Time Series Analysis in Rankings of American Universities.

Zuchen Li, Kun Cheng, Yishan Ren, Xiaoyu Sun

1. Introduction

1.1. Abstract

In our college research and those around us, we found that high school students were only relatively familiar with the current information on universities. However, little is known about historical information and future predictions. In our research, we found that there is no research on the current ranking information of American undergraduate universities. So, we collected data and proposed our research. We construct our model for prediction and summary of the past data. We selected some representative groups and specific schools to make our prediction and analysis.

We note that the closer the information is to the present, the greater the value of its parameters. It has no apparent function, but it does have some correlation. In addition, despite the long years of university ranking, it should not be regarded as a long-term process. As for the strength of a university, the demission or joining of several professors will not cause great changes to the overall strength of the university itself. Moreover, the reputation and popularity of universities cannot be changed in just a few years. Even reforms take a long time to show up. Therefore, although the university ranking data has a long span, its essence is closer to a short-term time series. In summary, these characteristics indicate that the exponential smoothing model can be used to analyze better and predict university ranking data. Finally, we analyze the historical trend and future predictions of many universities. This new information can help high school students who are doing school research.

1.2. Background Information

The world has realized that the quality of their education systems directly determines the nation's economic success and that the most influential factor of production is human capital expressed in the knowledge, skills, creative abilities, and moral qualities of individuals in society. Finding a good school plays a massive role in stepping toward success.

As the world's largest economy, the United States has many high-quality universities and colleges. Not only is the country concerned about the quality of education in universities, but people are even more concerned. Students conduct university research for their future development; parents seek university consultants for their children; companies cooperate with universities for better employees; professors choose higher-quality universities for better research. The most intuitive evaluation criterion for the quality of a university is its comprehensive ranking. The ranking is made by authoritative media using a specific algorithm. For the universities in the United States, the US News University ranking is the most authoritative and has the widest audience.

1.3. Background Research

In order to form an accurate conclusion, we carefully reviewed several previous research papers in the process of searching for data and methods.

Ouyang Liang and Pei Chen's academic ranking trend model of world universities has attracted our attention. This paper focuses on ranking the world's top 500 universities from 2014 to 2016 based on a linear combination of six indicators (Chang, 2018). This research paper gives us a glimpse into the methodology used to rank universities, so we can better measure and predict results.

Moreover, we also looked at how software can act as a tool to form a simulation of future university ranking (Siniksaran, 2020). WURS: a simulation software for university rankings— software review, by Enis Siniksaran M. Hakan Satman, introduces WURS (World university ranking Simulator) and three significant steps of creating university ranking: 1. Collect data from institutions and external databases. 2. Scale the data to obtain a score for each indicator ranging from 0 to 100. The most ranking system also uses "z-transformation" to the raw scores of the indicator, then ARWU scales all raw values based on the best performing entity. 3. Aggregate the scaled scores with their assigned weights. We obtained the composition of AR-WU and the author's train of thought from this research paper.

In addition, we reviewed relevant papers (Gardner, 1985) to help us better use the exponential smoothing model.

2. Data Collection and Pre-Processing

2.1. Data Collection

Since we want to investigate the ranking information of American universities, we need to collect as much historical data as possible correctly. Therefore, in order to pursue the stability and consistency of data, we choose US NEWS US University Rankings, a ranking system with a long history and little changed evaluation methods in the United States. We may get historical ranking information from online public databases.

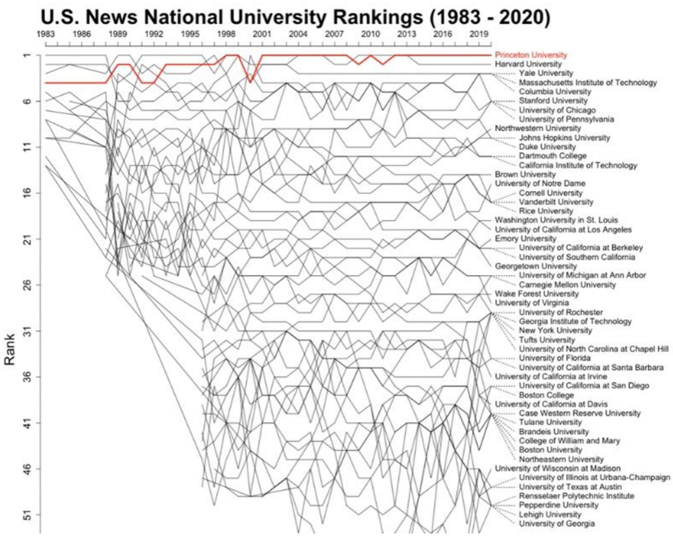

Figure 1 General University Rankings

We can see that the history of most of the schools at the top of the US rankings is relatively stable. However, historical information about some schools shows a trend. Some show increased trends; some show decreased trends.

