Computing a journal meta-ranking using paired comparisons and adaptive lasso estimators

Vana Gür, Laura; Hochreiter, Ronald; Hornik, Kurt

Published in:
Scientometrics

DOI:
10.1007/s11192-015-1772-6

Published: 01/01/2016

Document Version
Peer reviewed version

Computing a journal meta-ranking using paired comparisons and adaptive lasso estimators

Laura Vana*[1] Ronald Hochreiter[1] Kurt Hornik[1]

[1]WU Vienna University of Economics and Business

Abstract

In a "publish-or-perish culture", the ranking of scientific journals plays a central role in assessing the performance in the current research environment. With a wide range of existing methods for deriving journal rankings, meta-rankings have gained popularity as a means of aggregating different information sources. In this paper, we propose a method to create a meta-ranking using heterogeneous journal rankings. Employing a parametric model for paired comparison data we estimate quality scores for 58 journals in the OR/MS/POM community, which together with a shrinkage procedure allows for the identification of clusters of journals with similar quality. The use of paired comparisons provides a flexible framework for deriving an aggregated score while eliminating the problem of missing data.

Keywords: Adaptive lasso estimators, Journal lists, Meta-ranking, Operations research

Introduction

While the so-called "publish-or-perish culture" has been widely discussed and criticized mainly because of intense publication pressure (see for example Adler and Harzing, 2009; Frey, 2010; Willmott, 2011), an increasingly competitive research environment is in need of performance metrics. Generally, the reputation and the quality of the research outlets in which scholars publish their work along with received citations in peer-reviewed academic journals remain the main criteria for assessing research quality. The main publication outlets and means of knowledge dissemination in each business discipline are the academic journals in which research is published (Meredith et al., 2011; Fry and Donohue, 2013). In order to reflect the impact and quality of journals, rankings lie therefore at the core of research assessment. Whether it is used by universities for recruitment or promotion/tenure purposes, by editors to underline the importance of their journals or by publishers who aim at maximizing their revenue, a ranking supports decision making. Especially owing to this importance in academic life, the number of methods used to create rankings is increasing continuously.

Being faced with a large number of available rankings, institutions find it often hard to pick the most appropriate ranking for a special purpose. Different stakeholders often

*Correponding author: Laura Vana, Department of Finance, Accounting and Statistics, WU Vienna University of Economics and Business, Welthandelsplatz 1, D4/4, Vienna, Austria, Tel.: +43-1-31336-5032, E-mail: laura.vana@wu.ac.at.

prefer different rankings for various reasons, so that the creation of aggregated rankings (also referred to as meta-rankings) is a viable way to satisfy all parties involved. Moreover, the disadvantages of the traditional approaches, such as citation-based and survey-based methods, at creating journal rankings have been outlined in the literature (see e.g., Serenko and Bontis, 2013) and it has become clearer that one single measure might not be sufficient to describe citation patterns of scientific journals (Glänzel and Moed, 2002). In this light, meta-rankings have gained popularity because of their advantage of aggregating the available information and of building on, rather than against, previous work.

The problem of combining ranking results from different sources has been studied in many disciplines; most extensively, in the context of social choice theory, where individual inputs (e.g., preferences of individual agents) are combined into collective outputs (e.g., collective decisions, preferences) (List, 2013). In the context of Web search, Dwork et al. (2001) address the problem of computing a "consensus" ranking, given the individual preferences of several "judges" and call this the *rank aggregation problem*.

In this paper, we propose a novel method to aggregate heterogeneous journal rankings using adaptive lasso (least absolute shrinkage and selection operator) estimators. The parametric Bradley-Terry model (Bradley and Terry, 1952) for paired comparison data is employed for estimating journal quality scores. The use of paired comparisons provides a flexible framework for deriving an aggregated score using different sources, with no strict restriction on the number of journals and the number of rankings used in the analysis. In combination with a shrinkage procedure, the proposed approach allows the identification of clusters of journals with similar quality (Masarotto and Varin, 2012) and has the advantage of providing a partial ranking, that is, a ranking that allows ties.

Tüselmann et al. (2015) discuss several shortcomings of meta-studies, including arbitrary inclusion or datedness of journal rankings and journals, inadequate treatment of missing data and treatment of ordinal data as metric data. We address these shortcomings in the following way: (i) we base our analysis on an extensive list of 31 journal rankings that are used by stakeholders in decision making. An overview of the rankings is provided in the Section "Journals and quality assessment" and our descriptive statistics show that the significant correlations at a 5% significance level are positive. However, each of the rankings contains valuable information because, while having the common goal of assessing journal quality, they proceed differently in defining comparison measures. (ii) Given a set of journals, ranking lists are in general incomplete, as not all journals are rated in all the rankings. Our approach avoids the ensuing missing data issues by using paired comparisons in the respective ranking lists (similar to Cook et al., 2010; Theußl et al., 2014). (iii) The ordinal nature of rankings is kept in fact by the paired comparison data. In addition, the parametric model is adapted in order to appropriately handle ties.

We illustrate the application of the proposed method by computing a meta-ranking for 58 established journals in the Operations Research, Management Science and Production & Operations Management (OR/MS/POM) community. We also investigate the stability of the resulting meta-ranking by using a subset of the rankings, in our case only recent rankings published in 2013, and find that results remain mainly stable, with few significant changes.

This paper is organized as follows. In the next section we present an overview of existing ranking approaches. The third section presents the journals and journal ranking lists employed in the study. The Section "Methodological approach" introduces the methodology used for creating the journal meta-ranking. Results are discussed in the Section "Results" and the last section concludes.

Ranking approaches

Traditional approaches at assessing journal quality can be classified into subjective (survey-based), objective (citation-based) and a combination thereof (hybrid) (Tüselmann et al., 2015).

