Journal of Applied Research in Higher Edu

Statistical Assessment of Peer Opinions in Higher Education Rankings: The Case of U.S. Engineering Graduate Programs

Journal:	*Journal of Applied Research in Higher Education*
Manuscript ID	JARHE-09-2018-0196.R1
Manuscript Type:	Research Paper
Keywords:	University rankings, peer assessments, random parameters, linear regression

Statistical Assessment of Peer Opinions in Higher Education Rankings: The Case of U.S. Engineering Graduate Programs

Abstract

Unlike many other quantitative characteristics used to determine higher education rankings, opinion-based peer assessment scores, and the factors that may influence them, are not well understood. Using peer scores of U.S. colleges of engineering as reported annually in US News and World Report (USNews) rankings, this paper seeks to provide some insight into peer assessments by statistically identifying factors that influence them. With highly detailed data, a random parameters linear regression is estimated to statistically identify factors determining a college of engineering's average USNews peer assessment score. The findings show that a wide variety of college- and university-specific attributes influence average peer impressions of a university's college of engineering including the size of the faculty, the quality of admitted students, and the quality of the faculty measured by their citation data and other factors. Our estimation results show that average peer assessment scores can be readily and accurately predicted with observable data on the college of engineering and the university as a whole. In addition, the individual parameter estimates from our statistical modeling provide insights as to how specific college and university attributes can help guide policies to improve an individual college's average peer assessment scores and its overall ranking.

Keywords: University rankings; peer assessments; random parameters; linear regression

1. Introduction

The national and international demand for graduate degrees in higher education has continued to grow as many students seek to continue their education to advance their careers or to satisfy professional development. Many universities have developed policies to handle this influx of students including expanding their masters-degree offerings and placing an increased emphasis on their doctoral programs. In turn, competition among universities for the best graduate students has often focused on the university's placement in various university rankings (Dill and Soo, 2005; Williams and Van Dyke, 2008). With the information provided by these rankings as well as other sources, students seek admission at universities or academic programs that best suit their needs and abilities, often selecting the most highly ranked university to which they can be accepted. From the university perspective, university administrators increasingly use rankings as a marketing tool not only to attract the best students, but also to instill alumni pride (and subsequently increase donations) and to recruit better faculty members (Grewal et al., 2008). Moreover, university funding from governmental and private entities can be related to ranking results and this has affected the operating policies of universities (Hazelkorn, 2008; Lo, 2014). Thus, universities tend to pay very close attention to their position in the rankings and, due to a wide variety of internal and external pressures, they often alter policies to seek to improve their ranking whenever possible. This can have a long lasting and profound effect on the institutional direction of the university (Li, 2016).

There are a number of widely-recognized university rankings, but the annual US News and World Report (USNews) rankings is arguably the most popular among universities in the United States. In fact, over the years, a number of studies have investigated the effect of USNews rankings on university reputations. For example, Machung (1998) provided evidence that suggested that

USNews' rankings have had a significant impact on university admissions, donations, and so on. This study also showed that roughly two-thirds of parents found the information provided by USNews to be very useful. In terms of additional evidence of the influence of these rankings, Hansmann (1999) indicated that Yale University's admissions significantly increased when Yale reached the top position in the USNews ranking of law schools the year before. Bowman and Bastedo (2009) also found that moving up to the top USNews ranking list provided a substantial increase in the following year's admissions indicators. In other work, Luca and Smith (2013) found that a one-rank improvement in the USNews ranking led to one percent increase in the number of college applications. More recently, Yeung et al. (2018) found evidence that USNews university rankings were significantly correlated with university presidential salaries at public universities. And many other studies have explored various dimensions and nuances of USNews rankings and their effect on higher education (Dichev, 2001; Merideth, 2004; Pike, 2004; Sweitzer and Volkwein, 2009; Bougnol and Dula, 2015). All of these studies suggest that USNews rankings can potentially play an important role in guiding university policies.

Like most other university rankings, USNews ranking methodologies are not particularly transparent, meaning that the ranking formulas have not been published and universities are not able to directly predict their positions in the ranking. While a few studies have tried to develop predictive techniques (for example Gnolek et al. (2014) provided a simple method based on the weights provided by the USNews to predict the USNews rankings), success in unraveling the determinants of rankings has been limited.

