Statistical Modelling of Citation Exchange Among Statistics Journals

Cristiano Varin

Università Ca' Foscari, Venezia, Italy.

Manuela Cattelan

Università di Padova, Padova, Italy.

David Firth

University of Warwick, Coventry, UK.

Summary. Scholarly journal rankings based on citation data are often met with skepticism by the scientific community. Part of the skepticism is due to the discrepancy between the common perception of journals' prestige and their ranking based on citation counts. A more serious concern is the inappropriate use of journal rankings to evaluate the scientific influence of authors. This paper focuses on analysis of the table of cross-citations among a selection of Statistics journals. Data are collected from the Web of Science® database published by Thomson Reuters. Our results suggest that modelling the exchange of citations between journals is useful to highlight the most prestigious journals, but also that journal citation data are characterized by considerable heterogeneity, which needs to be properly summarized. Inferential conclusions require care in order to avoid potential over-interpretation of insignificant differences between journal ratings.

Keywords: Bradley-Terry Model; Citation Data; Impact Factor; Journal Ranking; Research Evaluation.

1. Introduction

The problem of ranking scholarly journals has arisen partly as an economic matter. When the number of scientific journals started to increase, librarians were faced with decisions as to which journal subscriptions should consume their limited economic resources; a natural response was to be guided by the relative importance of different journals according to a published or otherwise agreed ranking. Gross and Gross (1927) proposed the counting of citations received by journals as a direct measure of their importance. Garfield (1955) suggested that the number of citations received should be normalized by the number of citable items published by a journal. This idea is at the origin of the *Impact Factor*, the best known index for ranking journals. Published since the 1960s, the Impact Factor is 'an average citation rate per published article' (Garfield, 1972).

The Impact Factor of the journals where scholars publish has been employed — improperly, many might argue — in appointing to academic positions, in awarding grants, and in ranking universities and their departments. The *San Francisco Declaration on Research Assessment* (DORA, 2013) and the IEEE Position Statement on *Appropriate Use of Bibliometric Indicators for the Assessment of Journals, Research Proposals, and Individuals* (IEEE Board of Directors, 2013) are just two of the most recent authoritative standpoints regarding the risks of automatic, metric-based evaluations of scholars. Typically, only a small fraction of all published articles accounts for most of citations received by a journal (Seglen, 1997). Single authors should ideally be evaluated on the basis of their own outputs, perhaps through direct citation counts or assessment by peers, and not through citations of other papers that have appeared in the journals where their papers have been published (Seglen, 1997; Adler *et al.*, 2009; Silverman, 2009). As stated in a recent *Science* editorial (Alberts, 2013), no automatic metric evaluation based on journal citations should substitute judgment based on reading each researcher's publications:

> '(...) the leaders of the scientific enterprise must accept full responsibility for thoughtfully analyzing the scientific contributions of other researchers. To do so in a meaningful way requires the actual reading of a small selected set of each researcher's publications, a task that must not be passed by default to journal editors'.

Rankings based on the Impact Factor often differ substantially from common perceptions of journal prestige. Various causes of such discrepancy have been pointed out. First, there is the phenomenon that more 'applied' journals tend to receive citations from other scientific fields more often than do journals that publish theoretical work. Second is the short time-period used for computation of the Impact Factor, which can be completely inappropriate for some fields, in particular for Mathematics and Statistics (van Nierop, 2009). Finally, there is the risk of manipulation, whereby authors might be asked by journal editors to add irrelevant citations to other papers published in their journal (Sevinc, 2004; Frandsen, 2007). It is not surprising, therefore, that the Impact Factor and other 'quantitative' journal rankings have given rise to substantial skepticism about the value of citation data.

Journal citation data are unavoidably characterized by substantial variability. Nevertheless, quantification of uncertainty is typically lacking in published rankings of journals. The main purpose of this paper is to illustrate the risks of over-interpretation of insignificant differences between journal ratings. In particular, we focus on the analysis of the exchange of citations among a relatively homogeneous list of journals. Following Stigler (1994), we model the table of cross-citations between journals in the same field by using a Bradley-Terry model (Bradley and Terry, 1952) and thereby derive a ranking of the journals' ability to 'export intellectual influence' (Stigler, 1994). Although the Stigler (1994) approach has desirable properties and is simple enough to be promoted also outside the statistics community, there have been rather few published examples of application of this model since its first appearance; Stigler *et al.* (1995) and Liner and Amin (2004) are two notable examples of its application to the journals of Economics.

In this paper, we pay particular attention to methods that summarize the uncertainty in a ranking produced through the Stigler (1994) model-based approach. Our focus on properly accounting for 'model-based uncertainty in making comparisons' is close in spirit to Goldstein and Spiegelhalter (1996). We suggest also the use of the ranking lasso penalty (Masarotto and Varin, 2012) when fitting the Stigler model, in order to combine the benefits of shrinkage with an enhanced interpretation arising from automatic presentational grouping of journals with similar merits.

The paper is organized as follows. Section 2 describes the data collected from the Web of Science® database compiled by Thomson Reuters. Section 3 illustrates the use of cluster analysis to identify groups of Statistics journals sharing similar aims and types of content. In Section 4, after a brief summary of journal rankings published by Thomson Reuters in the Journal Citation Reports®, the Stigler (1994) method is described and applied to the table of cross-citations among Statistics journals. Section 5 illustrates the use of the ranking lasso penalty in the context of such a model for journal cross-citations. Section 6 collects some concluding remarks.

2. The Web of Science® database

The database used for our analyses is the 2010 edition of the Web of Science® produced by Thomson Reuters. The citation data contained in the database are employed to compile the Journal Citation Reports®, whose Science Edition® summarizes citation exchange between more than 8,000 journals in science and technology. Within the Journal Citation Reports®, scholarly journals are grouped into 171 overlapping subject categories. In particular, in 2010 the *Statistics and Probability* category comprises 110 journals. The choice of the journals that are encompassed in this category is to some extent arbitrary. The Scopus® database, which is the main commercial competitor of Web of Science®, includes in its Statistics and Probability category 105 journals, but only about two thirds of them are classified in the same category within Web of Science®. The Statistics and Probability category contains also journals related to fields such as Econometrics, Chemistry, Computational Biology, Engineering and Psychometrics.

A severe criticism of the Impact Factor relates to the time period used for its calculation. The standard version of the Impact Factor considers citations received to articles published in the previous two years. This period is too short to reach the peak of citations of an article, especially in mathematical disciplines (Hall, 2009). van Nierop (2009) finds that articles published in Statistics journals typically reach the peak of their citations significantly later than three years after publication. As reported by the Journal Citation Reports®, the median age of the articles cited in this category is more than 10 years. Thomson Reuters acknowledges this issue and computes a second version of the Impact Factor using citations to papers published in the previous five years. Recent published alternatives to the Impact Factor, to be discussed in Section 4.1, also count citations to articles that appeared in the previous five years.

This paper considers citations of articles published in the last ten years, in order to capture the influence, over a more substantial period, of papers published in statistical journals. A key requirement for the methods described here, as well as in our view for any sensible analysis of citation data, is that the journals jointly analyzed should be as homogeneous as possible. Accordingly, analyses are conducted on a subset of the journals from the Statistics and Probability category, among which there is a relatively high level of citation exchange. The selection is obtained by discarding journals in Probability, Econometrics, Computational Biology and Engineering, and other journals not sufficiently related to the majority of the journals in the selection. Furthermore, journals recently established, and thus lacking a record of ten years of citable items, are dropped. The final selection consists of the 47 journals listed in Table 1. Obviously, the methods discussed in this paper can be similarly applied to other selections motivated by different interests. For example, a statistician interested in applications to Economics will likely consider a different selection with econometrical and methodological statistical journals, discarding instead journals oriented towards bio-medical applications.

The 2010 edition of the Journal Citation Reports® provides various summaries of journal citation data for papers published in 2010. For each statistics journal, Table 2 shows the citations made by papers published in each journal in 2010 to papers published in other journals in the decade 2001-2010, as well as the citations that the papers published in a

Table 1. List of selected Statistics journals, with acronyms used in the manuscript.

