CalPERS question

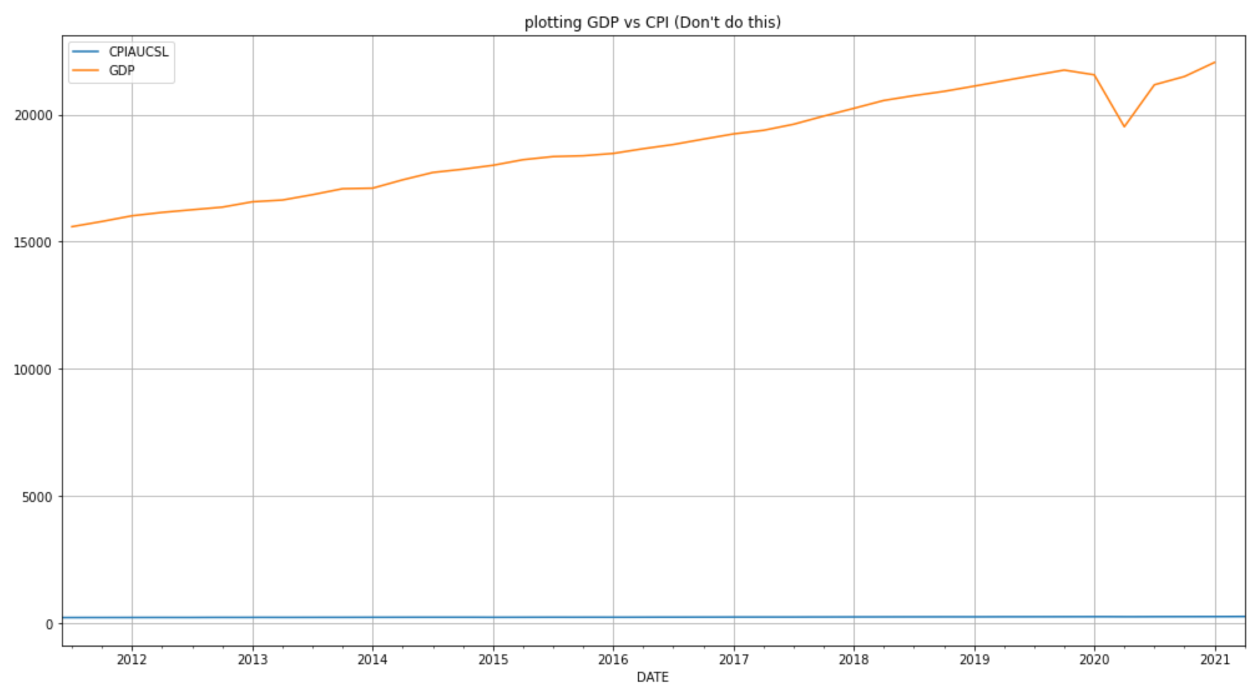
Question: Describe to me how you would test statistically for a long-run relationship and what pitfalls will you be looking out for in the results, and how might you remedy these pitfalls?

gdp and inflation

I used python code to pull historical data for US GDP and inflation from the St. Louis Federal Reserve

