

## ECO481 Assignment 2

Yuanhan Peng  
1005354055  
Pengyua9

### Exercise 1:

1. False.

Since in Baker, Bloom, Davis (QJE 2016), the author only counted papers that meet the following three conditions. Firstly, the paper contains at least have one of “uncertainty” or “uncertain”. Secondly, the paper has at least one of “economic” or “economy”. Thirdly, the paper should contain one of the following words: “Congress”; “deficit”; “Federal Reserve”; “legislation”; “regulation”; “White House” (including variants like “uncertainties”; “regulatory”; “the Fed”.

The statement in Q1 is false because it may not meet the previous two conditions, such as the paper containing word “deficit” but doesn’t contain words either “economy” or “economic” qualified the statement in Q1 but wouldn’t be counted according to the method in the Baker, Bloom, Davis.

2. It is useful to scale raw counts because for each newspaper and each month, there are different volume of articles. In this case, although two months may have the same counts from the same newspaper, those counts didn’t weight the same and may have different meanings. The authors used the total number of articles in the same newspaper and month to scale row counts. They standardized every monthly series from 1985 to 2009 in newspaper-level to unit standard deviation and then took average across all ten papers by month. Eventually, the series would be normalized to a mean of 100 from 1985 to 2009.
3. The authors used the following three methods to validate the measure of economic policy uncertainty obtained from counting newspaper articles.

Firstly, they used audit study that included 6 months in developing audit process and 18 months in holding large-scale human readings of newspaper articles. Then, using the human coding results, they identified 15 terms with the highest frequency in the newspapers and then by comparing the computer assignment for their permutation to human coding, they selected a P term set for the historical EPU index. Next, they compared human and computer-generated EPU indexes’ time-series movement. They also collected other four facts about policy uncertainty during the audit process.

Secondly, to further address the issue about the potential for political slant to skew newspaper coverage of EPU, the authors used media slant index to split those 10

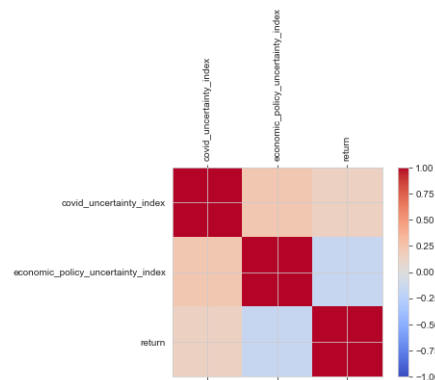
newspapers into the 5 most Republican and 5 most Democratic papers. Their finding showed that in EPU related newspapers, the political slant won't significantly change the variation over time; thus, it won't affect EPU index.

Thirdly, to evaluate the EPU index, the authors made comparison with other measures of uncertainty and policy uncertainty, such as other text sources and stock index. They analyzed the uncertainty indicators based on the Beige Book and found the correlation between it's normalized policy uncertainty count and the EPU index is 0.54. Also, they analyzed the S&P stock index from 1900 to 2012 and found the correlation between yearly count of daily stock market jumps influenced by policy news and the annually EPU index is 0.78.

4. In my opinion, I agree with the authors' method. I think the use of the index of economic policy uncertainty measured by the number of newspaper articles is better than an index constructed using a method based on Google trends because of the following reasons. Firstly, using the newspaper to measure the index, we could get more data of different countries and data for longer period of time. However, data collected from google trend could only start from 2004 to present. Moreover, in this paper, the author selected newspapers based on three conditions, but google trend could not select qualified data with the preconditions that the search need to at least have one of "uncertainty" or "uncertain" and one of "economic" or "economy". In other words, the newspaper-based indexes are easily to develop category-specific indexes with certain constrains; whereas the google trend data can only be categorized into certain categories they provided.

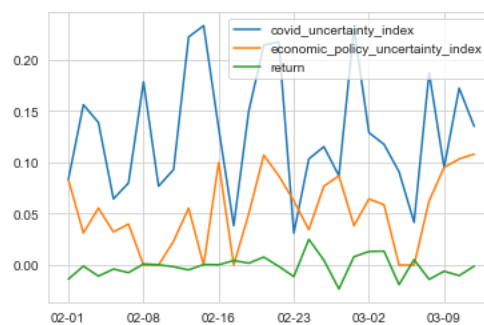
## Exercise 2:

1. Please see the python file.
2. Vocabulary of COVID-19 related words: 'COVID', 'Coronavirus', 'pandemic', 'vaccine', 'virus', 'quarantine', 'vaccination'
3. Please see the python file.
4. Please see the python file.
5. Please see the python file.
6. Here is the plot and simple correlation matrix of the covid uncertainty index, the economic policy uncertainty index, and the returns.



	covid_uncertainty_index	economic_policy_uncertainty_index	return
covid_uncertainty_index	1.000000	0.220230	0.134305
economic_policy_uncertainty_index	0.220230	1.000000	-0.153477
return	0.134305	-0.153477	1.000000

Here is the plot for the covid uncertainty index, the economic policy uncertainty index, and the returns. The y-axis is the value, and the x-axis is the date.



7. We can tell from the above plots that the correlation between the covid uncertainty index and the economic policy uncertainty index, and the correlation between the covid uncertainty index and the economic policy uncertainty index, and the returns are positive. However, the economic policy uncertainty index and the returns have a negative correlation. In general, all correlations between those three variables are pretty weak.

8. The lists below show the TOP 15 words with the highest frequency from the headlines that tend to be related to the returns. The left list shows the word in negative returns period. The right list shows the word in positive returns period.

[('biden', 92),	[('trump', 60),
('trump', 57),	('biden', 60),
('capitol', 30),	('capitol', 24),
('house', 26),	('trial', 23),
('new', 22),	('impeachment', 21),
('say', 22),	('say', 18),
('republican', 22),	('coronavirus', 18),
('riot', 21),	('vaccine', 18),
('u', 20),	('riot', 17),
('first', 19),	('house', 17),
('briefing', 19),	('democrat', 16),
('senate', 19),	('briefing', 16),
('coronavirus', 18),	('senate', 15),
('happened', 17),	('happened', 13),
('stimulus', 17)]	('today', 13)]

9. For both negative and positive returns categories, “trump”, “biden” and “capitol” are the most common words from the headlines. The only difference is that for the list of negative returns, the word “biden” has the highest frequency, whereas, in the list of positive returns, both words “biden” and “trump” occur the same times. Other than those three words, there are many other words that appear in both lists, such as “coronavirus”, “riot”, “happened”, “senate”, “briefing”, and “house”. However, the most significant difference between the words in two categories must be “republication” in the negative returns list and “democrat” in the positive returns. Although the word “coronavirus” appeared in both categories, it has the same frequency in both lists, which mean as the pandemic is ending, “coronavirus” may no longer be a factor that significantly affects the returns negatively anymore.

### Exercise 3:

Here is the 2-gram word cloud:



This text talks about a data processing system for financial assets, such as loans, to generate pricing information with loan level data. The pricing logic is illustrated in detail in the text, as well as the servicer and investor reporting logic, securitization logic, accounting logic, acquisition logic, committing logic, management logic, and delivery logic. This system can identify and track all cash flow and have a pricing engine to create interest rate and credit risk components.