Journal name	Acronym
American Statistician	AmS
Annals of Statistics	AoS
Annals of the Institute of Statistical Mathematics	AISM
Australian and New Zealand Journal of Statistics	ANZS
Bernoulli	Bern
Biometrical Journal	BioJ
Biometrics	Bcs
Biometrika	Bka
Biostatistics	Biost
Canadian Journal of Statistics	CJS
Communications in Statistics - Simulation and Computation	CSSC
Communications in Statistics - Theory and Methods	CSTM
Computational Statistics	CmpSt
Computational Statistics and Data Analysis	CSDA
Environmental and Ecological Statistics	EES
Environmetrics	Envr
International Statistical Review	ISR
Journal of Agricultural, Biological and Environmental Statistics	JABES
Journal of Applied Statistics	JAS
Journal of Biopharmaceutical Statistics	JBS
Journal of Computational and Graphical Statistics	JCGS
Journal of Multivariate Analysis	JMA
Journal of Nonparametric Statistics	JNS
Journal of Statistical Computation and Simulation	JSCS
Journal of Statistical Planning and Inference	JSPI
Journal of Statistical Software	JSS
Journal of the American Statistical Association	JASA
Journal of the Royal Statistical Society Series A	JRSS-A
Journal of the Royal Statistical Society Series B	JRSS-B
Journal of the Royal Statistical Society Series C	JRSS-C
Journal of Time Series Analysis	JTSA
Lifetime Data Analysis	LDA
Metrika	Mtka
Scandinavian Journal of Statistics	SJS
Stata Journal	StataJ
Statistica Neerlandica	StNee
Statistica Sinica	StSin
Statistical Methods in Medical Research	SMMR
Statistical Modelling	StMod
Statistical Papers	StPap
Statistical Science	StSci
Statistics	Stat
Statistics and Computing	StCmp
Statistics and Probability Letters	SPL
Statistics in Medicine	StMed
Technometrics	Tech
Test	Test

Statistics journal in 2001-2010 received from papers published in other journals in 2010. The same information is visualized in the back-to-back bar plots of Figure 1. Citations made and received are classified into three categories, namely self citations from a paper published in a journal to another paper in the same journal, citations to/from other journals in the list of selected Statistics journals, and citations to/from journals not in the selection.

The total numbers of citations reported in the second and fifth columns of Table 2 include citations given or received by all journals included in the Web of Science® database, not only those in the field of Statistics. The totals are influenced by the size of the journal and by the citation patterns of other categories to which journals are related. The number of references to articles published in 2001-2010 ranges from 275 for *Statistical Modelling*, which has a small size publishing around 350-400 pages per year, to 4,022 for *Statistics in Medicine*, which is a large journal with size ranging from 3,500 to 6,000 pages annually in the period examined. The number of citations from a journal to articles in the same journal is quite variable and ranges from 0.8% of all citations for *Computational Statistics* to 24% for *Stata Journal*. On average, 6% of the references in a journal are to articles appearing in the same journal and 40% of references are addressed to journals in the list. The *Journal of the Royal Statistical Society* Series A has the lowest percentage of citations to journals in the list, at only 15%. Had we kept the whole Statistics and Probability category of the Journal Citation Reports® that percentage would have risen by just 2 points to 17%; most of the references appearing in *Journal of the Royal Statistical Society* Series A are to journals outside the category.

The number of citations received ranges from 168 for *Computational Statistics* to 6,602 for *Statistics in Medicine*. Clearly, the numbers are influenced by the size of the journal. For example, the small number of citations received by *Computational Statistics* relates to only around 700 pages published per year by that journal. The citations received are influenced also by the pattern of citations of other categories. In particular, the number of citations received by journals oriented towards medical applications benefits from communication with a large field including many high-impact journals. For example, around 75% of the citations received by *Statistics in Medicine* come from journals outside the list of Statistics journals, mostly from medical journals. On average, 7% of the citations received by journals in the list come from the same journal and 40% are from other journals in the list.

As stated already, the Statistics journals upon which we focus have been selected from the Statistics and Probability category of the Journal Citation Reports®, with the aim of retaining those which communicate more. An extreme example of a journal discarded by our selection is *Utilitas Mathematica*, which makes just one citation to a journal in the Statistics and Probability category and which is cited only once by one of them in the study period. The inclusion of this journal in the category appears strongly questionable. As regards the other journals excluded, generally they do not exchange many citations with the journals retained. An important example is *Econometrica*, which is ranked in leading positions by all the published citation indices. *Econometrica* has only about 2% of its references addressed to other journals in our list, and receives only 5% of its citations from journals within our list.

Journal Citation Reports® also supplies detailed information about the citations exchanged between pairs of journals through the *Cited Journal Table* and the *Citing Journal Table*. The Cited Journal Table for journal i contains the number of times that articles published in journal j during 2010 cite articles published in journal i in previous years. Similarly, the Citing Journal Table for journal i contains the number of times articles published in journal j in previous years were cited in journal i during 2010. Those two tables

Table 2. Citations made (`Citing`) and received (`Cited`) in 2010 to/from articles published in 2001-2010. Columns are total citations (`Total`), proportion of citations that are self-citations (`%Self`), and proportion of citations that are to/from other Statistics journals (`%Stat.`). Journal acronyms are in Table 1.

Journal	Citing			Cited		
	Total	%Self	%Stat.	Total	%Self	%Stat.
AmS	380	0.11	0.43	648	0.07	0.29
AoS	1663	0.17	0.48	3335	0.09	0.47
AISM	459	0.04	0.36	350	0.05	0.57
ANZS	284	0.02	0.35	270	0.02	0.34
Bern	692	0.03	0.29	615	0.04	0.39
BioJ	845	0.07	0.50	664	0.08	0.42
Bcs	1606	0.12	0.49	2669	0.07	0.45
Bka	872	0.09	0.57	1713	0.04	0.60
Biost	874	0.06	0.41	1948	0.03	0.22
CJS	419	0.04	0.51	362	0.04	0.60
CSSC	966	0.03	0.43	344	0.08	0.48
CSTM	1580	0.06	0.41	718	0.13	0.59
CmpSt	371	0.01	0.33	168	0.02	0.38
CSDA	3820	0.13	0.45	2891	0.17	0.40
EES	399	0.10	0.34	382	0.10	0.23
Envr	657	0.05	0.27	505	0.06	0.27
ISR	377	0.05	0.21	295	0.07	0.32
JABES	456	0.04	0.26	300	0.05	0.27
JAS	1248	0.03	0.31	436	0.08	0.33
JBS	1132	0.09	0.33	605	0.16	0.33
JCGS	697	0.06	0.44	870	0.05	0.43
JMA	2167	0.09	0.49	1225	0.15	0.52
JNS	562	0.03	0.52	237	0.07	0.65
JSCS	736	0.04	0.43	374	0.09	0.45
JSPI	3019	0.08	0.44	1756	0.13	0.54
JSS	1361	0.07	0.21	1001	0.09	0.17
JASA	2434	0.10	0.41	4389	0.05	0.44
JRSS-A	852	0.05	0.15	716	0.05	0.24
JRSS-B	506	0.11	0.51	2554	0.02	0.42
JRSS-C	731	0.02	0.30	479	0.03	0.34
JTSA	327	0.08	0.32	356	0.07	0.41
LDA	334	0.06	0.57	247	0.09	0.59
Mtka	297	0.07	0.56	264	0.08	0.59
SJS	493	0.02	0.50	562	0.02	0.60
StataJ	316	0.24	0.36	977	0.08	0.11
StNee	325	0.01	0.24	191	0.02	0.31
StSin	1070	0.04	0.57	935	0.05	0.54
SMMR	746	0.04	0.33	813	0.03	0.18
StMod	275	0.03	0.41	237	0.03	0.35
StPap	518	0.03	0.35	193	0.08	0.42
StSci	1454	0.03	0.29	924	0.05	0.35
Stat	311	0.02	0.47	254	0.02	0.43
StCmp	575	0.04	0.46	710	0.03	0.24
SPL	1828	0.08	0.36	1348	0.11	0.46
StMed	4022	0.16	0.42	6602	0.10	0.24
Tech	494	0.09	0.37	688	0.06	0.38
Test	498	0.01	0.61	243	0.03	0.54

