

POLI 421 Spring 2015
Baumgartner
January 28, 2015

Results of your graphing assignment.

Figure 1 shows my results and all of your results.

Figure 1. My results (right scale), your results (left scale).

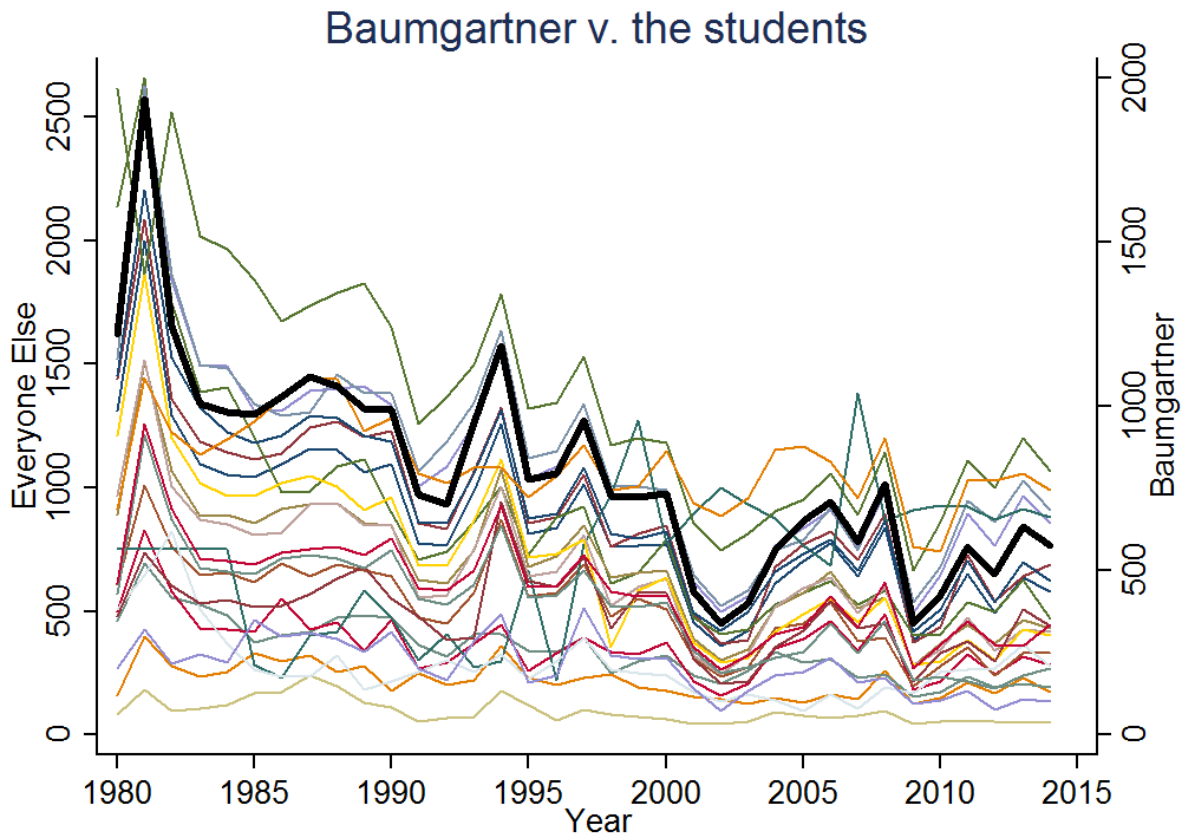


Table 1 shows all the keywords that you used. Clearly, many of you added lots of keywords to mine, generally causing you to get fewer hits than I got. But note that, for most of you, the trends over time were virtually identical. But not for all of you. A few of the lines in the graph above do not “load” with the others, or in other words follow the same pattern.

