

# Latent Variable Analysis: A New Approach to University Ranking<sup>1</sup>

CASSANDAR GUARINO, GREG RIDGEWAY,  
MARC CHUN and RICHARD BUDDIN

*This study applies a Bayesian latent variable analysis to the task of determining rankings of universities in the UK and US, on the basis of a set of quality-related measures. It estimates the degree of uncertainty in the rankings and permits the assessment of statistically significant differences across universities. It also provides a methodology for determining the weighting of various measures that is based on the patterns embedded in the data and compares the latent variable rankings with traditional weight-and-sum rankings. Overall, the methodology contributes to a better understanding of ranking efforts and illustrates the need for caution in interpreting distinctions published in traditional ranking systems.*

## Introduction

Interest in rankings of institutions of higher education has intensified over the course of the last two decades. Ranking efforts have expanded considerably since the publication of the first *U.S. News & World Report* annual ranking of *America's Best Colleges* in 1983. In addition to several popular ranking systems in the United States<sup>2</sup> and the United Kingdom,<sup>3</sup> many others exist in other countries.<sup>4</sup> Since 2003, two ranking systems that span international borders have emerged – the *Academic Ranking of World Universities* by the Shanghai Jiao Tong University (SJTU)<sup>5</sup> and the *World University Rankings* by *The Times Higher Education Supplement*.<sup>6</sup> The trend suggests that these systems will continue to proliferate.

<sup>1</sup> The authors performed this work in RAND Education, a unit of RAND Corporation. The research activity was funded from a contract from the State of Qatar Supreme Education Council. The authors wish to express their appreciation for this support. The present article derives from a research project aimed at identifying high-quality institutions of higher education, for the attention of Qatari students with government-sponsored scholarships. This paper does not represent the views of the State of Qatar or its Supreme Education Council.

<sup>2</sup> See *U.S. News & World Report's America's Best Colleges* at <<http://www.usnews.com/usnews/home.htm>>; *Barron's Profiles of American Colleges* at <<http://barronseduc.com/>>; *The Top American Research Universities*, published by TheCenter at the University of Florida at <<http://thecenter.ufl.edu/research2002.html>>, and Avery *et al.* (2004).

<sup>3</sup> See *The Times Good University Guide* at <<http://www.timesonline.co.uk/section/0,,716,00.html>>; *The Guardian University Guide* at <<http://education.guardian.co.uk/universityguide/0,10085,488282,00.html>>, and the *Sunday Times University Guide* at <<http://www.timesonline.co.uk/section/0,,8403,00.html>>.

<sup>4</sup> See, for example, *Maclean's Guide to Canadian Universities* at <<http://www.macleans.ca/universities/index.jsp>>; *Hobsons Good Universities Guide* at <<http://www.thegoodguides.com.au/>>, and the Melbourne Institute *Index of the International Standing of Australian Universities* at <[http://www.melbourneinstitute.com/publications/reports/aus\\_uni/mainpaper.pdf](http://www.melbourneinstitute.com/publications/reports/aus_uni/mainpaper.pdf)>, 22 May 2005>, for Australia.

<sup>5</sup> Available at <<http://ed.sjtu.edu.cn/ranking.htm>>.

<sup>6</sup> Available at <<http://www.thes.co.uk/worldrankings/>>.

Most ranking systems use a ‘weight-and-sum’ approach. They collect university-level data on a set of measures related to educational quality, such as the selectivity of the student body or ratings of prestige. They then assign weights to each measure, generally based on subjective opinions of the relative importance of each indicator, and sum the weighted measures for each institution. The weighted sums produce an ordinal ranking of institutions. These rankings, however, can be somewhat misleading in that they may ‘over-differentiate’ among institutions, assigning different rankings to institutions that may be more or less indistinguishable.

This paper proposes an alternate approach to ranking institutions along observable quality-related dimensions. The approach involves the use of a statistical procedure known as latent variable analysis. The result is a sequential ordering of individual institutions that carries with it information that allows us to assess whether differently ranked institutions are statistically distinguishable.

Our proposed approach can be used to address a number of issues related to university ranking systems. In this study, we answer the following questions:

1. How do we properly account for uncertainty in the assignment of ranks to different institutions?
2. When are measured quality differences between institutions statistically significant?
3. How important are different input measures to the overall ranking?

Answers to these questions have practical implications for potential students and the higher educational institutions themselves. Answers to the first two questions allow us to obtain a sense of the degree to which certain institutions are similar or dissimilar and also allow the possibility of differentiating among institutions in a meaningful way. An answer to the third question can reveal the nature of the factors creating similarity or dissimilarity.

The proposed latent variable approach is not a panacea, however. Like weight-and-sum ranking systems, the approach relies on observable ‘quality’ indicators and does not offset the deficiencies inherent in the indicators themselves or in the manner in which they are operationalized and measured. Better measures of what students gain from attending an institution would improve the precision of current ranking methods as well as the approach proposed here. Lacking better data, however, we illustrate the latent variable method with existing data from two prominent ranking systems, *The Times Good University Guide* and the *U.S. News & World Report*, and discuss important advantages of this approach over the standard weight-and-sum approach.

## Background

Ostensibly, rankings serve an important purpose in providing information to the general public and a means of fostering accountability. Shattock (2003: 5) states

There can be no doubt of the public interest in such assessments, nor that such interest has legitimacy, and any evaluation of university success must take their findings seriously.

Rankings are closely followed by university administrators and affect policy choices. Monks and Ehrenberg (1999) found, for example, that the admissions and pricing

policies of highly selective US institutions of higher education were linked to the rankings published by the *U.S. News & World Report*.

Existing ranking systems have been subject to a litany of criticisms. University administrators complain because they must struggle to deal with the consequences of shifts in annual rankings, and some<sup>7</sup> have questioned the integrity of the entire enterprise. At the most basic level, rankings have reframed higher education as a consumer good; doing so requires the savvy participant to think in terms of an abstract 'best model' as well as the best value for the dollar, as emphasized in the *U.S. News & World Report*. Recent studies lend validity to this notion. Brewer

*et al.* (1999), Hoxby (2001), and Black *et al.* (1997) find that the earnings of students from prestigious or highly ranked institutions are significantly higher than those of students with similar backgrounds who attend less prestigious colleges.

Setting aside the normative question of whether or not rankings *should* be done, rankings are handicapped at the outset by methodological concerns. First and foremost, no clear or universally agreed-upon measure of quality in higher education exists. Absent any consensus, each ranking system combines an available set of observable indicators – each of which serves as a rough proxy for some factor notionally tied to quality – in a formula that can be questioned both on the basis of its contributing elements and the manner in which the elements are combined.

Second, when considering multiple ranking systems, it becomes questionable whether or not meaningful differences between institutions exist. A strong and noteworthy criticism of ranking systems is that statistically significant differences generally do not emerge between adjacently ranked universities or even between universities that find themselves at a distance from one another the rankings (Clarke, 2002a, b).

Third, ranking systems are not neutral to the institutions they study but instead affect them. Many ranking systems heavily weight indicators of reputation, for example. Given that the rankings themselves play a prominent role in affecting reputation, the circular nature of these endeavours makes them a particularly strong self-perpetuating force. Thus, it is important to ask whether this force acts in the best interest of students and society. If universities are to be judged by the standards set by ranking systems and have strong incentives to conform to them, does moving in this direction take us closer to or further from true educational quality? And, if moving in this direction is inevitable, can we ensure that the rankings be carried out in a responsible manner?