2.2. Data Pre-Processing

The numerical meaning of the ranking information itself is not very accurate and will be artificially modified. In addition, because the ranking itself is a relative data. A school's strength may not change, but its ranking may change as the strength of other schools' changes. Therefore, through special data pretreatment, we pre-process the comprehensive ranking of each university into its comprehensive strength index.

$$C_i = (100 - R_i) \cdot 10$$

Where C_i represents the comprehensive strength index that shows the comprehensive strength of the i th university. R_i represents the original rankings of that university.

3. University Rankings Analysis and Prediction Model

The rankings are based on admission rates, research capabilities, peer reviews, and other information. The closer the information is to the present, the greater the value of its parameters. It has no apparent function, but it does have some correlation. In addition, despite the long years of university ranking, it should not be regarded as a long-term process. As for the strength of a university, the demission or joining of several professors will not cause significant changes to the overall strength of the university itself. Moreover, the reputation and popularity of universities cannot be changed in just a few years. Even reforms take a long time to show up. Therefore, although the university ranking data has a long span, its essence is closer to a short-term time series. In summary, these characteristics indicate that the exponential smoothing model can be used to analyse better and predict university ranking data.

3.1. Primary Exponential Smoothing

Assume time series to be $y_1, y_2, \cdots, y_t, \cdots$, α is the smoothing parameter, $0 < \alpha < 1$, the single exponential smoothing formula is given by:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha\left(y_t - S_{t-1}^{(1)}\right)$$

Which is modified from the moving Average formula. The formula of deriving the rolling average is:

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

We use $M_{t-1}^{(1)}$ as the best estimate of y_{t-N}, there is

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + \left(1 - \frac{1}{N}\right) M_{t-1}^{(1)}$$

Let $\alpha = \frac{1}{N}$..., use S_t to replace $M_t^{(1)}$, we get:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)}$$

In order to understand the essence of exponential smoothing, we expand and get

$$S_t^{(1)} = \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)S_{t-2}^{(1)}\right] = \cdots = \alpha \sum_{j=0}^{\infty}(1-\alpha)^j\, y_{t-j}$$

Which shows that $S_t^{(1)}$ is the weighted mean of all historical data, obviously, we get

$$\sum_{j=0}^{\infty} \alpha(1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1$$

Because weighing parameter follows the exponential law, and also applies to smooth data, it is known as exponential smoothing. Predictions based on this smooth value is Primary Exponential Smoothing. The prediction model is:

$$\hat{y}_{t+1} = S_t^{(1)}$$

Which can be also written as:

$$\hat{y}_{t=1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

which is using the *t* time period to predict *t*+1.

3.2. Quadratic Exponential Smoothing

The Primary exponential smoothing method overcomes the disadvantages of the moving average method. However, when the change in the time series shows a linear trend, the primary exponential smoothing method still shows obvious hysteresis error.

Therefore, it must also be amended. The correction method is the same as what to do with the trend movement average method, which contains another quadratic exponential smoothing, and uses the law of the hysteresis error to establish a linear trend model. This is called the quadratic exponential smoothing method. The calculation formula is:

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)}$$

In the formula, $S_t^{(1)}$ is the smoothed statistic in primary exponential smoothing, $S_t^{(2)}$ is the smoothed statistic of quadratic exponential smoothing:

$$\hat{y}_{t+T} = \alpha_t + b_t T, T = 1,2,\cdots$$

$$\begin{cases} \alpha_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right) \end{cases}$$

When time series $\{y_t\}$ starts to show a linear trend at one time period, just like Moving Average Method, can build the linear trend model to predict the rankings.

3.3. Cubic Exponential Smoothing

When the changes in the time series shows a quadratici curve trend, the cubic exponential smoothing is required. The cubic exponential smoothing is based on the secondary exponential smoothing, with the formula of

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha) S_{t-1}^{(3)} \end{cases}$$

where $S_t^{(3)}$ is the smoothed statistic of cubic exponential smoothing.

The cubic exponential smoothing prediction model is

$$\hat{y}_{t+T} = \alpha_t + b_t T + C_t T^2, T = 1,2,\cdots$$

Consequently, the Cubic Exponential Smoothing can be represented as:

$$\begin{cases} \alpha_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}\right] \\ c_t = \dfrac{\alpha}{2(1-\alpha)^2}\left(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right) \end{cases}$$

Where $S_t^{(1)}$ is the smoothed statistic in primary exponential smoothing, $S_t^{(2)}$ is the smoothed statistic of quadratic exponential smoothing, and $S_t^{(3)}$ is the smoothed statistic of cubic exponential smoothing.