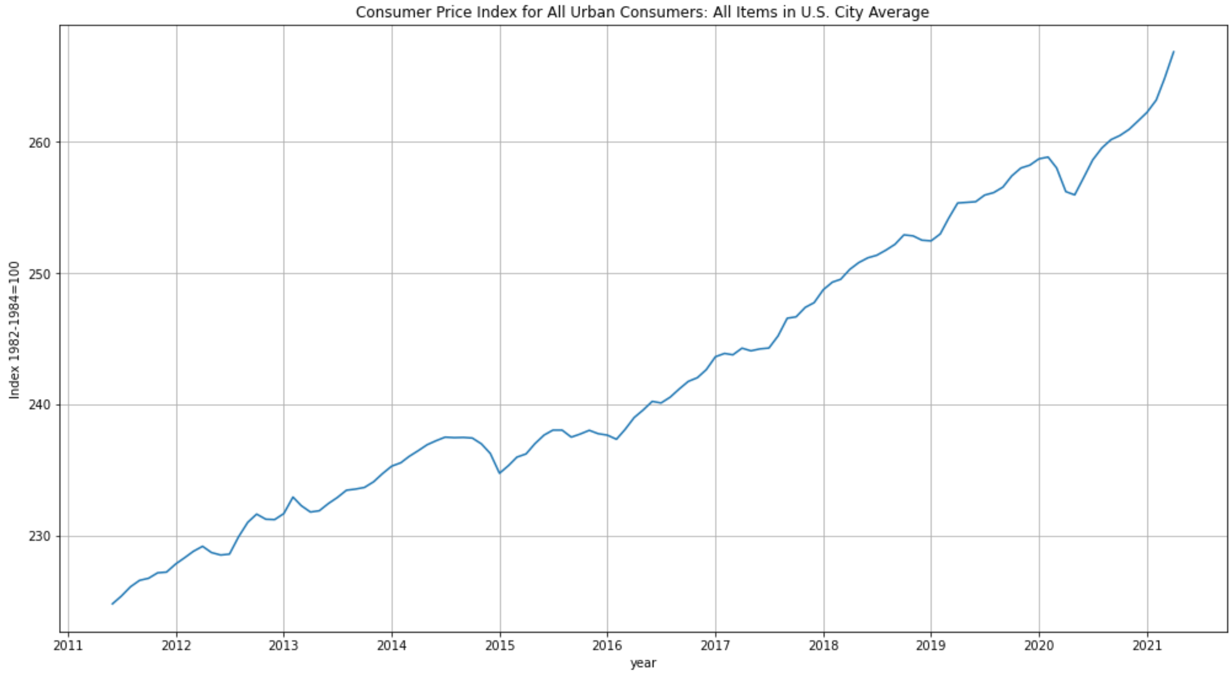
- Gross Domestic Product: Billions of Dollars, Seasonally Adjusted Annual Rate (Quarterly)
- CPI: Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (Index 1982-1984=100, Seasonally Adjusted, (Monthly))

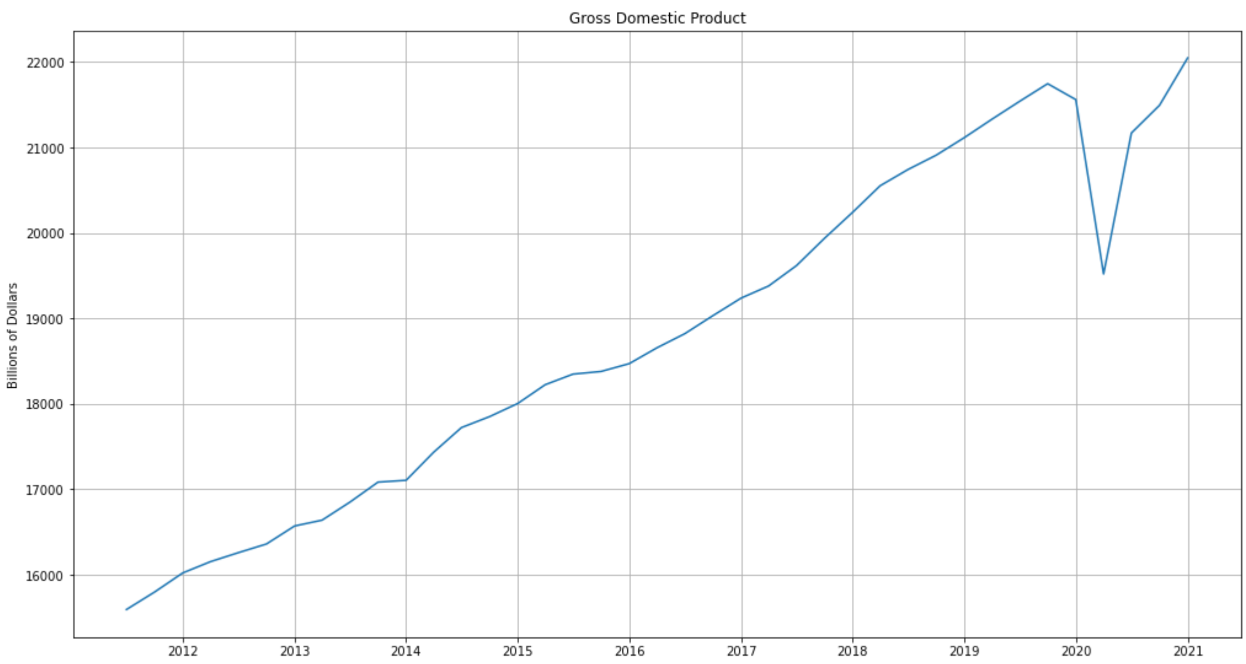
The first pitfall that I saw was that if you plot them on the same graph because they are between different values they plots will look weird.