Subjective or survey-based rankings are primarily created by universities or associations which ask scholars with different affiliations to rate journals on an (usually) ordinal scale. Academic research relying on survey-based quality assessment of OR/MS/POM journals has been conducted by Barman et al. (1991, 2001) (20 selected journals are ranked based on a questionnaire survey of the Decision Sciences Institute members) and Olson (2005), who surveyed faculty members of 25 business schools in the US in 2000 and 2002. One of the main advantages of the survey-based approach is that a journal's ranking position reflects a cumulative opinion of a representative group of its readers and contributors (Serenko and Dohan, 2011). The disadvantages, however, include subjectivity in the ranking process, familiarity bias and practitioner under-representation (for an extensive list of advantages and disadvantages of survey ranking methods, see Table I in Serenko and Bontis, 2013).

Objective rankings are based on different data-driven performance indicators. Performance metrics based on citations are mainly published by commercial providers like Thomson Reuters or Elsevier. The first academic studies to conduct a citation-based analysis of journals in the Operations Management (OM) field are Vokurka (1996) and Goh et al. (1996). Recent citation-based analyses focusing on the OR/MS field are Xu et al. (2011), who use Google's PageRank method to rank 31 OR/MS journals, and Cheang et al. (2014), who provide updated results for the study in Xu et al. (2011). The PageRank algorithm is a citation or link analysis algorithm which was first used to measure the impact of Web pages for Google's Web search (Page et al., 1999) and which has also been employed for measuring journal impact by considering both citation quantity and quality, in that it differentiates citations by source (i.e., citations from higher impact journals should outweigh citations from lower impact journals). Its origins lie in the work of Pinski and Narin (1976), who proposed an influence weighting methodology for determining citation-based influence measures for scientific journals in physics by relying on the basic idea that a citation from a highly prestigious journal should be given more weight than a citation from a less influential journal. Other PageRank-based indicators including the *Eigenfactor® Score* and the *SCImago Journal Rank* have been proposed for the ranking of scientific journals (see Section "Journals and quality assessment"). A metric that does not differentiate citations by source is proposed by Xu et al. (2015), who develop the sub-impact factor, which takes into account the distribution of the citations by dividing a given number of citations L by the smallest number of articles needed to cover at least L citations. While eliminating the subjectivity inherent in survey-based rankings, the citation-based metrics have been criticized because of the skewness of citation data (this means that a few papers generate a large number of citations), occasional mistakes, omissions and inconsistencies in journal databases, database coverage, etc. (Table II in Serenko and Bontis, 2013).

Hybrid rankings are generated by combining subjective and objective methods (Zhou et al., 2001). This type of rankings are usually used in research institutes for internal purposes such as staff promotion (Xu et al., 2015). Fry and Donohue (2014) take also the author affiliation into consideration and call this the author-based approach. Lozano and Salmerón (2005) use DEA to rank OR/MS journals based on two criteria: the delay in the reviewing and publication process (with data gathered from the editors of the journals through a survey) and the impact of journals in relation to the length of their articles. Holsapple and Lee-Post (2010) perform a behavior-based analysis by examining the publishing behavior of tenured OM researchers at leading research universities in the US.

Extensive comparisons across the different methods as well as discussions on their shortcomings can be also found in, e.g., Frey and Rost (2010), Serenko and Dohan (2011). Considering the drawbacks of the above approaches, the meta-ranking approach recommends itself naturally as it tries to reconcile a variety of methods by using the available information to build a composite journal ranking (Cook et al., 2010). This approach has been applied to journals from different fields. Cook et al. (2010) aggregate journal rankings for accounting journals, Theußl et al. (2014) propose an optimization-based consensus ranking approach for ranking marketing journals, while Halkos and Tzeremes (2011) employ DEA for the purpose of ranking 229 economics journals.

In the OM field, several meta-analyses have combined information from prior academic studies. Petersen et al. (2011) provide the first meta-analysis to examine the ranking of journals in the OM field by using 5 prior studies. Meredith et al. (2011) build journal rankings using official in-house journal lists of AACSB-accredited business schools and compare their results with 12 ranking studies published during 1990 to 2009. Fry and Donohue (2013) use the information from 15 previous OM journal ranking studies and provide a meta-analysis through DEA for assessing journal quality. A recent study is provided by Tüselmann et al. (2015), who base their study on 10 rankings from Harzing's Journal Quality Lists and on the *Impact Factor* (Garfield, 1972) and compute an aggregated ranking of journals in different subject areas by using DEA and random forests, with particular emphasis on the OR/MS/POM field.

Journals and quality assessment

We select 58 journals established in the OR/MS/POM community, which are also part of Thomson Reuters' Journal Citation Reports (Thomson Reuters, 2014). The complete list of journals can be found in Table 4. Furthermore, we collect an extensive list of 31 rankings still in use in 2013. We match the rankings to the list of journals using ISSN codes. All rankings used in the analysis are listed in Table 5. Out of 31 rankings, 13 were compiled in 2013. The other 18 are dated in the range from 2001 to 2012, but are still used in different contexts for assessing journal quality.

From Thomson Reuters' Journal Citation Reports we obtain citation-based journal metrics like the *Impact Factor*, *5-Year Impact Factor*, *Immediacy Index* and *Cited Half-Life*. The *Impact Factor* (Garfield, 1972) is one of the best known indexes and measures the number of citations received by a journal in the two preceding years normalized by the number of citable items in that journal. The *5-Year Impact Factor* extends the time period used for calculation to five years. The *Immediacy Index* measures the number of citations received by a journal in the same year, while *Cited Half-Life* measures the median age of a journal's cited articles in a certain year. These metrics have been criticized because they are not normalized with respect to different disciplines and fields. Moreover, the immediacy index is likely to be sensitive to publication delay, frequency of publication, speed of indexing or subject peculiarities (Glänzel and Moed, 2002).

Other citation metrics freely available are the *Eigenfactor® Score* and its normalized version, the *Article Influence® Score*. The Thomson Reuters citation data is used in the calculation of the Eigenfactor® metrics. Unlike the *Impact Factor*, these metrics assign different weights to different sources of citations received by a journal and exclude self-citations (Bergstrom et al., 2008). The time period used for the calculation is five years.