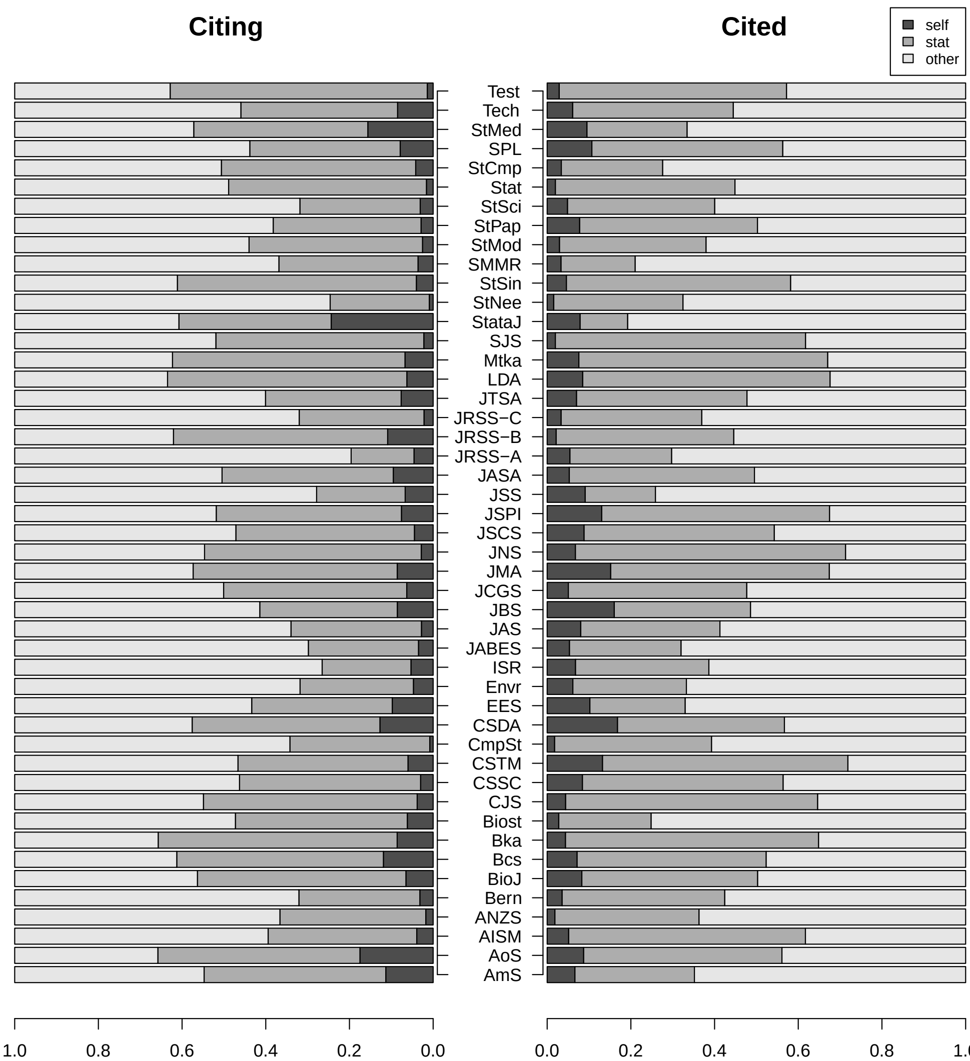

Fig. 1. Bar plots of citations made (`Citing`, left panel) and received (`Cited`, right panel) for the selected Statistics journals, as listed in Table 2, based on Journal Citation Reports® 2010. For each journal, the bar displays the proportion of self-citations (dark grey), the proportion of citations made/received to/from other Statistics journals in the list (mid grey), and to/from journals not in the list (light grey). Journal acronyms are in Table 1.

allow construction of the cross-citation table $\mathbf{C} = [c_{ij}]$, where c_{ij} is the number of citations from articles published in journal j in 2010 to papers published in journal i in the chosen time window ($i = 1, \ldots, n$). In our analyses, $n = 47$, the number of selected Statistics journals, and the time window is the previous ten years. Following Stigler (1994) and Stigler *et al.* (1995), summarizing the information contained in the cross-citation table will be the focus of the rest of the paper.

3. Clustering journals

Statistics journals have different stated objectives, and different types of content. Some journals emphasize applications and modelling, while others focus on theoretical and mathematical developments, or deal with computational and algorithmic aspects of statistical analysis. Applied journals are often targeted to particular areas, such as, for example, Statistics for medical applications, or for environmental sciences.

It is quite natural to wonder whether the cross-citations table $\mathbf{C}$ allows the identification of groups of journals with similar aims and types of content. To this end, the first step is to construct a measure of distance between journals. Consider the total number t_{ij} of citations exchanged between journals i and j,

$$t_{ij} = \begin{cases} c_{ij} + c_{ji}, & \text{for } i \neq j \\ c_{ii}, & \text{for } i = j. \end{cases}$$

The distance between two journals can be measured by $d_{ij} = 1 - \rho_{ij}$, where ρ_{ij} is the Pearson correlation coefficient of variables t_{ik} and t_{jk} $(k = 1, \ldots, n)$, i.e.,

$$\rho_{ij} = \frac{\sum_{k=1}^{n} (t_{ik} - \bar{t}_i)(t_{jk} - \bar{t}_j)}{\sqrt{\sum_{k=1}^{n} (t_{ik} - \bar{t}_i)^2 \sum_{k=1}^{n} (t_{jk} - \bar{t}_j)^2}},$$

with $\bar{t}_i = \sum_{k=1}^{n} t_{ik}/n$. Among the many available clustering algorithms, we consider a hierarchical agglomerative cluster analysis with complete linkage (Kaufman and Rousseeuw, 1990). The clustering process is visualized through the dendrogram in Figure 2. Visual inspection of the dendrogram suggests to cut it at height 0.6, thus obtaining eight clusters, two of which are singletons. The identified clusters are grouped in grey boxes in Figure 2 and listed in Table 3.

We comment first on the groups and later on the singletons, following the order of the journals in the dendrogram from left to right. The first group includes a large number of general journals concerned with theory and methods of Statistics, but also with applications. Among others, the group includes *Journal of Time Series Analysis*, *Journal of Statistical Planning and Inference*, and *Annals of the Institute of Statistical Mathematics*.

The second group contains the leading journals in the development of statistical theory and methods: *Annals of Statistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society* Series B and *Biometrika*. The group includes also other methodological journals such as *Bernoulli*, *Scandinavian Journal of Statistics* and *Statistica Sinica*. It is possible to identify some natural subgroups: *Journal of Computational and Graphical Statistics* and *Statistics and Computing*; *Biometrika*, *Journal of the Royal Statistical Society* Series B, and *Journal of the American Statistical Association*; *Annals of Statistics* and *Statistica Sinica*.

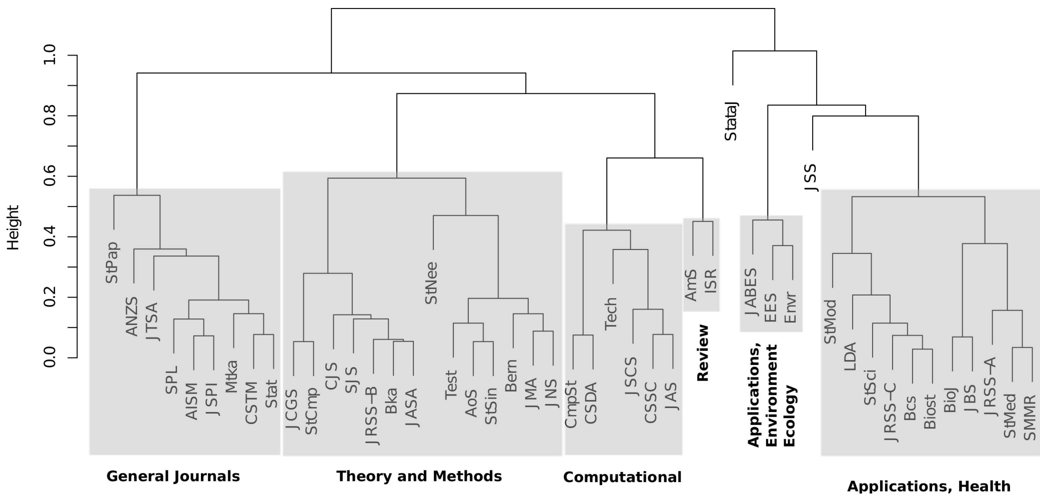

Fig. 2. Dendrogram of complete linkage hierarchical cluster analysis. Journals acronyms are in Table 1. Clusters obtained by cutting the dendrogram at height 0.6 are identified by grey boxes, see also Table 3.

The third group comprises journals mostly dealing with computational aspects of Statistics, such as *Computational Statistics and Data Analysis*, *Communications in Statistics – Simulation and Computation*, *Computational Statistics*, and *Journal of Statistical Computation and Simulation*. Other members of the group with a less direct orientation towards computational methods are *Technometrics* and *Journal of Applied Statistics*.

The fourth group includes just two journals both of which publish mainly review articles, namely *American Statistician* and *International Statistical Review*.

The fifith group comprises the three journals specializing in ecological and environmental applications: *Journal of Agricultural, Biological and Environmental Statistics*, *Environmental and Ecological Statistics* and *Environmetrics*.

The last group includes various journals emphasising applications, especially to health sciences and similar areas. It encompasses journals oriented towards biological and medical applications such as *Biometrics* and *Statistics in Medicine*, and also journals publishing papers about more general statistical applications, such as *Journal of the Royal Statistical Society* Series A and Series C. The review journal *Statistical Science* also falls into this group; it is not grouped together with the other two review journals already mentioned. Within the group there are some natural sub-groupings: *Statistics in Medicine* with *Statistical Methods in Medical Research*; and *Biometrics* with *Biostatistics*.