However, aside from the various quantitative measures used by USNews and other national and international rankings, almost all rely to some extent on peer evaluations of specific programs. In the USNews rankings of graduate engineering programs, peer assessment scores account for a quarter of the total ranking score. For USNews peer assessment scores, individuals from universities (typically engineering deans) are asked to rate the quality of specific programs on a scale from 1 (marginal) to 5 (distinguished). Unlike the overall rankings, which are determined as some function of observable data on productivity, faculty size, entrance requirements, and so on, peer assessment scores are based on individual impressions of program quality. These impressions are likely to be formed over many years from individuals' personal experiences with faculty and graduates of the schools they are evaluating, the school's overall reputation, and possibly the experiences of their colleagues and others that may influence their opinions (Collins and Park, 2016). Their impressions may also be influenced by their exposure to past rankings, and by the factual information used to support those past rankings. Thus, there is little doubt that, at some level, past rankings affect peer assessment scores which in turn affect current rankings.

Clearly one could argue that understanding the determinants of the peer assessment scores could be useful in guiding university programs to improve their overall rankings. As mentioned above, in USNews college of engineering rankings, peer assessment scores are one of the key elements for ranking colleges of engineering (in fact they carry more weight than any other individual measure). And, perhaps more importantly, they are the only element used to rank individual departments within the college of engineering.

The intent of the current paper is to provide some insight into the determinants of peer rankings in higher education by developing a statistical model of the average peer assessment scores of U.S. colleges of engineering, as reported in the 2018 Engineering Graduate program evaluations provided in USNews. The paper begins with a description of the methodological approach, followed by a discussion of the available data. Model estimation results are then presented and discussed in detail. The paper ends with a summary of the findings and the conclusions of the study.

2. Methodological Approach

We seek to develop a statistical model with the average peer program assessment score of U.S. graduate engineering programs as the dependent variable. This average score, reported by USNews in their annual program rankings, is based on the averages of an annual survey distributed to U.S. universities with engineering programs where university engineering representatives (Deans and Department Chairs) are asked to rate all university engineering graduate programs on a scale of 1 (marginal) to 5 (distinguished). USNews provides the resulting average peer assessment scores in tenths (such as 2.6, 2.7, 2.8, etc.). Thus the dependent variable is continuous and can be modeled with a standard regression approach as

$$y_n = \boldsymbol{\beta}\mathbf{X}_n + \varepsilon_n, \tag{1}$$

where, y_n is the dependent variable (average graduate program peer assessment score for university n's college of engineering), $\boldsymbol{\beta}$ is a vector of estimable parameters, $\mathbf{X}_n$ is a vector of explanatory variables for university n, and ε_n is a normally and independently distributed error term with zero mean and constant variance σ^2.

From a statistical modeling perspective, it is important to note here that the individual peer assessments (to which we do not have access) are bounded between 1 and 5, and could be appropriately modeled as an ordered discrete probability model (with discrete outcomes 1, 2, 3, 4 and 5), such as an ordered probit or logit model (see Washington et al., 2011). However, because we only have access to the average of all individual peer assessment scores, the average is not a discrete but instead a continuous variable (as noted above), but the average is still bounded between 1 and 5. This suggests that a regression model with both left and right censoring (at 1 and 5, respectively) might be considered. However, our estimation of a lower and upper censored tobit regression that accounted for this restriction (Tobin, 1958; Anastasopoulos et al., 2008; Washington et al., 2011) did not produce statistically different results relative to an uncensored regression represented in Equation 1. This is likely due to the fact that our average peer assessment score data has few observations near the lower and upper censored values (1.9 is the minimum average peer assessment score, and 4.9 is the maximum).

Moving forward with the uncensored regression in Equation 1, to account for heterogeneity (unobserved factors that may vary across universities), we allow for the possibility that each university may have its own parameter for one or more of the explanatory variables. This is referred to as a random parameters approach, and these models can be readily estimated with simulated maximum likelihood estimation methods (Washington et al., 2011; Mannering et al., 2016). For allowing the possibility of random parameters in Equation 1, we write estimable parameters (individual elements of the $\boldsymbol{\beta}$ vector) as,

$$\beta_n = \beta + \varphi_n \tag{2}$$

where β is the mean of the parameter estimate and φ_n is a randomly distributed term (for example a normally distributed term with mean zero and variance σ^2). We also allow for the possibility for the mean and variance of the parameter to be a function of explanatory variables so (Mannering, 2018),

$$\beta_n = \beta + \Theta\mathbf{Z}_n + \sigma_{kn} EXP\left(\boldsymbol{\omega}_{kn}\mathbf{W}_{kn}\right) + \varphi_n \tag{3}$$

where β is the mean parameter estimate, $\mathbf{Z}_n$ is a vector of explanatory variables that influence the mean of β_n, Θ is a vector of estimable parameters, $\mathbf{W}_{kn}$ is a vector of explanatory variables that captures heterogeneity in the standard deviation σ_{kn} , $\boldsymbol{\omega}_{kn}$ is the corresponding parameter vector, and φ_n is a randomly distributed term that captures unobserved heterogeneity across universities.