Due to the potentially significant consequences these rankings have in influencing student choices and institutional behaviour, the higher education community should encourage improved ranking methodologies and the collection of better measures of quality. In this paper, we describe a new methodological tool for examining quality differences among institutions and illustrate the efficacy of the tool using existing data measures. The new approach can supplement existing ranking methodologies and provide insights into the nature of distinctions among institutions.

---

<sup>7</sup> For example, President Gerhard Casper of Stanford University. Details at <<http://news-service.stanford.edu/news/1997/april23/usstatement.html>>.

TABLE 1. Indicators and weights used in *The Times Good University Guide 2005*

University feature	Weight	Description
Teaching	2.5	University-wide average of Teacher Quality Assessment scores in individual departments. Maximum possible score is 24
Research	1.5	University-wide average of Research Assessment Exercise scores in individual departments
Entry standards	1	Average A-level score (or Scottish Higher score) of new students under the age of 21. Maximum possible score is 30
Student: staff ratio	1	Number of student full-time equivalents (FTE) divided by total teaching FTE
Library and computer spending	1	Spending on library staff and holdings and computer hardware and software divided by student FTE
Facilities spending	1	Spending on facilities divided by student FTE
Percent high degrees	1	Percentage of graduates achieving first and upper second class degrees
Graduate destinations	1	Proportion of graduates that enter further study or a graduate track job
Completion rate	1	Length of time students take to complete degree compared with length of time they would be expected to study if they completed the course normally

Source: *The Times Good University Guide 2005*.

## Data

The data for this study are drawn from primary data used to construct rankings systems in the United Kingdom and the United States. The first dataset is from *The Times Good University Guide 2005*, and the second is from the *U.S. News & World Report*.

### *The Times Good University Guide Data*

*The Times* has been publishing institutional rankings since 1992. Nine indicators are used in the calculation of the rankings. They are based upon the data collected by the Higher Education Statistics Association to support the resource allocation decisions of the Higher Education Funding Councils. Table 1 describes the indicators and their assigned weights.<sup>8</sup> More detailed explanations regarding the indicators can be found in *The Times Good University Guide 2005*.<sup>9</sup>

Table 2 displays the means, standard deviations, and minimum and maximum values of the quality indicators pertaining to the 99 institutions of higher education represented in the dataset for the UK. A quick inspection of this table suggests that some of these variables have a higher degree of variability than others. The facilities, library, and research variables, in particular, exhibit a wide range of values.<sup>10</sup> It is worth noting that all variables are significantly correlated with one another, and many of the correlations are high. Only the library, facilities, and destinations variables show correlations with other variables that fall below 0.5.

<sup>8</sup> Data used with permission from *The Times*.

<sup>9</sup> In particular, see pp. 13–18.

<sup>10</sup> This can be seen from both the range and the coefficients of variation, i.e., the standard deviation divided by the mean. For the facilities, library, and research variables, the coefficients of variation are equal to 0.41, 0.38, and 0.38, respectively.

TABLE 2. Descriptive statistics for features of 99 universities in the UK\*

University features	Mean	Standard deviation	Minimum	Maximum
Teaching	21.48	0.82	18.8	23.1
Research	4.01	1.52	0.5	6.6
Entry standards	18.68	4.84	11.5	29.5
Percent high degrees	59.52	10.53	39.3	89.3
Destinations	71.57	6.87	54.9	89.8
Completion rate	85.29	6.53	65.0	98.0
Faculty: student ratio	0.06	0.02	0.04	0.14
Library	565.61	215.15	321	1,552
Facilities	196.07	79.46	11	460

\* Statistics computed using data from *The Times Good Universities Guide 2005*.

Source: The authors.

### ***U.S. News & World Report Data***

The *U.S. News & World Report* (henceforth, *U.S. News*) builds a dataset for its annual publication of *America's Best Colleges*.<sup>11</sup> The *U.S. News* provides separate rankings for several categories of higher educational institutions: national research universities, liberal arts colleges, master's granting institutions, and comprehensive colleges. For the sake of brevity, we utilize only data related to national research universities in this study. In addition, although the *U.S. News* provides data on 249 research universities, it ranks only the top 129 universities and reports fewer data for the others. Therefore, we have restricted our analyses to these 129 institutions.

The *U.S. News* rankings use 17 measures considered related to institutional quality. These are shown in Table 3.<sup>12</sup> Only 14 of these were available on the website. The *U.S. News* uses a twofold-weighting scheme, in which sub-factors are combined with assigned weights to form aggregate factors, and the aggregate factors are combined with assigned weights to form the total score.

Table 4 shows descriptive statistics for these measures for the subset of 129 research universities. A considerable range in the indicators across schools is evident.<sup>13</sup> As was the case with the variables associated with UK universities, most of these variables are highly and significantly correlated.

## **Methods**

We used a statistical technique to estimate a university's 'quality' based on a set of observable features. This technique is a Bayesian latent variable model. Bayesian methods have been used for ranking in other education-related studies. Lockwood *et al.* (2002: 255–270) discuss uncertainty in ranking teachers and schools from individual student test scores. Laird and Louis (1989: 29–46) use an empirical Bayes approach to rank schools by their ability to affect student achievement. Goldstein and Spiegelhalter

<sup>11</sup> Data retrieved November 30 2004, from <<http://www.usnews.com/usnews/home.htm>>. Data used with permission of the *U.S. News & World Report*.

<sup>12</sup> Details at <[http://www.usnews.com/usnews/edu/college/rankings/about/weight\\_brief.php#giving](http://www.usnews.com/usnews/edu/college/rankings/about/weight_brief.php#giving)>.

<sup>13</sup> The coefficients of variation for the percentage of students in the top 10 percent of their class and for the alumni giving rate are higher than 0.60.

TABLE 3. Indicators and weights used in the *U.S. News & World Report* university rankings

Aggregate category	Category weight (percent)	Sub-factor	Description	Sub-factor weight (percent)
Peer assessment survey	25		Survey of presidents, provosts, and deans of admissions in which schools are rated from 1 (lowest quality) to 5 (highest quality)	100
Graduation and retention rate	20	Actual six-year graduation rate	Proportion of students entering between 1994 through 1997 who graduated within six years	80
		Average freshman retention rate	Proportion of students entering in 1999–2000 who returned the following fall	20
		Proportion of small classes	Proportion of classes with fewer than 20 students	30
		Proportion of large classes	Proportion of class with 50 or more students	10
		Faculty compensation*	Average faculty pay plus benefits	35
Faculty resources	20	Percent faculty with top terminal degree*	Percent of faculty with a PhD or highest degree possible in their field	15
		Student: faculty ratio	Ratio of student FTE to faculty FTE	5
		Proportion full-time faculty	Proportion of full-time 2003–2004 FTE faculty	5
		SAT/ACT scores	25 <sup>th</sup> and 75 <sup>th</sup> percentile of the university distribution of Scholastic Aptitude Test (college entrance examination) scores of entering students	50
Student selectivity	15	Proportion of top 10 percent students	Proportion of first-year students who graduated in the top 10 percent of their secondary school class	40
		Acceptance rate	Ratio of students admitted to total applicants	10
Financial resources**	10		Average educational expenditures per student	100
Graduation rate performance	5		Difference between actual six-year graduation rates and predicted rate based on the characteristics of the institution and entering students.	100
Alumni giving rate	5			100
Total	100			

\* Measure not available.