Elsevier uses the Scopus® database in computing three alternative journal metrics: the *Source Normalized Impact per Paper* (SNIP) (Moed, 2011) is an "a priori" normalized metric which takes into account both the frequency of other papers in the reference lists of the papers providing citations and the coverage of the corresponding subject field in the database. It corrects for differences between fields by dividing the number of citations per paper in a journal by a normalized measure of the citation potential in the subject field covered by that journal (Glänzel and Moed, 2013). The *Impact per Publication* (IPP) is a metric similar to Thomson Reuters' *Impact Factor* and is defined as the ratio of citations in a year to scholarly papers published in the three previous years divided by the number of scholarly papers published in those same years. A third metric is the *SCImago Journal Rank* (SJR), which is inspired by the PageRank algorithm and differentiates citations by source. Self-citations are limited to 33% of total citations and the citation time window is set to three years (González-Pereira et al., 2010). The SJR index is a normalized, size-independent indicator and it ranks journals by their "average prestige per article". Detailed information on the methodology of the Journal Metrics ® of Elsevier can be found on the website `www.journalmetrics.com`.

An important source of journal rankings is the Harzing Journal Quality List (JQL). The 52nd edition of the JQL data (Harzing, 11 February 2014) contains 22 rankings provided by different (research) institutions or scientific studies. Most of the JQL rankings are on an ordinal scale (see Table 5) and are survey-based. Rankings like the ones produced by Cranfield University School of Management (Cra12), Le Centre national de la recherche scientifique (Cnrs13), Erasmus Research Institute of Management (EJL12) or the ABS Journal Quality Guide are hybrid rankings, as they are based on both expert opinions and bibliometric information.

To measure similarity between rankings, we need a measure which can conveniently handle the presence of ties and missing data. We thus employ the rank correlation coefficient proposed by Emond and Mason (2002), which corresponds to the unique association measure of Kemeny and Snell (1962) and satisfies four basic axioms that should apply to any distance measure between two weak orderings (where ties are allowed). The metric is denoted by τ_x and is an extension of Kendall's τ, which handles ties more appropriately (see also Hornik et al., 2007):

$$\tau_x^{(k_1,k_2)} = 1 - \frac{\sum_{i=1}^{N_c}\sum_{j=1}^{N_c} |a_{ijk_1} - a_{ijk_2}|}{N_c(N_c-1)},$$

where N_c is the number of journals rated in at least one of the rankings R_{k_1} and R_{k_2} and a_{ijk} is defined as:

$$a_{ijk} = \begin{cases} 1 & \text{if ranking } R_k \text{ rates journal } J_i \text{ higher than or tied with journal } J_j, \\ 0 & \text{if } i = j \text{ or if } R_k \text{ does not rate } J_i \text{ or/and } J_j, \\ -1 & \text{otherwise.} \end{cases}$$

We present the τ_x correlation coefficients in Table 1, together with the number of journals rated in the sample for each ranking. For each rank correlation coefficient, we compute p-values using 1000 bootstrap samples. Column "MLE" presents the correlation of the ranking produced using the method described in the "Methodological approach" section with the other 31 rankings. The citation-based indexes of Thomson Reuters, the *Eigenfactor® Score*, the *Article Influence® Score* and the Elsevier metrics are mostly significantly correlated. Not surprisingly, the metric that is not significantly correlated with the rest is *Cited Half-Life*, as it is the only metric based on the age of cited articles. We observe that at 5% significance level the significant rank correlations are positive, but the degree of the association varies, indicating that the quality assessment of journals differs across different metrics or sources.

In Table 2 the percentage of times each journal appeared in the 31 rankings is reported in column "% rated". The journal that is rated most often is *Operations Research (OR)*, followed by *Production and Operations Management (POM)* and *Decision Support Systems (DSS)*. In total, the sample contains 16452 pairwise comparisons between the 58 journals and each pair of journals has been compared 10 times on average, with a standard deviation of 6. Journal pair *International Transactions in Operational Research (ITOR)* and *Military Operations Research (MILOR)*, as well as pair *MILOR* and *Journal of Simulation (JSIM)* are compared twice (minimum in the sample). Pair *OR* and *POM* is the most often compared pair in the sample (28 comparisons).

Methodological approach

Consider a total number of K journal rankings $\mathscr{R} = \{R_1, \ldots, R_K\}$ of N journals $\mathscr{J} = \{J_1, \ldots, J_N\}$. Each journal in the set $\mathscr{J}$ is ranked in at least one of the rankings in $\mathscr{R}$.

The aim of the analysis is to suitably aggregate the information available from the K different ranking sources by finding a criterion that reflects the overall quality of the journals in $\mathscr{J}$.

Bradley-Terry model for paired comparison data

Each ranking R_k naturally induces a collection of paired comparisons for the journals included in the ranking. I.e., for each pair J_i and J_j of journals included in the ranking we can determine whether J_i is ranked higher/lower than or the same as J_j. A popular statistical model for such paired comparison data is the Bradley-Terry model (Bradley and Terry, 1952), a logistic regression model, where the probability of journal J_i being better than journal J_j is given by $\exp(\mu_i - \mu_j)/(1 + \exp(\mu_i - \mu_j))$, with μ_i and μ_j being the ability (quality) parameters for journals J_i and J_j, respectively. Note that N need not be larger than K.

As many journals in the analysis fall into the same ranking class, we employ a modification of the original Bradley-Terry model which can properly handle ties and treats ties as one-half of a success and one-half of a failure. The log-likelihood is then given by:

$$\ell(\boldsymbol{\mu}) = \sum_{i<j}^{N} \sum_{k}^{K} y_{ijk}(\mu_i - \mu_j) - \log(1 + \exp(\mu_i - \mu_j)), \quad (1)$$

Table 1: Number of rated journals (N_R) and τ_x association measure; p-values based on 1000 bootstrap samples.