Estimation of the random parameters regression is undertaken by simulated maximum likelihood approaches because the required integration for parameter estimation is not closed form. Previous studies have shown that Halton draws provide a more efficient distribution of simulation draws than purely random draws (McFadden and Ruud, 1994; Bhat, 2003). We use 1,000 Halton draws in our simulated likelihood functions, a number that has been shown to be more than sufficient to provide accurate parameter estimates (Halton, 1960; Bhat, 2003; Anastasopoulos and Mannering, 2009). In this study, the normal distribution has been used in estimation of the explanatory variables because it provided the best overall statistical fit (other distributions such as the log-normal, uniform and exponential were not found to produce statistically better results than the normal distribution).

3. Data

The average peer assessment scores are gathered from USNews' 2018 rating of U.S. engineering graduate programs (published in 2017). Along with these peer assessment scores, we extract other USNews-related data provided with their 2018 graduate program rankings including: engineering graduate student enrollment; total number of engineering faculty; number of engineering faculty in the National Academy of Engineering; average math Graduate Record Exam score of accepted engineering students; percent of engineering applications accepted; number of doctoral students graduated in the past year; and research expenditures per faculty. In addition, we supplement this USNews data with information on university membership in the American Association of Universities (AAU), and whether the university is public or private. We also gather information on many other factors for the university as a whole, such as national merit scholars admitted, incoming students' average SAT scores, number post-doctoral appointees, and so on (these data are available in Lombardi et al., 2017). Finally, we gather citation information of engineering faculty and university faculty as a whole from Google Scholar including total citations, number of documents, h-index, and number of papers cited 10 times or more. We attempted to gather all of these data for the 147 universities included in the 2018 USNews engineering graduate program rankings. However, we were unable to obtain all data eventually used in model estimation (see below) for 8 of these 147 universities, leaving us with a total of 139 university observations with complete data.

4. Model Estimation Results

Summary statistics for the variables found to be statistically significant in the model are presented in Table 1, and random parameter linear regression estimation results are presented in Table 2 for the 139 universities with complete data available. Table 2 shows that a total of 13 variables were found to produce statistically significant parameter estimates. Of these, 3 were found to vary across observations (random parameters as shown in Equation 2). Interestingly, we found no statistically significant evidence of heterogeneity in means and variances (see Equation 3). Depending on the structure of unobserved heterogeneity in the data, this is not an uncommon finding (Mannering, 2018). The overall statistical fit of the model is excellent with an R-squared value of 0.933 and an adjusted R-squared value of 0.926, which accounts for the number of estimated parameters in the model (Washington et al., 2011). These relatively high R-squared values likely reflect the nature of the ranking process used by USNews. That is, with the exception of peer assessments, much of the ranking is determined by quantitative measures. Over time, peers observe USNews rankings from previous years (which are driven by these quantitative measures) and tend to adjust their peer assessment to more closely align with these past rankings. Thus, quantitative measures, such as those included as explanatory variables in our model, become a very strong predictor of peer assessments, and this is reflected in our relatively high R-squared values.

To illustrate the overall statistical fit of the model, Figure 1 presents the observed versus predicted values, showing the strong overall statistical fit of the model. It should be noted that the predicted values are determined based on individual-level parameters computed for those parameters determined to be random (varying across observations). In the case of random parameters, the individual university parameters are computed using the simulated Bayesian approach described in Greene (2004).

Table 2 shows that all parameter estimates are statistically significant from zero at the 99% confidence level (except one which is at the 95% confidence level) using a two-tailed *t*-test. Please note that potential multicollinearity among variables in the model does not present an issue in model estimation. In the presence of multicollinearity, parameter estimates retain all desirable estimation properties with the only adverse consequence being high standard errors (Washington et al., 2011). Given the high t-statistics (reflecting low standard errors) of the variables included in our model multicollinearity is not playing a significant role in model estimation.

