Effect of Economic Recessions on Educational Activity

Blake Stanford <bstanfor@bu.edu>, Vincent Wahl <vinwah@bu.edu>, Chenyang Yu <cyu1221@bu.edu>

1. Project Task

With increasing tuition rates in the US, the question of whether college is affordable and worthwhile becomes more pressing. Over several decades, tuition rates have increased far more than inflation [1], and within the same period, several economic recessions have occurred. If the price of education is at the limit of what is affordabable, one would expect to see a decrease in educational activity as a result of the financial uncertainty that accompanies recession periods.

The goal of our project is to find out how economic recessions affect educational activity. Particulary, we want to look at how recessions affect enrollment rates, graduation rates, provision of financial aid, student loans, and research expenses. Furthermore, we want to look at what schools are most affected and what factors contribute to a school being affected.

Our project reposetory can be found at: https://github.com/cyu1221/CS506

2. Related Work

There exist previous research on relationship between economic factors and higher education, many of which focuses on the increasing tuition rates. However, as the increasing tuition rates has become a prominent phenomena, some schools have started taking steps towards limiting the increase in total fees, such as Middlebury College who, for the past 5 years, have used an upper bound of CPI+1% [1].

Although there are good reasons for why the tuition rate increases at this rate, such as institution based financial aid, as addressed in “The Challenging Economics of Higher Education” [1], the students not covered by institutional financial aid still must be willing and able to cover the increased charges, which may negatively affect the aggregated educational activity in the country.

3. Datasets

The primary datasources used in this project are: The National Center for Education Statistics (NCES), Yahoo! Finance (Yahoo), and the Federal Reserve Bank of St. Louis (FRED). NCES is a part of the United States Department of Education, and it provides per institution, yearly, educational statistics. The data of interest from NCES are number of newly enrolled, number of newly graduated, amount of financial aid, amount of student loans, resources allocated to research, revenue, and expenditures. Yahoo is used to retrieve data on four major stock indices and bond yields. The stock indices used in this project are: Dow Jones Industrial Average (DJIA), Standard & Poor’s 500 (SNP), NASDAQ (NSD), and Russell 2000 (RUT). To reflect bond yields the 10-year US treasury yields (TNX) will be used. The St. Louis Federal Reserve has provided the unemployment rate for people above age 25 and with a bachelor’s degree or higher as well as the Consumer Price Index (CPI). The CPI indicates the inflation rate in the country for a normal consumer, and is thus best suited to adjust for inflation for consumer choices such as the one in question.

So far, we have extracted all the data listed above. For the data associated with student aid (total student loans, federal grants, state and local grants, institutional grants), we have access to data from 3900 US post-secondary schools in the time period 1999-2016. For the data associated with the particular institution (assets, expenses, revenue, and research expenses), we have access to data from 3728 schools in the time period 2001-2016. For the data associated with enrollment (grand total men and women enrolled), we have accessed 4309 institutions in the time period 1997-2016 and 2188 institutions for graduation rate during the time period 1997-2016.

Yahoo’s financial data makes available much more specific data than needed, both intra and inter day, but we have chosen to only extract the opening price of each week for the period 1992-2018, since we are only looking for extended periods of depressions in the market. For both unemployment and CPI, the data available is sampled at a monthly frequency.

In total, the extracted data we will be using extends across 9 csv files covering a time period of at least 2001 to 2016 and 20 csv files from NCES dated from 1997 to 2016. The csv file format is chosen because we have no nested data structures and it seamlessly works with python libraries like numpy and pandas.

4. Approach

Using the data described in section 3, we will first graph our data to get a sense of the correlation between yearly aggregated educational data and financial data. We will do this by graphing each variable as a function of time, and, by visual examination of the financial data, determine where major financial recessions hit. We will try to correlate these periods with any prominent changes in the educational data by further inspecting these knowing where the major economic declines already occur.

After getting this preliminary sense of the data, we will adjust all USD values in the data using CPI conversion with base year being 2017. This way we can better understand the actual changes experienced by the average consumer. For example, to convert $100USD_{2000}$ to USD_{2017}, we get the following formula:

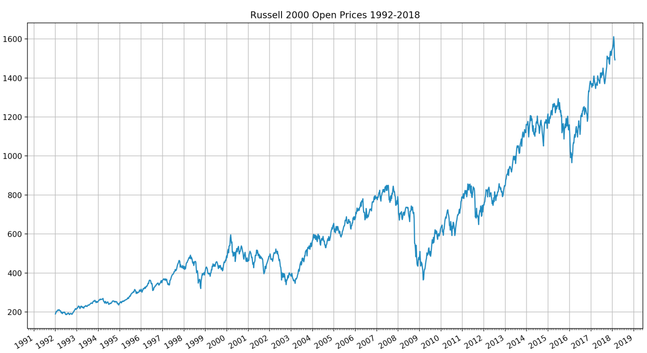
$$\frac{CPI_{2017}}{CPI_{2000}} * 100USD_{2000}$$