	N_R	IF5Y	IMMI	CHL	EFS	AIS	SNIP	IPP	SJR	Wie01	Vhb03	Bjm04	Hkb05	Theo05	Ejis07	EjisCI07	UQ07
IF	58	0.74***	0.51***	−0.04	0.48***	0.42***	0.09***	0.11***	0.08**	0.02	0.02	−0.02	0.00	0.00	0.05	0.07*	0.05*
IF5Y	56		0.45***	−0.02	0.48***	0.56***	0.13***	0.14***	0.12***	0.03	0.05	0.02	0.00	0.00	0.06	0.07	0.05
IMMI	56			0.02	0.36***	0.31***	0.06	0.08*	0.08*	0.03	0.03	0.01	−0.01	0.00	0.06	0.06	0.02
CHL	57				0.17	0.12	0.00	0.02	0.02	0.06*	0.04	0.03	0.04	0.00	0.04	0.05	0.02
EFS	58					0.51***	0.06*	0.06*	0.08**	0.03	0.05	0.00	0.01	0.00	0.09**	0.12**	0.04*
AIS	56						0.12***	0.10**	0.12***	0.04	0.12***	0.04	0.04	0.01	0.11**	0.12**	0.03
SNIP	24							0.77***	0.61***	0.44***	0.34	0.44**	0.37	0.28	0.35*	0.35*	0.37
IPP	24								0.60***	0.44***	0.34	0.44**	0.36	0.28	0.35*	0.35*	0.37
SJR	24									0.44***	0.34	0.44**	0.37	0.28	0.34*	0.35*	0.37
Wie01	22										0.32	0.17	0.27	0.02	0.25	0.23	0.19
Vhb03	28											0.10	0.23	0.01	0.47***	0.44***	0.18
Bjm04	16												0.13	−0.03	0.06	0.06	0.16
Hkb05	21													0.04	0.22	0.20	0.21
Theo05	6														0.02	0.02	0.11
Ejis07	31															0.82***	0.23

	N_R	Ast08	Wie08	ABS10	Den11	HEC11	UQ11	Vhb11	Aeres12	Cra12	EJL12	Abdc13	Cnrs13	Ess13	Fneg13	MLE
IF	58	0.03	0.01	0.09**	0.09*	0.03	0.02	0.04	0.03	0.04*	0.06*	0.10***	0.05	0.04	0.00	0.62***
IF5Y	56	0.04	0.04	0.10**	0.10**	0.05	0.03	0.06	0.03	0.06**	0.09**	0.12***	0.05	0.05	0.01	0.63***
IMMI	56	0.02	0.00	0.06	0.07	0.05	0.00	0.03	0.03	0.02	0.10**	0.08*	0.06	0.02	0.01	0.49***
CHL	57	0.02	0.04	0.00	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.00	0.03	0.02	0.01	0.22**
EFS	58	0.04*	0.02	0.05	0.06	0.04	0.02	0.05	0.06**	0.02	0.05	0.10**	0.09***	0.03	0.01	0.68***
AIS	56	0.04	0.05	0.10**	0.13***	0.07*	0.04	0.09**	0.03	0.07**	0.09**	0.10**	0.08*	0.06*	0.01	0.60***
SNIP	24	0.38*	0.36*	0.33	0.32*	0.38*	0.38*	0.32	0.38*	0.40*	0.34	0.33	0.36*	0.33	0.33	0.10***
IPP	24	0.38*	0.36*	0.33	0.32*	0.38*	0.38*	0.33	0.38*	0.40*	0.34	0.33	0.36*	0.33	0.34	0.10***
SJR	24	0.38*	0.36*	0.33	0.32*	0.38*	0.38*	0.33	0.38*	0.40*	0.35	0.32	0.36*	0.33	0.34	0.11***
Wie01	22	0.18	0.30	0.16	0.19	0.25	0.18	0.17	0.30	0.16	0.23	0.18	0.24	0.21	0.23	0.05**
Vhb03	28	0.11	0.36**	0.27	0.22	0.23	0.17	0.40***	0.25	0.19	0.22	0.19	0.29	0.25	0.18	0.09***
Bjm04	16	0.23	0.13	0.07	0.04	0.18	0.19	0.09	0.10	0.10	0.15	0.02	0.16	0.16	0.28	0.01
Hkb05	21	0.14	0.25	0.14	0.08	0.23	0.21	0.20	0.22	0.22	0.17	0.06	0.19	0.17	0.28	0.04*
Theo05	6	0.03	0.02	0.02	0.01	0.01	0.11	0.07	0.12	0.06	0.02	0.01	0.07	0.07	0.35	0.01
Ejis07	31	0.10	0.25	0.21	0.30*	0.22	0.14	0.33**	0.22	0.15	0.21	0.23	0.30	0.17	0.08	0.14***
EjisCI07	32	0.12	0.30	0.23	0.37**	0.21	0.13	0.34**	0.22	0.16	0.23	0.25	0.29	0.22	0.08	0.14***
UQ07	21	0.25	0.22	0.18	0.10	0.20	0.47***	0.23	0.22	0.25	0.19	0.23	0.25	0.17	0.09	0.05**
Ast08	19		0.20	0.19	0.09	0.22	0.27	0.16	0.22	0.30	0.19	0.23	0.22	0.19	0.18	0.05**
Wie08	25			0.28	0.26	0.21	0.22	0.28	0.30	0.24	0.24	0.20	0.24	0.36	0.27	0.03*
ABS10	28				0.32*	0.20	0.18	0.22	0.21	0.25	0.28	0.29	0.23	0.48***	0.11	0.09***
Den11	34					0.12	0.13	0.21	0.21	0.12	0.25	0.34**	0.20	0.24	0.06	0.12**
HEC11	22						0.15	0.23	0.18	0.26	0.30	0.20	0.27	0.29	0.22	0.06**
UQ11	19							0.17	0.22	0.25	0.14	0.20	0.23	0.17	0.12	0.03*
Vhb11	26								0.29	0.23	0.26	0.17	0.32	0.24	0.21	0.09***
Aeres12	24									0.22	0.21	0.26	0.65***	0.25	0.15	0.05**
Cra12	18										0.22	0.22	0.23	0.25	0.20	0.06***
EJL12	26											0.24	0.34	0.27	0.15	0.10***
Abdc13	29												0.27	0.19	0.06	0.12***
Cnrs13	26													0.29	0.15	0.11***
Ess13	25														0.19	0.05*
Fneg13	11															0.02**

