -297.3

Df Residuals: 50 BIC: -295.3

Df Model: 1

Covariance Type: nonrobust

==========

coef std err t P>|t| [0.025 0.975]

IncomePct40_chg 1.0070 0.037 27.504 0.000 0.933 1.081

==========

Omnibus: 5.175 Durbin-Watson: 2.569

Prob(Omnibus): 0.075 Jarque-Bera (JB): 4.203

Skew: 0.677 Prob(JB): 0.122

Kurtosis: 3.381 Cond. No. 1.00

==========

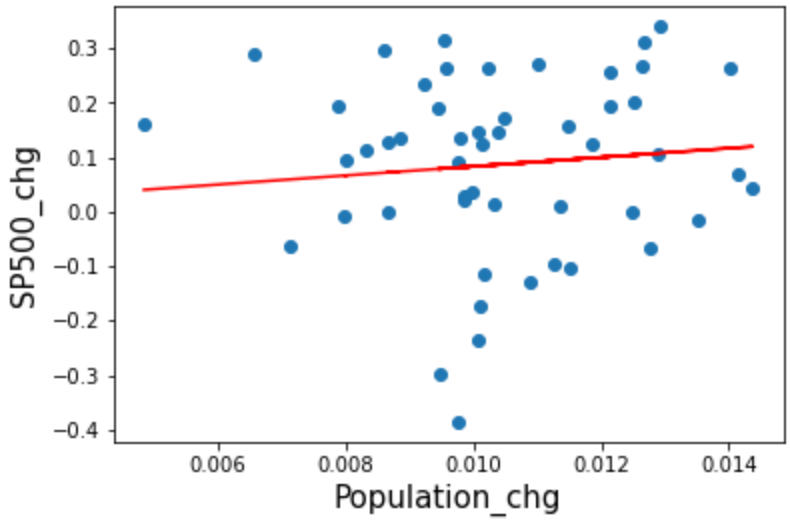
Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

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

The R2 value for the annual rates of change for the 40th percentile vs. 20th percentile of income is 0.938, making the 40th percentile 93.8% accurate in predicting the 20th percentile.

d. the annual change in the stock market (S&P 500 index)



['Population_chg']

OLS Regression Results

==========

Dep. Variable: SP500_chg R-squared (uncentered): 0.229

Model: OLS Adj. R-squared (uncentered): 0.213

Method: Least Squares F-statistic: 14.83

Date: Tue, 21 Nov 2023 Prob (F-statistic): 0.000335

Time: 17:50:07 Log-Likelihood: 20.249

No. Observations: 51 AIC: -38.50

Df Residuals: 50 BIC: -36.57

Df Model: 1

Covariance Type: nonrobust

==========

coef std err t P>|t| [0.025 0.975]

Population_chg 8.3423 2.166 3.851 0.000 3.992 12.693

==========

Omnibus: 5.043 Durbin-Watson: 2.035

Prob(Omnibus): 0.080 Jarque-Bera (JB): 4.400

Skew: -0.717 Prob(JB): 0.111

Kurtosis: 3.121 Cond. No. 1.00

==========

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

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

The regression for population vs. SP500 has an R2 value of 0.229. This is the highest R2 value out of all the other independent variables, but it is still a poor predictor of SP500.