\*\* ‘Expenditures per student’ not available on the website, but the university rank on this measure was.

Source: *U.S. News & World Report* 2005.

TABLE 4. Descriptive statistics for features of 129 research universities in the US\*

University features	Mean	Standard deviation	Minimum	Maximum
Peer assessment	3.5	0.6	2.5	4.9
Acceptance rate	0.57	0.22	0.10	0.93
Proportion of top 10 percent students	0.53	0.26	0.16	0.99
SAT score at 25 <sup>th</sup> percentile	1136	116	840	1460
SAT score at 75 <sup>th</sup> percentile	1335	104	1170	1590
Proportion full-time faculty	0.90	0.07	0.69	1.00
Faculty-student ratio	0.09	0.04	0.05	0.33
Proportion classes smaller than 20	0.48	0.14	0.19	0.75
Proportion classes smaller than 50	0.12	0.07	0.01	0.31
Average freshman retention rate	0.89	0.06	0.77	0.98
Actual six-year graduation rate	0.74	0.12	0.48	0.98
Graduation rate performance	0.73	0.12	0.50	0.94
Financial resources rank	75.40	50.3	1.0	200.0
Alumni giving rate	0.21	0.10	0.05	0.61

\* Statistics computed using data from the *USNWR, America's Best Colleges 2005*.

Source: The authors.

(1996: 385–443) advocate Bayesian analysis for ranking individuals and institutions and use Markov Chain Monte Carlo integration to compute uncertainty in the ranks. These studies all involve ranking units on a single measure (e.g., teacher effects). In contrast, we rank universities simultaneously incorporating several measures. Our latent variable model essentially maps these measures to a single factor from which we can produce ranks.

The latent variable technique differs from standard weight-and-sum rankings because it (a) determines the relative importance of different university features using information embedded in the data rather than subjective opinion; and (b) simultaneously determines the degree of uncertainty that surrounds the ranks. If an institution is clearly superior to all others on virtually all measured features, our method will find very little variance in its rank and will rank it as number one with a high degree of certainty. Alternatively, if an institution has features that weakly distinguish it from other institutions (e.g., its graduation rate is average), then there will be many similar institutions. Since we allow for variability in how influential particular features are on the measure of overall school quality, slight changes in the graduation rate influence can permute the ordering of the schools near the middle. Analogously, in weight-and-sum approaches many universities end up being clustered in the middle, all with very similar overall scores. Slightly perturbing the weight for important variables can swing the rank of a university in this group by twenty or more universities. In such a situation, our method would supply a ranking but the ranking would be accompanied by a great deal of uncertainty. Thus, it might be ranked, say 50<sup>th</sup>, but it would not be possible for us to assert that it was statistically superior to the 51<sup>st</sup> ranked university. In the next few paragraphs we provide a concise technical explanation of the methodology.

In the model, we suppose that each university has an unobserved, latent ‘quality’ feature denoted  $[z_i]$ . While we do not observe  $[z_i]$  of each university directly, its observable features express the university quality. More precisely, we use a linear

model to associate the observed features,  $[x_1, \dots, x_J]$ , with  $[z_i]$ . For the UK universities, for which we have nine measurements on each school, we will have a system of nine regression models of the form:

$$\begin{aligned} \text{teaching}_i &= \beta_0^{(1)} + \beta_1^{(1)} z_i + \varepsilon_i^{(1)}, & \varepsilon_i^{(1)} &\sim N(0, \sigma_1^2) \\ \text{research}_i &= \beta_0^{(2)} + \beta_1^{(2)} z_i + \varepsilon_i^{(2)}, & \varepsilon_i^{(2)} &\sim N(0, \sigma_2^2) \\ &\vdots \\ \text{facilities}_i &= \beta_0^{(9)} + \beta_1^{(9)} z_i + \varepsilon_i^{(9)}, & \varepsilon_i^{(9)} &\sim N(0, \sigma_9^2) \end{aligned}$$

Note that there are nine different intercept terms,  $[\beta_0\text{'s}]$ , nine different coefficients on  $[z_i]$ , the  $[\beta_1\text{'s}]$ , and each regression model has its own residual variance. There is no need to shift and rescale the observed features, since this is built into the model. This model is similar to a factor analysis with a single factor. We take a Bayesian approach, obtaining a posterior distribution for the latent  $[z_i\text{'s}]$  and compute the distribution of the university ranks.

Briefly, the Bayesian approach aims to compute the joint distribution of all the unknown quantities – the regression parameters and the  $[z_i\text{'s}]$  – conditional on the observed quantities – the university features. This posterior distribution captures all of the uncertainty about the unknown quantities. This framework is particularly helpful for ranking institutions. Not only will we be able to describe the uncertainty in estimating the  $[z_i\text{'s}]$ , but also the uncertainty in their ranks relative to one another.

To complete the model specification, we put non-informative priors on the regression parameters and a Gaussian prior on the  $[z_i\text{'s}]$ , with mean 0 and variance 1. Note that the location and scale of the  $[z_i\text{'s}]$  is not identifiable – nor of interest – since any shift in the  $[z_i\text{'s}]$  can be offset by a shift in  $[\beta_0]$  and any rescaling of the  $[z_i\text{'s}]$  can be offset by scaling  $[\beta_1]$ .

While the complete joint posterior distribution of the  $\{z_i\text{'s}\}$ ,  $[\beta\text{'s}]$ , and  $[\sigma^2\text{'s}]$  is complex, the posterior mean of  $[z_i]$ , given all the other parameters and observed data, is a simple linear combination of university  $[i\text{'s}]$  observed features. The posterior mean of  $[z_i]$ , given all other parameters, is:

$$E(z_i|\cdot) = a + b \sum_{j=1}^J \frac{\beta_1^{(j)}}{\sigma_j} \frac{x_j - \mu_j}{\sigma_j} \quad (1)$$

where  $[a]$  and  $[b]$ , respectively, shift and scale, so that, collectively, the  $[z_i\text{'s}]$  have mean  $[0]$  and variance  $[1]$ . Note that the absolute size of  $[z_i]$  is unimportant – only its position relative to other universities matters. The estimated mean of  $[z_i]$  is thus computed using a weight-and-sum method with each standardized feature  $[j]$  weighted with  $[w_j = \beta_1^{(j)} / \sigma_j]$ . Thus, as opposed to other weight-and-sum approaches that assign subjectively determined weights to the observable features, this method simultaneously estimates ranks and weights. We report estimates of the  $[w_j\text{'s}]$ , which inform us as to the relative contribution of the particular feature in determining the rank.