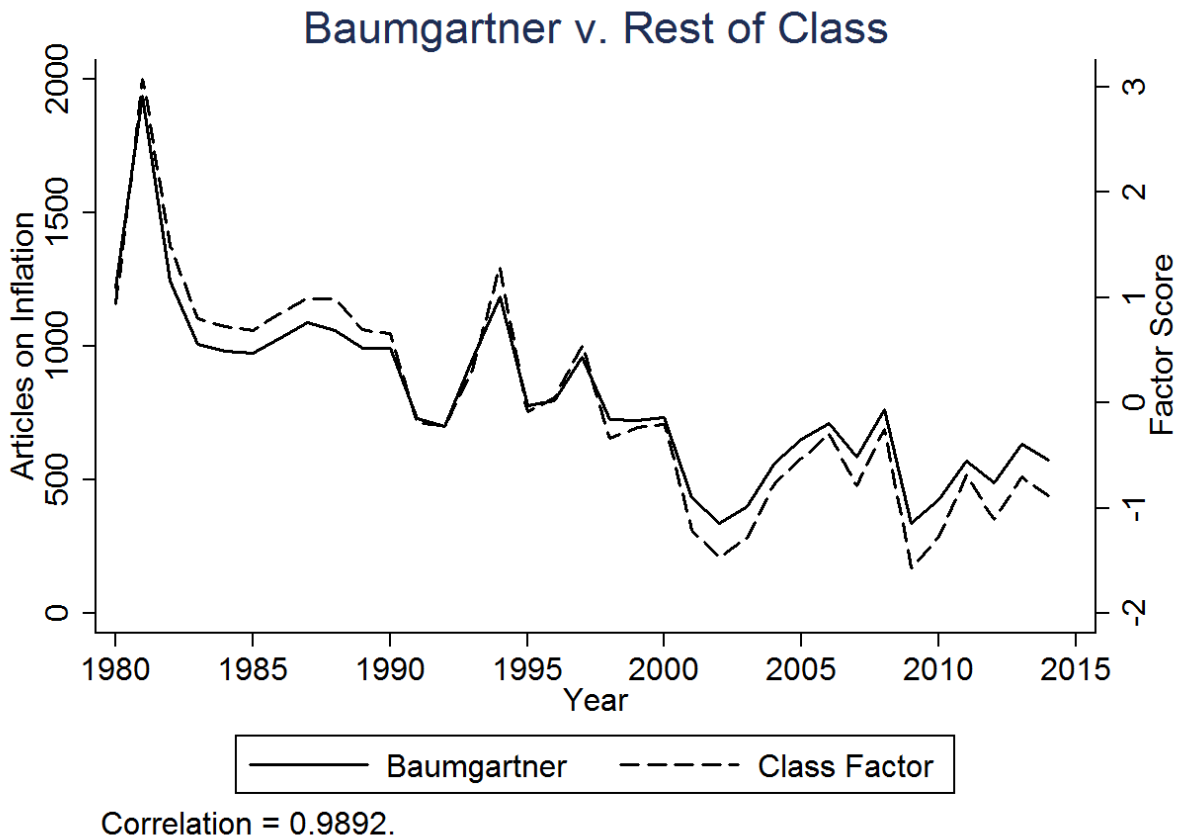
Table 1. Keywords used

Name	Keywords:
Baumgartner	inflation AND price AND ("United States" OR US)
Shiro	inflation AND (price OR "price increase" OR "increasing prices" OR "price increases" OR CPI OR "consumer price index") AND (United States OR US)
Hildebrand	inflation AND price AND ("United States" OR US OR America OR American) OR (increase OR growth OR hike OR expansion) AND NOT deflation
Parker	US OR U.S. OR America OR United States AND inflation OR consumer price index AND NOT Canada AND NOT Latin America AND NOT Mexico AND NOT China AND NOT Europe AND NOT India AND NOT Venezuela AND NOT Japan AND NOT Brazil AND NOT Portugal AND NOT adjusted for inflation AND NOT penalty inflation AND NOT after inflation AND NOT inflation-adjusted
Hayes	inflation AND economy AND value AND ("US" OR "United States")
Zi	("United States" OR allcaps(US) OR allcaps(USA) OR "United States of America" OR singular(America)) AND (infla! w/3 rate OR defla! w/3 rate) OR (percen! w/s change w/s price OR price w/s change) OR ("consumer price index" OR allcaps(CPI) OR "consumer prices") OR (price w/15 increas! OR price w/15 decreas! OR price w/15 declin! OR price w/15 fall! OR price w/15 r*s!) OR (pressure w/5 inflation w/5 price)
Buso	inflation AND price AND (demand OR cost OR "economic growth") AND (unemployment OR jobs OR oil) AND ("United States" OR US)
Koppenhofer	inflation AND (price OR "interest rate" OR demand OR GDP) AND ("United States" OR US OR America OR American)
Waters	inflation AND price AND dollar AND increase AND value AND ("United States" OR US)
Franklin	inflation AND price OR prices AND interest AND percent
Davidson	(Markets OR Economy) AND Inflation AND ("United States" OR US OR America)
Berry	Inflation AND "Economic Crisis" AND "United States"
Williams	inflation AND ("United States" OR US) AND rate AND ("Consumer Price Index" OR CPI OR price)
Vaughn	inflation AND price AND rise OR increase AND ("United States" OR US)
Harrison	nflation AND price AND ("United States" OR US) AND NOT Asia AND NOT China AND NOT "adjusted for" AND NOT ("automobile tire" OR "car tire")
Whitehead	inflation AND economic AND unemployment AND ("United States")
Funk	inflation OR CPI AND price OR "consumer price index" OR "money supply" AND ("United States" OR US)
Coto	inflation AND ("federal reserve" OR fed) AND ("United States" OR America* OR US)
Wagoner	united states AND (CPI OR inflation OR deflation OR "consumer price index")

