International Journal of
Information
and
Management
Sciences

International Journal of Information and Management Sciences
29 (2018), 35-56. DOI:10.6186/IJIMS.2018.29.1.2

Trend Models on the Academic Ranking of World Universities

Farrah Pei-Chen Chang and Liang-Yuh Ouyang

Tamkang University

Abstract

The Academic Ranking of World Universities (ARWU) has provided annual global rankings of universities since 2003, making it the earliest of its kind. ARWU draws on six indicators to measure the academic performance of universities. Top 500 universities are ranked each year since 2004 by linear combinations of the six indicators. This paper uses a natural log regression model, called the Score-Rank Model, to present the relationship between the score variable and the rank variable for each year from 2004 to 2016. This paper also presents the Trend Model, built by a two-stage process; first, a linear regression model between two parameters (a_t and b_t in year t) is established; and second, an ARIMA model is built to obtain the value of b_t. The Trend Model can be used to forecast the overall score of a particular rank, or the rank of a specific overall score for future years. It is shown that the Trend Model is valid in an empirical study using ranking data from 2005 to 2015 to forecast the overall scores of the top 500 ranks in 2016. When comparing the forecast results with the real ranking outcomes of 2016 in a graph, it presents two very similar and almost overlapping curves.

Keywords: Academic Ranking of World Universities, natural log regression model, ARIMA model, trend model, coefficient of determination.

1. Introduction

Higher education has become a growing international market in recent years. Stakeholders of higher education – from students and parents to universities and governments – have also grown increasingly intrigued by the positions of higher education institutions in this global competition. As a result, rankings of world universities are leading a powerful trend that is likely to continue in the future (Marginson [11], Yonezawa [16]). Currently, the major world university rankings include the Academic Ranking of World Universities (ARWU, also known as Shanghai Ranking; since 2003), the QS World University Rankings (by Quacquarelli Symonds, since 2004) and the Times Higher Education (THE) World University Rankings (since 2010; Times Higher Education had collaborated with Quacquarelli Symonds to publish the joint THE-QS World University

Rankings from 2004 to 2009 before they ended collaboration and split into two separate rankings systems). The three ranking systems are similar in scopes and purposes, while vary in methodologies (Luque-Martnez and del Barrio-Garca [9]). ARWU emphasizes on high extraordinary research achievement (Huang [6]); and although QS and THE World University Rankings also focus on research outcomes, a large portion of their data relies on global surveys of university reputations. ARWU's indicators consider the number of alumni and staff winning Nobel Prizes and Fields Medals, the number of highly cited researchers, the number of articles published in journals of Nature and Science, the number of articles indexed in Science Citation Index-Expanded (SCIE) and Social Sciences Citation Index (SSCI), and per capita performance of a university (ARWU [2]). QS World University Rankings' indicators examine a university's academic and employer reputations, the student-to-faculty ratio, citations per faculty, and international faculty and student ratios (QS Top Universities [11]). THE World University Rankings' indicators focus on a university's teaching (including reputation survey, staff-to-student ratio, doctorate-to-bachelor's ratio, doctorates-awarded-to-academic-staff ratio, and institutional income), research (including reputation survey, research income and productivity), citations, international outlook (including international-to-domestic-student/staff ratios, and international collaboration), and industry income (Times Higher Education [13]).

As Huang [6] noted, there are two major approaches in research evaluation, namely peer review evaluation and bibliometric evaluation. Peer review evaluation is often criticized for its subjectivity; bibliometric is widely used for its objectivity and operability. Two reasons support the objectivity of bibliometric evaluation as follows. First, results from bibliometric evaluation can be scientifically verified in replication; it is free from possible reviewer prejudice and bias. Second, the publications and citations based bibliometric evaluations may be viewed as a form of peer review - the totality of multi-layered and bottom-up indirect peer reviews. Since data obtained from reputation surveys account for 50% of QS World University Rankings' total weights (QS Top Universities [11]), and 33% for the THE World University Rankings (Times Higher Education [13]), "The high percentage of peer review can easily bias the ranking toward universities of international visibility" (Huang [6]). For the purposes of exploring ranking trends and conducting objective analyses, this paper focuses on ARWU, which employs the bibliometric approach in its data collection process.

ARWU was first compiled and published by the Center for World-Class Universities (CWCU) at the Graduate School of Education of Shanghai Jiao Tong University, China. It has provided annual global rankings of universities since 2003, making it the earliest of its kind. Since 2009, ARWU has been published by ShanghaiRanking Consultancy, an independent organization on higher education intelligence. ARWU is now one of the best known international ranking of universities (Dehon et al. [4]). While its initial purpose was to ascertain the relative position of Chinese universities internationally, ARWU has since attracted much interest from around the world, as its announcement now receives considerable press attention annually. The ARWU website features the scores used to compute for the rankings; however, raw data are not available. The missing data indicates that the results cannot be reproduced, which is unfortunate, as Razvan [12] noted, since reproducibility is a principal requirement for any scientific method. Despite its popularity, ARWU has come under some criticism regarding both its methodology and choice of variables (Liu et al. [8], Van Raan [15]). ARWU relies solely on research indicators, and is heavily weighted toward institutions whose alumni and staff have won Nobel Prizes and Fields Medals. Given the criterion and indicators ARWU uses in its methodology, it is evident that universities specializing in science, and whose research published in English-speaking journals, are naturally favored. Any ranking is controversial, and no ranking is absolutely objective (Liu [7]).

2. ARWU Indicators

ARWU draws on six different indicators and the relative percentage weights to measure the academic performance of universities. After computing each of the six scores, universities then receive an overall score (a weighted average of individual indicator scores), and are ranked by the overall score they obtain. The best performing university of a particular ranking year is given a score of 100, and the scores of other (following) universities are measured accordingly. Since 2004, universities are ranked by linear combinations of ARWU's six indicators, namely, alumni and staff (with 10% and 20% weights, respectively) winning Nobel Prizes and Fields Medals, highly cited researchers in 21 broad subject categories (with a 20% weight), papers published in Nature and Science (with a 20% weight), papers indexed in major citation indices (Science Citation Index-Expanded (SCIE) and Social Science Citation Index (SSCI); with a 20% weight), and the per capita academic performance of an institution (with a 10% weight). Overall, ARWU measures four criteria with these six indicators. The quality of education of universities is measured by the number of Nobel prizes and Fields medals won by a university's alumni (coded Alumni). The quality of faculty is measured by the number of Nobel prizes and Fields medals won by staff (coded Award), and the number of highly cited researchers (coded HiCi). Universities' research output is reflected in papers published in Nature and Science (coded N&S) and papers indexed in SCIE and SSCI (coded PUB). And universities' per capita performance (coded PCP), the sixth indicator, is a weighted average of the scores obtained in the previous five categories, divided by the number of current full-time equivalent academic staff members. See Table 1 for ARWU's criteria, indicators, codes and weights.

Scores for each indicator are weighted (as shown in Table 1) to arrive at a final overall score (coded Score) for an institution. The highest scoring institution is assigned a score of 100, and the other institutions are calculated as a percentage of the top score. This can be formulated into the following equation, using the 2013 top score as an example. In 2013, the highest scoring institution, Harvard University, had an overall score of 97.25; the scores for the rest of the institutions on the league table were calculated as follows,

$$\begin{aligned}\text{Score} &= (0.1\,\text{Alumni}+0.2\,\text{Award}+0.2\,\text{HiCi}+0.2\,\text{N\&S}+0.2\,\text{PUB}+0.1\,\text{PCP}) \times 100/97.25\\ &= 0.1028\,\text{Alumni}+0.2056\,\text{Award}+0.2056\,\text{HiCi}+0.2056\,\text{N\&S}+0.2056\,\text{PUB}\end{aligned}$$

Table 1: Criteria, indicators and weights for ARWU (2004−2016).