Finally, and perhaps not surprisingly, the two singletons are the software-oriented *Journal of Statistical Software* and *Stata Journal*. The latter is, by some distance, the most remote journal in the list according to the measure of distance used here.

4. Ranking journals

4.1. *Impact Factor, Eigenfactor and Article Influence Score*

The Journal Citation Reports® annually publishes various rating indices, the best known being the already discussed Impact Factor. Thomson Reuters also publishes the *Immediacy*

Table 3. Clusters of Statistics journals. Journals and groups are listed in the same order as in the dendrogram plotted in Figure 2.

Journals	Characteristics
Statistical Papers	General Journals
Australian and New Zealand Journal of Statistics	
Journal of Time Series Analysis	
Statistics and Probability Letters	
Annals of the Institute of Statistical Mathematics	
Journal of Statistical Planning and Inference	
Metrika	
Communications in Statistics - Theory and Methods	
Statistics	
Journal of Computational and Graphical Statistics	Theory and Methods
Statistics and Computing	
Canadian Journal of Statistics	
Scandinavian Journal of Statistics	
Journal of the Royal Statistical Society Series B	
Biometrika	
Journal of the American Statistical Association	
Statistica Neerlandica	
Test	
Annals of Statistics	
Statistica Sinica	
Bernoulli	
Journal of Multivariate Analysis	
Journal of Nonparametric Statistics	
Computational Statistics	Computational
Computational Statistics and Data Analysis	
Technometrics	
Journal of Statistical Computation and Simulation	
Communications in Statistics - Simulation and Computation	
Journal of Applied Statistics	
American Statistician	Review
International Statistical Review	
Stata Journal	
Journal of Agricultural, Biological and Environmental Statistics	Applications Environment/ Ecology
Environmental and Ecological Statistics	
Environmetrics	
Journal of Statistical Software	
Statistical Modelling	Applications, Health
Lifetime Data Analysis	
Statistical Science	
Journal of the Royal Statistical Society Series C	
Biometrics	
Biostatistics	
Biometrical Journal	
Journal of Biopharmaceutical Statistics	
Journal of the Royal Statistical Society Series A	
Statistics in Medicine	
Statistical Methods in Medical Research	

Index, which describes the average number of times an article is cited in the year of its publication. The Immediacy Index is very unsuitable for evaluating Statistics journals, but it could be worthy of attention in fields where citations occur very quickly, for example some areas of life sciences.

It is well known in the bibliometric literature that the calculation of the Impact Factor contains some important inconsistencies (Glänzel and Moed, 2002). The numerator of the Impact Factor includes citations to all items, while the number of citable items in the denominator excludes letters to the editor and editorials; such letters are an important element of some journals, notably medical journals. The inclusion of self citations, defined as citations from a journal to articles in the same journal, exposes the Impact Factor to possible manipulation by editors. Indeed, Sevinc (2004) and Frandsen (2007) report instances in which authors were asked to add irrelevant references to their articles, presumably with the aim of increasing the Impact Factor of the journal. Journal self-citations can also be a consequence of authors' preferring to cite papers published in the same journal instead of equally relevant papers published elsewhere, particularly if they perceive such self-citation as likely to be welcomed by the journal's editors. Nevertheless, the potential for such behaviour should not lead to the conclusion that self-citations are always unfair. Many self-citations are likely to be genuine, especially since scholars often select a journal for submission of their work according to the presence of previously published papers on related topics.

The *Eigenfactor Score*® and the derived *Article Influence Score*® (Bergstrom, 2007; West, 2010) have been proposed to overcome the limitations of the Impact Factor. Both the Eigenfactor® and the Article Influence Score® are computed using a five-year time period and removing self-citations in order to eliminate possible sources of manipulation. The idea underlying the Eigenfactor Score® is that the importance of a journal relates to the time spent by scholars in reading that journal. As stated by Bergstrom (2007), it is possible to imagine that a scholar starts reading an article selected at random. Then, the scholar randomly selects another article from the references of the first paper and reads it. Afterwards, a further article is selected at random from the references included in the previous one and the process may go on *ad infinitum*. In such a process, the time spent in reading a journal might reasonably be regarded as an indicator of that journal's importance.

Apart from modifications needed to account for special cases such as journals that do not cite any other journal, the Eigenfactor® algorithm is summarized as follows. The Eigenfactor® is computed from the normalized citation matrix $\tilde{\mathbf{C}} = [\tilde{c}_{ij}]$, whose elements are the citations c_{ij} from journal j to articles published in the previous five years in journal i divided by the total number of references in j in those years, $\tilde{c}_{ij} = c_{ij}/\sum_{i=1}^{n} c_{ij}$. The diagonal elements of $\tilde{\mathbf{C}}$ are set to zero, to discard self-citations. A further ingredient of the Eigenfactor® is the vector of the normalized number of articles $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$, with a_i being the number of articles published by journal i during the five-year period divided by the number of articles published by all considered journals. Let $\boldsymbol{e}^{\mathrm{T}}$ be the row vector of ones, so that $\boldsymbol{a}\boldsymbol{e}^{\mathrm{T}}$ is a matrix with all identical columns $\boldsymbol{a}$. Then

$$\mathbf{P} = \lambda\tilde{\mathbf{C}} + (1-\lambda)\boldsymbol{a}\boldsymbol{e}^{\mathrm{T}}$$

is the transition matrix of a Markov process that assigns probability λ to a random movement in the journal citation network, and probability $1-\lambda$ to a random jump to any journal; for jumps of the latter kind, destination journal attractiveness is proportional to size.

The damping parameter λ is set to 0.85, just as in the *PageRank*® algorithm at the basis of the Google search engine; see Brin and Page (1998). The leading eigenvector $\boldsymbol{\psi}$ of $\mathbf{P}$

corresponds to the steady-state fraction of time spent reading each journal. The Eigenfactor Score® EF_i for journal i is defined as 'the percentage of the total weighted citations that journal i receives'; that is,

$$\mathrm{EF}_i = 100\frac{[\tilde{\mathbf{C}}\boldsymbol{\psi}]_i}{\sum_{i=1}^n [\tilde{\mathbf{C}}\boldsymbol{\psi}]_i}, \quad i = 1, \ldots, n,$$

where $[\mathbf{x}]_i$ denotes the ith element of vector $\mathbf{x}$. See `http://www.eigenfactor.org/methods.pdf` for more details about the methodology behind the Eigenfactor® algorithm.

The Eigenfactor® 'measures the total influence of a journal on the scholarly literature' (Bergstrom, 2007) and thus it depends on the number of articles published by a journal. The Article Influence Score® AI_i of journal i is instead a measure of the per-article citation influence of the journal obtained by normalizing the Eigenfactor® as follows:

$$\mathrm{AI}_i = 0.01\frac{\mathrm{EF}_i}{a_i}, \quad i = 1, \ldots, n.$$

The rankings of the selected Statistics journals according to Impact Factor, five-year Impact Factor, Immediacy Index, Eigenfactor®, and Article Influence Score® are reported in columns two to six of Table 4.

The substantial variation among those five rankings is the first aspect that leaps to the eye; those different published measures clearly do not yield a common, unambiguous picture of the journals' relative standings.

A diffuse opinion within the statistical community is that the four most prestigious Statistics journals are (in alphabetic order) *Annals of Statistics*, *Biometrika*, *Journal of the American Statistical Association*, and *Journal of the Royal Statistical Society* Series B. See, for example, the survey about how statisticians perceive Statistics journals described in Theoharakis and Skordia (2003). Accordingly, a minimal requirement for a ranking of acceptable quality is that the four most prestigious journals should occupy prominent positions. Following this criterion, the least satisfactory ranking is, as expected, the one based on the Immediacy Index, which ranks the *Journal of the American Statistical Association* only 22nd and *Biometrika* just a few positions ahead at 19th.

Impact Factors computed at two and five years are highly correlated with one another. Indeed, the Kendall τ rank correlation of the rankings based on these two indices is equal to 0.81 with 95% confidence interval $(0.73, 0.89)$; see Table 5. In both the versions of Impact Factor ranking, *Journal of the Royal Statistical Society* Series B occupies first position, *Annals of Statistics* is second according to the two-year and sixth according to the five-year Impact Factor, *Journal of the American Statistical Association* is eighth and fourth, *Biometrika* only eleventh and tenth, respectively. The two software journals have excellent Impact Factors. *Journal of Statistical Software* is ranked fourth according to the two-year Impact Factor and second according to the five-year Impact Factor, while *Stata Journal* is ninth and seventh using the two- and five-year versions, respectively. Other journals ranked highly according to the Impact Factor measures are *Biostatistics* and *Statistical Science*.