*** – p value < 0.001, ** – p value < 0.01, *– p value < 0.05.

where the y_{ijk} is defined as:

$$y_{ijk} = \begin{cases} 1 & \text{if ranking } R_k \text{ rates journal } J_i \text{ higher than journal } J_j, \\ 0.5 & \text{if ranking } R_k \text{ rates journal } J_i \text{ and journal } J_j \text{ the same,} \\ 0 & \text{if ranking } R_k \text{ rates journal } J_i \text{ lower than journal } J_j. \end{cases}$$

The parameter of interest is the vector of journal abilities $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\top$, which can be further used for ranking the N journals. Only pairwise differences $\mu_i - \mu_j$ are identifiable in this model, hence a restriction needs to be imposed on the vector $\boldsymbol{\mu}$. We choose to set the ability parameter μ_1 of the first journal *4OR* equal to zero. An alternative to treating ties as one-half of a success and one-half of a failure would be using the cumulative link Bradley-Terry model (Agresti, 2010). We choose the former approach as it is computationally more efficient.

The Bradley-Terry model offers the possibility of quantifying the uncertainty in the journal abilities. Since the model is identified through pairwise differences, uncertainty quantification requires the complete variance matrix of $\hat{\boldsymbol{\mu}}$. A way to avoid the computation of the complete variance matrix is through quasi-variances (Firth and De Menezes, 2004), which are constructed in order to allow approximate calculation of any variance of a difference as the sum of the quasi-variances of the components:

$$\text{var}(\mu_i - \mu_j) \approx \text{qvar}_i + \text{qvar}_j.$$

This approach facilitates the approximate inference on the significance of the difference between any two journals. The estimation of the quasi-variances is performed using the `qvcalc` package (Firth, 2015) in R (R Core Team, 2015).

Maximum likelihood estimation of the Bradley-Terry model assumes that, given the estimated journal abilities $\hat{\mu}$, the paired comparisons are realizations of independent binomial experiments. This assumption is questionable in our application as (1) the same pair of journals is repeatedly compared in different rankings, and (2) because the same entity rates more journals. Correlations between the journal abilities could potentially lead to the phenomenon of over-dispersion (the presence of higher variability in a data set compared to what would be expected based on a given statistical model), which means that the standard errors of the estimated parameters might be underestimated. In a logistic regression setting, over-dispersion cannot be formally estimated. Alternative models could be employed to capture the correlation structure of the parameters, but the problem of modeling dependent paired comparison data is still an open research question (for an extensive discussion, see Cattelan, 2012). While we do not formally correct for over-dispersion, the interpretation of the reported quasi-standard errors (square root of the estimated quasi-variances) will need to be performed with care.

Clustering of journals using the ranking lasso

In order to avoid potential over-interpretation of insignificant differences between journal abilities and to obtain a clustering of journals with similar ability parameters, Masarotto and Varin (2012) propose the ranking lasso technique, which computes the solution for the modified Bradley-Terry model by maximizing the log-likelihood and imposing an $\mathscr{L}_1$ penalty on the pairwise ability differences $\mu_i - \mu_j$:

$$\widehat{\boldsymbol{\mu}}_\lambda = \arg\min \left\{ -\ell(\boldsymbol{\mu}) + \lambda \sum_{i<j}^{N} w_{ij} |\mu_i - \mu_j| \right\},$$

where $\ell(\boldsymbol{\mu})$ is the log-likelihood function in (1), λ is the lasso penalty and the w_{ij} are pair-specific weights.

The lasso penalty λ is a tuning parameter which controls the "strength" of the $\mathscr{L}_1$ penalty. The standard maximum likelihood solution is obtained for $\lambda = 0$, whereas the fitting is penalized as λ increases to infinity. In the limiting case all journal abilities are estimated to have the same value (i.e., all journals are assigned to one cluster) for λ large enough. The value of the lasso penalty can be suitably chosen by the minimization of an information criterion, such as the Akaike Information Criterion (AIC):

$$\mathrm{AIC}(\lambda) = -2\ell(\widehat{\boldsymbol{\mu}}_\lambda) + 2\mathrm{df}(\lambda)$$

or the Bayesian information criterion (BIC):

$$\mathrm{BIC}(\lambda) = -2\ell(\widehat{\boldsymbol{\mu}}_\lambda) + \log(n)\mathrm{df}(\lambda),$$

where n is the total number of paired comparisons and the effective degrees of freedom $\mathrm{df}(\lambda)$ are estimated as the number of distinct groups formed with a certain λ (in line with Tibshirani, 2011; Masarotto and Varin, 2012). Compared to the AIC, the BIC favors sparser models with a smaller number of groups.

In the standard lasso problem, equal weights $w_{ij} \equiv w$ are assigned and the same penalty is applied to all pairwise differences $\mu_i - \mu_j$. In this setting, several studies, including Fan and Li (2001) and Zou (2006), have shown that the value of λ required for variable selection consistency, that is, for the correct identification of the "true" zero pairwise differences as n increases, overshrinks the nonzero differences, which in turn leads to asymptotically biased estimates of the nonzero terms $\mu_i - \mu_j$. One possibility to overcome this drawback is to apply a weighted penalty (suggested also by Masarotto and Varin, 2012). We do this by choosing the pair-specific weights w_{ij} to be inversely proportional to the maximum likelihood estimates (MLE):

$$w_{ij} = |\hat{\mu}_i^{(\mathrm{MLE})} - \hat{\mu}_j^{(\mathrm{MLE})}|^{-1}.$$

This method is called the adaptive ranking lasso and was proposed by Zou (2006). Small differences in the MLE estimates are penalized stronger than bigger differences. This has in general the effect of enforcing a stronger clustering of the journal abilities compared to the standard lasso method.