Criterion	Indicator	Code	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	Alumni	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	Award	20%
	Highly cited researchers in 21 broad subject categories	HiCi	20%
Research Output	Papers published in Nature and Science	N&S	20%
	Papers indexed in Science Citation Index-expanded and Social Science Citation Index	PUB	20%
Per Capita Performance	Per capita academic performance of an institution	PCP	10%
Total			100%

Source: ARWU [1]

$$+0.1028\,\mathrm{PCP}.$$

According to ARWU's methodology, the data of the number of alumni and staff of an institution winning Nobel Prizes and Fields Medals are used for almost 100 years (for indicators Alumni and Award, respectively); the data of highly cited researchers (HiCi) selected by Thomson Reuters for the 2003 to 2015 rankings are used from the last five years, but for the 2016 ranking, only the data from the previous year (2015) are used. The Institute for Scientific Information (ISI) defines a "highly cited researcher" as one of 250 most cited authors of journal papers in 21 subject areas; citation data of a journal paper for 20 years in the old list (2001-2013) and for 10 years in the new list (2014-2016) are used (Bauer et al. [3]). Papers published in Nature and Science (N&S) are used for five years for the indicator N&S. PUB adopts papers indexed in SCIE and SSCI in the previous year. And, PCP combines scores of the above indicators.

3. Trend Analysis

The ARWU website publishes the complete score information for each of the six indicators and the overall scores for institutions ranked top 1 to 100. Thereafter, the ranks are featured in groups of 50 from top 101 to 200, and groups of 100 from top 201 to 500; while the score information for each indicator remains public, the overall scores are omitted. The data used for this paper are collected directly from the website (ARWU [1]), including the overall scores, ranks, and scores of each indicators. As previously mentioned, the website only presents the overall scores and ranks for the top 100 institutions; for institutions ranking from 101 to 500, only their rank groups and indicator scores are presented. For the purpose of this research, the overall scores for institutions ranking from 101 to 500 from 2004 to 2016 were recomputed using ARWU's scoring method.

Examining the relationship between the overall scores and their ranks, Figure 1 presents the overall scores of universities ranking from 1 to 500, which shows thirteen very stable trend curves from 2004 to 2016. The overall scores of universities ranking from 1 to 100 are on a very rapid downgrade, which indicates that top ranks have a greater difference in scores; and ranks from 401 to 500 show a rather slow downgrade, which indicates a small difference in scores.

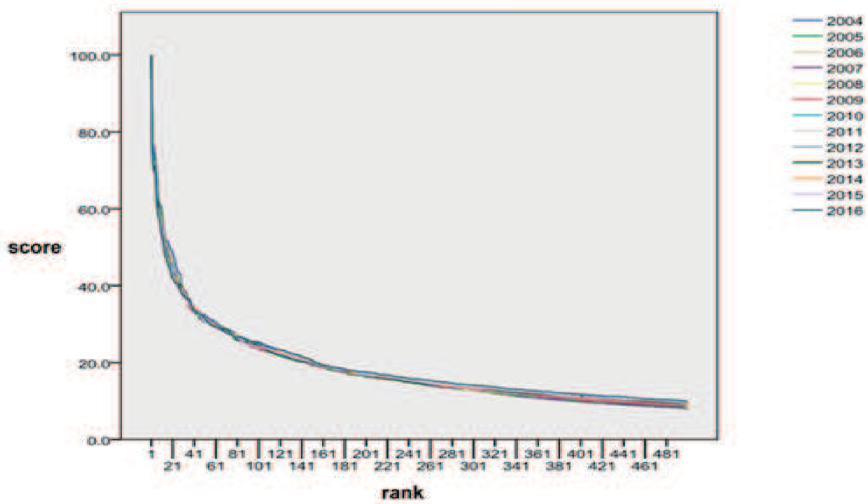

Figure 1: Curves of score by rank.

The steady relationship between the scores and ranks, and the similar curves they produce over the period of thirteen years, could be due to the ARWU indicators and the relative weights. ARWU places heavy weights (a total of 30%) toward institutions whose alumni and staff have won Nobel Prizes and Fields Medals (the Alumni and Award indicators), and these data are used consecutively for about 100 years. In the recent 100 years, there were a total of 211 institutions whose alumni won Nobel Prizes or Fields Medals, 146 institutions whose faculty won Nobel Prizes and Fields Medals, and 116 institutions whose alumni and faculty won both Nobel Prizes and Fields Medals. Once institutions with scores in Alumni and Award enter the top 100, they are likely to stay on the league table in the years that follow, given that they continue to earn good scores for the other three indicators. For institutions whose alumni or staff have not won Nobel Prizes and Fields Medals (without scores for Alumni and Award), it was very difficult to enter the top 100 between 2004 and 2016, unless they had earned high scores in the HiCi, N&S and PUB indicators simultaneously for the ranking year. There were a total of 127 institutions that entered the top 100 between 2004 and 2016, and only 12 of these institutions did not have alumni or staff winning Nobel Prizes and Fields Medals. Among these 12 institutions without having any scores in Alumni and Award, University of California at Davis was an especially noteworthy case, where it was able to maintain top performance and staying between ranks 41 to 58 during 2004 to 2016 by relatively

Table 2: Top 1 to 11 Universities in 2004 and their ranges of ranks.

University	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Range
Harvard University	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Stanford University	2	3	3	2	2	2	3	2	2	2	2	2	3	1
University of Cambridge	3	2	2	4	4	4	5	5	5	5	5	5	4	3
University of California, Berkeley	4	4	4	3	3	3	2	4	4	3	4	4	3	2
Massachusetts Institute of Technology	5	5	5	5	5	5	4	3	3	4	3	3	5	2
California Institute of Technology	6	6	6	6	6	6	6	6	6	6	7	7	8	2
Princeton University	7	8	8	8	8	8	7	7	7	7	6	6	6	2
University of Oxford	8	10	10	10	10	10	10	10	10	10	9	10	7	3
Columbia University	9	7	7	7	7	7	8	8	8	8	8	8	9	2
University of Chicago	10	9	8	9	9	8	9	9	9	9	10	9	10	2
Yale University	11	11	11	11	11	11	11	11	11	11	11	11	11	0

Source: ARWU [1]

high scores of other indicators. The other 11 institutions have the same situation in the years when they entered the top 100.

The steady score-rank relationship and the similar curves are also reflected in the institutions' ranges of ranks. A range of ranks is calculated from the different ranks that each university obtained from 2004 to 2016. Table 2 presents the universities ranked top 1 to 11 from 2004 to 2016 with their ranges of ranks. Harvard University ranks top 1 during the entire duration of the rankings; its range of ranks is 0; Stanford University ranks either top 2 or 3, the range of ranks is 1; University of Cambridge has a range of ranks of 3, which has obtained ranks in the top 2 to 5; Yale University ranks top 11 throughout the entire duration; its range of ranks is also 0. These top 11 institutions change ranks very slowly and steadily over the years; their ranges of ranks from 2004 to 2016 is only 3 at most.

Institutions that appear lower on the league tables tend to have larger ranges of ranks. As seen in Table 3, for universities ranking from 78 to 100 in 2004, their ranges of ranks from 2004 to 2016 are rather widespread, ranging from 11 to 101, with almost 80% of the universities with ranges of ranks from 18 to 47. This indicates that universities with lower ranks (towards the bottom of the league table) change their positions in the ranking more drastically.

As previously mentioned, data used for this research are collected directly from the ARWU website. The steady relationship between the overall scores and ranks from 2004 to 2016 generate thirteen stable trend curves. ARWU indicators and the relative weights could have contributed to this stable trend.

Table 3: Top 78 to 100 universities in 2004 and their ranges of ranks.

