

Demystifying the Carnegie Classifications: A Sensitivity Analysis

Paul Harmon

Adviser: Dr. Mark Greenwood
Montana State University

Spring Semester 2017

Abstract

The Carnegie Classifications of research activity are used to compare like institutions in higher education. In 2015, the newest update of the Carnegie Classifications were released, with Montana State University moving from the top-tier category of "Highest Research Activity" to second highest tier, "Higher Research Activity." The classification system is based on two separate indices calculated using principal components analysis. The first index is based on a set of aggregate covariates and the other on a set of per-capita metrics. This analysis re-creates the calculation of the classifications and examines how sensitive they are to changes in the underlying characteristics of a given institution. Care should be taken when interpreting the results of this analysis; a static analysis of a dynamic process can illuminate the most sensitive aspects of the classifications but cannot forecast the way that other institutions will grow with respect to Montana State University. That being said, this analysis informs how difficult it would be to move from the R-2 status to R-1 in the future, but more importantly, it further illuminates the reasons for why Montana State was placed in the "Higher Research Activity" classification in 2015.

1 Introduction

The Carnegie Classifications of Institutions of Higher Education are released every 5 years. They are intended to be used for institutions to identify other schools which are similar in size, research production, and academic goals so that meaningful comparisons can be made between institutions. They are unfortunately often mistaken as a ranking system; however, the classifications of each institution are not meant to identify schools as being better or worse than institutions in other classifications.

In 2015, the Center for Postsecondary Research at the Indiana University School of Education took over the formulation of the classifications. When

the 2015 updates were released, Montana State University - among a cohort of several institutions - moved from the "Very High Research Activity" to the "Moderately High Research Activity" category.

This analysis seeks to recreate the classifications produced by the researchers at Indiana University. Further, I analyze the sensitivity to minor perturbations in the underlying indices used to calculate each score for in order to determine which variables most strongly affect the score for a given institution. Moreover, I create an app that demonstrates where Montana State would end up relative to the other institutions in the dataset if it experienced these slight marginal changes. Administrators at Montana State University (and other institutions like it) have made obtaining R1 status an institutional goal; this sensitivity analysis shows that the path from the current classification to the higher one would be at least somewhat arduous.

2 The Data

The data are obtained from both the Montana State University Office of Planning and Analysis but can be obtained more generally from the Carnegie Classifications website. In either case, the data contain data pertaining to many levels of institutions; only those that grant doctoral degrees are of interest. The data are therefore processed in order to remove the non-doctoral granting institutions. The data that are reported come from a variety of sources, including IPEDS, CCIHE, and the NSF. In the final dataset, 335 institutions granted doctoral degrees during the period of interest. However, some of the smaller institutions do not report expenditures for STEM-related fields. Therefore, I removed 59 institutions from the dataset, all of which were classified as "high research activity." Doing so leaves a count of 276 schools on which the classifications were calculated.

3 Replicating the Classifications

The classifications are calculated from two different indices, one on the aggregate counts for each institution, and the other on a per-capita basis. The classifications themselves do not count the per-capita measures based on student populations at each school; rather, they focus on the size of the faculty at each institution in terms of a raw headcount of tenurable and non-tenurable faculty. In general, institutions with relatively large values in one index are likely to be relatively large in the other index. In 2010 the correlation between the two indices was a strong positive 0.83. In 2015 the correlation was 0.84.

3.1 Aggregate Index

The aggregate index includes PhD degrees awarded in humanities, professional fields, social sciences, and stem fields; however, it also includes the research

staff, stem expenditures, and non-stem expenditures from the Per Capita Calculation. The formula for the aggregate index is given below for the i th institution: $AggregateIndex_i = HumanitiesDoctorates_i + StemDoctorates_i + OtherDoctorates_i + SocialScienceDoctorates_i + StemExpenditures_i + NonStemExpenditures_i + ResearchStaff_i$

3.2 Per Capita Index

The per-capita index considers only three variables: research staff, stem expenditures, and non-stem research expenditures divided by the size of the faculty at the given institution. The formula for the per-capita index is given below for the i th institution: $PerCapitaIndex_i = \frac{ResearchStaff_i + StemExpenditures_i + NonStemExpenditures_i}{FacultySize_i}$

3.3 Combining the Indices

After calculating the individual per-capita and aggregate indices, the two are combined with a single plot. The per capita ranking is plotted along the X axis and the aggregate ranking is plotted along the y-axis. The Carnegie Classifications are then based on the natural spacing that occurs in the data; in some years the data are fairly well separated into three clusters and in other years they are not. In 2015, no obvious clusters were formed; further, there were no obvious spaces of separation that would define the two borders between the three groups. The classifications were found by dividing the per-capita index roughly into three groups and finding the areas of greatest separation between points in that region. In the 2010 update, the lines that divided groups were hand drawn; however, in 2010 the lines were drawn by calculating circles with uniform radius from the origin. Using circles allows for institutions with outlying scores in a single index to be more likely to end up in the highest classification.

3.4 Method For Ties

In replicating the classification done by the Carnegie Institute, only a handful of decisions must be made. The most important choice comes in the ranking step. When ranking institutions that are tied, there are several methods for calculating rank that have important consequences for an analysis, especially where an institution can move up or down in rank in subsequent years. In the statistical software package R (R Core Team), the default setting of the rank function is to take the average of the ranks. For instance, if the first five institutions have the same value for Stem Expenditures, each institution would receive a rank value of 2.5. Alternative methods include taking the first, last, minimum, or maximum of the ranks. The first and last methods result in taking a permutation of each of the potential ranks with first involving increasing values and last involving decreasing values. The minimum rank method refers to the more commonly known ranking method used in sports; in the previous example, each of the five institutions would be ranked 1st. Similarly, the maximum rank gives the largest rank to all of the observations; in the previous example, each

institution would be ranked fifth. The literature on the Carnegie Classifications does not specify exactly the method that the institute used for tied ranks, but comparing each method to the final results indicates a clear picture. Below is a table of standardized loadings for each method along with the loadings generated from the actual classifications.

Moreover, the plots generated by each of the methods confirm that the minimum ranking was used in the calculation of the classifications. Other methods generate plots where institutions are misclassified; further, the separation for the minimum ranking is better than for any of the other methods. For institutions that seek to improve their classification - or for institutions that plan to avoid dropping into the lower category - this is of particular importance: Breaking ties can lead to large changes if many institutions are tied, compared to other methods of handling ties. For instance, Montana State University is tied with 61 institutions that have 0 social science doctorates. Under the average ranking method, each of those institutions would be ranked as the average rank; however, using the minimum rank, an institution with a single social science PhD would improve its position by 61 points. To be clear, the use of the minimum rank method for dealing with ties indicates that Montana State could gain substantial ground simply by going from having 0 Social Science PhDs to a single one in the next iteration of the classifications. Moreover, increasing the number of Stem PhDs is unlikely to have the same impact as increasing Social Science PhDs because Montana State is tied with fewer institutions for STEM PhDs.

#here's where I'll include the plot of the different ranking methods

4 Sensitivity of the Rankings

4.1 Shiny App

4.2 Other Methods

5 Discussion

5.1 Montana State as an R-2 Institution

Montana State University had previously been classified in the highest tier of "Very High Research Activity" in the 2010 Carnegie Classification update. However, it may not have been well-classified.

5.2 Moving Up: How Hard Would It Be?

5.3 Limitations of Static Analysis

This analysis considers changes at Montana State, holding the values at other institutions constant. Certainly, this assumption