McDowell	increase AND price And goods OR ("Bureau Labor Statistics" OR BLS) AND ("United States" OR US)
Boykin	inflation AND price AND economy AND ("United States" OR US) AND NOT Canada AND NOT China
Johnson	inflation AND economy AND (price OR "purchasing power" OR "consumer price index" OR cpi OR "price bubble")AND("united states" OR usa OR us)
Gooding	inflation AND price AND (US or "United States") AND (CPI OR index OR Consumer Price Index) AND (rise OR rising)

We can do a “factor analysis” to see if there is an underlying pattern to all the data, and, also, whether and to what extent each individual series is correlated with the others. Figure 2, for example, is my series compared to the “factor” or weighted average of all the series combined.

Figure 2. Baumgartner’s results compared to the weighted average of all the other results.



Now, we can also see the degree to which each of those lines in Figure 1 correlated, or loaded, on the common factor. Table 2 shows these scores.

Table 2. How each student loads on the first statistical factor.

baumgartner	0.9892
shiro	0.9911
hildebrand	0.9869
parker	0.8853
hayes	0.8128
zi	0.9357
buso	0.8839
koppenhofer	0.9776
waters	0.7674
franklin	0.9673
davidson	0.9714
berry	-0.4966
williams	0.9944
vaughn	0.9915
harrison	0.9787
whitehead	0.6753
funk	0.9907
coto	0.7897
wagoner	0.8558
mcdowell	0.7552
boykin	0.9639
johnson	0.9742
gooding	0.7906

Clearly, a few of the students used key-words that caused them to produce series that were quite different from the rest. Those loading below 0.8 are listed in Table 3.

Table 3. Cases loading below 0.80 on the first factor.

waters	0.7674
berry	-0.4966
whitehead	0.6753
coto	0.7897
mcdowell	0.7552
gooding	0.7906

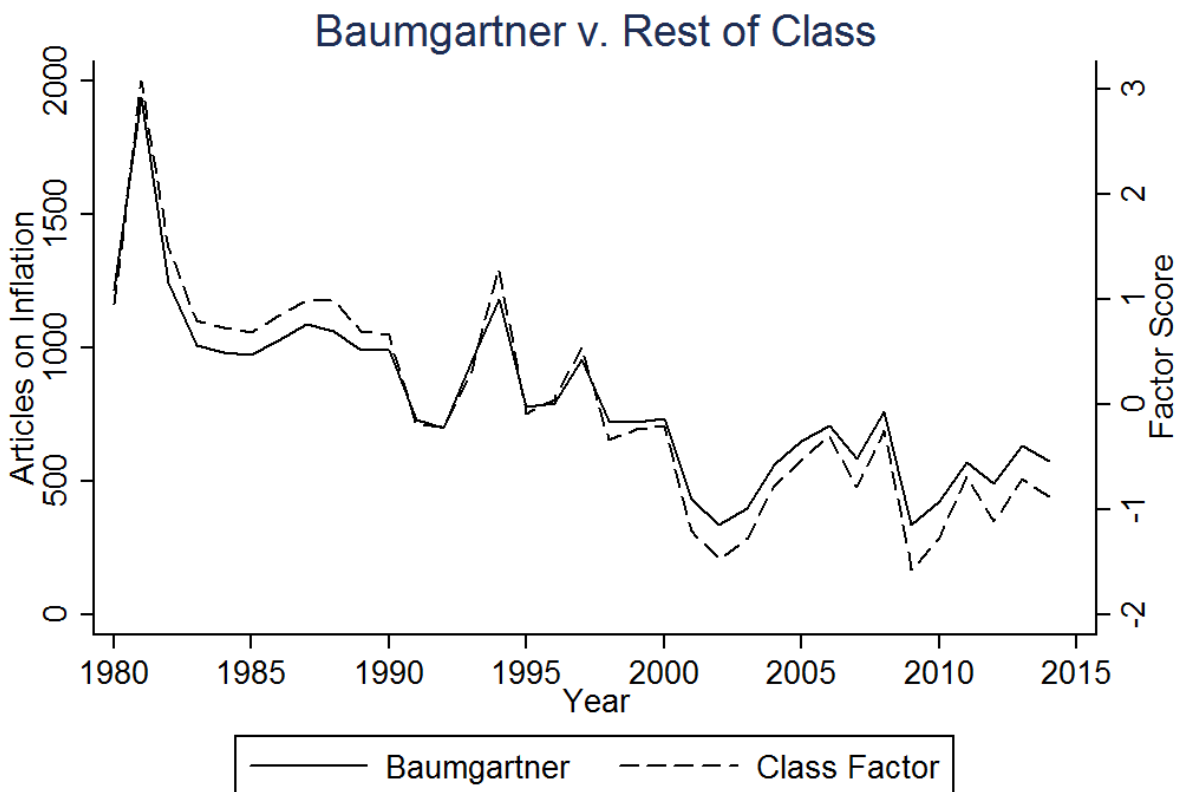
And here are the keywords that those students used. You can see from this why the results might have been different.