The regression parameters and the  $[z_i\text{'s}]$  are estimated jointly. The estimation algorithms iterate between computing the conditional distribution of the  $[z_i\text{'s}]$  and computing the conditional distribution of the regression parameters. If the  $[z_i\text{'s}]$  are



known then the remaining parameters are easily estimable with standard linear regression. Both the EM algorithm and the Gibbs sampler operate in this fashion. We used a Gibbs sampler as implemented in OpenBUGS<sup>14</sup> to draw samples from the distribution of the  $[z_i, s]$  given the observed university features. With each draw from this distribution we obtain a set of  $[z_i, s]$ , and by ordering these  $[z_i, s]$ , we can derive ranks. This process is repeated many times. From 10,000 draws, we can get an accurate estimate of the distribution of the rank of the  $[z_i]$  for each university. For each university, we report the median rank as well as the ranks at the 2.5 and 97.5 percentiles, computed over the 10,000 draws. The 2.5 and 97.5 percentiles bound a 95 percent posterior probability interval, similar to a confidence interval.

## Findings

This section describes our analyses ranking universities in the UK and US. The analysis is structured in three steps for each country:

1. The latent variable measure is used to rank institutions and reveal the degree of uncertainty or variance associated with each ranking.
2. The results show the relative importance of different quality input measures for the overall latent variable ranking.
3. The new latent variable ranking is compared with existing rankings based on the same data measures.

## Results for Universities in the UK

Figure 1 displays the new rankings by university for the 99 universities in the UK. The horizontal axis represents the rank obtained using the latent variable method. The universities are positioned on the vertical axis according to their rank, with the top-ranked school at the top of the chart. Thus, each point corresponding to the name of a university represents the median rank in that university posterior distribution of ranks, and the lines extending on either side of that point represent the interval within which the rank would be expected to fall with 95 percent probability.

The solid vertical line drawn at rank 10 helps illustrate the uncertainty associated with these rankings. Only four universities have ranks that are statistically superior – based on the specific measures used to assess quality in this analysis – to 10. This can be seen on the plot by noting that only four schools to the left of the vertical line have probability intervals that do not intersect it. If we wish to compare particular schools, say Sussex, Swansea, and Northumbria, we can check to see whether their intervals overlap. From the dotted vertical lines that enclose the interval related to Swansea, we see that Sussex is statistically higher in rank to both Swansea and Northumbria, but the latter two universities are not statistically distinguishable. These examples illustrate the point that it is generally not possible to assert that an institution in an ordinal ranking system is of higher quality than all those ranked below.

Also of interest is the size of the probability intervals. At the very top of the rankings, we find that the intervals are noticeably smaller than toward the middle or

---

<sup>14</sup> See <mathstat.helsinki.fi/openbugs/>.

bottom of the rankings. This tells us that the rankings of those schools are more certain than those in the middle and bottom. Uncertainty expands as we proceed down the rankings and then begins to narrow again slightly towards the bottom.

Table 5 shows the relative influence of the various measures used as inputs in determining the ranking. The table displays the estimated weights and t-statistics associated with the coefficients in the regression equations linking each input measure to the latent variable. We see that the entry standards measure exerts far more weight than other measures.

Figure 2 illustrates the relationship between the new rankings and the original rankings reported in *The Times Good University Guide*. In this figure, the horizontal axis again represents the rankings, but this time, the universities are positioned on the vertical axis according to their ranking in *The Times*. The new rankings are again identified by the points and associated intervals, while *The Times* rankings are identified by the short vertical lines (some of these short lines are longer than others to indicate ties). Had the new rankings been identical to those of *The Times*, all points would have fallen on the lines.

It is immediately evident that the two ranking systems are quite different. Nearly all points and several entire probability intervals in the new ranking system remain off the lines. The difference between the two sets of rankings is due to the different weighting schemes used.

## Results for Research Universities in the US

The next set of figures replicates the above analysis, this time using data on the 129 ‘top’ US national research universities ranked by the *U.S. News*. Figure 3 plots the ranks and probability intervals produced by the latent variable analysis, and reveals a pattern of uncertainty similar to that seen for the UK universities. Due to the size of the plot, we have left off the names. These can be found in the Appendix. Again we see that the intervals for universities at the top of the rankings are tighter than those for universities in the middle with some narrowing towards the bottom.

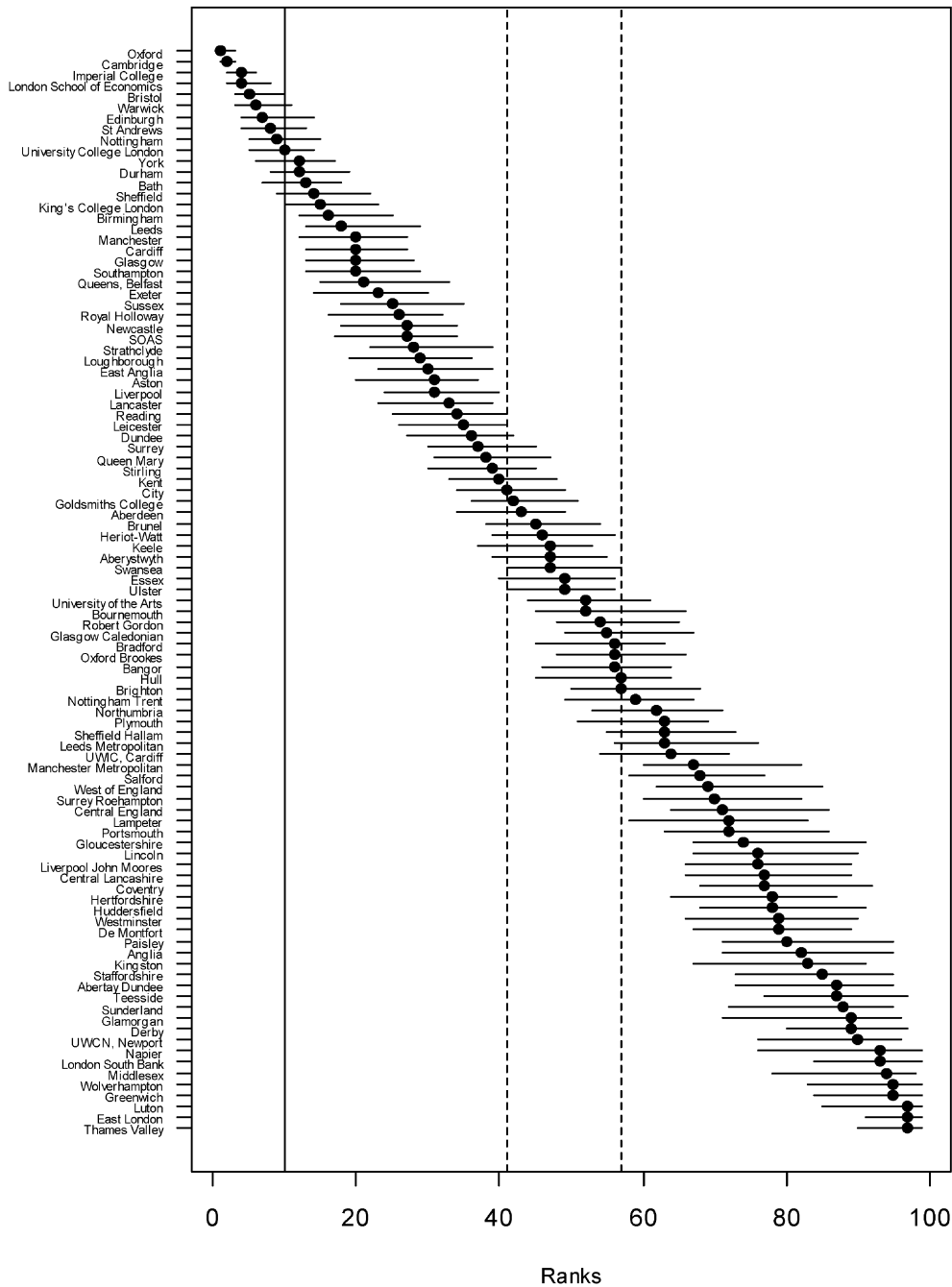
Table 6 indicates the relative importance of the various input measures in determining the overall rankings. The SAT scores are the most powerful drivers of the rankings. Together, student selectivity measures (SAT scores, rejection rate, and the proportion of top 10 percent students) account for a large percentage of the weight.