As indicated in Table 2, the variables found to be statistically significant in determining college of engineering average peer assessment scores are classified into broad categories of attributes relating to the college of engineering, and those related to the university as a whole. These are discussed in separate sections below.

4.1 College of Engineering Attributes

Turning to college of engineering attributes first, Table 2 shows that the average math Graduate Record Examination (GRE) scores of incoming graduate students positively increases the average peer evaluation. The parameter estimate shows that an 8 point increase in the math GRE scores of graduate students would result in roughly a 0.1 point increase in a college of engineering's average peer assessment score. It is important to note that even a 0.1 point increase can be important in separating out universities. For example, the 2018 data had 17 universities with an average peer assessment score of 2.2, and a 0.1 increase for any one of the schools would lift them above this grouping.

Following on in Table 2, we find that universities that admit more than 31% of their graduate applicants tend to have average peer assessment scores that are 0.10396 less than universities that admit 31% or less. During estimation this 31% threshold produced the most statistically significant results, and is likely an outgrowth of the admission policies of universities during this specific graduate-school admission cycle. This result clearly reflects the importance of admission standards in the determination of average peer assessment scores.

The number of doctoral students graduated was also found to be statistically significant in determining average peer assessment scores. The parameter estimate for this variable suggests that graduating 34 more doctoral students would result in a 0.1 increase in the average peer assessment score. This seems to reflect the national trend that relates quantity of doctoral students graduated with the program quality, and the number of graduated doctoral students is used by USNews as an input to their college of engineering rankings.

Faculty size, in terms of total number of faculty, was also found to positively influence average peer assessment scores, with larger faculty size more likely to result in a higher score.

However, our estimation results show that there is a distinction between public and private universities. For public universities, the model's parameter estimate suggests that nearly 60 faculty would need to be added for a 0.1 increase in the average peer assessment score. In contrast, for private universities, the model's parameter estimate suggests that about 35 faculty would need to be added for a 0.1 increase in the average peer assessment score. This difference likely reflects the different expectations between public and private college of engineering faculty sizes, but may also reflect differences in overall faculty quality, areas of expertise, publicity-related matters, and so on.

The model estimation results also show that a college of engineering faculty-size threshold of 200 significantly influences average peer assessment scores. This smaller-faculty indicator variable (1 if the total number of engineering faculty is less than 200, and 0 otherwise) shows that faculty sizes less than 200 tend to have an adverse effect on peer assessment scores, but that the effect of this indicator variable varies significantly across universities that are in this group (as indicated by the statistically significant standard deviation of the normally distributed parameter density function). In estimating individual parameters using the simulated Bayesian approach suggested by Greene (2004), we find this smaller-faculty indicator is negative for roughly 68% of the universities and positive for 32%. This indicator variable must be considered along the total number of faculty variables discussed above. The fact that having fewer than 200 faculty can increase average peer assessment scores for 32% of universities (relative to having more than 200 faculty) reflects the unobserved heterogeneity among existing college of engineering faculty. In other words, it is not just about size, but about quality as well. And, while we have some measures of quality relating citation data, a finding such as this (with both positive and negative effects) suggest that unobserved heterogeneity is playing an important role (Mannering et al., 2016).

Increasing the ratio of graduate student enrollment per engineering faculty member was found to have a negative effect on average peer assessment scores. This variable produced a statistically significant random parameter indicating significant variance in its effect across universities. In estimating individual parameters for this variable, we find this parameter to be negative for all universities, but the fact that the values vary across universities suggests that unobserved factors relating to student and faculty quality are affecting the influence that this ratio has on average peer assessment scores.

As is well known, colleges of engineering have focused on recruiting faculty who are members of the National Academy of Engineering (NAE), which is an important input into USNews college of engineering rankings. It is thus important to explore how NAE membership affects average peer assessment scores. The estimation results in Table 2 show that the effect of this variable varies across universities (as indicated by the statistically significant standard deviation of the parameter estimate, indicating a random parameter) with the percent of the college of engineering's faculty in the academy has a mean parameter estimate of 0.02182 (meaning that increasing NAE membership of the faculty by 4.5%, on average, will result in a 0.1 increase in average peer assessment score). However, the parameter has a large standard deviation of 0.0204. In estimating individual parameters, we find this variable to be positive for all universities (implying that adding an NAE faculty member always has a positive effect), but the effect ranges from a 0.1 increase in the average peer assessment score due to a 1.5% increase in NAE membership, to virtually no effect on the average peer assessment score. This finding likely reflects a number of factors. For example, the effect of NAE membership is likely to be variable because various NAE members may be at different stages in their career, thus contributing intellectually at different levels. In addition, there is considerable variance in the academic impact of NAE members (academic credentials are not all that is considered for academy membership), and this may also explain the variation in this parameter estimate. All of this means that adding NAE members is likely to have a differential effect across universities with regard to average peer assessment scores; although, the effect will always be positive.