Table 4. Keywords for the cases not loading as expected.

Name	Keywords:
Waters	inflation AND price AND dollar AND increase AND value AND ("United States" or US)
Berry	Inflation AND "Economic Crisis" AND "United States"
Whitehead	inflation AND economic AND unemployment AND ("United States")
Coto	inflation AND ("federal reserve" OR fed) AND ("United States" OR America* OR US)
McDowell	increase AND price And goods OR ("Bureau Labor Statistics" OR BLS) AND ("United States" OR US)
Gooding	inflation AND price AND (US or "United States") AND (CPI OR index OR Consumer Price Index) AND (rise OR rising)

So, now if I exclude those cases and make another index only on the cases remaining, all of which load highly on the factor, we get these results:

Figure 3. Baumgartner v. rest of class, excluding some.



Correlation = 0.994. Excludes cases loading below 0.80 on first factor.

One thing you might notice is that it actually did not make much of a difference. The correlation moves from .98 to .99 but the trends remain virtually identical. Here is how each individual series corresponds not to my results, but to the factor.

Figure 4. Factor 1 v. the students.

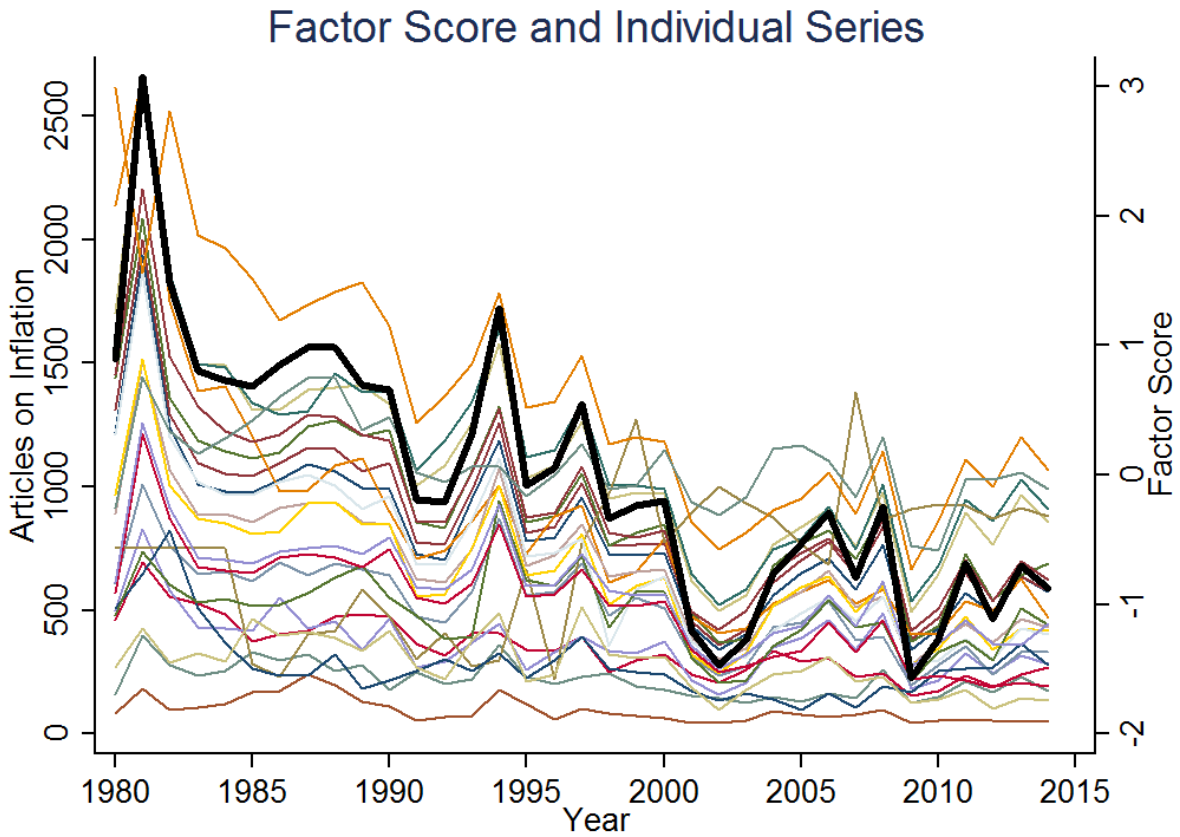
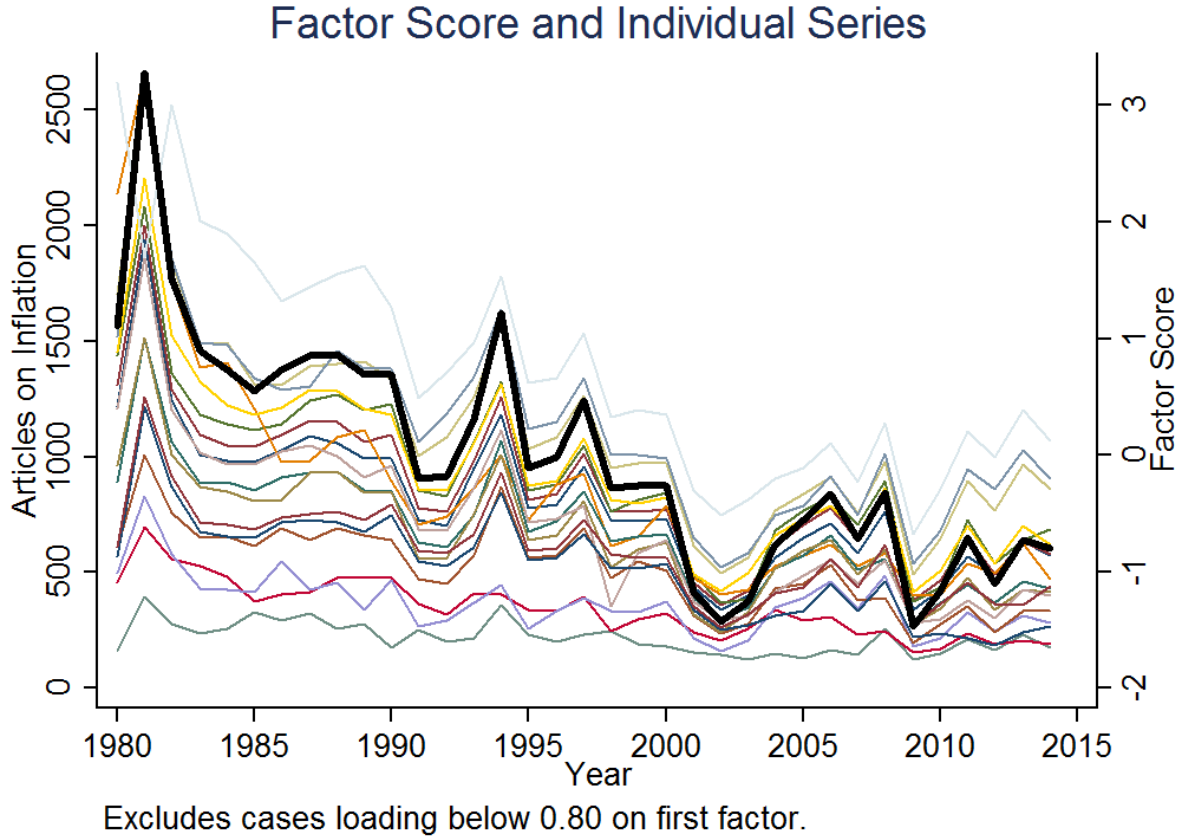


Figure 5. Factor 1 v. the students, excluding some.



So what were you supposed to learn from this?

- If you are interested in tracing trends over time, many combinations of key-words will work.
- It's not hard to do this at all. In fact, once you get the keywords debugged, it is very fast.
- You do, however, have to read through your results and determine, based on reading many of your cases, that your keywords really are working to generate a set of articles that are "true hits" not false ones.

Other things to consider:

- Inflation is a relatively easy case. Some topics lend themselves to this type of keyword searching, but others are a nightmare.
- Further, sometimes when you go back in time over many decades, the vocabulary can change. Think of race and the words used to describe Black Americans over time, or what we call now climate change but used to be global warming, the ozone hole, etc..
- Solutions: pick a topic that is feasible, take care in your keywords, and be sensitive to how they may change over time.