University	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Range
University of Manchester	78	53	50	48	40	41	44	38	40	41	38	42	35	43
University of Göttingen	79	84	85	87	90	90	93	86	107	103	110	111	102	32
Michigan State University	80	80	79	81	82	83	84	85	86	83	103	102	114	35
University of Nottingham	80	80	80	80	83	86	86	92	96	92	105	99	126	46
University of Melbourne	82	86	85	70	71	69	65	65	65	67	75	75	91	26
University of Strasbourg	82	82	96	99	112	108	105	103	105	97	95	87	116	34
Brown University	82	82	78	79	73	75	62	60	57	54	44	44	40	42
École normale suprieure - Paris	85	93	99	83	73	70	71	69	73	71	68	72	87	31
University of Vienna	86	86	159	177	183	178	187	180	170	167	159	165	158	101
Boston University	87	81	81	85	83	74	77	76	71	75	72	74	75	16
University of Freiburg	88	88	90	87	89	91	88	89	92	92	90	96	85	11
McMaster University	88	88	93	94	96	102	105	102	99	100	105	106	132	44
Hebrew University of Jerusalem	90	90	60	64	65	64	52	57	53	59	70	69	89	38
University of Basel	91	91	81	82	87	85	86	89	85	83	90	89	105	24
Lund University	92	99	90	97	97	101	104	109	114	111	125	119	137	47
Sapienza University of Rome	93	98	90	92	91	94	99	107	110	111	122	130	127	40
University of Birmingham	93	97	100	109	115	120	126	127	127	127	151	151	165	72
University of Utah	95	95	94	93	79	80	82	79	82	85	87	95	100	21
Stockholm University	97	97	84	86	86	88	79	81	81	82	81	80	81	18
Nagoya University	98	102	98	94	103	82	79	94	96	101	108	79	72	36
Tufts University	99	113	105	99	97	98	93	94	103	103	94	97	101	20
University of Bonn	78	53	50	48	40	41	44	38	40	41	38	42	35	43

Source: ARWU [1]

4. Score-Rank Models

According to ARWU scoring rule, the highest scoring institution is assigned a score of 100, and the other institutions are calculated as a percentage of the top score. Therefore, only the overall scores of top 2 to 500 in each year's ranks will vary. The thirteen score-rank curves from 2004 to 2016 (as seen in Figure 1) of each year's ranks from top 2 to 500 fit well to the natural log regression model (called log-mode): For each $t = 2004, 2005, \dots, 2016$,

$$\text{Score}_{ti} = a_t + b_t \cdot \ln(\text{Rank}_{ti}) + e_t, \qquad i = 2, 3, \dots, 500, \tag{4.1}$$

Table 4: Values of a_t, b_t and R_t^2 in log-models.

Year (t)	a_t			b_t			R_t^2
	estimate	standard error	p-value	estimate	standard error	p-value	
2004	82.467	0.380	0.000	-12.240	0.071	0.000	0.983
2005	79.207	0.387	0.000	-11.682	0.073	0.000	0.981
2006	78.418	0.361	0.000	-11.498	0.068	0.000	0.983
2007	78.284	0.371	0.000	-11.495	0.070	0.000	0.982
2008	77.984	0.356	0.000	-11.396	0.067	0.000	0.983
2009	77.400	0.361	0.000	-11.303	0.068	0.000	0.982
2010	77.314	0.368	0.000	-11.246	0.069	0.000	0.982
2011	77.715	0.355	0.000	-11.307	0.067	0.000	0.983
2012	77.045	0.375	0.000	-11.187	0.071	0.000	0.981
2013	76.562	0.377	0.000	-11.093	0.071	0.000	0.980
2014	76.125	0.351	0.000	-10.981	0.066	0.000	0.982
2015	75.309	0.351	0.000	-10.835	0.066	0.000	0.982
2016	75.026	0.332	0.000	-10.692	0.062	0.000	0.983

let $y_{ti} = \text{Score}_{ti}$ and $x_{ti} = \ln(\text{Rank}_{ti})$, then Eq. (4.1) can be rewritten as:

$$y_{ti} = a_t + b_t x_{ti} + e_t, \qquad i = 2, 3, \ldots, 500. \tag{4.2}$$

Using the least square estimation, the estimates of a_t and b_t (for simplicity, still use the notations a_t and b_t) are given as:

$$b_t = \frac{\sum_{i=2}^{500} (x_{ti} - \bar{x}_t)(y_{ti} - \bar{y}_t)}{\sum_{i=2}^{500} (x_{ti} - \bar{x}_t)^2}, \tag{4.3}$$

$$a_t = \bar{y}_t - b_t \bar{x}_t \tag{4.4}$$

where Rank_{ti} = Rank i in year t, Score_{ti} = Score of Rank i in year t, $\bar{x}_t = \sum_{i=2}^{500} x_{ti}/499$, and $\bar{y}_t = \sum_{i=2}^{500} y_{ti}/499$.

For each year $t = 2004, 2005, \ldots, 2016$, the values of a_t, b_t, standard errors, p-values and coefficient of determination R_t^2 are calculated by SPSS statistical package and shown in Table 4. The high R_t^2 values indicate that the log-model for each year fits well.

Because the error term e_t can be 0 in Eq. (4.1), the overall score (Score_t) of a particular rank (Rank_t) at year t can be estimated by

$$\text{Score}_t = a_t + b_t \cdot \ln(\text{Rank}_t) \tag{4.5}$$

for $\text{Rank}_t = 2, 3, \ldots, 500$. And, the rank ($\text{Rank}_t$) of a particular overall score (Score_t) can be estimated by:

$$\text{Rank}_t = \exp\left(\frac{\text{Score}_t - a_t}{b_t}\right) \tag{4.6}$$

for $0 < \text{Score}_t < 100$, where $t = 2004, 2005, \ldots, 2016$.

After constructing the Score-Rank Models as Eq. (4.5) and Eq. (4.6) and determining the values of a_t and b_t, we can use Eq. (4.5), called score-by-rank model, to estimate score by rank, and use Eq. (4.6), called rank-by-score model to estimate rank by score. In the first score-by-rank example, University of Minnesota Twin Cities was ranked 30 in 2014, using $a_{2014} = 76.125$ and $b_{2014} = -10.981$ in Table 4, our model estimates an overall score of 38.776 (see calculation below):

$$\text{Score}_{2014} = 76.125 - 10.981 \cdot \ln(30) = 38.776.$$

The real overall score was 39.3; the difference between the estimated score and the real score is $-0.524 (= 38.776 - 39.3)$. In another example, University of Copenhagen and University of Illinois at Urbana-Champaign were both ranked at 30 in 2016, using $a_{2016} = 75.026$ and $b_{2016} = -10.692$ in Table 4, our model estimates an overall score of 38.660 (see calculation below):

$$\text{Score}_{2016} = 75.026 - 10.692 \cdot \ln(30) = 38.660.$$

The real overall score was 37.7; the difference between the estimated score and the real score here is 0.960 $(= 38.660 - 37.7)$. Although errors are observed between the estimated scores and the real scores in the above examples, both errors are considered small (-0.524 and 0.960, respectively).

Next, we also used the models for rank-by-score estimations in two examples. In 2013, Cornell University earned an overall score of 50, its rank placement estimated by our model, where $a_{2013} = 76.562$ and $b_{2016} = -11.093$ in Table 4, is 11 (see calculation below):

$$\text{Rank}_{2013} = \exp\left(\frac{50 - 76.562}{-11.093}\right) = 10.963 \approx 11.$$

The actual rank was 13; the difference between the estimated rank and the actual rank is 2. In another example, the University of Tokyo had an overall score of 42, using $a_{2015} = 75.309$ and $b_{2015} = -10.835$ in Table 4, the estimated rank is 22 (see calculation below):

$$\text{Rank}_{2015} = \exp\left(\frac{42 - 75.309}{-10.835}\right) = 21.632 \approx 22.$$

The actual rank was 21; the difference between the estimated rank and the actual rank here is 1. From the above examples, we conclude that estimations by Score-Rank Models have produced close and reliable outcomes, with only small errors.