Figure 4 illustrates the divergence between the new rankings and the *U.S. News* for the top 50 institutions in the *U.S. News* ranking. Again, the new rankings diverge noticeably from the old, and several intervals do not intersect the vertical lines. Thus, we find that the weighting scheme determined by the latent variable method places many universities in different positions from those they occupied in the *U.S. News* rankings.

## Discussion

Using the latent variable method, we have been able to compress the information contained in a set of measures into a single number that captures the relationship of these measures to an underlying ‘quality’ construct. This number is the best

FIGURE 1. Distribution of ranks of UK universities



Source: The authors.

TABLE 5. Relative importance of U.K. university features in determining the ranking\*

Variable	Weight ( $w_j$ )	95 percent interval	$\beta_1/\text{se}(\beta_1)$
Entry standards	0.778	(0.098,3.818)	56.41
Percent high degrees	0.053	(0.043,0.064)	19.24
Research	0.051	(0.042,0.063)	18.54
Completion rate	0.033	(0.025,0.040)	12.05
Teaching	0.023	(0.017,0.030)	8.42
Student-faculty ratio	0.022	(0.016,0.028)	8.11
Graduate destinations	0.018	(0.012,0.024)	6.69
Facilities	0.012	(0.006,0.017)	4.38
Library	0.010	(0.003,0.018)	3.39

\* Latent variable analysis using data from *The Times Good Universities Guide 2005*.

Source: The authors.

expression of the interrelationship of all the measures in the set. The model yields a number for each institution and a set of weights such that if we lost the actual values of the observable measures for a particular institution, we could optimally reconstruct them.

In a weight-and-sum approach, the inclusion of several highly correlated measures of the same underlying construct may inflate the importance of that construct and distort the overall ranking if the total weight assigned to these redundant measures is too large. The latent variable method deals with these correlations differently – in essence, assigning the weights to the various input measures in its simultaneous estimation of their coefficients. Thus, it is less subject to the accusation of ‘stacking the deck,’ so to speak, with the addition of redundant measures.

The method shows that student selectivity is a major component of ‘quality,’ as expressed in the set of measures used for ranking universities in the UK and US. The weight given to selectivity is noticeably higher than the weights allowed by *The Times* and *U.S. News*. Both rankings systems down weight the importance of a high-achieving student body and its overall relationship to the other quality measures. Both the US and the UK have selective university systems, however. It is possible that in a country with a less selective university system (e.g., Germany), that this type of measure would not have a high weight in the latent variable analysis.

It might be argued that certain measures *should* be down-weighted. If we have subjective information on the value of particular measures in relation to true quality, the Bayesian framework can readily accommodate such information by encoding it in its choice of prior distributions on the coefficients that form an essential part of the weights.<sup>15</sup> Thus, were there reason to down weight the influence of, say, selectivity, the proposed method could accomplish this. As a normative question, however, it is important to ask on what theoretical basis these prior suppositions should be made.

## Conclusions

The above application of the latent variable approach to the task of differentiating among institutions of higher education provides insights that lead to a greater

<sup>15</sup> Recall that our analysis used non-informative priors.

FIGURE 2. A comparison of the new ranks with *The Times* ranks for UK universities

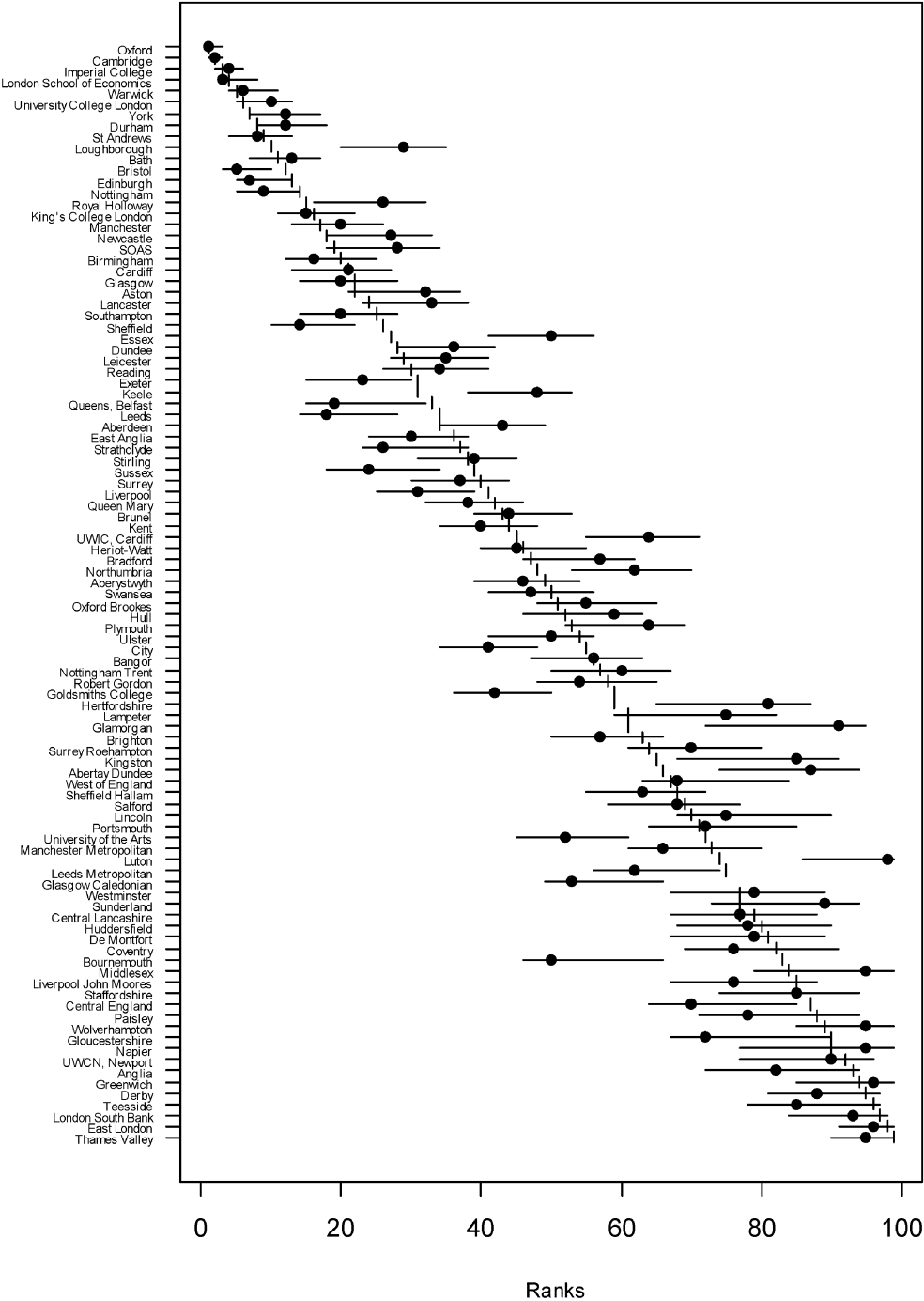
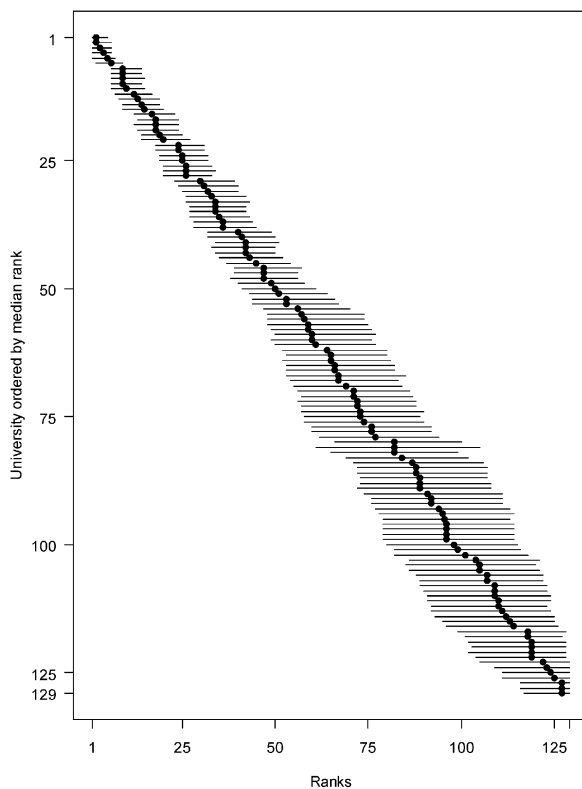