4. Results

4.1. Comprehensive Analysis of Top10 Universities

We decided to start with this year's top 10 universities. Because the advantage of exponential smoothing is that it can make clever and scientific use of historical data. Compared with simple linear regression, it can make more rational use of past data. Using a specific exponential smoothing model, we found that the top 10 universities' rankings were stable. The ranking information does not show a significant linear or nonlinear trend. Therefore, we choose immediate exponential smoothing to analyze the comprehensive strength index of these universities.

In order to better compare the effects of specific weighting coefficients, we chose three possible weighting coefficient values to help us in our final model selection. Also, we plotted the raw data, a = 0.2, a = 0.4, and a = 0.6, on the same chart.

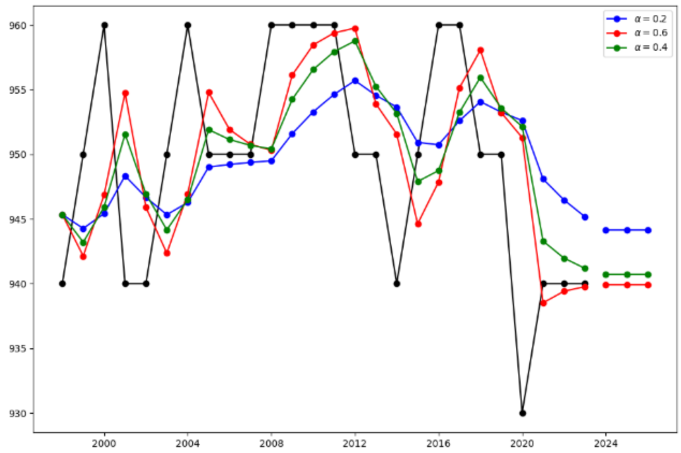

Figure 2 Stanford University Result

where the horizontal axis is the specific year, the vertical axis is the comprehensive strength Index C_i . The above figure is the result of Stanford University.

After applying the calculation process to all top 10 universities in 2022, we got our results.

