AEA Papers and Proceedings 2022, 112: 630–634
https://doi.org/10.1257/pandp.20221064

Statistical Uncertainty in the Ranking of Journals and Universities[†]

By Magne Mogstad, Joseph Romano, Azeem Shaikh, and Daniel Wilhelm*

Economists are obsessed with rankings of institutions, journals, or scholars according to the value of some feature of interest. These rankings are invariably computed using estimates rather than the true values of such features. As a result, there may be considerable uncertainty concerning the ranks. In this paper, we consider the problem of accounting for such uncertainty by constructing confidence sets for the ranks. We consider the problem of constructing marginal confidence sets for the rank of, say, a particular journal as well as simultaneous confidence sets for the ranks of all journals.

The purpose of this paper is to review the approach to the construction of such confidence sets by Mogstad et al. (2020) and then apply their methods to rankings of economics journals and universities by impact factors.[1]

I. Confidence Sets for Ranks

For concreteness, consider the ranking of $j = 1, \ldots, p$ journals according to their impact factors. Denote by P_j the distribution of data for journal j, by $\theta(P_j)$ the population ("true") impact factor of journal j, by $\hat{\theta}_j$ an estimate of $\theta(P_j)$, and by $\hat{se}_j$ the corresponding standard error (assumed available). The population rank of journal j is defined as $r_j(P) \equiv 1 + \sum_{k \neq j} \mathbf{1}\{\theta(P_k) > \theta(P_j)\}$ and $P \equiv (P_1, \ldots, P_p)$.

A. *Marginal Confidence Sets*

The goal in this subsection is to construct a two-sided confidence set $R_{n,j}$ for the rank of a particular journal j that satisfies

$$(1) \quad \liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P\{r_j(P) \in R_{n,j}\} \geq 1 - \alpha$$

for some "large" set of distributions $\mathbf{P}$ and some prespecified confidence level $1 - \alpha$. The construction is based on simultaneous confidence sets for the differences of impact factors, as in Mogstad et al. (2020) and Bazylik et al. (2021).[2] For concreteness, we explain one particular approach based on the parametric bootstrap, but other constructions are possible; see Mogstad et al. (2020). To this end, consider the confidence set

$$(2) \quad C_{\mathrm{symm},n,j,k} \equiv \left[\hat{\theta}_j - \hat{\theta}_k \pm \hat{se}_{j,k} c^{1-\alpha}_{\mathrm{symm},n,j}\right],$$

where $\hat{se}^2_{j,k}$ is an estimate of the variance of $\hat{\theta}_j - \hat{\theta}_k$ and $c^{1-\alpha}_{\mathrm{symm},n,j}$ is the $(1 - \alpha)$-quantile of

$$\max_{k: k \neq j} \frac{|\hat{\theta}_j - \hat{\theta}_k - (\theta(P_j) - \theta(P_k))|}{\hat{se}_{j,k}}.$$

This quantile could, for instance, be approximated by the bootstrap. Since, in our applications, we do not have access to the microdata that were used to compute the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_p$, a nonparametric bootstrap is not feasible. However, it is reasonable to assume that the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_p$ are approximately normally distributed and independent. Based

* Mogstad: University of Chicago (email: magne.mogstad@gmail.com); Romano: Stanford University (email: romano@stanford.edu.) Shaikh: University of Chicago (email: amshaikh@uchicago.edu); Wilhelm: University College London (email: d.wilhelm@ucl.ac.uk). The second author acknowledges support from the National Science Foundation (MMS-1949845). The third author acknowledges support from the National Science Foundation (SES-1530661). The fourth author acknowledges support from the European Research Council (Starting Grant No. 852332). We are grateful to D. Stern and C. Zimmermann for sharing their data with us.

[†] Go to https://doi.org/10.1257/pandp.20221064 to visit the article page for additional materials and author disclosure statement(s).

[1] There is a considerable body of academic work on the ranking of journals and institutions. See, e.g., Kalaitzidakis, Mamuneas, and Stengos (2003); Stern (2013); Ham, Wright, and Ye (2021); and references therein.

[2] Another proposal for confidence sets for ranks that satisfy (1) is Klein, Wright, and Wieczorek (2020); a comparison of the two approaches can be found in Mogstad et al. (2020).

on this assumption, we set $\hat{se}_{j,k}^2 = \hat{se}_j^2 + \hat{se}_k^2$ and use a parametric bootstrap to approximate $c_{\text{symm},n,j}^{1-\alpha}$ as follows. Generate R draws of normal random vectors $Z \equiv (Z_1, \ldots, Z_p)' \sim N\left(0, diag\left(\hat{se}_1^2, \ldots, \hat{se}_p^2\right)\right)$. The desired quantile $c_{\text{symm},n,j}^{1-\alpha}$ can then be approximated by the empirical $(1-\alpha)$-quantile of the R draws of $\max_{k:k\neq j}|Z_j - Z_k|/\hat{se}_{j,k}$.