CalPERS question

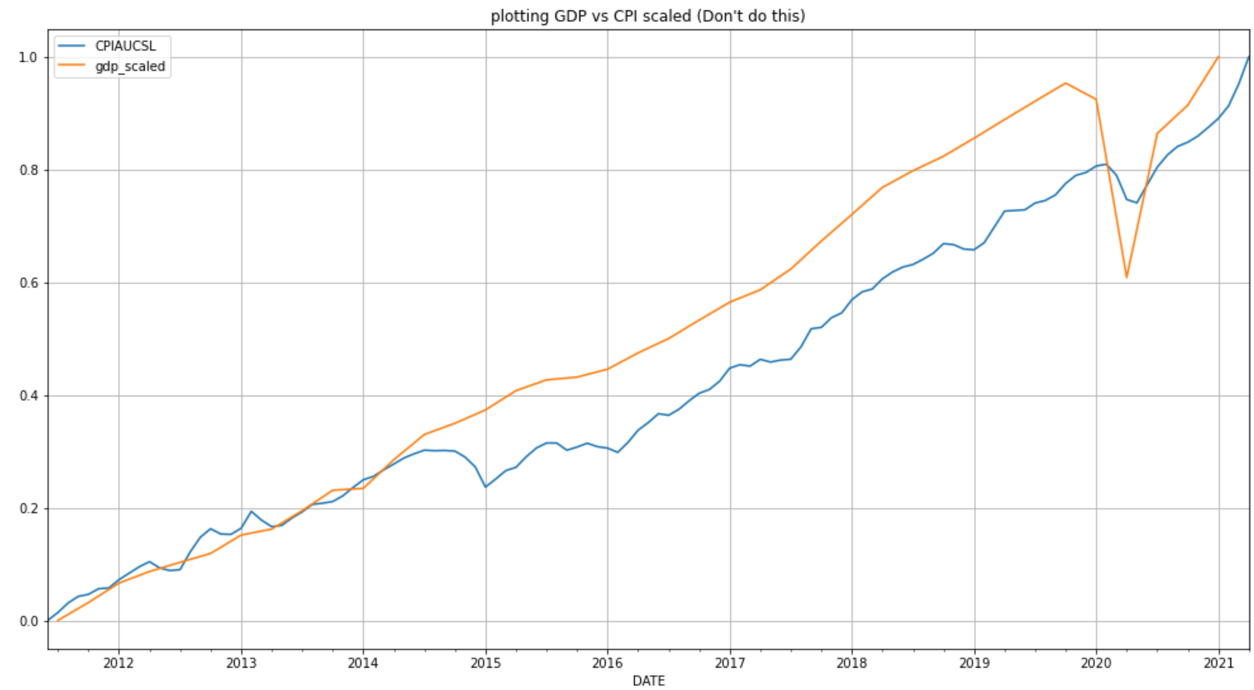
Instead its better to plot the graphs on top of each other