The Eigenfactor® performs somehow better than the Impact Factor in ranking the four top journals, with *Journal of the American Statistical Association*, *Annals of Statistics*, *Journal of the Royal Statistical Society* Series B, and *Biometrika* ranked at positions 1, 3, 5 and 7, respectively. The Eigenfactor® rewards journals that publish many papers per year, with *Statistics in Medicine* ranked 2nd, *Computational Statistics and Data Analysis* 4th, *Journal of Statistical Planning and Inference* 8th, *Journal of Multivariate Analysis*

Table 4. Rankings of selected Statistics journals based on Journal Citation Reports® edition 2010. Columns correspond to Impact Factor `IF`, five-year Impact Factor `IF5`, Immediacy Index `II`, Eigenfactor® `EF`, Article Influence® `AI`, and the Stigler model `SM`. Braces indicate groups identified by ranking lasso with BIC selection. Journal acronyms are in Table 1.

Pos.	IF	IF5	II	EF	AI	SM
1	JRSS-B	JRSS-B	JSS	JASA	JRSS-B	JRSS-B
2	AoS	JSS	Biost	StMed	StSci	AoS
3	Biost	StSci	SMMR	AoS	JASA	Bka
4	JSS	JASA	StCmp	CSDA	AoS	JASA
5	JRSS-A	Biost	AoS	JRSS-B	Bka	Bcs
6	StSci	AoS	EES	Bcs	Biost	JRSS-A
7	StMed	StataJ	JRSS-B	Bka	StataJ	Bern
8	JASA	SMMR	JCGS	JSPI	StCmp	SJS
9	StataJ	JRSS-A	StMed	Biost	JRSS-A	Biost
10	StCmp	Bka	BioJ	JMA	JSS	JCGS
11	Bka	StCmp	CSDA	SPL	Bcs	Tech
12	SMMR	StMed	StSci	Bern	Bern	AmS
13	Bcs	Bcs	JRSS-A	StSci	JCGS	JTSA
14	EES	Tech	StSin	JCGS	SMMR	ISR
15	Tech	JCGS	JBS	StSin	Tech	AISM
16	BioJ	EES	StataJ	JRSS-A	SJS	CJS
17	JCGS	CSDA	Bcs	JSS	StMed	StSin
18	CSDA	SJS	Envr	StataJ	Test	StSci
19	JBS	AmS	Bka	StCmp	CJS	LDA
20	Test	JBS	JMA	SJS	StSin	JRSS-C
21	JMA	Bern	Tech	BioJ	JRSS-C	StMed
22	Bern	JRSS-C	JASA	Tech	AmS	ANZS
23	AmS	BioJ	JRSS-C	CSTM	JMA	StCmp
24	AISM	JABES	ISR	SMMR	EES	StataJ
25	StSin	JMA	JNS	CJS	JTSA	SPL
26	LDA	CJS	Test	AmS	LDA	StNee
27	ISR	Test	Bern	JBS	BioJ	Envr
28	SJS	StMod	JABES	JRSS-C	StMod	JABES
29	Envr	StSin	JSPI	JTSA	CSDA	Mtka
30	JABES	LDA	SJS	Envr	JABES	StMod
31	StMod	Envr	AmS	JSCS	AISM	JSPI
32	JSPI	JTSA	AISM	AISM	ANZS	SMMR
33	CJS	ISR	StMod	Test	ISR	BioJ
34	JTSA	ANZS	Mtka	CSSC	JSPI	JMA
35	JRSS-C	JSPI	StNee	LDA	Envr	EES
36	ANZS	AISM	StPap	EES	JBS	CSDA
37	StPap	Stat	SPL	JAS	StNee	JNS
38	Mtka	Mtka	ANZS	Mtka	CmpSt	CmpSt
39	Stat	CmpSt	LDA	JNS	JNS	Stat
40	CmpSt	StNee	JTSA	JABES	Stat	Test
41	JSCS	JSCS	JSCS	ANZS	Mtka	CSTM
42	JNS	StPap	CJS	CmpSt	JSCS	JSS
43	SPL	SPL	CmpSt	Stat	StPap	JBS
44	CSTM	JNS	CSTM	StPap	SPL	JSCS
45	CSSC	JAS	Stat	ISR	CSTM	CSSC
46	StNee	CSTM	JAS	StNee	CSSC	StPap
47	JAS	CSSC	CSSC	StMod	JAS	JAS

Table 5. Kendall τ rank correlations with 95% confidence intervals between Impact Factor `IF`, five-years Impact Factor `IF5`, Immediacy Index `II`, Eigenfactor® `EF`, Article Influence® `AI`, number of articles `ART`, and the ranking based on the Stigler model `SM`. Computations are based on Journal Citation Reports® edition 2010.

	IF	IF5	II	EF	AI	ART
IF5	0.81					
	(0.73, 0.89)					
II	0.68	0.63				
	(0.59, 0.76)	(0.52, 0.74)				
EF	0.48	0.47	0.39			
	(0.35, 0.62)	(0.33, 0.62)	(0.26, 0.52)			
AI	0.72	0.79	0.52	0.49		
	(0.61, 0.82)	(0.70, 0.88)	(0.39, 0.66)	(0.34, 0.64)		
ART	-0.07	-0.11	-0.04	0.36	-0.14	
	(-0.28, 0.15)	(-0.32, 0.11)	(-0.23, 0.16)	(0.15, 0.56)	(-0.35 0.07)	
SM	0.40	0.42	0.23	0.36	0.55	-0.10
	(0.21, 0.59)	(0.22, 0.61)	(0.02, 0.44)	(0.19, 0.54)	(0.38, 0.72)	(-0.30, 0.11)

10th, and *Statistics and Probability Letters* 11th. In fact, the Kendall τ rank correlation between the Eigenfactor® and the number of articles is 0.36 with 95% confidence interval $(0.15, 0.56)$, while all of the other ranking indices published in Journal Citation Reports® have statistically non-significant rank correlation with the number of articles; see Table 5.

Among the indices produced by Thomson Reuters, the Article Influence Score® yields the most satisfactory ranking with respect to the four leading journals, which stand within the first five positions. The normalization used to construct the AI index from the Eigenfactor® is effective in removing the influence of the number of articles, as summarized by Kendall τ falling to the value -0.14 with 95% confidence interval $(-0.35, 0.07)$.

All of the indices discussed in this section are constructed by using the complete Web of Science® database, thus counting citations among Statistics journals as well as citations from journals in other fields.

4.2. The Stigler model

Stigler (1994) considers the export of intellectual influence from a journal in order to determine its importance. The export of influence is measured by the citations received by the journal. Stigler assumes that the log-odds that journal i exports to journal j rather than vice-versa is equal to the difference of the *export scores*

$$\text{log-odds}\,(\text{journal } i \text{ is cited by journal } j) = \mu_i - \mu_j, \tag{1}$$

where μ_i is the export score of journal i. With Stephen Stigler's words 'the larger the export score, the greater the propensity to export intellectual influence'.

The Stigler model (1) has some attractive features:

(a) *Journal self-citations are not counted.* In contrast to the Impact Factor, the rankings based on journal export scores μ_i are not affected by the risk of manipulation through journal self-citations;

(b) *Only citations between journals under comparison are counted.* If the Stigler model is applied to the list of 47 Statistics journals, then only citations among these journals

Table 6. Characteristics of journal rankings derived from Journal Citation Reports®.

	IF	IF5	II	EF	AI	SM
Omission of journal self-citations	no	no	no	yes	yes	yes
Exclusion of external citations	no	no	no	no	no	yes
Journal size correction	yes	yes	yes	no	yes	yes

are counted. Such an application of the Stigler model thus aims unambiguously to measure influence within the research field of Statistics, rather than combining that with potential influence on other research fields.

(c) *The size of the journals is not important.* Rankings based on the Stigler model are not affected by the number of papers published. As shown by Stigler (1994, pg. 102), if two journals are merged into a single journal then the odds in favour of that 'super' journal against any third journal is a weighted average of the odds for the two separate journals against the third one.

As summarized in Table 6, none of the ranking indices published by Thomson Reuters has all of the three features above.

The Stigler model is an example of the Bradley-Terry model (Bradley and Terry, 1952; Agresti, 2002) for paired comparison data. Maximum likelihood estimation of the vector of journal export scores $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ can be obtained through standard software for fitting generalized linear models. Alternatively, specialized software such as the `R` package `BradleyTerry2` (Turner and Firth, 2012) is available through the CRAN repository at URL `http://cran.r-project.org/web/packages/BradleyTerry2`. Since the Stigler model is specified through pairwise differences of export scores $\mu_i - \mu_j$, model identification requires a constraint, such as, for example, a 'reference journal' constraint $\mu_1 = 0$, or the sum constraint $\sum_{i=1}^{n} \mu_i = 0$. Without loss of generality we use the latter constraint in what follows.