Under weak conditions, the confidence sets for the differences simultaneously cover all true differences involving journal j:

$$(3)\quad \liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{\Delta_{j,k}(P) \in C_{\text{symm},n,j,k} \text{ for all } k \text{ with } k \neq j\} \geq 1-\alpha,$$

where $\Delta_{j,k}(P) \equiv \theta(P_j) - \theta(P_k)$. Collect the journals k whose differences with j have a confidence set $C_{\text{symm},n,j,k}$ that lies entirely below zero,

$$N_j^- \equiv \left\{k : k \neq j \text{ and } C_{\text{symm},n,j,k} \subseteq \mathbf{R}_-\right\}$$

and, similarly,

$$N_j^+ \equiv \left\{k : k \neq j \text{ and } C_{\text{symm},n,j,k} \subseteq \mathbf{R}_+\right\}.$$

Thus, N_j^- contains all journals k that have a significantly higher impact factor than j, while N_j^+ contains all the journals k that have a significantly lower impact factor than j. If the true impact factors of journals k in N_j^- (N_j^+) were indeed all higher (lower) than that of journal j, then the rank of journal j cannot be better than $|N_j^-|+1$ and cannot be worse than $p - |N_j^+|$. Thus, the set

$$(4)\qquad R_{n,j} \equiv \left\{|N_j^-|+1, \ldots, p - |N_j^+|\right\}$$

would contain the true rank of journal j. Of course, the confidence sets for the differences cover the true differences only with probability approximately no less than $1-\alpha$, so $R_{n,j}$ covers the true rank of journal j only with probability approximately no less than $1-\alpha$. In conclusion, for the construction described in this subsection, (3) implies that $R_{n,j}$ is a confidence set for the rank $r_j(P)$ satisfying (1) as desired.

It is possible to improve the simple construction of $R_{n,j}$ above by inverting hypotheses tests of

$$H_{j,k} : \theta(P_j) - \theta(P_k) = 0$$

versus its negation for all k that are not equal to j. After testing this family of hypotheses, one then counts the number of hypotheses that were rejected in favor of $\theta(P_j) < \theta(P_k)$ and in favor of $\theta(P_j) > \theta(P_k)$. The first number plus one is then used as lower endpoint, and the second number subtracted from p is then used as upper endpoint for $R_{n,j}$. This confidence set satisfies (1) provided that the procedure used to test the family of hypotheses controls the mixed directional familywise error rate (mdFWER) at α; i.e.,

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} \text{mdFWER}_P \leq \alpha,$$

where mdFWER is the probability of making any mistake, either a false rejection or an incorrect determination of a sign; see Mogstad et al. (2020) for details.

B. *Simultaneous Confidence Sets*

A small modification of the above construction of a marginal confidence set for the rank of a single journal delivers two-sided confidence sets $R_{n,j}$ for the ranks of all journals $j = 1, \ldots, p$ such that all true ranks are covered simultaneously; i.e.,

$$(5)\quad \liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{r_j(P) \in R_{n,j} \text{ for all } j = 1, \ldots, p\} \geq 1-\alpha.$$

We start with confidence sets for the differences $C_{\text{symm},n,j,k}$ as in (2) except that the critical value $c_{\text{symm},n,j}^{1-\alpha}$ is now defined as the $(1-\alpha)$-quantile of

$$\max_{(j,k):k\neq j} \frac{|\hat{\theta}_j - \hat{\theta}_k - \left(\theta(P_j) - \theta(P_k)\right)|}{\hat{se}_{j,k}},$$

where the max is taken over all pairs (j,k) such that $j \neq k$, so the critical value is independent of j. As above, this critical value can be approximated by the $(1-\alpha)$-quantile of the R draws of $\max_{(j,k):k\neq j}|Z_j - Z_k|/\hat{se}_{j,k}$. Then, the confidence set for journal j, $R_{n,j}$, is computed as in (4) using the definitions of N_j^-, N_j^+ as above except that the confidence sets for the differences, $C_{\text{symm},n,j,k}$, are replaced by the new ones described here.

Stepwise methods can be used to improve this simple construction of simultaneous confidence sets similarly to the stepwise improvements described for the marginal confidence sets.