Results

We proceed with the presentation of the results from applying the method described in the previous section to the dataset of 58 journals and 31 rankings. We then repeat the analysis but use only rankings from 2013 and investigate the stability of the results.

All rankings

Table 2 presents the journal abilities estimated by maximum likelihood (MLE) and the abilities estimated by employing the adaptive lasso (ALASSO) estimators for which the shrinkage parameter λ was optimally chosen according to AIC and BIC. The shrinkage procedure groups journals with similar ability parameters. For λ chosen by minimizing AIC, 24 clusters of journals are identified. Using the BIC for choosing the optimal value of λ leads to a sparser solution with 17 clusters. In addition, the quasi-standard errors (QSE) are reported. An important feature that can be observed from QSE is that some of the estimated abilities are not significantly different from one another. This implies that the ranking lasso method should be employed for shrinking the differences of similar coefficients to zero. Note that, in the light of dependent data, the standard errors might be underestimated. Higher standard errors would only support the shrinkage procedure more strongly. The correlation between the MLE and the rankings used in the analysis is positive (see column "MLE" in Table 1). The metric exhibiting the highest correlation with the MLE ranking of the modified Bradley-Terry model is the *Eigenfactor® Score*. The same holds for the partial rankings corresponding to the AIC- and BIC-based shrinkage estimates.

Figure 1 illustrates the ability parameters estimated by the ALASSO estimators for each journal. The journal with the highest ability parameter is *Management Science* followed by *Operations Research*, *Mathematical Programming (MP)*, *Transportation Research Part B: Methodological* and *Journal of Operations Management (JOM)*. The ranking lasso groups the latter four journals into one cluster because of their similar MLEs. The journal with the lowest score is *Military Operations Research (MILOR)*.

Petersen et al. (2011), Fry and Donohue (2013), Cheang et al. (2014) and Tüselmann et al. (2015) also identify *Management Science* as the top outlet for OR/MS/POM research. When regarding *only* the intersection of the sets of journals, *JOM* and *POM* (position 5 and 10 in the MLE ranking) also make it in the top ten lists of Meredith et al. (2011) (position 1 and 2, respectively), Petersen et al. (2011) (position 3 and 4), Fry and Donohue (2013) (position 3 and 4). Seven out of the top 10 journals of Cheang et al. (2014) are also found in our top 10 list. *OR, MP, TS, EJOR* are in the top ten of Fry and Donohue (2013), Tüselmann et al. (2015). *IJPR* receives a top ten position in Meredith et al. (2011), Petersen et al. (2011) and Fry and Donohue (2013), but in our analysis it is ranked only in position 18 (and position 11 in Tüselmann et al., 2015).

In order to investigate the extent to which the incomplete journal list design relates to the estimated journal abilities used for deriving the meta-ranking, we plot in Figure 2 the percentage of times each journal is rated in the sample against the maximum likelihood estimates. Spearman's correlation coefficient is 0.67 and it holds in the sample that journals that are present in more journal lists will tend to have higher positions in the meta-ranking corresponding to the maximum likelihood estimates of the Bradley-Terry model. This finding is not surprising, since journals rated in more than 50% of the rankings are also considered by experts as being core journals with the main focus on OR/MS/POM research. Moreover, previous research has extensively included the journals clustered on the right side of Figure 2 in the journal lists used for assessing research quality in the OR/MS/POM field. For example, Xu et al. (2011) include the majority of these journals in their core journal list, needed to derive the PageRank quality index. Journals on the left side of Figure 2 are regional journals, journals relating to a specific sub-field (e.g., military operations research) or journals from adjacent disciplines like optimization, simulation or computing. One can observe that through the lasso shrinkage procedure, this feature is reduced, as clusters (according to AIC) contain journals with different ranking coverage.

Using 2013 rankings only

Journal quality and the perception in the community are likely to change over time. An aspect that can be verified in our proposed framework is whether new rankings (i.e., in our case dated 2013) lead to a different meta-ranking than the one obtained using the whole sample of rankings. Rankings compiled in 2013 could provide a more up-

Table 2: Results on estimated journal abilities using MLE, quasi-standard errors (QSE) and adaptive lasso estimators based on AIC and BIC for the modified Bradley-Terry model; column "Pos" indicates the position in the ranking.