Standard maximum likelihood estimation of the Stigler model assumes that citation counts c_{ij} are realizations of independent binomial variables C_{ij}. The assumption could be questionable, as the number of citations from journal j to journal i is not independent of the number of citations from journal j to another journal k. See Cattelan (2012) for a general discussion on handling dependence in paired comparison modelling. The presence of dependence between citation counts may lead to the well-known phenomenon of overdispersion. A simple way to deal with overdispersion is provided by the method of quasi-likelihood (Wedderburn, 1974). Accordingly, we consider the 'quasi-Stigler' model

$$\mathrm{E}(C_{ij}) = t_{ij}\pi_{ij} \quad \text{and} \quad \mathrm{var}(C_{ij}) = \phi\, t_{ij}\pi_{ij}(1-\pi_{ij}),$$

where $\pi_{ij} = \exp(\mu_i - \mu_j)/\left\{1 + \exp(\mu_i - \mu_j)\right\}$ and $\phi > 0$ is the dispersion parameter. The estimate of the dispersion parameter obtained here, for the model applied to Statistics journal cross-citations between 2001 and 2010, is $\hat{\phi} = 1.76$, thus suggesting presence of overdispersion. The quasi-likelihood estimated export scores of the Statistics journals are reported in Table 7.

4.2.1. Estimation uncertainty

Estimation uncertainty is unexplored in relation to the various published journal rankings. Despite this lacuna, many academics produce vibrant critiques of 'statistical citation analyses', although such analyses are actually rather non-statistical. A key advantage of the Stigler model over other ranking methods is a straightforward quantification of the uncertainty in journal export scores.

Since the Stigler model is identified through pairwise differences, uncertainty quantification requires the complete variance matrix of $\hat{\boldsymbol{\mu}}$. Routine reporting of such a large variance matrix is impracticable for space reasons. A solution is provided through the presentational device of quasi-variances (Firth and de Menezes, 2005), constructed in such a way as to allow approximate calculation of the variance of the differences $\text{var}(\hat{\mu}_i - \hat{\mu}_j)$ as if $\hat{\mu}_i$ and $\hat{\mu}_j$ were independent:

$$\text{var}\,(\hat{\mu}_i - \hat{\mu}_j) \simeq \text{qvar}_i + \text{qvar}_j, \quad \text{for all choices of } i \text{ and } j.$$

Reporting the estimated export scores with their quasi-variances, then, is an economical way to allow approximate inference on the significance of the difference between any two journals' export scores. The quasi-variances are computed by minimizing a suitable penalty function of the distance between the true variances, $\text{var}\,(\hat{\mu}_i - \hat{\mu}_j)$, and their quasi-variance representation $\text{qvar}_i + \text{qvar}_j$. See Firth and de Menezes (2005) for details.

Table 7 reports the estimated journal export scores computed under the sum constraint $\sum_{i=1}^{n} \mu_i = 0$ and the corresponding quasi-standard errors, defined as the square root of the quasi-variances. Quasi-variances are calculated using the `R` (R Core Team, 2013) package `qvcalc` (Firth, 2012) available through the CRAN repository at URL `http://cran.r-project.org/web/packages/qvcalc`. For illustration, consider testing whether the export score of *Biometrika* is significantly different from that of the *Journal of the American Statistical Association*. The z test statistic as approximated through the quasi-variances is

$$z \simeq \frac{\hat{\mu}_{\text{Bka}} - \hat{\mu}_{\text{JASA}}}{\sqrt{\text{qvar}_{\text{Bka}} + \text{qvar}_{\text{JASA}}}} = \frac{1.29 - 1.26}{\sqrt{0.08^2 + 0.06^2}} = 0.30.$$

The 'usual' variances for those two export scores in the sum-constrained parameterization are respectively 0.0376 and 0.0344, and the covariance is 0.0312; thus the 'exact' value of the z statistic in this example is

$$z = \frac{1.29 - 1.26}{\sqrt{0.0376 - 2\,(0.0312) + 0.0344}} = 0.31,$$

so the approximation based upon quasi-variances is quite accurate. The z statistic suggests that there is insufficient evidence to rule out the possibility that *Biometrika* and *Journal of the American Statistical Association* have the same ability to 'export intellectual influence' within the 47 Statistics journals in the list.

4.2.2. Results

We proceed now with the interpretation of the ranking based on the Stigler model. It is reassuring that the four leading Statistics journals are ranked in the first four positions. *Journal of the Royal Statistical Society* Series B is ranked first with a remarkably larger export score than the second ranked journal, *Annals of Statistics*: the approximate z statistic for the significance of the difference of their export scores is 5.44. The third position is occupied by *Biometrika*, closely followed by *Journal of the American Statistical Association*.

The fifth-ranked journal is *Biometrics*, followed by *Journal of the Royal Statistical Society* Series A, *Bernoulli*, *Scandinavian Journal of Statistics*, *Biostatistics*, *Journal of Graphical and Computational Statistics*, and *Technometrics*.

Table 7. Journal ranking based on the Stigler model using data from Journal Citation Reports® edition 2010. Columns correspond to quasi-likelihood estimates (`Quasi`), quasi-standard errors (`QSE`), ranking lasso estimates with shrinkage parameter selected by Akaike information criterion (`AIC`) and Schwartz information criterion (`BIC`).

				Lasso	
Pos.	Journal	Quasi	QSE	AIC	BIC
1	Journal of the Royal Statistical Society Series B	2.09	0.11	1.95	1.87
2	Annals of Statistics	1.38	0.07	1.24	1.17
3	Biometrika	1.29	0.08	1.16	1.11
4	Journal of the American Statistical Association	1.26	0.06	1.16	1.11
5	Biometrics	0.85	0.07	0.72	0.65
6	Journal of the Royal Statistical Society Series A	0.70	0.19	0.44	0.31
7	Bernoulli	0.69	0.15	0.44	0.31
8	Scandinavian Journal of Statistics	0.66	0.12	0.44	0.31
9	Biostatistics	0.66	0.11	0.44	0.31
10	Journal of Computational and Graphical Statistics	0.64	0.12	0.44	0.31
11	Technometrics	0.53	0.15	0.40	0.31
12	American Statistician	0.40	0.18	0.14	0.04
13	Journal of Time Series Analysis	0.37	0.20	0.14	0.04
14	International Statistical Review	0.33	0.25	0.14	0.04
15	Annals of the Institute of Statistical Mathematics	0.32	0.16	0.14	0.04
16	Canadian Journal of Statistics	0.30	0.14	0.14	0.04
17	Statistica Sinica	0.29	0.09	0.14	0.04
18	Statistical Science	0.11	0.11	-0.02	-0.04
19	Lifetime Data Analysis	0.10	0.17	-0.02	-0.04
20	Journal of the Royal Statistical Society Series C	0.09	0.15	-0.02	-0.04
21	Statistics in Medicine	0.06	0.07	-0.02	-0.04
22	Australian and New Zealand Journal of Statistics	0.06	0.21	-0.02	-0.04
23	Statistics and Computing	0.04	0.15	-0.02	-0.04
24	Stata Journal	0.02	0.33	-0.02	-0.04
25	Statistics and Probability Letters	-0.09	0.09	-0.04	-0.04
26	Statistica Neerlandica	-0.10	0.25	-0.04	-0.04
27	Environmetrics	-0.11	0.18	-0.04	-0.04
28	Journal of Agricultural Biological and Environmental Statistics	-0.16	0.23	-0.04	-0.04
29	Metrika	-0.18	0.17	-0.04	-0.04
30	Statistical Modelling	-0.22	0.21	-0.04	-0.04
31	Journal of Statistical Planning and Inference	-0.33	0.07	-0.32	-0.31
32	Statistical Methods in Medical Research	-0.35	0.16	-0.32	-0.31
33	Biometrical Journal	-0.40	0.12	-0.35	-0.31
34	Journal of Multivariate Analysis	-0.45	0.08	-0.42	-0.36
35	Environmental and Ecological Statistics	-0.48	0.25	-0.42	-0.36
36	Computational Statistics and Data Analysis	-0.52	0.07	-0.42	-0.36
37	Journal of Nonparametric Statistics	-0.53	0.15	-0.42	-0.36
38	Computational Statistics	-0.64	0.22	-0.42	-0.36
39	Statistics	-0.65	0.18	-0.42	-0.36
40	Test	-0.70	0.15	-0.45	-0.36
41	Communications in Statistics - Theory and Methods	-0.74	0.10	-0.53	-0.36
42	Journal of Statistical Software	-0.80	0.19	-0.53	-0.36
43	Journal of Biopharmaceutical Statistics	-0.83	0.16	-0.53	-0.36
44	Journal of Statistical Computation and Simulation	-0.92	0.15	-0.55	-0.36
45	Communications in Statistics - Simulation and Computation	-1.26	0.14	-1.04	-0.88
46	Statistical Papers	-1.35	0.20	-1.04	-0.88
47	Journal of Applied Statistics	-1.41	0.15	-1.08	-0.88

The 'centipede' plot in Figure 3 visualizes the estimated export scores along with the 95% comparison intervals with limits $\hat{\mu}_i \pm 1.96\,\mathrm{qse}(\hat{\mu}_i)$, where qse indicates the quasi-standard error. The centipede plot highlights the outstanding position of *Journal of the Royal Statistical Society* Series B, and indeed of the four top journals whose quasi-confidence intervals are well separated from those of the remaining journals. However, the most striking general feature is the substantial uncertainty in most of the estimated journal scores. Many of the small differences that appear among the estimated export scores are not statistically significant.