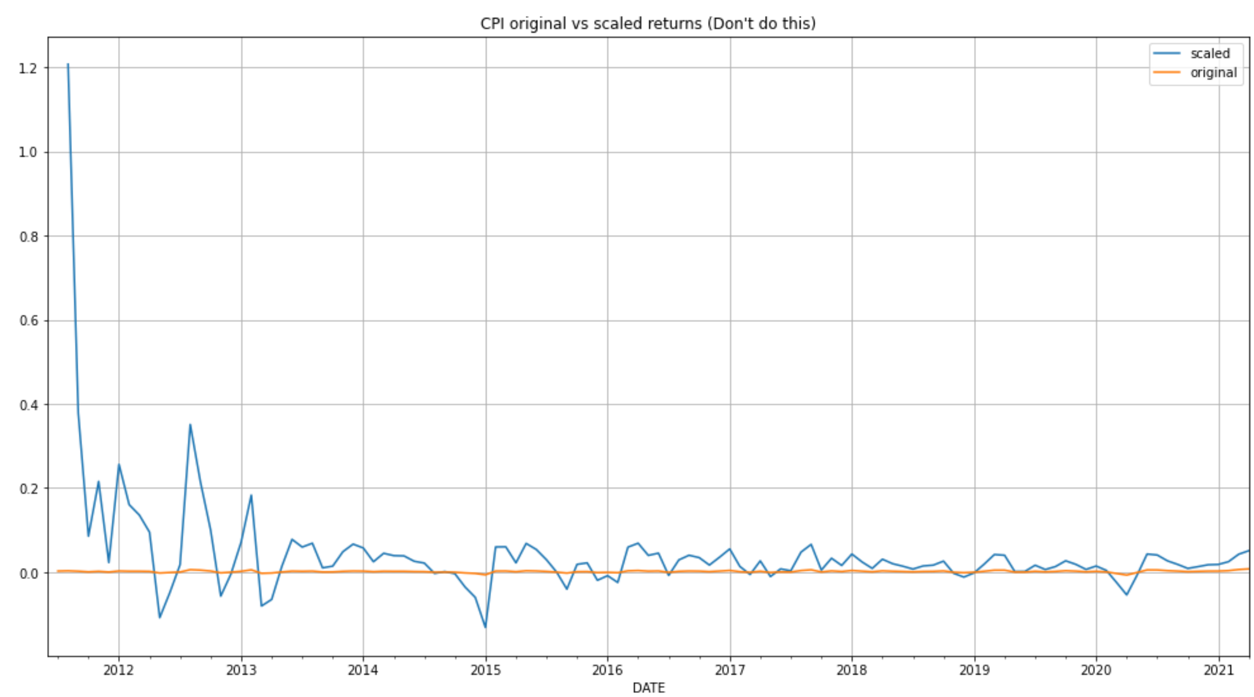


CalPERS question

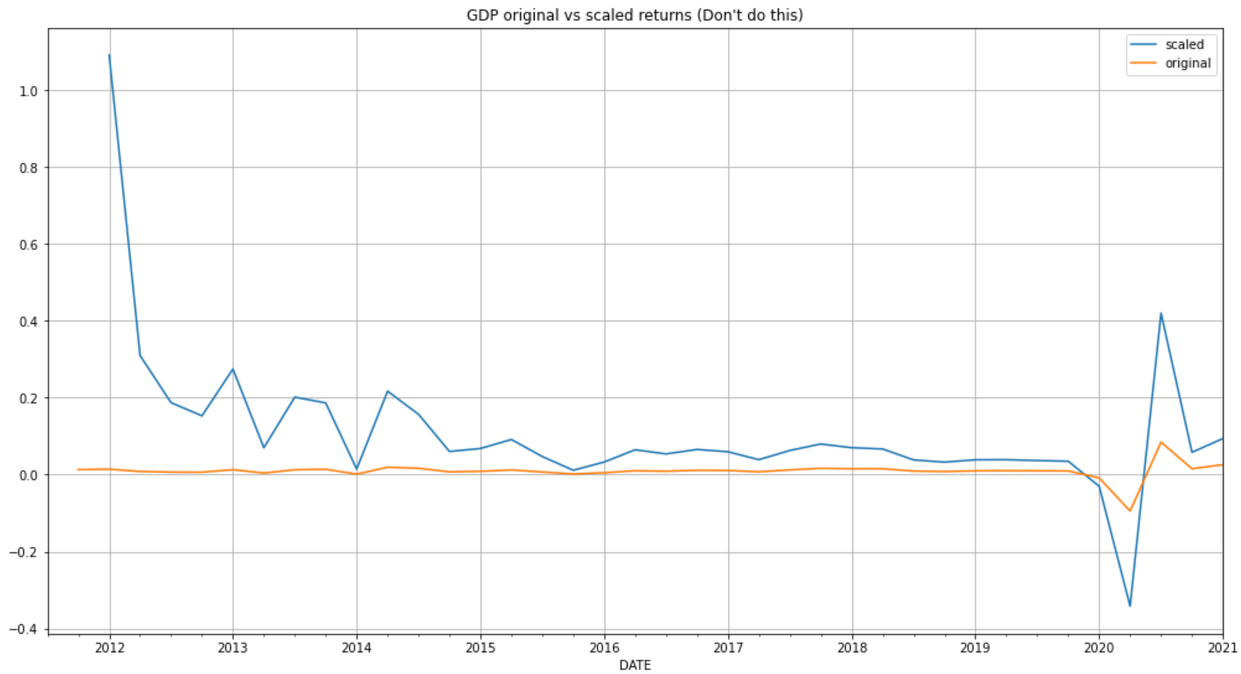
Another “remedy” that wouldn’t work is to scale the data. The problem with scaling is that it brings the peaks and troughs smaller or bigger. If we normalize the data it looks like.