FIGURE 3. Distribution of ranks of US research universities



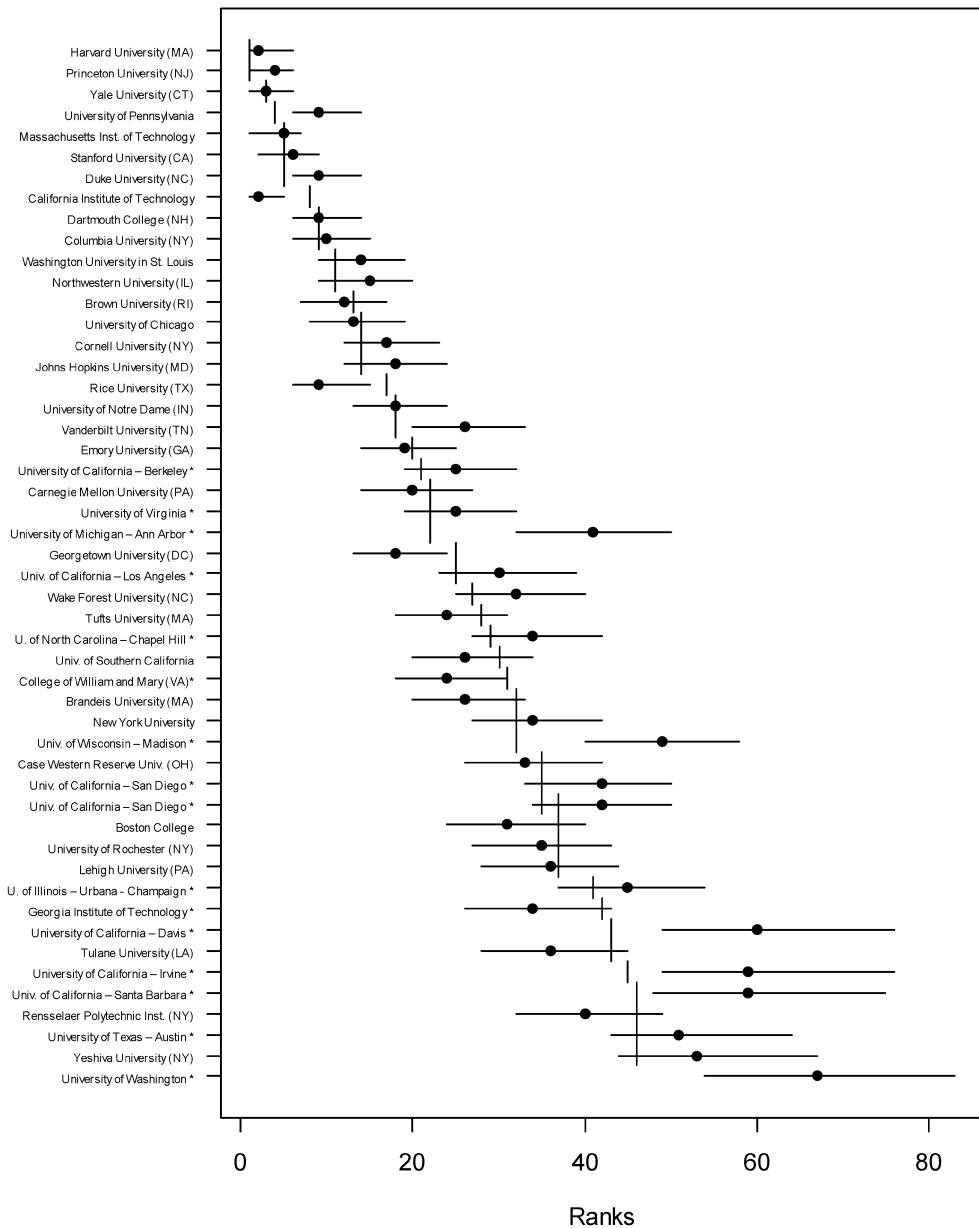
Source: The authors.

TABLE 6. Relative importance of US university features in determining the ranking

Variable	Weight [ $w_j$ ]	95 percent interval	$[\beta_1]/\text{se}[\beta_1]$
SAT at 75 <sup>th</sup> percentile	0.264	(0.204,0.351)	51.44
SAT at 25 <sup>th</sup> percentile	0.165	(0.138,0.195)	38.74
Freshman retention	0.099	(0.082,0.117)	19.89
Rejection rate	0.097	(0.081,0.114)	19.75
Actual six-year graduation rate	0.095	(0.078,0.112)	19.40
Proportion top 10 percent students	0.077	(0.063,0.091)	16.37
Peer assessment	0.073	(0.060,0.088)	15.16
Alumni giving rate	0.056	(0.044,0.067)	11.61
Faculty-to-student ratio	0.051	(0.040,0.063)	10.61
Proportion of small classes	0.044	(0.033,0.056)	8.89
Financial	-0.013	(-0.022,-0.005)	-3.60
Graduation rate performance	-0.007	(-0.016,0.003)	-1.40
Proportion of large classes	-0.006	(-0.015,0.004)	-1.19
Proportion of full-time faculty	0.005	(-0.004,0.014)	1.08

Source: The authors.

understanding of the heterogeneity that exists among institutions and perhaps to a greater degree of caution in asserting that certain institutions are of higher quality than others.

FIGURE 4. A Comparison of the new ranks with the *U.S. News* ranks for the top 50 US research universities

Source: The authors.