Finally for the college of engineering attributes, Table 2 shows that the Google Scholar *h*-index of the 10th most cited engineering faculty member has a positive effect on average peer assessment scores. By definition, an author's *h*-index value indicates that the author has published *h* papers, each of which has been cited at least *h* times. The use of the 10th most cited faculty member was found to provide the best statistical fit, with faculty 10 spots down on the college's citation list capturing some measure of the depth of the scholarly output of the college of engineering, without unfairly penalizing colleges with smaller faculty numbers (as using the 30th or 40th most-cited faculty would). While many colleges of engineering in the U.S. have been slow to adopt citation information available in sources such as Google Scholar, Scopus, and Web of Science, for faculty evaluation, it is clear based on this estimation result that such information has a statistically significant effect on average peer assessment scores.

Please note that many of the college of engineering attributes found to be significant are actual metrics that are used along with average peer assessment scores to determine actual USNews graduate college of engineering rankings. While using the actual USNews ranking in the current year (or even in previous years) as an explanatory variable in our statistical model would violate fundamental regression assumptions (because it would be an endogenous variable setting up a correlation between X and corresponding error term ε), using the data as we have avoids this modeling issue (Washington et al., 2011). However, the finding that many of the factors that determine average peer scores are also used to determine overall college ranking in USNews formulas shows the fact that past USNews rankings are likely influencing current peer impressions. It is important to keep this dynamic relationship between peer impressions and actual USNews rankings in mind when formulating strategies to improve average peer assessment scores.

4.2 Overall University Attributes

Colleges of engineering do not operate in a vacuum, and it is quite understandable that peer impressions of the college will be influenced by their overall impression of the university. This "halo" effect has the reputation of the university inexorably influencing the reputation of the college of engineering within that university. Thus one would expect highly prestigious universities to have engineering programs ranked above less prestigious ones even if all of their engineering metrics were identical. The question becomes how to best define this "halo" effect with measurable university attributes.

Table 2 estimation results show that a university's membership in the American Association of Universities (AAU) results in an improved peer assessment score of 0.1376 holding all else constant. AAU membership requires that a wide variety and high level of research and scholarly output be achieved by the university. While there are a few universities that may have been grandfathered in when the selection criteria were less rigorous, membership in the AAU is an excellent indicator of university reputation, and the statistical significance of this variable confirms the importance of AAU recognition.

The estimation results in Table 2 also show that universities having 5 or more faculty at the university, as a whole, with more than 50,000 Google scholar citations enjoy higher peer rankings (an average 0.15848 higher peer assessment scores) than those universities that do not have as many of these faculty. The 50,000 citation threshold seems to be the best statistical measure of elite scholars in highly-citable fields that contribute to the overall "halo" effect of the institution.

Interestingly, the combination of AAU membership and the highly-cited faculty indicator means that two identical colleges of engineering (in terms of other attributes) will have roughly a 0.3 higher peer score (0.296) in a university that is a member of the AAU and has 5 or more faculty with more than 50,000 Google Scholar citations relative to a university without these attributes. This shows a substantial "halo" effect and, depending on the college of engineering's overall peer score, a 0.3 increase in the peer score could move the college past the average peer scores of 30 or more universities.

The number of post-doctoral students at a university reflects the breadth and depth of its highest-level research programs and seems to be the hallmark of prestigious universities. We find that the average number of annual postdoctoral researchers employed at the university as a whole (averaged over the past 15 years) positively influences the average peer assessment score of the college of engineering. It should be noted that the number of post-doctoral students is a factor in AAU membership and other measures of university reputations.