The Kendall τ rank correlation between the ranking based on Stigler model and the number of papers in a journal is $\tau = -0.10$ with 95% confidence interval $(-0.30, 0.11)$, thus confirming that the analysis is unrelated to the size of the journals. Among the rankings published by Thomson Reuters, the one most strongly correlated with the Stigler-model ranking is that provided by the Article Influence Score®, $\tau = 0.55$ with 95% confidence interval $(0.38, 0.72)$.

5. Ranking in groups with lasso

It is well known that shrinkage estimation offers notable improvement over standard maximum likelihood estimation when the target is simultaneous inference on a vector of mean parameters. See, for example, Morris (1983). It seems natural to consider shrinkage estimation also for the Stigler model. Masarotto and Varin (2012) fit Bradley-Terry models with a lasso-type penalty (Tibshirani, 1996) that, in our application here, forces journals with close export scores to be estimated at the same level. The method, termed ranking lasso, has the twofold advantage of shrinkage and enhanced interpretation, because it avoids over-interpretation of small differences between estimated journal export scores.

For a given value of a bound parameter $s \geq 0$, the ranking lasso method fits the Stigler model by maximizing the log-likelihood

$$\ell(\boldsymbol{\mu}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} [c_{ij}(\mu_i - \mu_j) - t_{ij} \ln\{1 + \exp(\mu_i - \mu_j)\}]$$

with an L_1 penalty on all the pairwise differences of export scores,

$$\hat{\boldsymbol{\mu}}^{(s)} = \operatorname*{argmax}_{\boldsymbol{\mu} \in \mathbb{R}^n} \ell(\boldsymbol{\mu}), \quad \text{subject to} \quad \sum_{i<j}^{n} w_{ij}|\mu_i - \mu_j| \leq s \quad \text{and} \quad \sum_{i=1}^{n} \mu_i = 0, \tag{2}$$

where w_{ij} are data-dependent weights discussed below. Maximum likelihood estimation is obtained for a sufficiently large value of the bound s. As s decreases to zero, the L_1 penalty causes journal export scores that differ little to be estimated at the same value, thus producing a ranking in groups. The ranking lasso method can be interpreted as a generalized version of the fused lasso (Tibshirani *et al.*, 2005).

Many authors (e.g., Fan and Li, 2001; Zou, 2006) have observed that lasso-type penalties may be too severe, thus yielding inconsistent estimates of the non-zero effects. In the ranking lasso context, this means that if the weights w_{ij} in (2) are all identical, then the pairwise differences $\mu_i - \mu_j$ whose 'true' value is non-zero might not be consistently estimated. Among various possibilities, an effective way to overcome the drawback is to resort to the adaptive lasso method (Zou, 2006), which imposes a heavier penalty on small effects. Accordingly,

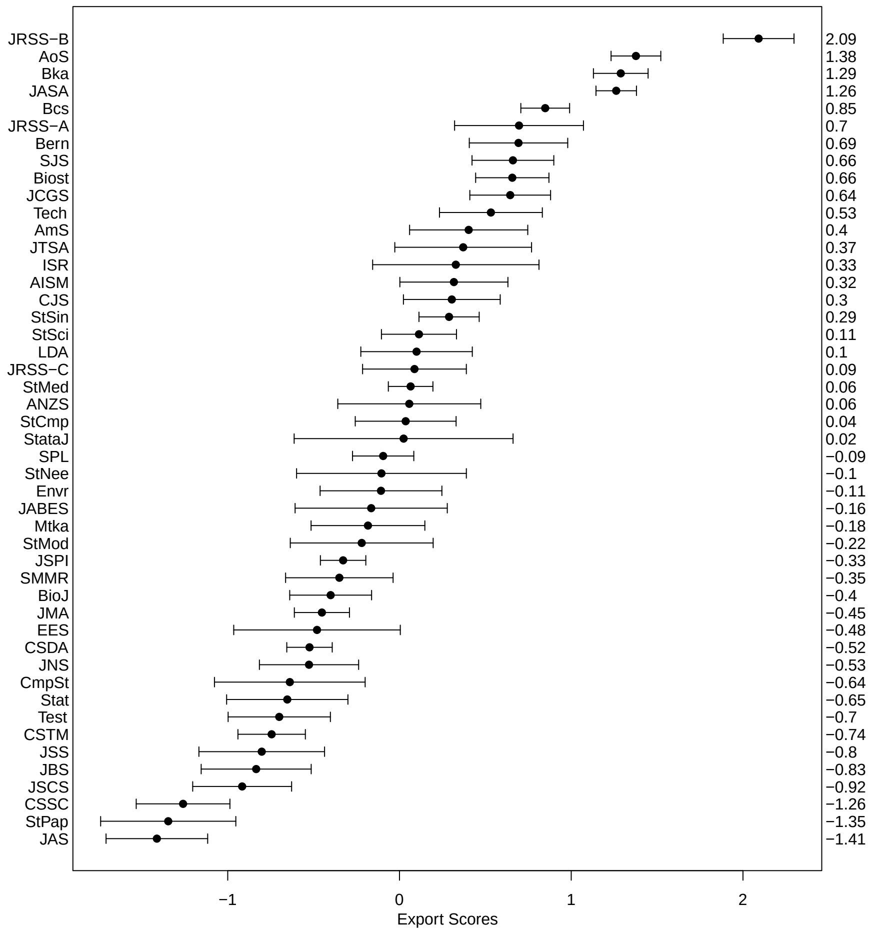

Fig. 3. Centipede plot of 95% comparison intervals of the estimated journal export scores based on Journal Citation Reports® edition 2010. The error-bar limits are $\hat{\mu}_i \pm 1.96$ qse$(\hat{\mu}_i)$, with the estimated export scores $\hat{\mu}_i$ marked by solid circles. Journal abbreviations are as in Table 1.

the adaptive ranking lasso employs weights equal to the reciprocal of a consistent estimate of $\mu_i - \mu_j$, such as

$$w_{ij} = |\hat{\mu}_i^{(\text{mle})} - \hat{\mu}_j^{(\text{mle})}|^{-1},$$

with $\hat{\mu}_i^{(\text{mle})}$ being the maximum likelihood estimate of the export score for journal i. Masarotto and Varin (2012) compute estimates $\hat{\boldsymbol{\mu}}(s)$ of the adaptive ranking lasso by using an augmented Lagrangian algorithm (Nocedal and Wright, 2006) for a sequence of bounds s ranging from complete shrinkage ($s = 0$) — *i.e.* all journals have the same estimated export score — to the maximum likelihood solution. The optimal value for s can be chosen by minimization of information criteria, such as the Akaike information criterion

$$\text{AIC}(s) = -2\,\ell(\hat{\boldsymbol{\mu}}^{(s)}) + 2\,\text{enp}(s),$$

or the Schwartz information criterion

$$\text{BIC}(s) = -2\,\ell(\hat{\boldsymbol{\mu}}^{(s)}) + \ln(n)\,\text{enp}(s),$$

where the effective number of parameters enp(s) is estimated as the number of distinct groups formed with bound s.

Figure 4 displays the path plot of the ranking lasso, while Table 7 reports the estimated export scores corresponding to the solutions identified by AIC and BIC. See also Table 4 for a comparison with the Thompson Reuters rankings. The path plot of Figure 4 visualizes how the estimates of the export scores vary as the degree of shrinkage increases, *i.e.*, as bound s decreases. The plot confirms the outstanding position of *Journal of the Royal Statistical Society* Series B, the leader in the ranking at any level of shrinkage. Also *Annals of Statistics* keeps the second position for more than half of the path before joining the paths of *Biometrika* and *Journal of the American Statistical Association*. *Biometrics* is solitary in the fifth position for almost the whole of its path. AIC identifies a total of 17 groups, while BIC supports a sparser solution with only 10 groups. According to AIC, the five top journals are followed by a group of five further journals, namely *Journal of the Royal Statistical Society* Series A, *Bernoulli*, *Scandinavian Journal of Statistics*, *Biostatistics*, and *Journal of Computational and Graphical Statistics*. The use of BIC would include also *Technometrics* in the latter group.