	Rankings 2001–2013						Rankings 2013					
Journal	Pos	% rated	MLE	QSE	AIC	BIC	Pos	% rated	MLE	QSE	AIC	BIC
MS	1	83.87	3.50	0.18	2.95	2.79	1	76.92	3.79	0.16	2.73	2.62
OR	2	96.77	2.42	0.15	1.90	1.77	8	92.31	2.05	0.15	1.40	1.32
MP	3	64.52	2.39	0.16	1.90	1.77	4	61.54	2.53	0.17	1.68	1.59
TRBM	4	58.06	2.34	0.16	1.90	1.77	3	61.54	2.78	0.20	1.88	1.78
JOM	5	87.10	2.30	0.15	1.90	1.77	2	76.92	2.85	0.17	1.88	1.78
TS	6	61.29	2.06	0.15	1.66	1.55	5	69.23	2.25	0.16	1.40	1.32
SCL	7	35.48	1.79	0.16	1.27	1.12	6	46.15	2.22	0.20	1.40	1.32
EJOR	8	80.65	1.65	0.14	1.24	1.12	7	69.23	2.05	0.17	1.40	1.32
TRELT	9	64.52	1.48	0.15	1.01	0.89	11	92.31	1.69	0.16	1.20	1.14
POM	10	93.55	1.40	0.14	1.01	0.89	10	100.00	1.75	0.15	1.20	1.14
MOR	11	77.42	1.32	0.14	0.95	0.85	16	92.31	1.21	0.15	0.77	0.71
DSS	12	90.32	1.30	0.14	0.95	0.85	14	76.92	1.31	0.16	0.77	0.71
MSOM	13	64.52	1.17	0.15	0.85	0.77	13	100.00	1.36	0.16	0.77	0.71
JOTA	14	51.61	1.16	0.15	0.85	0.77	12	53.85	1.45	0.16	0.77	0.71
OMEGA	15	87.10	1.15	0.14	0.85	0.77	9	69.23	1.98	0.18	1.40	1.32
JGO	16	19.35	0.92	0.17	0.53	0.44	18	46.15	0.97	0.17	0.42	0.35
IIE	17	80.65	0.91	0.14	0.53	0.44	23	84.62	0.62	0.15	0.37	0.34
IJPR	18	87.10	0.87	0.14	0.53	0.44	19	76.92	0.89	0.16	0.38	0.34
JSCH	19	61.29	0.86	0.14	0.53	0.44	28	84.62	0.48	0.15	0.37	0.34
INFORMS	20	61.29	0.84	0.15	0.53	0.44	22	84.62	0.75	0.16	0.38	0.34
EXSA	21	54.84	0.74	0.15	0.48	0.42	20	69.23	0.81	0.16	0.38	0.34
AOR	22	83.87	0.71	0.14	0.48	0.42	21	69.23	0.76	0.16	0.38	0.34
COR	23	77.42	0.69	0.14	0.48	0.42	15	69.23	1.25	0.16	0.77	0.71
ORS	24	64.52	0.60	0.14	0.44	0.42	26	69.23	0.53	0.16	0.37	0.34
JORS	25	83.87	0.59	0.14	0.44	0.42	25	69.23	0.56	0.16	0.37	0.34
COA	26	19.35	0.57	0.17	0.44	0.42	24	46.15	0.61	0.17	0.37	0.34
ORL	27	74.19	0.51	0.14	0.44	0.42	32	69.23	0.11	0.16	0.00	0.00
NSE	28	29.03	0.49	0.16	0.44	0.42	27	69.23	0.51	0.16	0.37	0.34
TNV	29	74.19	0.41	0.14	0.39	0.41	17	69.23	1.10	0.16	0.72	0.68
NETW	30	38.71	0.36	0.15	0.39	0.41	31	53.85	0.24	0.17	0.27	0.27
OMS	31	29.03	0.30	0.16	0.39	0.41	30	69.23	0.31	0.16	0.27	0.27
NAVRL	32	70.97	0.11	0.14	0.00	0.00	36	92.31	−0.20	0.15	−0.28	−0.27
JMS	33	54.84	0.04	0.15	0.00	0.00	29	69.23	0.33	0.16	0.27	0.27
FODM	34	29.03	0.01	0.16	0.00	0.00	33	69.23	0.01	0.16	0.00	0.00
4OR	35	29.03	0.00	0.00	0.00	0.00	34	69.23	0.00	0.00	0.00	0.00
EO	36	29.03	−0.06	0.16	0.00	0.00	35	69.23	−0.06	0.16	0.00	0.00
PPC	37	80.65	−0.23	0.14	−0.30	−0.27	39	76.92	−0.38	0.16	−0.28	−0.27
QS	38	35.48	−0.28	0.16	−0.30	−0.27	38	46.15	−0.29	0.18	−0.28	−0.27
OPT	39	19.35	−0.28	0.17	−0.30	−0.27	37	46.15	−0.28	0.17	−0.28	−0.27
INTER	40	77.42	−0.31	0.15	−0.30	−0.27	43	53.85	−0.63	0.18	−0.33	−0.28
OCAM	41	19.35	−0.42	0.17	−0.30	−0.27	40	46.15	−0.42	0.17	−0.28	−0.27
OL	42	29.03	−0.48	0.16	−0.32	−0.27	41	69.23	−0.50	0.16	−0.30	−0.27
ITOR	43	64.52	−0.55	0.15	−0.34	−0.27	47	53.85	−0.84	0.17	−0.33	−0.28
DEDS	44	19.35	−0.62	0.17	−0.34	−0.27	42	46.15	−0.62	0.17	−0.33	−0.28
DO	45	29.03	−0.62	0.16	−0.34	−0.27	44	69.23	−0.64	0.16	−0.33	−0.28
IJITD	46	19.35	−0.70	0.18	−0.34	−0.27	45	46.15	−0.71	0.17	−0.33	−0.28
OE	47	29.03	−0.78	0.17	−0.34	−0.27	46	69.23	−0.80	0.16	−0.33	−0.28
MMOR	48	67.74	−0.92	0.15	−0.73	−0.66	49	84.62	−1.23	0.16	−0.85	−0.80
TOP	49	29.03	−1.19	0.17	−0.87	−0.75	48	69.23	−1.22	0.17	−0.85	−0.80
PJO	50	29.03	−1.23	0.17	−0.87	−0.75	50	69.23	−1.26	0.17	−0.85	−0.80
CEJOR	51	29.03	−1.25	0.17	−0.87	−0.75	51	69.23	−1.29	0.17	−0.85	−0.80
INFOR	52	16.13	−1.41	0.20	−0.87	−0.75	52	38.46	−1.43	0.31	−0.85	−0.80
JSIM	53	22.58	−1.65	0.20	−1.19	−1.05	53	53.85	−1.69	0.19	−1.14	−1.07
ASMB	54	29.03	−1.79	0.19	−1.22	−1.05	55	69.23	−1.83	0.18	−1.14	−1.07
RAIRO	55	19.35	−1.80	0.20	−1.22	−1.05	54	46.15	−1.83	0.34	−1.14	−1.07
IMAJMM	56	29.03	−2.00	0.19	−1.34	−1.16	56	69.23	−2.04	0.18	−1.27	−1.18
APJOR	57	19.35	−2.86	0.26	−2.07	−1.87	57	46.15	−2.89	0.23	−1.98	−1.88
MILOR	58	12.90	−5.16	0.73	−3.23	−2.88	58	30.77	−5.20	0.75	−3.05	−2.89

to-date picture of the current research environment. We address this topic by redoing the analysis using only the 13 rankings published in 2013. The MLE and ALASSO estimated abilities are presented in Table 2. ALASSO estimated abilities based on the AIC are visualized in Figure 3. The number of clusters is slightly lower than for the whole sample of rankings (20 clusters according to AIC and 18 clusters according to BIC).