5. Trend Models

As demonstrated in previous examples, the scores and ranks from 2004 to 2016 of ARWU can be estimated by using the Score-Rank Models. In this section, we want to apply the same models to forecast future ranking scores and ranks. To distinguish predictions from estimations, models used for forecasting are named the Trend Models here. Different from constructing the Score-Rank Models for the years from 2004 to 2016, there are no available data y_t (Score of year t) and x_t (natural log of rank in year t) to be used to estimate parameters a_t and b_t for the Trend Models for future year $t = 2017, 2018, \ldots$. In order to build the Trend Models, it is essential to estimate the a_t and b_t for the future year t. The estimations are made in the following two-stage process: stage one is building a linear regression relationship between a_t and b_t from known data, that is for $t = 2004, 2005, \ldots, 2016$; stage two is building a forecast model to predict the value of parameter b_t for future year $t = 2017, 2018, \ldots$, then using the linear regression model built in stage one to obtain the value of parameter a_t for future year $t = 2017, 2018, \ldots$.

In the first stage, looking at known paired data (a_t, b_t), where $t = 2004, 2005, \ldots, 2016$, in Table 4 as a sample of the virtual bivariate (a, b), we obtained the linear regression by SPSS statistical package as:

$$a = 28.214 - 4.363b \tag{5.1}$$

with a high R^2 of 0.993. Using the above equation, when $b = b_t$ is estimated, we can find its corresponding estimate of a by $a_t = 28.214 - 4.363b_t$, for $t = 2017, 2018, \ldots$.

In the second stage, to find the estimate b_t for $t = 2017, 2018, \ldots$, we used the b_t values in Table 4 starting from 2005 as time series data, which provided a 10% higher stationary R^2 and R^2 values than starting from 2004 (on average), and then simulated ARIMA models by SPSS statistical package to find suitable ARIMA models of b_t. As a result, there were 29 models (as shown in Table 5) that had stationary R^2 and R^2 values ranging from 0.975 to 0.983, which were the 29 highest and all rounded off to 0.98. This indicates that all of the models fit very well. (For details on the formulation of ARIMA models, see Appendix.)

The presented model fitting ARIMA statistics include goodness-of-fit measures by SPSS statistical software, namely stationary R^2, R^2, root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), maximum absolute percentage error (MaxAPE), maximum absolute error (MaxAE) and Normalized Bayesian Information Criterion (Normalized BIC). Bayesian Information Criterion (BIC) is closely related to the Akaike information criterion (AIC), where the outcomes are the smaller the better. Stationary R^2 and R^2 are looking for outcomes not larger than 1, and the closer they are to 1, the better. RMSE, MAPE, MAE, MaxAPE, and MaxAE are different measures of distance between the predicted and the actual values of the models and should, as such, be as close to 0 as possible.

Stationary R^2 is to measure whether difference of series data would be better than the original series data; that is, to check that the series data is stationary or not, instead

Table 5: Suitable ARIMA models and their Stationary R^2 and R^2 (in descending order).

Model	Stationary R^2	R^2
ARIMA(8,0,0)	.983	.983
ARIMA(5,0,4)	.982	.982
ARIMA(7,0,1)	.982	.982
ARIMA(9,0,0)	.982	.982
ARIMA(5,0,3)	.981	.981
ARIMA(6,0,3)	.981	.981
ARIMA(7,0,0)	.981	.981
ARIMA(5,0,2)	.980	.980
ARIMA(6,0,2)	.980	.980
ARIMA(7,0,2)	.980	.980
ARIMA(5,0,0)	.979	.979
ARIMA(5,0,1)	.979	.979
ARIMA(6,0,0)	.979	.979
ARIMA(8,0,1)	.979	.979
ARIMA(6,0,1)	.978	.978
ARIMA(4,0,5)	.978	.978
ARIMA(1,0,6)	.978	.978
ARIMA(2,0,6)	.978	.978
ARIMA(3,0,6)	.978	.978
ARIMA(0,0,8)	.978	.978
ARIMA(0,0,9)	.978	.978
ARIMA(2,0,7)	.977	.977
ARIMA(2,0,5)	.976	.976
ARIMA(3,0,5)	.976	.976
ARIMA(0,0,5)	.975	.975
ARIMA(1,0,5)	.975	.975
ARIMA(0,0,6)	.975	.975
ARIMA(0,0,7)	.975	.975
ARIMA(1,0,8)	.975	.975

of testing unit root. A large stationary R^2 means that the original series data will be better than difference of series data; that is, the series data can be accepted as stationary. R^2 is a coefficient of determinant to measure model fitting; and confident interval length is to measure accuracy of forecast value. In general, the larger the R^2, the smaller the mean square error (MSE) and root mean square error (RMSE), and similarly, the smaller the MAPE, MAE, MaxAPE, and MaxAE.

When choosing the most ideal ARIMA model, high stationary R^2 and R^2 values and short confidence interval lengths are to be considered as criteria. Both higher stationary R^2 and R^2 values indicate a better model fitting; a shorter confidence interval length indicates a more precise estimating, while half of the confidence interval length is the standard error. Since the stationary R^2 and R^2 values of the 29 ARIMA models in Table 5 were very close and all rounded off to 0.98, which led to good fitting of those ARIMA models, we then chose the model with the shortest confidence interval length as the most ideal model for forecasting future years. For our research, we wanted to forecast scores and ranks for the next ten years (Yonezawa [16]); hence, we compared the models that had the ten shortest confidence interval lengths for each year from 2017 to 2026. Table 6 shows the results of the 10 best models of each future year.

As previously mentioned, to choose the most ideal model from Table 6, the confidence interval lengths were used as candidate criteria. As shown in Table 6, it just happened that the models with the first ten shortest confidence interval lengths were the same in each year from 2017 to 2026, although not in the same order. Among these ten models (shown in Figure 2), the model ARIMA$(8, 0, 0)$ had the shortest confidence interval length each year from 2017 through 2026. Therefore, ARIMA$(8, 0, 0)$ was the most ideal model for forecasting b_t. Once the values of b_t were obtained, the values of a_t followed from Eq. (5.1). From Table 5 (very high Stationary R^2 and R^2) and Table 7 (very small RMSE, MAPE, MAE, MaxAPE, MaxAE and Normalized BIC), we find that the proposed 10 candidates in Table 6 of the most suitable Trend model all fit very well. As a result, we have obtained the most ideal Trend Models for the next ten years, as presented in Table 8.

Table 6. Best ARIMA models for 2017-2026 to forecast b_t.

