Name for contribution: Arthur Spirling, John L Loeb Associate Professor of the Social Sciences, Harvard University, Cambridge, Massachusetts, United States of America.
Email: arthur.spirling@gmail.com
Words: 385

Motivation and Measurement

Like any measurement problem, ranking journals involves judgements about the reliability and validity of the metric. On the former, more information about how rankings might change (suddenly?) over time and how sensitive they are to a few 'big' outlier articles in a given year is warranted. On validity, the fact that the authors' measurement strategy is deemed successful in part because it returns the 'big four' statistics journals at the head of the rank order should be interpreted with some caution to ensure that we do not reward methods for simply reproducing our priors with the data.

The Bradley-Terry model seems straightforward to fit and has obvious benefits; but the decision over *which* journals to compare is consequential, disputable and requires domain expertise. There is some danger that the original problem of "*how* should we compare journals?" is replaced with "*what* journals should we compare?"—which may be a very thorny issue! Related to this, is there not a way to use the GLM linear predictor, or indeed a random effect, to 'control' for discipline or subject matter in estimating the impact of a journal?

The authors' motivation is partly "economic" insofar as librarians must make choices as to the journals they subscribe to. As more journals move towards an 'open access' model, this justification seems less important, although perhaps the information will still be helpful for submitting authors.

Extensions

The authors' efforts can certainly be used in other fields: Professor Carter and I attempted something similar a few years ago.[1] In that case, we noted that discipline-specific practices—such as a focus on books along with articles as a venue for information transmission—can make ranking journals less helpful as a way to assess research output. Furthermore, journal-specific practices in some disciplines—including the encouragement away from long literature reviews for certain outlets—can cause obvious problems for an 'import-export' metric that assumes 'fair trade' across journals.

On the statistical side, a more explicitly Bayesian approach might allow for (a) a cost function to penalize 'incorrect' decisions by e.g. librarians, given the ranking uncertainty, (b) the addition

---

[1] "Under the Influence? Intellectual Exchange in Political Science". David Carter and Arthur Spirling. PS: Political Science and Politics, Vol. 41, No. 2 (Apr., 2008), pp. 375-378

of discipline random effects without the need for numerical integration methods and (c) a natural way to incorporate (expert) prior information on the relative prestige of outlets.