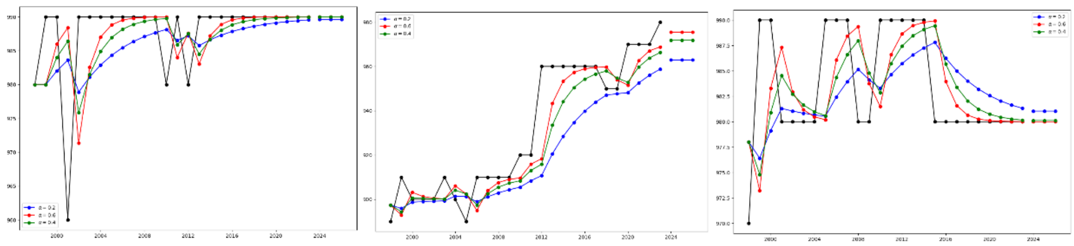

Figure 3 Princeton Columbia Harvard

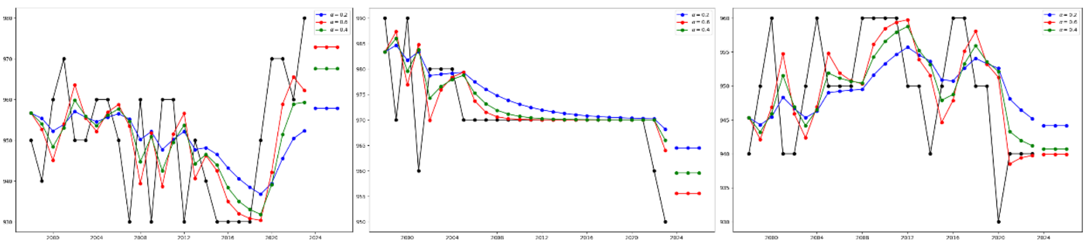

Figure 4 MIT Yale Stanford

Figure 5 Uchi Upenn Caltech

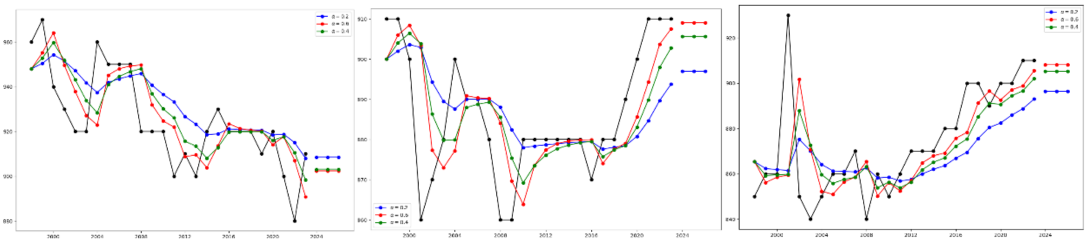

Figure 6 Duke JHU Northwestern

	R	C	a=0.2	new rank	dif	a=0.4	new rank	dif
Princeton University	1	990	989.6	1	0	989.9	1	0
Columbia University	2	980	963	4	-2	971.8	3	-1
Harvard University	2	980	981	2	0	980.1	2	0
Massachusetts Institute of Technology	2	980	957.8	5	-3	967.6	4	-2
Yale University	5	950	964.5	3	2	959.6	5	0
Stanford University	6	940	944.1	7	-1	940.7	7	-1
University of Chicago	6	940	947	6	0	945.7	6	0
University of Pennsylvania	8	920	925.2	8	0	922.9	8	0
California Institute of Technology	9	910	902.4	10	-1	901.4	12	-3
Duke University	9	910	908.4	9	0	902.9	11	-2
Johns Hopkins University	9	910	896.3	12	-3	905.2	10	-1
Northwestern University	9	910	897	11	-2	905.6	9	0

Figure 7 Top10 Results

We can see that the historical ranking information of the top universities is relatively stable. However, some universities are not seen as promising in the future. This is because exponential smoothing is a weighted average. The further the data is, the less weight it has on the future. So even though a school's current ranking is high, its final predicted value will be low due to its low historical data.

4.2. Analysis of Special Universities

Figure 8 CMU Result

For schools like Carnegie Mellon that have a history of high and low rankings. Exponential smoothing can better reflect its real information. In addition, we note that the ranking information may contain a certain linear relationship. So, we use quadratic exponential smoothing to analyze and make prediction.

Figure 9 NYU UCSB USC Results

In addition, we selected several universities with strong linear trends in ranking information: New York University, UNIVERSITY of California, Santa Barbara, and University of Southern California. Since it has a strong linear upward trend, we do a second exponential smoothing analysis on it based on the first smoothing.

	R	C	Primary: a=0.4	Quadratic a=0.4
Carnegie Mellon University	25	750	748.6	715.7
New York University	28	720	706.6	745.6
University of California-Santa Barbara	28	720	692.1	743.6
University of Southern California	28	720	754.6	716.3

Figure 10 CMU NYU UCSB USC Results

We can notice that after taking linear trends into account using quadratic exponential smoothing. Our model takes into account not only the different weights of its historical data, but also the overall upward and downward trend of its ranking information. We can see that both Carnegie Mellon and USC have a downward trend. NYU and UC Santa Barbara are on the rise.

5. Discussion

The first difficulty encountered in the research was the difficulty finding relevant and comprehensive data when collecting it. In order to get a less biased and variable database, we wanted to get all the data since US News established the university ranking project. However, after many searches, we only got the university ranking data. Moreover, for another planned aspect of the research, we can only find nearly 20 years of data on college acceptance rates and enrollment numbers, which disrupts our initial plans and consumes a lot of our time and energy.

Our model may have limitations on different aspects of data, adapting different models for scope inference. At the same time, in reality, university performances are affected by many factors, such as government policies, environmental safety, research results, and high-level positions in the university. Therefore, the forecasting model we make may be limited to stable and controllable situations. It may fail to predict precisely if there are extraordinary achievements and changes.

For the future of this research, we can start with the relevant data of the College of Arts and Sciences and more data on other aspects, such as majors' rankings and student satisfaction. In addition, we can try to account for changes in the model when dealing with extraordinary situations and improve the accuracy of the predictions.

After all, as the first degree displayed when applying for jobs, the ranking and influence of the university are essential to be performing well and stable. University applicants may want to see the future of a university, not only for the quality of education in four years but also for the feeling of being proud of their university's progress or retention when applying for a job in the future.

6. Conclusion

In conclusion, using historical data as a crucial factor, we build Exponential Smoothing models to predict American universities' ranking in 2023. We first collected the officially released ranking of American universities in past decades. Out of hundreds of universities, we select those ranking from top 1 to top 10 in the most recent ranking and those which exhibit notable trends for their rankings' prediction in 2023. Our result suggests that universities that show constant rankings, that is, varying within constant amplitude on their rankings from 1997 to 2022, are recommended to be applied. Universities exhibiting an uplifting trend in their rankings are risky, and application for which requires prudent consideration. Universities that show a decreasing trend in their rankings are not recommended. Exponential Smoothing emphasizes the impact of historical data, and this method benefits students in college application since it evokes attention to the historical richness of universities.

7. Reference

[1] Chang, F. P. C., & Ouyang, L. Y. (2018). Trend models on the academic ranking of world universities. International Journal of Information and Management Sciences, 29(1), 35-56.

[2] Siniksaran, E., & Satman, M. H. (2020). WURS: a simulation software for university rankings—software review. Scientometrics, 122(1), 701-717.

[3] Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. Journal of forecasting, 4(1), 1-28.

[4] https://i.imgur.com/rXZOH8e.mp4 (https://www.reddit.com/r/ApplyingToCollege/comments/emz9yu/us_news_college_rankings_1983_2020_top50_schools/)

[5] https://andyreiter.com/datasets/