Finally, Table 2 shows that the past 10-year average of math plus verbal (1600 maximum) Scholastic Aptitude Test (SAT) scores of students entering the university positively influence the average peer assessment scores of engineering programs. The parameter estimates for this variable show that a 150 point increase in the average SAT score of all university students will result in a 0.1 increase in the average peer assessment score of the university's college of engineering. The average SAT scores of admitted university students are obviously a strong indicator of student potential and the university's overall reputation. And, this university "halo" effect significantly influences average peer impressions of engineering college quality.

5. Summary and Conclusions

Many universities in the United States use USNews annual program rankings to advertise and promote their programs (using current rankings and or ranking trajectories over the years), and administrators from the university president down to deans and department chairs often use these annual rankings as a measure of their and their program's success. While the quantitative data submitted to USNews to rank colleges of engineering are clearly important in determining overall ranking, peer assessment scores (the impressions of others of a university's program) invariably play an important role in overall program ranking.

This paper explored the factors influencing the USNews average peer assessment scores of U.S. engineering programs by estimating a random parameters linear regression using data extracted from a number of sources (USNews data, the Center for Measuring University Performance, Google Scholar). Our estimation results showed that college-level factors such as total number of faculty, average math Graduate Record Examination scores, the percent of applicants accepted, number of doctoral students graduated, total number of engineering faculty, graduate student enrollment per faculty, percent of faculty in the National Academy of Engineering, and Google Scholar h-index all influence average peer assessment scores. In addition, university-level factors such as university membership in the American Association of Universities (AAU), number of very highly cited faculty, number of post-doctoral appointees, and Scholastic Aptitude Test (SAT) scores of incoming university students were also influential in determining college of engineering peer assessment scores.

Interestingly, although research funding is needed to produce many of the variables that were found to be statistically significant in the model (such as number of doctoral students graduated, graduate student enrollment, and number of post-doctoral students), research funding itself was found to be statistically insignificant. This is a wake-up call to the many universities who blindly pursue research dollars without carefully thinking about the scholarly productivity that such dollars can potentially produce.

Over the years, because peer assessments are based on opinions and impressions, U.S. engineering colleges have tended to use promotional brochures, website enhancements, and other means to influence how peers assess them. However, our statistical analysis suggests that peer assessments are rooted much deeper in the measurable accomplishments of the college's faculty, the quality of its students, and the quantifiable achievements of its university as whole. In fact, the model estimation results presented herein provide valuable insights as to how universities might improve their peer assessment scores. While our statistical findings are strong and our model's predictive capabilities are as well (see Figure 1), it is important to keep in mind that the determinants of average peer sores are likely to change over time, as measures that influence peer impressions vary in importance over time and new measures become more widely accepted. Thus, it is important to undertake future statistical assessments in an effort to better understand the factors that affect peer assessments, and how the influence of these factors may change over time.

References

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41(1), 153-159.

Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40(2), 768-775.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37(9), 837-855.

Bougnol, M.-L., Dula, J., 2015. Technical pitfalls in university rankings. Higher Education 69(5), 859-866.

Bowman, N., Bastedo, M., 2009. Getting on the front page: Organizational reputation, status signals, and the impact of US News and World Report on student decisions. Research in Higher Education 50(5), 415-436.

Collins, F., Park, G.-S., 2016. Ranking and the multiplication of reputation: reflections from the frontier of globalizing higher education. Higher Education 72(1), 115-129.

Dichev, I., 2001. News or noise? Estimating the noise in U.S. News university rankings. Research in Higher Education 42(3), 237-266.

Dill, D., Soo, M., 2005. Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. Higher Education 49(4), 495-533.

Gnolek, S., Falciano, V., Kuncl, R., 2014. Modeling Change and Variation in US News and World Report College Rankings: What would it really take to be in the Top 20? Research in Higher Education 55(8), 761-779.

Greene, W., 2004. Interpreting estimated parameters and measuring individual heterogeneity in random parameter models. New York University Working Paper No. EC-04-08. New York, NY.

Grewal, R., Dearden, J., Llilien, G., 2008. The university rankings game: Modeling the competition among universities for ranking. The American Statistician 62(3), 232-237.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84-90.

Hansmann, H., 1999. Higher education as an associative good. Yale Centre for International Finance, Working paper no. 99-13, Yale Law School, Yale University, New Haven, CT.

Hazelkorn, E. 2008. Learning to live with league tables and ranking: The experience of institutional leaders. Higher Education Policy 21(2), 193-215.

Luca, M., Smith, J., 2013. Salience in quality disclosure: evidence from the USNews college rankings. Journal of Economics and Management Strategy 22, 58-77.