C. *Confidence Sets for the τ-Best*

In this section, we are interested in constructing confidence sets for the τ-best journals, defined as

$$R_0^{\tau-\text{best}}(P) \equiv \{j \in \{1, \ldots, p\} : r_j(P) \leq \tau\}.$$

The goal is to construct a (random) set $R_n^{\tau-\text{best}}$ satisfying

$$(6) \qquad \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{R_0^{\tau-\text{best}}(P) \subseteq R_n^{\tau-\text{best}}\} \geq 1-\alpha.$$

To this end, let $R_{n,j}$, $j = 1, \ldots, p$, be simultaneous lower confidence bounds on the ranks of all journals; i.e., each $R_{n,j}$ has upper endpoint equal to p and (5) is satisfied. Such one-sided confidence sets for the ranks can be constructed similarly to the two-sided confidence sets described in Section IIB except that the two-sided confidence sets for the differences are replaced by one-sided confidence sets; see online Appendix A for details. Then,

$$R_n^{\tau-\text{best}} \equiv \{j \in J : \tau \in R_{n,j}\}$$

is a confidence set satisfying (6). Mogstad et al. (2020) propose a different, more direct approach to constructing confidence sets for the τ-best, which in simulations has been shown to produce shorter confidence sets but is computationally more demanding.

Confidence sets for the τ-worst can be constructed as confidence sets for the τ-best among $-\theta(P_1), \ldots, -\theta(P_p)$.

II. Ranking Academic Journals by Impact Factors

In this section, the methods from Section I are applied to the ranking of economics journals by impact factors, using the data from Stern (2013). The original dataset comprises estimated impact factors and their standard errors for 230 journals. The impact factor for a given journal is computed in 2011 as the average number of Web of Science citations for all articles published in that journal from the years 2006 to 2010. For more details, see the original paper. The impact factors and standard errors are plotted in Figures A1 and A2 in the online Appendix.

Figure A4 in the online Appendix shows the marginal confidence sets for the ranks of all 230 journals, ordered such that the journals with the highest impact factors (lowest ranks) appear at the bottom. Figure 1 shows the marginal confidence sets for the ranks among only the 30 journals that were identified by Kalaitzidakis, Mamuneas, and Stengos (2003) as the "top 30" and were reanalyzed in Stern (2013, Figure 2).[3] The corresponding simultaneous confidence sets are shown in Figures A5 and A6 in the online Appendix. We compute the marginal confidence sets by the stepwise procedure described in Section IA with $R = 1{,}000$ bootstrap draws. Since more comparisons have to be performed among all 230 journals than among only 30 journals, the confidence sets for the 30 journals in Figure 1 are shorter than the corresponding ones in Figure A4.

The broad pattern in Figure A4 shows that confidence sets for the ranks are relatively more informative at the bottom and the top of the ranking compared to the middle. In addition, there are some journals with extremely wide confidence sets. For instance, the journal ranked twenty-fourth (*Experimental Economics*) has a marginal confidence set for the rank ranging from 4 to 230. The ranking of the 30 journals in Figure 1 is much more informative, in the sense that confidence sets are relatively narrow, especially at the top and bottom of the ranking. For instance, with 95 percent confidence, the rank of the *Journal of Economic Literature* (*JEL*) is between 1 and 2, and that of *Economics Letters* is equal to 30.

Due to the importance of "top-five" publications in the economics discipline, we compute 95 percent confidence sets for the 5 best and for the 25 worst among the 30 journals in Figure 1. We employ the method described in Section IC, where a stepwise procedure is used to compute the simultaneous confidence sets. The confidence set for the five best contains ten journals: *JEL*, the *Quarterly Journal of Economics* (*QJE*), the *Journal of Economic Perspectives*, the *Journal of Financial Economics*, the *Journal of Political Economy*, *Econometrica*, the *American Economic Review*, the *Review of*

[3] Kalaitzidakis, Mamuneas, and Stengos (2003) selected the "top 30" based on citations data from 1994 to 1998. We take this selection as given and do not take into account that it was based on data.