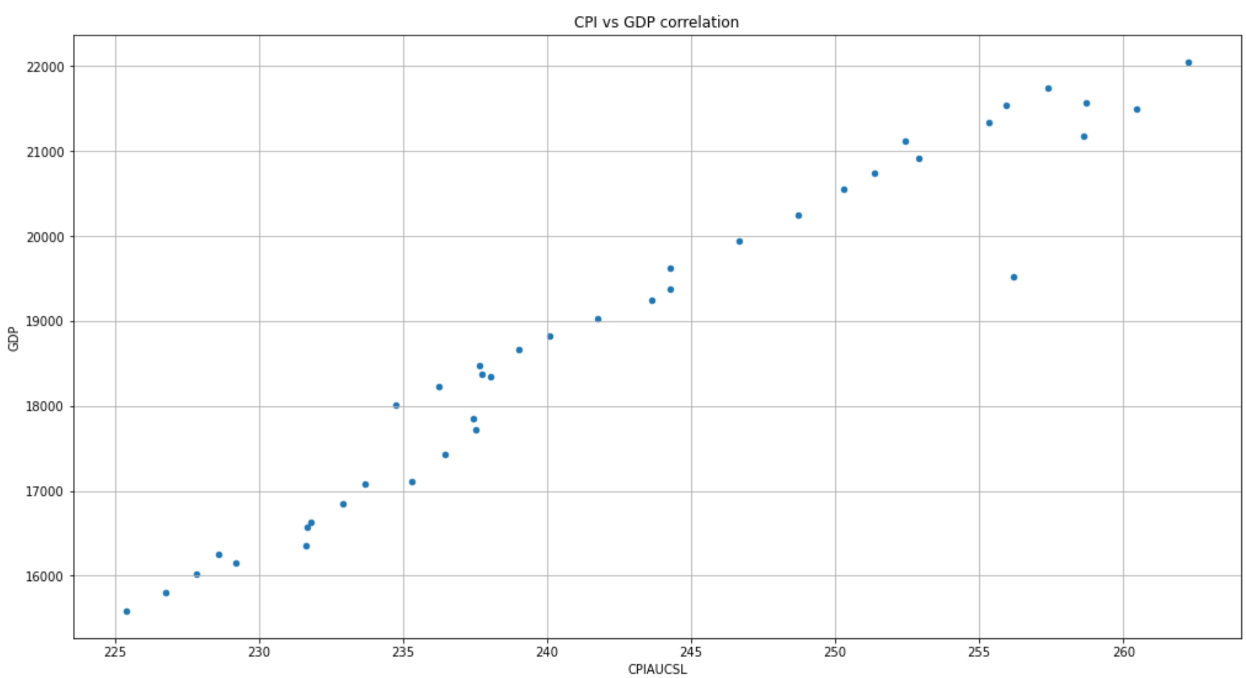
It is evident that the scaled vs normal time series definitely alters the data if we look at the percent changes.