Year	Model	Forecast b_t	95% Confidence Interval		Interval length=UCL-LCL
			UCL	LCL	
2017	ARIMA(8,0,0)	-10.711	-10.674	-10.749	0.075
	ARIMA(7,0,1)	-10.728	-10.635	-10.820	0.185
	ARIMA(7,0,0)	-10.728	-10.633	-10.823	0.190
	ARIMA(5,0,0)	-10.666	-10.560	-10.772	0.212
	ARIMA(5,0,1)	-10.690	-10.564	-10.816	0.252
	ARIMA(6,0,0)	-10.674	-10.546	-10.802	0.256
	ARIMA(5,0,2)	-10.695	-10.564	-10.826	0.262
	ARIMA(9,0,0)	-10.712	-10.581	-10.843	0.262
	ARIMA(0,0,5)	-10.672	-10.519	-10.825	0.306
	ARIMA(6,0,1)	-10.671	-10.505	-10.837	0.332
2018	ARIMA(8,0,0)	-10.659	-10.620	-10.697	0.077
	ARIMA(7,0,0)	-10.652	-10.554	-10.751	0.197
	ARIMA(7,0,1)	-10.657	-10.558	-10.756	0.198
	ARIMA(5,0,0)	-10.641	-10.534	-10.747	0.213
	ARIMA(6,0,0)	-10.650	-10.521	-10.780	0.259
	ARIMA(5,0,2)	-10.648	-10.516	-10.780	0.264
	ARIMA(9,0,0)	-10.667	-10.534	-10.800	0.266
	ARIMA(5,0,1)	-10.654	-10.520	-10.788	0.268
	ARIMA(0,0,5)	-10.636	-10.474	-10.798	0.324
	ARIMA(6,0,1)	-10.646	-10.480	-10.812	0.332
2019	ARIMA(8,0,0)	-10.555	-10.516	-10.593	0.077
	ARIMA(5,0,0)	-10.599	-10.489	-10.709	0.220
	ARIMA(7,0,0)	-10.568	-10.442	-10.695	0.253
	ARIMA(6,0,0)	-10.608	-10.475	-10.742	0.267
	ARIMA(9,0,0)	-10.559	-10.425	-10.693	0.268
	ARIMA(5,0,1)	-10.611	-10.475	-10.748	0.273
	ARIMA(7,0,1)	-10.566	-10.420	-10.712	0.292
	ARIMA(0,0,5)	-10.576	-10.413	-10.739	0.326
	ARIMA(5,0,2)	-10.611	-10.442	-10.780	0.338
	ARIMA(6,0,1)	-10.608	-10.430	-10.786	0.356
2020	ARIMA(8,0,0)	-10.518	-10.467	-10.569	0.102
	ARIMA(5,0,0)	-10.557	-10.443	-10.671	0.228
	ARIMA(6,0,0)	-10.564	-10.426	-10.703	0.277
	ARIMA(5,0,1)	-10.568	-10.425	-10.710	0.285
	ARIMA(7,0,0)	-10.545	-10.400	-10.691	0.291
	ARIMA(7,0,1)	-10.533	-10.384	-10.681	0.297
	ARIMA(0,0,5)	-10.513	-10.348	-10.677	0.329
	ARIMA(5,0,2)	-10.563	-10.389	-10.737	0.348
	ARIMA(9,0,0)	-10.529	-10.346	-10.713	0.367
	ARIMA(6,0,1)	-10.568	-10.378	-10.757	0.379

Table 6. Best ARIMA models for 2017-2026 to forecast b_t- Contd..

Year	Model	Forecast b_t	95% Confidence Interval		Interval length=UCL-LCL
			UCL	LCL	
2021	ARIMA(8,0,0)	-10.486	-10.435	-10.537	0.102
	ARIMA(5,0,0)	-10.511	-10.397	-10.625	0.228
	ARIMA(6,0,0)	-10.516	-10.377	-10.655	0.278
	ARIMA(5,0,1)	-10.519	-10.375	-10.662	0.287
	ARIMA(7,0,0)	-10.537	-10.390	-10.685	0.295
	ARIMA(7,0,1)	-10.530	-10.363	-10.697	0.334
	ARIMA(0,0,5)	-10.434	-10.264	-10.603	0.339
	ARIMA(5,0,2)	-10.521	-10.346	-10.696	0.350
	ARIMA(9,0,0)	-10.501	-10.317	-10.685	0.368
	ARIMA(6,0,1)	-10.518	-10.327	-10.708	0.381
2022	ARIMA(8,0,0)	-10.372	-10.315	-10.428	0.113
	ARIMA(5,0,0)	-10.374	-10.239	-10.509	0.270
	ARIMA(7,0,0)	-10.357	-10.210	-10.505	0.295
	ARIMA(6,0,0)	-10.371	-10.207	-10.536	0.329
	ARIMA(7,0,1)	-10.368	-10.200	-10.535	0.335
	ARIMA(5,0,1)	-10.356	-10.185	-10.527	0.342
	ARIMA(5,0,2)	-10.358	-10.170	-10.545	0.375
	ARIMA(9,0,0)	-10.384	-10.183	-10.584	0.401
	ARIMA(0,0,5)	-10.333	-10.131	-10.535	0.404
	ARIMA(6,0,1)	-10.383	-10.167	-10.600	0.433
2023	ARIMA(8,0,0)	-10.308	-10.238	-10.378	0.140
	ARIMA(5,0,0)	-10.227	-10.091	-10.363	0.272
	ARIMA(7,0,0)	-10.260	-10.112	-10.408	0.296
	ARIMA(6,0,0)	-10.225	-10.060	-10.390	0.330
	ARIMA(7,0,1)	-10.275	-10.105	-10.446	0.341
	ARIMA(5,0,1)	-10.230	-10.058	-10.402	0.344
	ARIMA(5,0,2)	-10.234	-10.044	-10.423	0.379
	ARIMA(0,0,5)	-10.255	-10.053	-10.457	0.404
	ARIMA(6,0,1)	-10.240	-10.024	-10.456	0.432
	ARIMA(9,0,0)	-10.282	-10.055	-10.509	0.454
2024	ARIMA(8,0,0)	-10.234	-10.164	-10.305	0.141
	ARIMA(5,0,0)	-10.118	-9.976	-10.260	0.284
	ARIMA(7,0,0)	-10.149	-9.998	-10.299	0.301
	ARIMA(7,0,1)	-10.158	-9.988	-10.329	0.341
	ARIMA(6,0,0)	-10.122	-9.948	-10.295	0.347
	ARIMA(5,0,1)	-10.133	-9.953	-10.313	0.360
	ARIMA(0,0,5)	-10.177	-9.975	-10.379	0.404
	ARIMA(5,0,2)	-10.122	-9.909	-10.334	0.425
	ARIMA(9,0,0)	-10.205	-9.978	-10.433	0.455
	ARIMA(6,0,1)	-10.125	-9.896	-10.354	0.458

Table 6. Best ARIMA models for 2017-2026 to forecast b_t- Contd..

Year	Model	Forecast b_t	95% Confidence Interval		Interval length=UCL-LCL
			UCL	LCL	
2025	ARIMA(8,0,0)	-10.038	-9.967	-10.110	0.143
	ARIMA(5,0,0)	-10.036	-9.886	-10.186	0.300
	ARIMA(7,0,0)	-10.030	-9.879	-10.182	0.303
	ARIMA(7,0,1)	-10.022	-9.850	-10.195	0.345
	ARIMA(6,0,0)	-10.046	-9.863	-10.230	0.367
	ARIMA(5,0,1)	-10.058	-9.868	-10.248	0.380
	ARIMA(0,0,5)	-10.099	-9.897	-10.301	0.404
	ARIMA(5,0,2)	-10.052	-9.833	-10.272	0.439
	ARIMA(9,0,0)	-10.011	-9.777	-10.246	0.469
	ARIMA(6,0,1)	-10.037	-9.795	-10.280	0.485
2026	ARIMA(8,0,0)	-9.915	-9.838	-9.992	0.154
	ARIMA(5,0,0)	-9.958	-9.807	-10.108	0.301
	ARIMA(7,0,0)	-9.952	-9.800	-10.103	0.303
	ARIMA(7,0,1)	-9.933	-9.760	-10.106	0.346
	ARIMA(6,0,0)	-9.972	-9.789	-10.155	0.366
	ARIMA(5,0,1)	-9.981	-9.791	-10.171	0.380
	ARIMA(0,0,5)	-10.021	-9.820	-10.223	0.403
	ARIMA(5,0,2)	-9.968	-9.746	-10.191	0.445
	ARIMA(6,0,1)	-9.965	-9.721	-10.208	0.487
	ARIMA(9,0,0)	-9.922	-9.675	-10.169	0.494

After constructing the Trend Models and determining the values of a_t and b_t, we demonstrate using the models to forecast score-by-rank and rank-by-score for the future

Table 7: Goodness-of-fit measures except stationary R^2 and R^2.