The latent variable approach is useful for several reasons. It highlights the degree of uncertainty that exists in the ordinal ranking of universities and permits testing for statistically significant differences among institutions. It highlights the relative importance of particular input measures in the determination of the overall rank. In its divergence from weight-and-sum rankings, it reveals the degree to which different weighting schemes affect the rankings.

The methodology is also subject to limitations. Like standard weight-and-sum ranking systems, it is dependent upon a set of observable quality indicators that may be flawed or incomplete. While this approach may not move us closer a true assessment of the relative quality of education offered at various institutions of higher education, it represents an improvement over traditional systems in its ability to reveal the uncertainty behind rankings and identify where meaningful distinctions can be drawn between one institution and another on the basis of a given set of measures. Thus, it remains a useful tool for those who publish rankings to gain greater insights into the nature of the distinctions they promulgate.

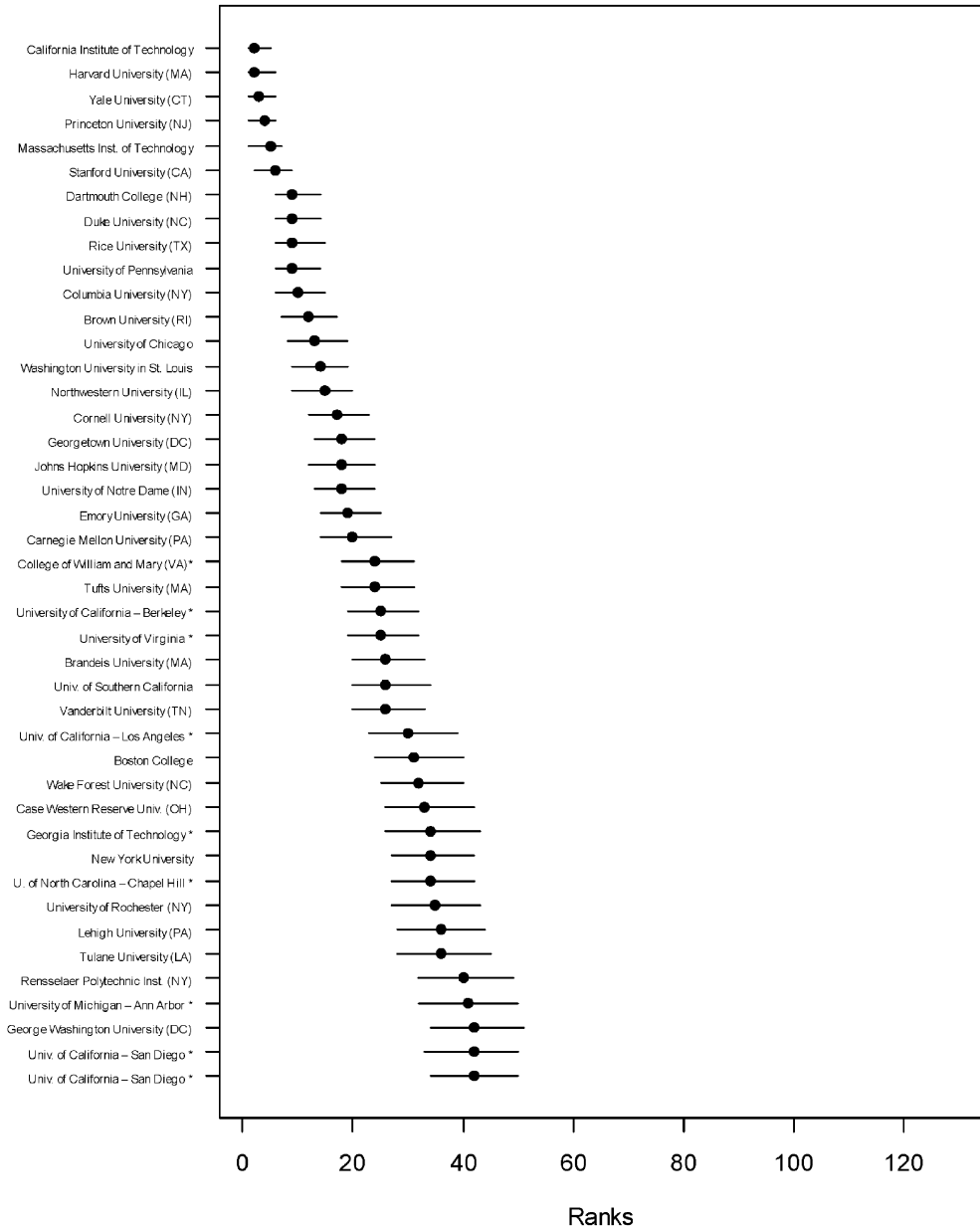
## References

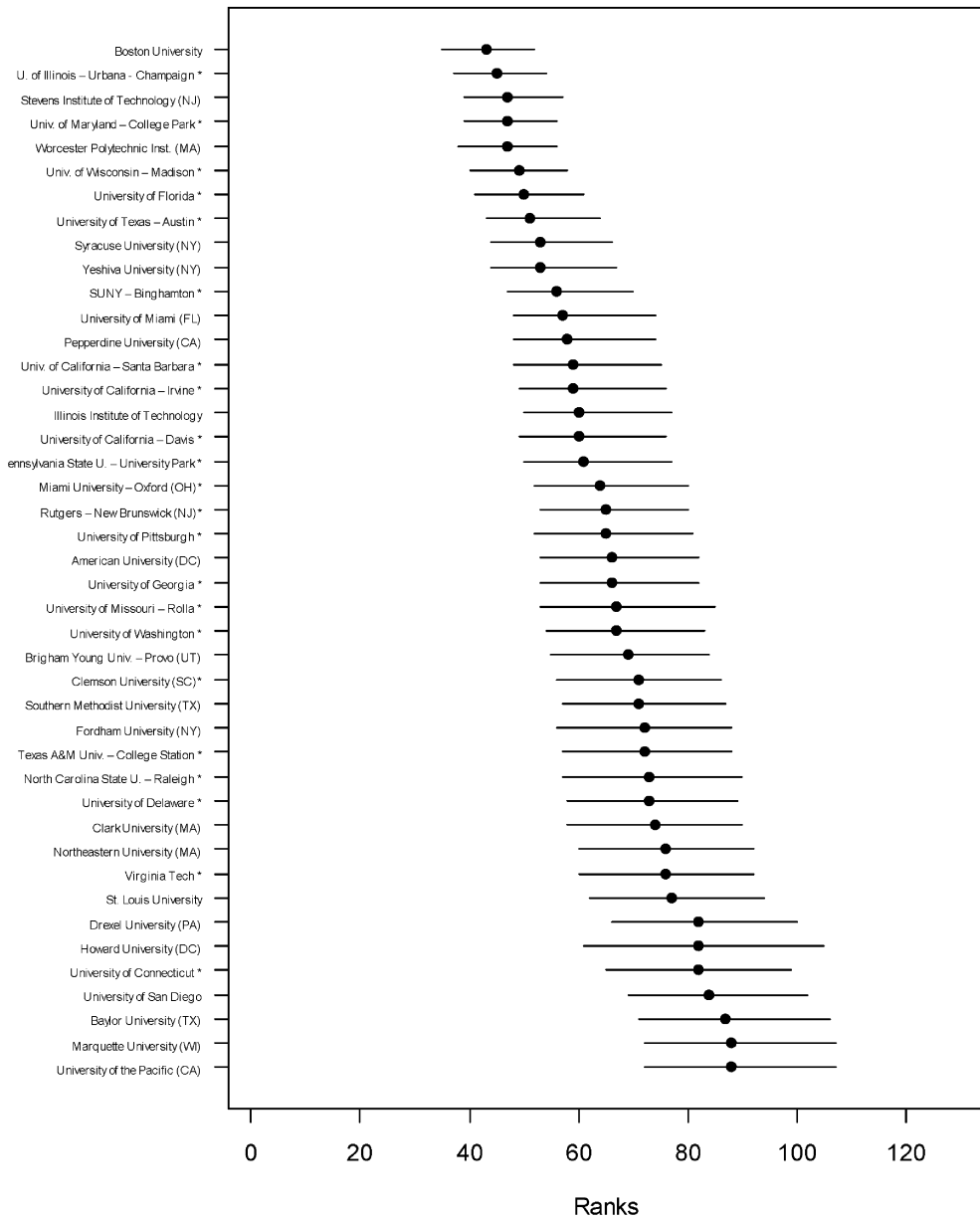
- AVERY, C., GLICKMAN, M., HOXBY, C. and METRICK, A. "A Revealed Preference Ranking of US Colleges and Universities". National Bureau of Economic Research Working Paper 10803, Cambridge, MA, 2004.
- BLACK, D., DANIEL, K. and SMITH, J. "College Quality and the Wages of Young Men". Unpublished manuscript, University of Maryland, 2003.
- BREWER, D., EIDE, E.R. and EHRENBERG, R.G. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings", *The Journal of Human Resources* 34 1 (1999): 104–123.