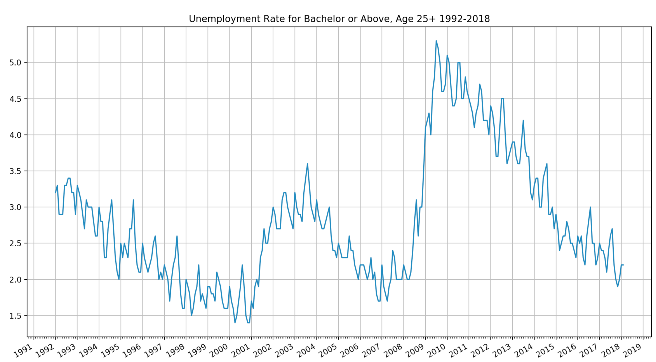
After adjusting all the data for inflation, we will further examine the financial data for smaller declines in specific periods (quarterly) and outlier detection in the educational data. For the financial data, this will be done using a sliding window with moving average and deviation from mean from a larger window. This way, we can find periods with greater financial instability, and further specify how abnormal the particular periods are compared to other periods. This information will be used together with preliminary results to find recessionary periods as well as generally unstable periods [2]. For the educational data in these particular periods, we will use an appropriate clustering algorithm to categorize which schools behave similarly in these periods. To get information about behavior, the values of variables in these periods will be given by change rather than absolute values. We believe the Gaussian Mixture Model (GMM) will provide an appropriate clustering algorithm for this task since it soft-assignes classes. We can then retrieve information on whether a school is definately in the group or somewhat of an outlier. This is important as it will allow us to focus better on the representative elements of the assignment when determining what factors affect clustering.

Hopefully, we will have time to test some more advanced anomaly detection algorithms to find muliti-variable anomalies in the educational data, which can further provide insight into what can be catgorised as an anomaly and also better determine recessions', as well as other variables', effects on educational activity. Anomaly detection approaces we would want to look at are Fast Fourier Transform [3], Markov Chains [4], and feed-forward-nural-network. Other outlier algorithms we might consider can be found at [5].

5. Preliminary Results

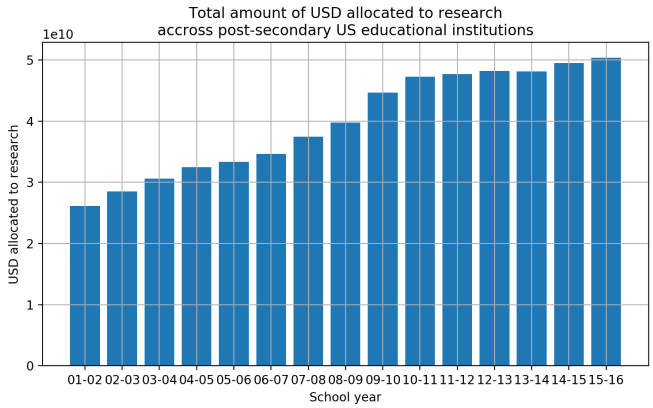
Thus far, we have extracted all the data we expect to use and graphed the financial data as well as some of the educational. From all the financial graphs, we can clearly see the financial recessions that occurred in 2000 and 2008. In particular, the unemployment data shows dramatic changes in these two periods. Further more, we can see that the period 2010 to 2016 has been more volatile than previous periods. It will be interesting to see if, when the negative periods line up with the decision time for selecting schools (April-May), they have any effect on the enrollment rate the following fall.





From the two graphs above, Russel index and unemployment, we were not surprised to find that unemployment and the general stock market were closly negatively correlated when considering the same time periods. We will continue to investigate this correlation further, hopefully finding some link between unemployment rates at the time of an individual's admission to a university and that individual's likelihood of completing their degree.

On the graph below, you can see the total amount of yearly allocated USD by post-secondary schools. On the graph you can see that research allocation actualy increased in the time period 08-011. We did not expect this, but one possible explenation for this is that the financial crisis in 2008, which was quite sever, was itself research by universities. We will further look into this, and hope to find a reasonable explenation



6. Timeline

Task	Deadline
Correlate individual variables between financial and educational data. Adjust data for inflation. Prepare educational data for clustering.	03/24/18
Find financially unstable periods and determine extent and how abnormal they are.	03/28/18
Cluster financial data for schools in periods of interest.	04/01/18
Analyze results and write up report	04/05/18
Receive input from report on 04/05/18, and make appropriate changes. Try out more advanced statistical analysis on educational data. Write up finalized report.	Before: 04/26/18

References

1) https://digitalcommons.csbsju.edu/cgi/viewcontent.cgi?article=1144&amp=&context=admin_pubs&amp=&sei-redir=1&referer=https%253A%252F%252Fscholar.google.com%252Fscholar%253Fas_ylo%253D2014%2526q%253Dtuition%252Bvs%252Binflation%2526hl%253Den%2526as_sdt%253D0%252C22#search=%22tuition%20vs%20inflation%22
2) https://www.datascience.com/blog/python-anomaly-detection
3) https://bugra.github.io/work/notes/2014-03-31/outlier-detection-in-time-series-signals-fft-median-filtering/
4) http://bugra.github.io/work/notes/2014-04-26/outlier-detection-markov-chain-monte-carlo-via-pymc/
5) https://blog.statsbot.co/time-series-anomaly-detection-algorithms-1cef5519aef2