Model	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC
ARIMA(8,0,0)	0.089	0.225	0.026	0.829	0.095	-2.777
ARIMA(7,0,1)	0.091	0.241	0.027	0.847	0.097	-2.723
ARIMA(7,0,0)	0.076	0.255	0.029	0.847	0.097	-3.290
ARIMA(5,0,0)	0.063	0.287	0.032	0.854	0.098	-4.093
ARIMA(5,0,1)	0.069	0.292	0.033	0.834	0.096	-3.698
ARIMA(6,0,0)	0.070	0.289	0.033	0.845	0.097	-3.670
ARIMA(5,0,2)	0.078	0.267	0.030	0.850	0.098	-3.243
ARIMA(9,0,0)	0.130	0.231	0.026	0.847	0.097	-1.804
ARIMA(0,0,5)	0.068	0.306	0.035	0.777	0.089	-3.934
ARIMA(6,0,1)	0.082	0.297	0.034	0.861	0.099	-3.127

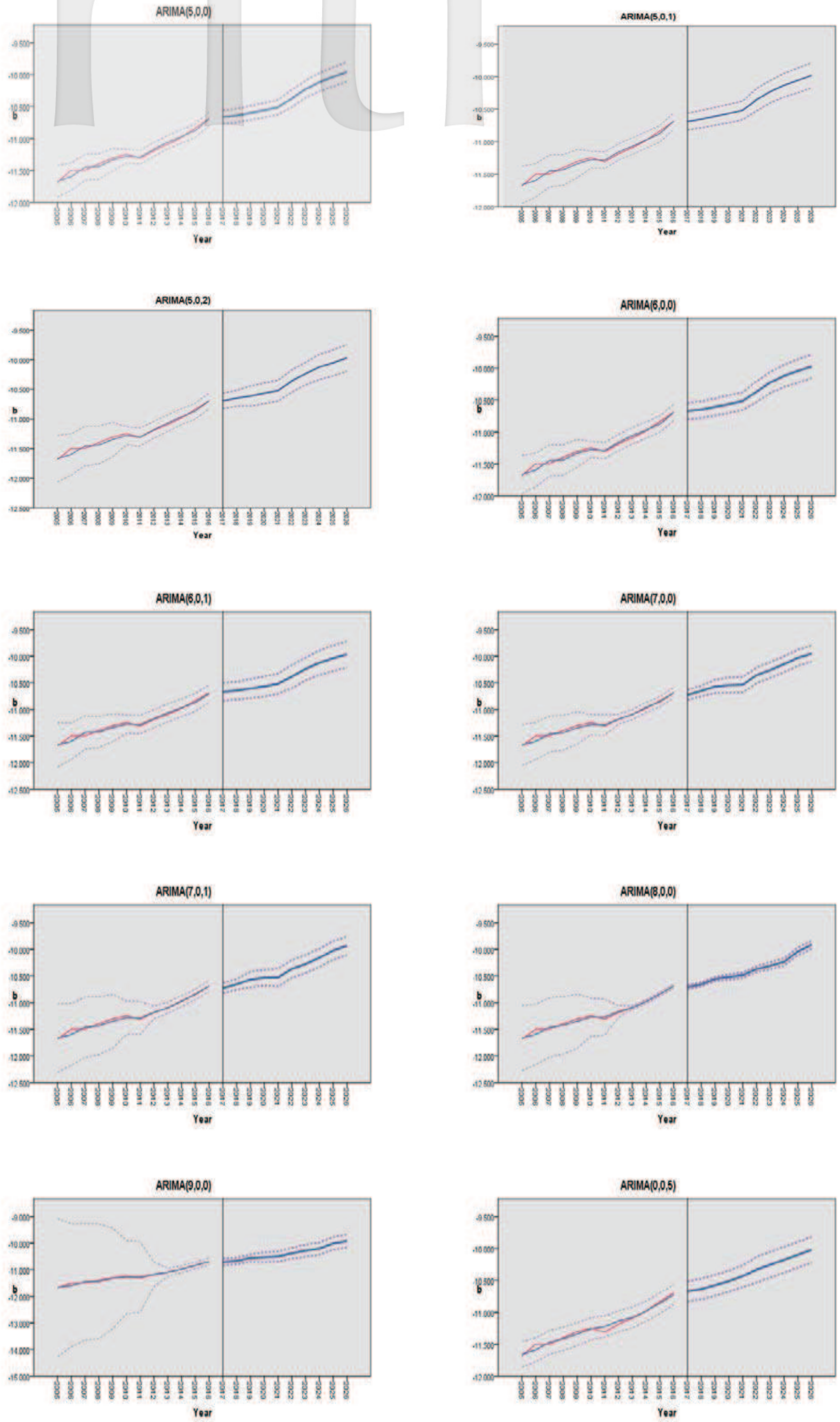

Figure 2: Confidence interval of b_t in ten ARIMA models in 2017-2026.

Table 8: Values of parameters a_t and b_t, and Trend Models from 2017 to 2026.

Year (t)	b_t^*	a_t	Trend Model
2017	-10.711	74.946	$\text{Score}_{2017} = 74.946 - 10.711 \cdot \ln(\text{Rank}_{2017})$ $\text{Rank}_{2017} = \exp\left(\frac{\text{Score}_{2017} - 74.946}{-10.711}\right)$
2018	-10.659	74.719	$\text{Score}_{2018} = 74.719 - 10.659 \cdot \ln(\text{Rank}_{2018})$ $\text{Rank}_{2018} = \exp\left(\frac{\text{Score}_{2018} - 74.719}{-10.659}\right)$
2019	-10.555	74.265	$\text{Score}_{2019} = 74.265 - 10.555 \cdot \ln(\text{Rank}_{2019})$ $\text{Rank}_{2019} = \exp\left(\frac{\text{Score}_{2019} - 74.265}{-10.555}\right)$
2020	-10.518	74.104	$\text{Score}_{2020} = 74.104 - 10.518 \cdot \ln(\text{Rank}_{2020})$ $\text{Rank}_{2020} = \exp\left(\frac{\text{Score}_{2020} - 74.104}{-10.518}\right)$
2021	-10.486	73.964	$\text{Score}_{2021} = 73.964 - 10.486 \cdot \ln(\text{Rank}_{2021})$ $\text{Rank}_{2021} = \exp\left(\frac{\text{Score}_{2021} - 73.964}{-10.486}\right)$
2022	-10.372	73.467	$\text{Score}_{2022} = 73.467 - 10.372 \cdot \ln(\text{Rank}_{2022})$ $\text{Rank}_{2022} = \exp\left(\frac{\text{Score}_{2022} - 73.467}{-10.372}\right)$
2023	-10.308	73.188	$\text{Score}_{2023} = 73.188 - 10.308 \cdot \ln(\text{Rank}_{2023})$ $\text{Rank}_{2023} = \exp\left(\frac{\text{Score}_{2023} - 73.188}{-10.308}\right)$
2024	-10.234	72.865	$\text{Score}_{2024} = 72.865 - 10.234 \cdot \ln(\text{Rank}_{2024})$ $\text{Rank}_{2024} = \exp\left(\frac{\text{Score}_{2024} - 72.865}{-10.234}\right)$
2025	-10.038	72.01	$\text{Score}_{2025} = 72.01 - 10.038 \cdot \ln(\text{Rank}_{2025})$ $\text{Rank}_{2025} = \exp\left(\frac{\text{Score}_{2025} - 72.01}{-10.038}\right)$
2026	-9.915	71.473	$\text{Score}_{2026} = 71.473 - 9.915 \cdot \ln(\text{Rank}_{2026})$ $\text{Rank}_{2026} = \exp\left(\frac{\text{Score}_{2026} - 71.473}{-9.915}\right)$

$^*b_t$ is forecasted by model ARIMA(8,0,0).

years in the following examples. In the first score-by-rank example, to predict the overall score of rank 100, 300 and 500 in 2017, with $a_{2017} = 71.946$ and $b_{2017} = -10.711$, our model predicts rank 100 with an overall score of 25.620 (see calculation): $\text{Score}_{2017} = 74.946 - 10.711 \cdot \ln(100) = 25.620$; the predicted overall score of rank 300 in 2017 would be 13.853: $\text{Score}_{2017} = 74.946 - 10.711 \cdot \ln(300) = 13.853$; and the predicted overall score of rank 500 in 2017 would be 8.381: $\text{Score}_{2017} = 74.946 - 10.711 \cdot \ln(500) = 8.381$. If an institution aims to enter ARWU top 100 in 2017, it should have an overall score of a_t least 25.620. To be ranked at 300, the score would be 13.853; and a score of at least 8.381 would be needed to be ranked at 500 in 2017.