Li, J., 2016. The global ranking regime and the reconfiguration of higher education: Comparative case studies on research assessment exercises in China, Hong Kong, and Japan. Higher Education Policy 29(4), 473-493.

Lo, W., 2014. The ranking phenomenon and the experience of academics in Taiwan. Higher Education Policy 27(2), 259-277.

Lombardi, J., Phillips, E., Abbey, C., Craig, D., 2017. The top American research universities, 2016 Annual Report, The Center for Measuring University Performance, University of Massachusetts, Amherst, MA.

Machung, A., 1998. Playing the ranking game. Change: The magazine of higher learning 30, 12-16.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1-13.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1-16.

McFadden, D., Ruud, P., 1994. Estimation by simulation. The Review of Economics and Statistics 76 (4), 591-608.

Merideth, M. 2004. Why do universities compete in the ratings game? An empirical analysis of the effects of the U.S. News and World Report college rankings. Research in Higher Education 45(5), 443-461.

Pike, G. 2004. Measuring quality: A comparison of U.S. News rankings and NSSE benchmarks. Research in Higher Education 45(2), 193-208.

Sweitzer, K., Volkwein, J., 2009. Prestige among graduate and professional schools: Comparing the U.S. News' graduate school reputation rating between disciplines. Research in Higher Education 50(8), 812-836.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26(1), 24-36.

Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton, FL, Second Edition.

Williams, R., Van Dyke, N., 2008. Reputation and reality: ranking major disciplines in Australian universities. Higher Education 56(1), 1-28.

Yeung, R., Gigliotti, P., Nguyen-Hoang, P., 2018. The impact of U.S. News college rankings on the compensation of college and university presidents. Research in Higher Education, https://doi.org/10.1007/s11162-018-9501-7.

Table 1. Descriptive statistics of key variables included in the model of average peer assessment score.

Variable description	Mean	Standard Deviation	Minimum	Maximum
College of Engineering Attributes				
Average math Graduate Records Examination (GRE) score	161.57	3.47	157	169
Percent of graduate applications accepted indicator (1 if greater than 31%, 0 otherwise)	0.626	0.486	0	1
Number of engineering doctoral students graduated in previous year	78.31	73.82	0	398
Total number of engineering faculty if public university, 0 if private university	109.22	5.27	0	500
Total number of engineering faculty if private university, 0 if public university	32.33	64.42	0	376
Smaller-faculty indicator (1 if the total number of engineering faculty is less than 200, 0 otherwise)	0.833	0.374	0	1
Graduate student enrollment per engineering faculty member	8.95	5.77	2.5	36.3
Percent of faculty in the National Academy of Engineering	2.86	4.18	0	20
Google Scholar h-index of the 10th most cited College of Engineering faculty member	31.26	20.13	1	85
Overall University Attributes				
American Association of Universities (AAU) indicator (1 if the university is a member of the AAU, 0 otherwise)	0.361	0.482	0	1
Highly-cited faculty members indicator (1 if the number of faculty members within the entire university with more than 50,000 citations is 5 or greater, 0 otherwise)	0.320	0.468	0	1
Average annual number of post-doctoral appointees over the last 15 years	286.31	491.74	0	4890
Past 10-year average university Scholastic Aptitude Test (SAT) score (math plus verbal, 1600 maximum)	1204.95	136.63	744	1521

Table 2. Random parameters linear regression of average peer assessment score (all random parameters are normally distributed).

Variable description	**Parameter estimate**	***t*-statistic**
Constant	-0.9921	-1.46
College of Engineering Attributes		
Average math Graduate Records Examination (GRE) score	0.01181	3.18
Percent of graduate applications accepted indicator (1 if greater than 31%, 0 otherwise)	-0.10396	-4.45
Number of engineering doctoral students graduated in previous year	0.00295	6.80
Total number of engineering faculty if public university, 0 if private university	0.00170	5.27
Total number of engineering faculty if private university, 0 if public university	0.00285	6.91
Smaller-faculty indicator (1 if the total number of engineering faculty is less than 200, 0 otherwise) (*standard deviation of parameter distribution*)	-0.02279 (*0.11619*)	-0.050 (*12.60*)
Graduate student enrollment per engineering faculty member (*standard deviation of parameter distribution*)	-0.00269 (*0.00466*)	-1.57 (*5.78*)
Percent of faculty in the National Academy of Engineering (*standard deviation of parameter distribution*)	0.02182 (*0.0204*)	6.36 (*11.57*)
Google Scholar h-index of the 10th most cited College of Engineering faculty member	0.00155	1.97
Overall University Attributes		
American Association of Universities (AAU) indicator (1 if the university is a member of the AAU, 0 otherwise)	0.13760	5.01
Highly-cited faculty members indicator (1 if the number of faculty members within the entire university with more than 50,000 citations is 5 or greater, 0 otherwise)	0.15849	5.64
Average annual number of post-doctoral appointees over the last 15 years	0.00013	6.08
Past 10-year average university Scholastic Aptitude Test (SAT) score (math plus verbal, 1600 maximum)	0.00063	5.48
Number of observations	139	
R-squared	0.933	
Adjusted R-squared	0.926	