- CLARKE, M. "Some Guidelines for Academic Quality Rankings", *Higher Education in Europe* 27 4 (2002a): 443–459.
- CLARKE, M. "Quantifying Quality: What Can the *U.S. News & World Report* Rankings Tell Us about the Quality of Higher Education?", *Education Policy Analysis Archives*, 10 16 (2002b)
- GOLDSTEIN, H. and SPIEGELHALTER, D. "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance (with discussion)", *Journal of the Royal Statistical Society Series A* 159 (1996).
- HOXBY, C.M. "The Return to Attending a More Selective College: 1960 to the Present", *Forum Futures. Exploring the Future of Higher Education*, 2000 Papers, Forum Strategy Series, Vol. 3, Jossey-Bass, 2001.
- LAIRD, N. and LOUIS, T. "Bayes and Empirical Bayes Ranking Methods", *Journal of Educational Statistics* 14 (1989).
- LOCKWOOD, J., LOUIS, T. and MCCAFFREY, D. "Uncertainty in Rank Estimation: Implications for Value-Added Modeling Accountability Systems", *Journal of Educational and Behavioral Statistics* 27 3 (2002).
- MONKS, J. and EHRENBERG, R.G. "The Impact of *U.S. News & World Report* College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions". National Bureau of Economic Research Working Paper 7227, Cambridge, MA, 1999.
- SHATTOCK, M. *Managing Successful Universities*. Berkshire: Open University Press, 2003.
- The Times Good University Guide 2005*. London: Times Books in Association with PricewaterhouseCoopers, 2004.
- U.S. News & World Report. America's Best Colleges 2004*.

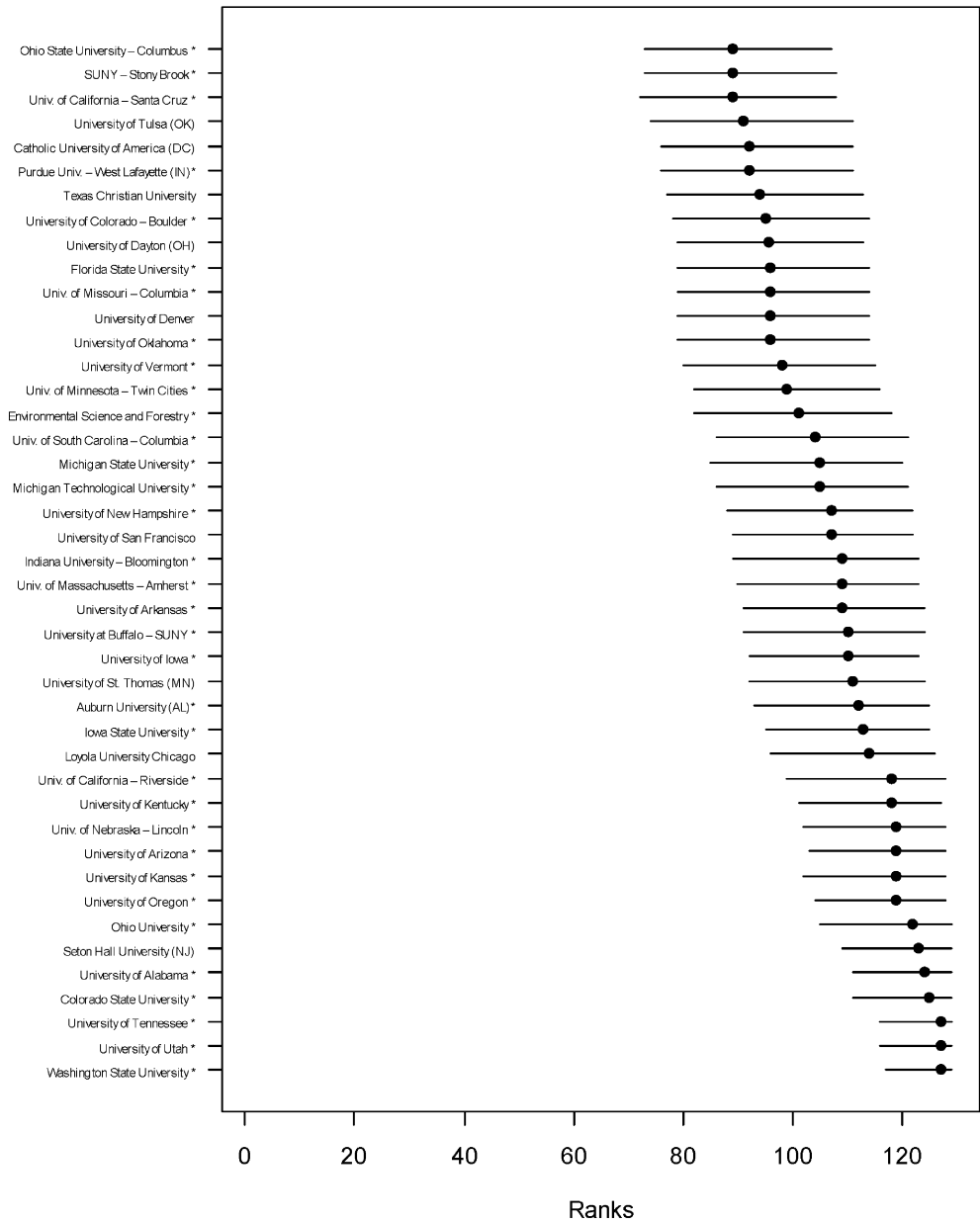


## Appendix

### The Complete Listing of the Ranks of 129 US Research Universities







Source: The authors.

Copyright of Higher Education in Europe is the property of Routledge, Ltd.. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.