Similarly, to predict the scores needed to be ranked at 500 in 2018, 2019 and 2020, using the particular parameters a_t and b_t for each year, our models forecast the scores in the following calculations: $\text{Score}_{2018} = 74.719 - 10.659 \cdot \ln(500) = 8.478$; $\text{Score}_{2019} = 74.265 - 10.555 \cdot \ln(500) = 8.670$; $\text{Score}_{2020} = 74.104 - 10.518 \cdot \ln(500) = 8.739$. If an institution aims to enter ARWU at 500 in 2018, its overall score should be at least 8.478; to be ranked at 500 in 2019, the overall score should be at least 8.670; and to be ranked at 500 in 2020, the overall score should be at least 8.739.

In the rank-by-score examples, if an institution earns an overall score 30 in the next four years, from 2017 to 2020, using the particular parameters a_t and b_t for each year, our models forecast the rank placement to be 66 in the following calculations:

$$\begin{aligned}
\text{Rank}_{2017} &= \exp\left(\frac{\text{Score}_{2017} - 74.946}{-10.711}\right) = \exp\left(\frac{30 - 74.946}{-10.711}\right) = 66.4;\\
\text{Rank}_{2018} &= \exp\left(\frac{\text{Score}_{2018} - 74.719}{-10.659}\right) = \exp\left(\frac{30 - 74.719}{-10.659}\right) = 66.4;\\
\text{Rank}_{2019} &= \exp\left(\frac{\text{Score}_{2019} - 74.265}{-10.555}\right) = \exp\left(\frac{30 - 74.265}{-10.555}\right) = 66.3;\\
\text{Rank}_{2020} &= \exp\left(\frac{\text{Score}_{2020} - 74.104}{-10.518}\right) = \exp\left(\frac{30 - 74.104}{-10.518}\right) = 66.2.
\end{aligned}$$

From the above examples, we conclude that predictions by Trend Models can be utilized as useful reference for institutions wishing to have a better understanding of their ranking performance and placement in future years.

6. Empirical Study

To verify the validity of our Trend Models, we used ranking data from 2005 to 2015 to forecast the overall scores of the top 500 ranks in 2016. Following the two-stage process, mentioned in the previous section, we could find the parameters values of a_t and b_t for 2016. In the first stage, using known paired parameters (a_t, b_t), $t = 2005, 2006, \ldots, 2015$, from Table 4 as a sample of the bivariate (a, b), we found a very good relationship between a_{2016} and b_{2016} with the following linear regression:

$$a_{2016} = 26.474 - 4.517 b_{2016}. \tag{6.1}$$

In the second stage, we found the value of parameter b_{2016} for the Trend Model by looking up the b_t values in Table 4 from 2005 to 2015 as time series data, and then simulate ARIMA models by SPSS statistical package to find suitable ARIMA models of b_{2016}. As a result, there were 10 models that had high stationary R^2 and R^2 values, ranging from 0.970 to 0.974; all of them fit very well. Among these 10 models, the ARIMA(7, 0, 0)

Table 9: The most ideal Trend Model for 2016.

Year (t)	b^*_{2016}	a_{2016}	Trend Model
2016	-10.716	74.878	$\text{Score}_{2016} = 74.878 - 10.716 \cdot \ln(\text{Rank}_{2016})$ $\text{Rank}_{2016} = \exp\left(\frac{\text{Score}_{2016} - 74.878}{-10.716}\right)$

$^*b$ is forecasted by model ARIMA(7,0,0).

model had the shortest confidence interval length (0.351) for 2016. Hence, ARIMA(7,0,0) was the most ideal model for forecasting b_{2016}. Once the value of b_{2016} was obtained, the value of a_{2016} followed from Eq. (5.1). The Trend Model for 2016 is shown in Table 9. By using this Trend model, we obtained the estimated scores of all 500 ranks in 2016, and generated a curve of score by rank (as shown in dotted red in Figure 3). A curve of score by rank of the real results from all 500 ranks in 2016 is also shown in Figure 3 (in solid black).

As shown in Figure 3, the two curves are very similar in shapes and patterns, and are almost overlapping. This indicates that our Trend Model produces outcomes that closely represent the real results. Therefore, we conclude that our Trend Model is valid.

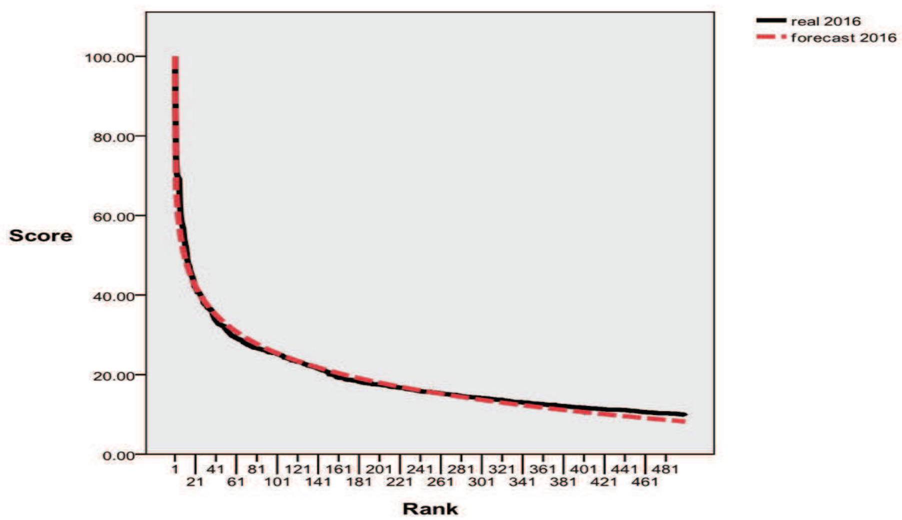

Figure 3: Curves of real scores and forecast scores of top 500 ranks in 2016.

7. Discussion and Conclusion

ARWU is one of the best known international ranking of universities that draws on six different indicators to measure the academic performance of higher education institutions. Examining the relationship between ARWU's overall scores and ranks from

2004 to 2016, it generates thirteen stable trend curves that fit well to the natural log regression model, which we have constructed into the Score-Rank Models. This paper shows that estimations of scores and ranks by the Score-Rank Models have produced close and reliable outcomes for the 2004-2016 league tables. Furthermore, based on data patterns observed from 2004 to 2016, we have also built the Trend Models to predict ranks and scores in the next 10 years.