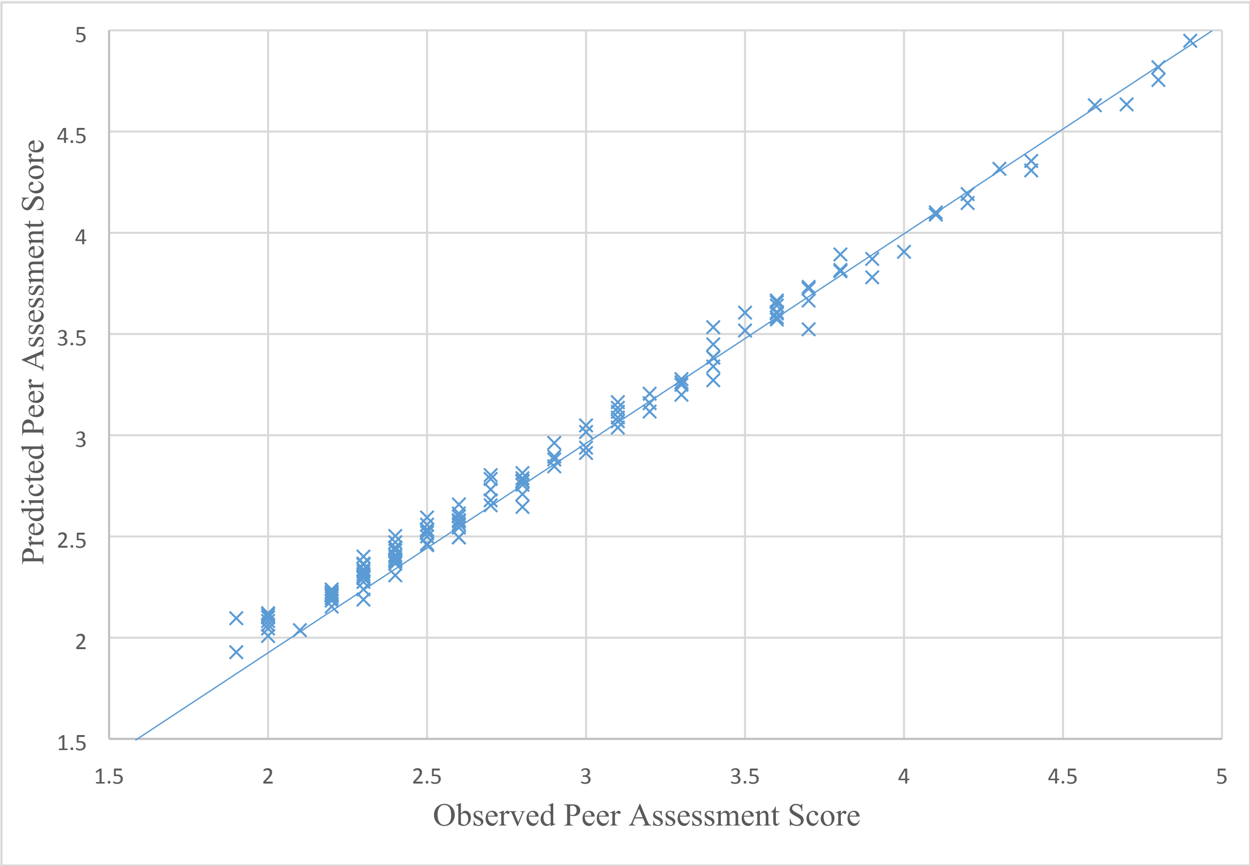

Figure 1. Comparison of observed vs. predicted average peer assessment rankings.