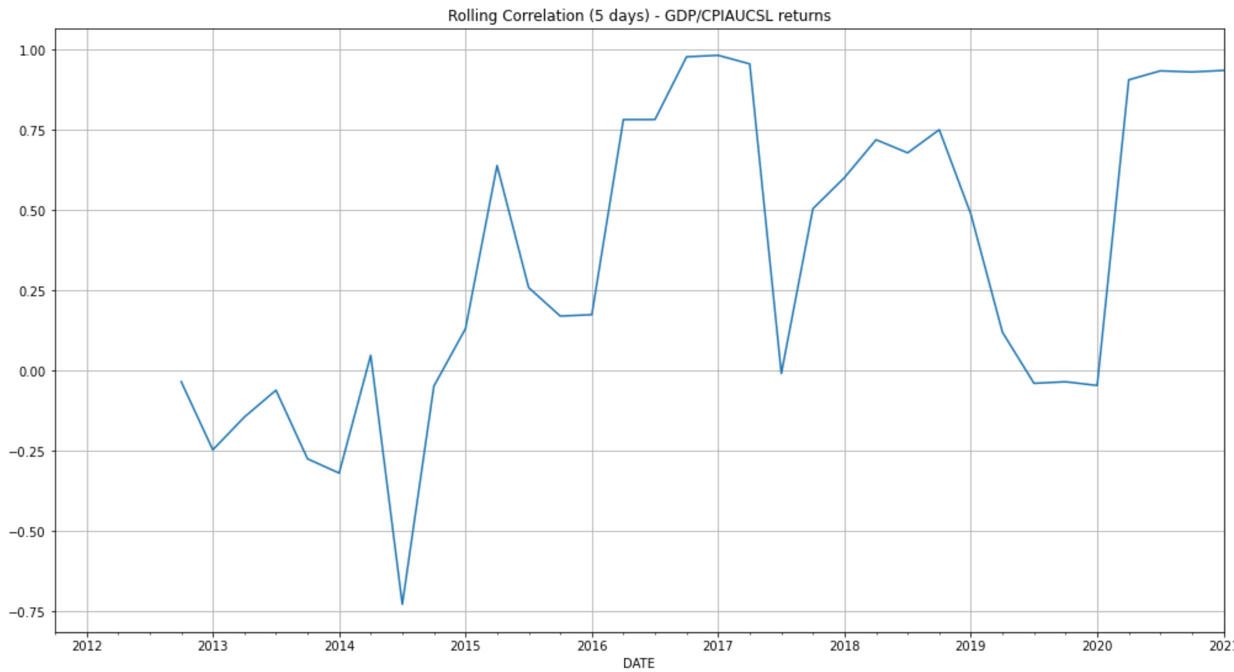
CalPERS question



We can test the correlation by using a scatter plot



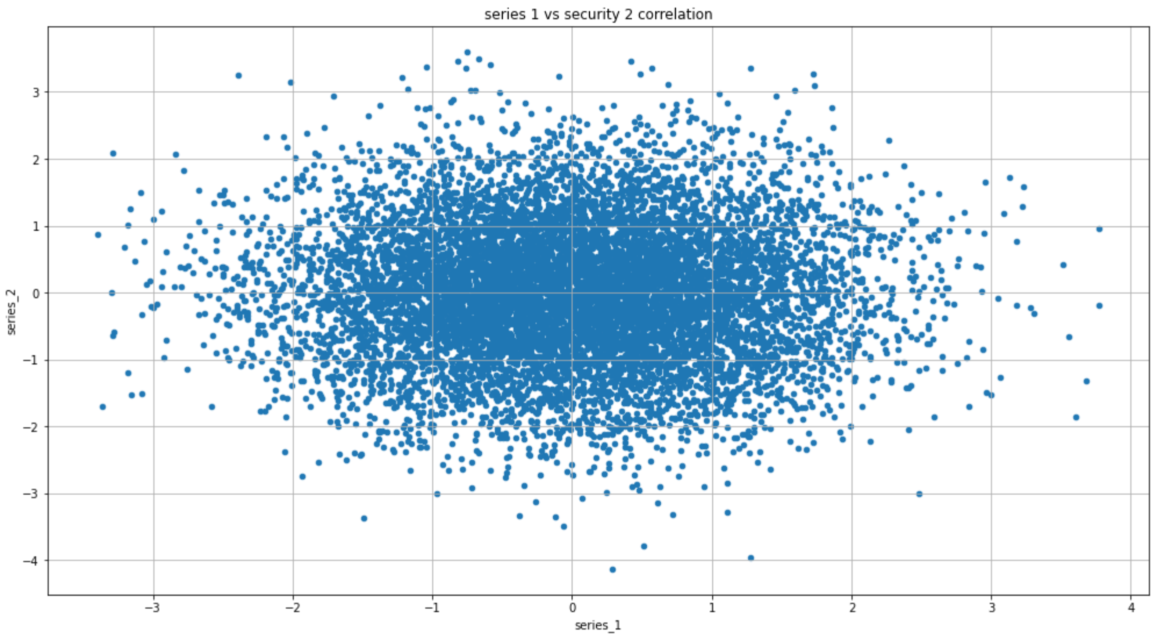
We can see that there is positive correlation, but to make this dataframe we only kept points that matched on the index. Because CPI and GDP are quoted at different frequencies (GDP = quarterly, CPI = monthly), we essentially made CPI data quarterly. This isn't technically wrong but less desirable. We can always look at the rolling correlation over the day, in this case I used 5 days.