According to statistical theory, when using regression models or time series models from times series data to forecast the future, the outcomes fit better to the time periods that are closer to the time of prediction. This phenomenon is illustrated in Figure 2, where all of the confidence bands have a narrower area in the time periods that are closer to the present (current) time of prediction. This is also shown in the empirical study, which uses ranking data from 2005 to 2015 to forecast the overall scores of the top 500 ranks in 2016. When comparing the forecast results with the real ranking outcomes of 2016 in a graph, it presents two very similar and almost overlapping curves. Therefore, we suggest that after the most recent data have been announced, for a more precise forecast, institutions could rebuild the Trend Model for the next immediate year by using data from 2005 to present (current) year.

To excel in the global competition in higher education, and to increase global visibility by appearing on ARWU, we suggest institutions of higher education using the Trend Models as valuable references in setting higher performance goals in their long-term strategic planning. The models can be used as practical tools to obtain better outcome forecasts, for ideal placement in ranking scores and ranks. Rankings of world universities have become increasingly popular and useful references for stakeholders in the global higher education market. Therefore, it is crucial for decision makers in higher education to have a clear understanding of their institutions' performance and placement in the global competition.

Acknowledgements

The authors greatly appreciate the anonymous referees for their valuable and helpful comments and suggestions on an earlier version of the manuscript.

Appendix

There are three models in linear time series: (1) Autoregressive (AR) models, (2) Integrated (I) models, and (3) Moving average (MA) models. If a model is a combination of model (1) and model (3), it is then called an autoregressive moving average (ARMA) model; and if it is a combination of models (1), (2) and (3), then it is called a autoregressive integrated moving average (ARIMA) model. ARIMA(p, r, q) indicates that a model is a combination of the model AR(p) and the model MA(q) for the r^{th} order difference of series data, the I(r) series data; where AR(p) is the autoregressive model of lag p, MA(q) is moving average model of order q, and I(r) is the integrated model of order r. Moreover, ARMA(p, q) is a combination of AR(p) and MA(q), which is called the ARMA model. Other combinations can be denoted as follows: ARIMA$(p,0,q)$ = ARMA(p,q), ARIMA$(p,0,0)$ = ARMA$(p,0)$ =AR(p), and ARIMA$(0,0,q)$ = ARMA$(0,q)$ = MA(q). The series data $\{b_t\}$ from 2005 to 2016 has only 12 data points; therefore, its sample size is only 12. All the other ARIMA(p,r,q) models not indicated in Table 5 either have low stationary R^2 (for $r > 0$ and $p+r+q \leq 9$) or lower R^2 (for $r = 0$, $p < 5$ and $q < 5$), or cannot even be built due to not having enough data points (for $p+r+q > 9$).

The presented model fitting ARIMA statistics include goodness-of-fit measures by SPSS statistical software, namely stationary R^2, R^2, root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), maximum absolute percentage error (MaxAPE), maximum absolute error (MaxAE) and Normalized Bayesian Information Criterion (Normalized BIC). Bayesian Information Criterion (BIC) is closely related to the Akaike information criterion (AIC), where the outcomes are the smaller the better. Stationary R^2 and R^2 are looking for outcomes not larger than 1, and the closer they are to 1, the better. RMSE, MAPE, MAE, MaxAPE, and MaxAE are different measures of distance between the predicted and the actual values of the models and should, as such, be as close to 0 as possible.

Stationary R^2 is to measure whether difference of series data would be better than the original series data; that is, to check that the series data is stationary or not, instead of testing unit root. A large stationary R^2 means that the original series data will be better than difference of series data; that is, the series data can be accepted as stationary. R^2 is a coefficient of determinant to measure model fitting; and confident interval length is to measure accuracy of forecast value. In general, the larger the R^2, the smaller the mean square error (MSE) and root mean square error (RMSE), and similarly, the smaller the MAPE, MAE, MaxAPE, and MaxAE.

The goodness-of-fit measures also include information concerning residuals, namely autocorrelation function (ACF) and partial autocorrelation function (PACF), the Ljung-Box Q test. Since the time series data $\{b_t\}$ from 2005 to 2016 has only 12 data points (a sample size of only 12), such short time series data does not generate much meaning in using autocorrelation function (ACF), partial autocorrelation function (PACF), or the Ljung-Box Q test. Instead, we found all possible ARIMA(p,r,q) models with $p+r+q \leq 9$ by simulation and kept 29 models in comparison that had the highest stationary R^2 and R^2 values ranging from 0.975 to 0.983, which were all rounded off to 0.98.

References

[1] ARWU. (2003-2016). Academic Ranking of World Universities 2003 - Academic Ranking of World Universities 2016. Retrieved from http://www.shanghairanking.com/ARWU2003.html - http://www.shanghairanking.com/ARWU2016.html

[2] ARWU. (2016). Academic Ranking of World Universities 2016 - Methodology. Retrieved from http://www.shanghairanking.com/ARWU-Methodology-2016.html

[3] Bauer, J., Leydesdorff, L. and Bornmann, L. (2015). *Highly-cited papers in Library and Information Science (LIS) : Authors, Institutions, and Network Structures*, Journal of the Association for Information Science and Technology, Vol.67, No.12, 3095-3100.

[4] Dehon, C., McCathie, A. and Verardi, V. (2010). Uncovering excellence in academic rankings: a closer look at the Shanghai Ranking. Scientometrics, Vol.83, No.2, 515-524.

[5] Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, London, UK.

[6] Huang, M. (2011). *A Comparison of Three Major Academic Rankings for World Universities: From a Research Evaluation Perspective*, Journal of Library and Information Studies, Vol.9, No.1, 1-25.

[7] Liu, N. C. (2009). The Story of Academic Ranking of World Universities. International Higher Education, 54(Winter), 2-3.

[8] Liu, N. C., Cheng, Y. and Liu, L. (2005). *Academic ranking of world universities using scientometrics - A comment to the "Fatal Attraction"*, Scientometrics, Vol.64, No.1, 101-109.

[9] Luque-Martnez, T. and del Barrio-Garca, S. (2016). *Constructing a synthetic indicator of research activity*, Scientometrics, Vol.108, No.3, 1049-1064.

[10] Marginson, S. (2017). *Do Rankings Drive Better Performance*? International Higher Education, Vol.89(Spring), 6-8.

[11] QS Top Universities. (2017). QS World University Rankings - Methodology. Retrieved from https://www.topuniversities.com/qs-world-university-rankings/methodology.

[12] Razvan, F. (2007). *Irreproducibility of the Results of the Shanghai Academic Ranking of World Universities*, Scientometrics, Vol.72, No.1, 25-32.

[13] Times Higher Education. (2016). World University Rankings 2016-17 methodology, Retrieved from https://www.timeshighereducation.com/world-university-rankings/methodology-world- university-rankings-2016-2017

[14] Van Raan, A. (2005a). *Fatal Attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods*, Scientometrics, Vol.62, No.1, 133-143.

[15] Van Raan, A. (2005b). Challenges in Ranking of Universities. Invited paper for the First International Conference on World Class Universities. Shanghai, China: Shanghai Jaio Tong University (June 16-18).

[16] Yonezawa, A. (2015). *Will the Ranking Game Continue After a Decade*? International Higher Education, Vol.80(Spring), 19-20.

[17] Zitt, M. and Filliatreau, G. (2007). Big is (made) Beautiful: Some comments about the Shanghai ranking of world-class universities. In J. Sadlack, & N. C. Liu (Eds.), The World Class University and Ranking: Aiming beyond Status (141-160). Romania: UNESCO-CEPES, Cluj University Press.

Department of Management Sciences, Tamkang University, 25137, Taiwan.

E-mail: farrah@mail.tku.edu.tw

Major area(s): Teaching evaluation, educational technology.

Department of Management Sciences, Tamkang University, 25137, Taiwan.

E-mail: liangyuh@mail.tku.edu.tw

Major area(s): Inventory management, supply chain management, mathematical programming.

(Received June 2017; accepted November 2017)