6. Concluding remarks

In his Presidential Address at the *2011 Institute of Mathematical Statistics Annual Meeting* about controversial aspects of measuring research performance through bibliometrics, Professor P. G. Hall concluded that

> ‘As statisticians we should become more involved in these matters than we are. We are often the subject of the analyses discussed above, and almost alone we have the skills to respond to them, for example by developing new methodologies or by pointing out that existing approaches are challenged. To illustrate the fact that issues that are obvious to statisticians are often ignored in bibliometric analysis, I mention that many proponents of impact factors, and other aspects of citation analysis, have little concept of the problems caused by averaging very heavy tailed data. (Citation data are typically of this type.) We should definitely take a greater interest in this area’ (Hall, 2011).

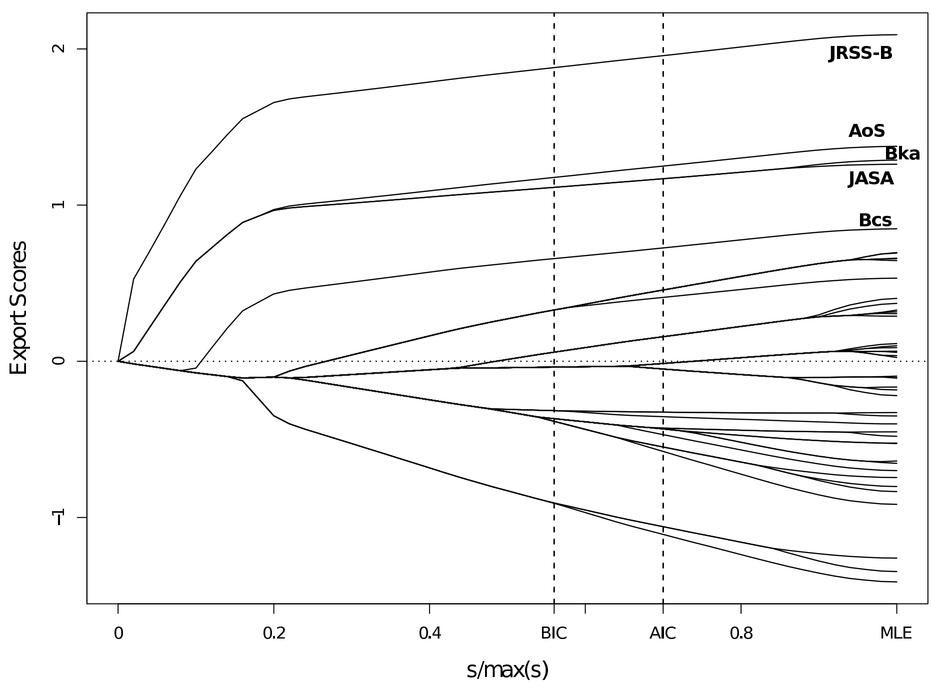

Fig. 4. Path plot of adaptive ranking lasso analysis based on Journal Citation Reports® edition 2010. Journals acronyms are given in Table 1.

The model-based approach to journal ranking discussed in this paper is a contribution in the direction that Professor Hall recommended. Explicit statistical modelling of citation data has two important merits. First, transparency, since model assumptions need to be clearly stated and can be assessed through standard diagnostic tools. Secondly, the evaluation and reporting of uncertainty in statistical models can ba based upon well established methods.

Many journals' websites report the latest journal Impact Factor and the journal's corresponding rank in its category. Very small differences in the reported Impact Factor often imply large differences in the corresponding rankings of Statistics journals. Statisticians should naturally be concerned about whether such differences are significant. Our analyses conclude that most of the apparent differences among estimated export scores are insignificant, and thus differences in journal ranks are often not reliable. The clear difficulty of discriminating between journals based on citation data is further evidence that the use of journal rankings for evaluation of individual researchers will often — and perhaps always — be inappropriate.

Journal homogeneity is a minimal prerequisite for a meaningful statistical analysis of citation data (Lehmann *et al.*, 2009). The aforementioned *Science* editorial entitled *Impact Factor Distortions* (Alberts, 2013) reports that

> '(...) in some nations, publication in a journal with an impact factor below 5.0 is officially of zero value.'

In the last edition (2012) of the Journal Citation Reports®, the very highest Impact Factor reported in the category *Statistics and Probability* was 4.91, achieved by the *Journal of Statistical Software*. The category *Mathematics* achieved still lower Impact Factors, with the highest value there in 2012 being 3.57 for the *Journal of the American Mathematical Society*. Although perhaps obvious, it should be stressed that comparisons between different research fields will rarely make sense, and that such comparisons should be avoided. Research fields differ very widely, for example in terms of the frequency of publication, the typical number of authors per paper and the typical number of citations made in a paper, as well as in the sizes of their research communities.

References

Adler, R., Ewing, J. and Taylor, P. (2009). Citation statistics (with discussion and rejonder). *Statistical Science* **24**, 1–14.

Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.

Alberts, B. (2013). Impact Factor distortions. *Science* **340**, 787.

Bergstrom, C. (2007). Eigenfactor: Measuring the value and the prestige of scholarly journals. *College & Research Libraries News* **68**, 314–316.

Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39**, 324–345.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* **30**, 107–117.

Cattelan, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science* **27**, 412–433.

Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Firth, D. (2012). `qvcalc`: Quasi variances for factor effects in statistical models. R package version 0.8-8. http://CRAN.R-project.org/package=qvcalc

Firth, D. and de Menezes, R. X. (2005). Quasi-variances. *Biometrika* **91**, 65–80.

Frandsen, T. F. (2007). Journal self-citations - Analysing the JIF mechanism. *Journal of Informetrics* **1**, 47–58.

Garfield, E. (1955). Citation indices for Science. *Science* **122**, 108–111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science* **178**, 471–479.

Glänzel, W. and Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics* **53**, 171–193.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society Series A* **159** (3), 385–443.

Gross, P. L. K. and Gross, E. M. (1927). College libraries and chemical education. *Science* **66**, 385–389.

Hall, P. G. (2009). Comment: Citation statistics. *Statistical Science* **24** (1), 25–26.

Hall, P. G. (2011). 'Ranking Our Excellence,' or 'Assessing Our Quality,' or Whatever.... *Institute of Mathematical Statistics Bulletin* **40** (6), 12–14.

IEEE Board of Directors (2013). IEEE position statement on 'Appropriate use of bibliometric indicators for the assessment of journals, research proposals, and individuals'.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Lehmann, S., Lautrup, B. E., and Jackson, A. D. (2009). Comment: Citation statistics. *Statistical Science* **24** (1), 17–20.

Liner, G. H. and Amin, M. (2004). Methods of ranking economics journals. *Atlantic Economic Journal* **32**, 140–149.

Masarotto, G. and Varin, C. (2012). The ranking lasso and its application to sport tournaments. *Annals of Applied Statistics* **6**, 1949–1970.

Morris, C.N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* **78**, 47–65.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. 2nd edition. Springer.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`.

San Francisco Declaration on Research Assessment (DORA) (2013). URL `http://am.ascb.org/dora/`

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* **314**, 497.

Sevinc, A. (2004). Manipulating impact factor: An unethical issue or an editor's choice? *Swiss Medical Weekly* **134**, 410–410.

Silverman, B. W. (2009). Comment: Citation statistics. *Statistical Science* **24** (1), 21–24.

Stigler, G. J., Stigler, S. M. and Friedland, C. (1995). The journals of economics. *The Journal of Political Ecomony* **103**, 331–359.

Stigler, S. M. (1994). Citation patterns in the journals of statistics and probability. *Statistical Science* **9**, 94–108.

Theoharakis, V. and Skordia, M. (2003). How do statisticians perceive statistics journals? *The American Statistician* **57**, 115–123.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* **58**, 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Kneight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B* **67**, 91–108.

Turner, H. and Firth, D. (2012). Bradley-Terry models in R: The `BradleyTerry2` package. *Journal of Statistical Software* **48** (9), 1–21.

van Nierop, E. (2009). Why do statistics journals have low impact factors? *Statistica Neerlandica* **63**, 52–62.

Wedderburn, R. W. M. (1974). Quasi-likelihood, generalized linear models, and the Gauss-Newton method. *Journal of the Royal Statistical Society Series B* **61**, 439–447.

West, J. D. (2010). Eigenfactor: Ranking and mapping scientific knowledge. Ph.D. Dissertation. University of Washington

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.