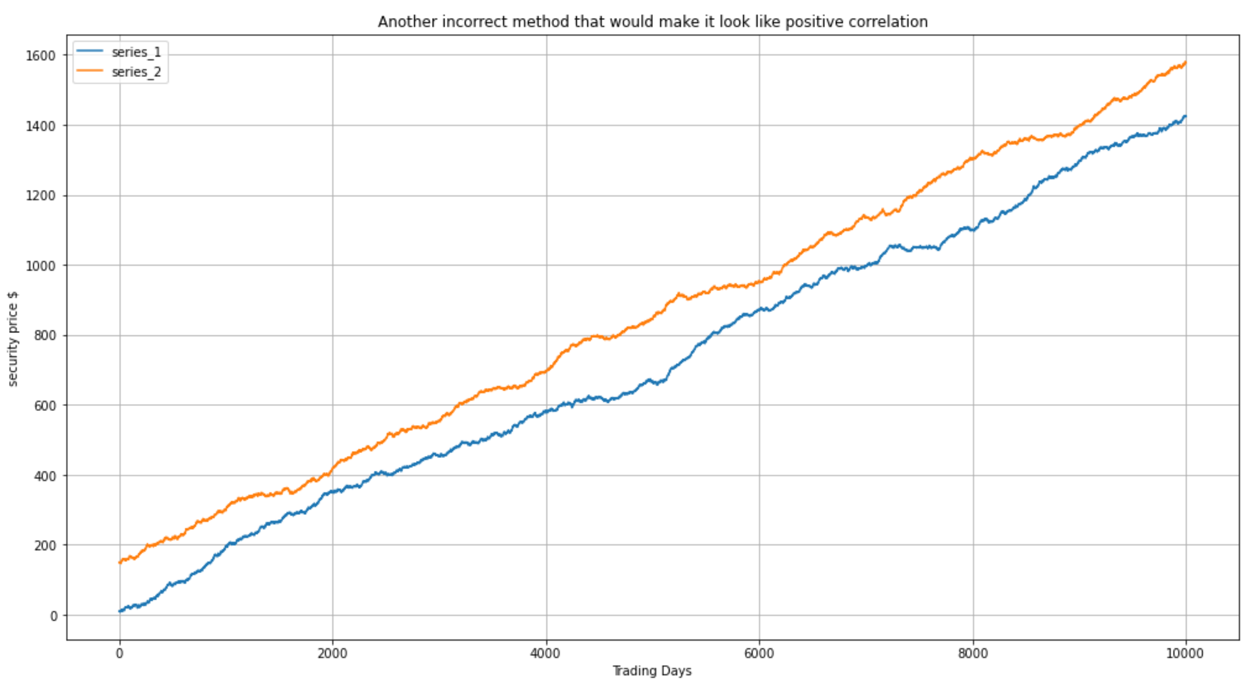
One common pitfall that I see is that if we draw two random time series prices may not have any correlation because those time series tend to be non-stationary. When we take the percent change we can see that the time series are become stationary and it looks like there is a positive correlation.

CalPERS question

Here are the two randomly generated series



- But when we look at the returns they look positively correlated even though they aren't



- The workaround would to be cognizant on whether or not the time series stationary vs non-stationary. Sometimes it comes down to keeping track, but a foolproof method would be to the Augmented Dickey-Fuller test which is in the statsmodel API.