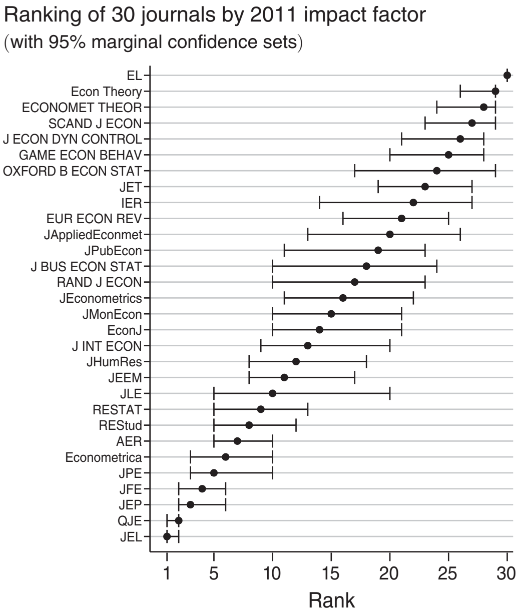

FIGURE 1. RANKING OF THE TOP 30 JOURNALS OF KALAITZIDAKIS, MAMUNEAS, AND STENGOS (2003)

Notes: The dots show the estimated ranks, and the horizontal lines represent the 95 percent marginal confidence sets for the ranks of each journal. Journals are ordered by their impact factors, with those with the highest impact factors appearing at the bottom (small ranks) and those with the smallest appearing at the top (large ranks). Names of journals are as in Stern (2013).

Economic Studies, the *Review of Economics and Statistics*, and the *Journal of Labor Economics*. The confidence set for the 25 worst includes all journals except the *QJE*. In conclusion, in terms of impact factor as defined here, 10 of the 30 journals cannot be rejected to be among the top 5 journals, and only 1, the *QJE*, can be rejected to be among the worst 25.

Note that confidence sets for the ranks in Figure 1 reveal similar patterns in the uncertainty pertaining to the ranks of each journal as the "range of ranks" constructed in Figure 2 of Stern (2013). For instance, both methods indicate little uncertainty at the very top and the very bottom of the ranking and more uncertainty in the middle of the ranking. However, the advantage of our confidence sets is that they satisfy the formal coverage guarantee discussed in Section IA; i.e., they cover the true ranks approximately, with probability no less than a prespecified level.

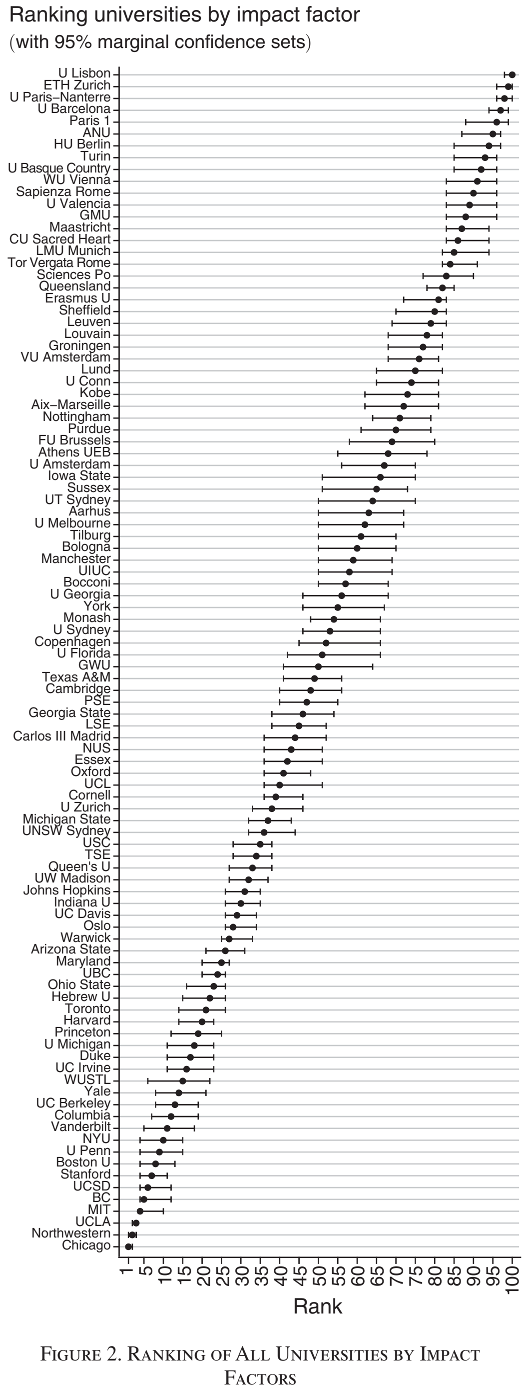

FIGURE 2. RANKING OF ALL UNIVERSITIES BY IMPACT FACTORS

Notes: Each horizontal line represents the 95 percent marginal confidence set for the rank of a university, where universities are ordered by their impact factor, with those with the highest impact factors appearing at the bottom (small ranks) and those with the smallest appearing at the top (large ranks). The dots show the estimated ranks of each university.

III. Ranking Universities by Impact Factors

In this section, the methods from Section I are applied to the ranking of universities by impact factors, using data on 662,604 articles by 40,496 authors that were deposited on Research Papers in Economics (RePEc) (Zimmermann 2013) in July 2021. We remove authors whose affiliation is missing or who have multiple affiliations but did not specify weights for them. For each of the remaining authors, all of their publications' impact factors (defined as the impact factor of the journal in which the article was published) are assigned to each of their affiliations after multiplying them by the specified weights of the affiliations. The average impact factor of publications assigned to an institution then form the basis for the league tables of institutions that are created. We remove all institutions that are not universities, and from the remaining ones, we select the 100 universities that are ranked 100 or better according to the average impact factor. The resulting dataset of 100 impact factors and standard errors are plotted in online Appendix Figure A3.

Note that the weights according to which impact factors of publications are apportioned to affiliations are the weights recorded on RePEc in July 2021. Since researchers move between institutions, we do not necessarily assign impact factors to the institutions at which the corresponding publications were created, but rather those with which the authors were affiliated in July 2021. Therefore, the estimated impact factor could be interpreted as a measure of the average "stock of impact" that the collection of researchers at a university has accumulated up until July 2021. This can be viewed as a noisy estimate of the expected stock of impact of a university.

Figure 2 shows the marginal confidence sets for the ranks of all 100 universities, ordered such that the universities with the highest impact factors (lowest ranks) appear at the bottom. We use the stepwise procedure described in Section IA with $R = 1{,}000$ bootstrap draws. The corresponding simultaneous confidence sets are shown in Figure 9 in the online Appendix.

Interestingly, the broad pattern in Figure 2 reveals that the confidence sets for the ranks are fairly informative throughout the entire ranking, but particularly so at the bottom and the top. For instance, with 95 percent confidence, the rank of Chicago is either one or two, and that of UCLA is either two or three.

REFERENCES

Bazylik, Sergei, Magne Mogstad, Joseph P. Romano, Azeem M. Shaikh, and Daniel Wilhelm. 2021. "Finite- and Large-Sample Inference for Ranks Using Multinomial Data with an Application to Ranking Political Parties." Centre for Microdata Methods and Practice Working Paper CWP40/21.

Ham, John C., Julian Wright, and Ziqiu Ye. 2021. "New Rankings of Economics Journals: Documenting and Explaining the Rise of the New Society Journals." https://ap5.fas.nus.edu.sg/fass/ecsjkdw/journalrankings_20211230.pdf.

Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos. 2003. "Rankings of Academic Journals and Institutions in Economics." *Journal of the European Economic Association* 1 (6): 1346–66.

Klein, Martin, Tommy Wright, and Jerzy Wieczorek. 2020. "A Joint Confidence Region for an Overall Ranking of Populations." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69 (3): 589–606.

Mogstad, Magne, Joseph P. Romano, Daniel Wilhelm, and Azeem M. Shaikh. 2020. "Inference for Ranks with Applications to Mobility across Neighborhoods and Academic Achievement across Countries." Centre for Microdata Methods and Practice Working Paper CWP10/20.

Stern, David I. 2013. "Uncertainty Measures for Economics Journal Impact Factors." *Journal of Economic Literature* 51 (1): 173–89.

Zimmermann, Christian. 2013. "Academic Rankings with RePEc." *Econometrics* 1 (3): 249–80.

This article has been cited by:

1. Wojciech Charemza, Michał Lewandowski, Łukasz Woźny. 2024. On journal rankings and researchers' abilities. *Journal of Informetrics* **18**:3, 101559. [Crossref]
2. Reza Aminghafouri, Liping Fu. 2024. Joint Confidence Region Approach to Ranking Hotspot Locations Considering Uncertainty in Expected Risk Estimates. *Transportation Research Record: Journal of the Transportation Research Board* **2678**:1, 555-567. [Crossref]
3. Wojciech Charemza, Michal Lewandowski, Lukasz Patryk Wozny. On Journal Rankings and Researchers' Abilities **14**, . [Crossref]
4. Boudewijn de Bruin. 2023. Ranking philosophy journals: a meta-ranking and a new survey ranking. *Synthese* **202**:6. . [Crossref]
5. Fulya Y. Ersoy, Jennifer Pate. 2023. Invisible hurdles: Gender and institutional differences in the evaluation of economics papers. *Economic Inquiry* **61**:4, 777-797. [Crossref]
6. John A. List, Azeem M. Shaikh, Atom Vayalinkal. 2023. Multiple testing with covariate adjustment in experimental economics. *Journal of Applied Econometrics* **38**:6, 920-939. [Crossref]