The resulting estimated journal ability parameters are similar to the ones using the whole sample of 31 rankings and the meta-ranking remains highly stable. However, a few differences can be identified. On the one hand, *OR* is downgraded from the second cluster to the fourth cluster. Significantly reduced ALASSO estimated abilities are observed for *ORL*. On the other hand, *JOM*, *OMEGA*, *TNV* and *COR* have higher estimated ability parameters and higher positions in the ranking.

Comparison with other rank-aggregation methods

In addition to assessing how similar the derived meta-ranking is to the individual rankings (see Table 1), it is of interest to investigate how the proposed methodology compares with other rank-aggregation methods. We consider the following additional methods:

- *average rank positions*: journals are ordered according to the average of their ranks in the ranking in which they are contained;
- *Borda's method* (de Borda, 1781): for each ranking, each journal receives a score equal to the number of journals it outranks; scores are then summed over all lists to create an aggregated score called the Borda count, which is subsequently used for ordering the journals;
- Kemeny-Young method (Young, 1986): this determines an ordering which minimizes the total number of discrepancies among the rankings in their pairwise preferences between all journals. Finding such an ordering is computationally hard (NP hard) with complete enumeration often not feasible. Hence, one typically employs greedy minimization heuristics. For calculation we use the `C++` Program for Kemeny-Young Preference Aggregation by William H. Press.

Table 3 presents the average τ_x association measure as well as the corresponding average rank correlation coefficient (Kendall's τ) of each aggregated ranking vis-à-vis the individual rankings, which are used as criteria for the representation quality of the aggregated rankings.

Table 3: Average τ_x association measure and average rank correlation coefficient of the different aggregated rankings and the individual rankings

Aggregated measure	τ_x	Kendall's τ
MLE	0.163	0.536
ALASSO (AIC)	0.165	0.547
ALASSO (BIC)	0.166	0.555
Average rank	0.101	0.289
Borda's method	0.141	0.432
Kemeny-Young	0.157	0.495

One can see that the Bradley-Terry estimators (both MLE and ALASSO) have, on average, higher rank correlations with the individual rankings than other methods.

Conclusion

We use the ranking lasso method proposed by Masarotto and Varin (2012) together with the Bradley-Terry model for paired comparison data in order to aggregate existing journal rankings and to estimate journal quality scores which are subsequently used for deriving a journal meta-ranking. This approach overcomes the issue of missing data by using paired comparisons. Moreover, the shrinkage procedure based on adaptive lasso estimators delivers a partial ranking by identifying clusters of journals with similar quality scores.

As an illustration, we apply this method to 58 OR/MS/POM journals and 31 ranking lists and conclude that *Management Science* maintains its position as the leading OR/MS/POM outlet, followed by *OR, MP, TRBM, JOM, TS, SCL, EJOR, TRELT* and *POM*. We investigate the stability of the resulting meta-ranking when using only 2013 rankings and find increased quality scores for *OMEGA* and *TNV*. The meta-rankings obtained from employing the Bradley-Terry estimators (both MLE and ALASSO) are closest to the ranking obtained by the *Eigenfactor*® *Score* and exhibit higher rank correlations with the individual rankings than other methods of rank aggregation.

The presented method is a flexible way to integrate a wide range of heterogeneous rankings. The resulting graph provides both insight into the relative standing of individual journals within a discipline and guidance on how journals of similar quality can be suitably grouped.

Acknowledgments The authors thank the two reviewers for their insightful comments.

References

Nancy J Adler and Anne-Wil Harzing. When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning & Education*, 8(1):72–95, 2009.

Alan Agresti. *Analysis of ordinal categorical data*. John Wiley & Sons, 2 edition, 2010.

Samir Barman, Richard J Tersine, and M Ronald Buckley. An empirical assessment of the perceived relevance and quality of POM-related journals by academicians. *Journal of Operations Management*, 10(2):194–212, 1991.

Samir Barman, Mark D Hanna, and R Lawrence LaForge. Perceived relevance and quality of POM journals: a decade later. *Journal of Operations Management*, 19(3): 367–385, 2001.

Carl Bergstrom, D West Jevin, and Marc A Wiseman. The Eigenfactor (TM) metrics. *The Journal of Neuroscience*, 28(45):11433–11434, 2008.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Manuela Cattelan. Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science*, 27(3):412–433, 2012.

Brenda Cheang, Samuel Kai Wah Chu, Chongshou Li, and Andrew Lim. OR/MS journals evaluation based on a refined PageRank method: an updated and more comprehensive review. *Scientometrics*, 100(2):339–361, 2014.

Wade D Cook, Tal Raviv, and Alan J Richardson. Aggregating Incomplete Lists of Journal Rankings: An Application to Academic Accounting Journals. *Accounting perspectives*, 9(3):217–235, 2010.

Jean-Charles de Borda. Mémoire sur les elections au scrutin. Histoire de l'Académie Royale des Sciences, 1781.

Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

Edward J Emond and David W Mason. A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, 11(1):17–28, 2002.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456): 1348–1360, 2001.

David Firth. *qvcalc: Quasi Variances for Factor Effects in Statistical Models*, 2015. URL http://CRAN.R-project.org/package=qvcalc. R package version 0.8-9.

David Firth and Renée X De Menezes. Quasi-variances. *Biometrika*, 91(1):65–80, 2004.

Bruno S Frey. Withering academia? *CESifo Working Paper Series No. 3209. Available at SSRN: http://ssrn.com/abstract=1694448*, 2010.

Bruno S Frey and Katja Rost. Do rankings reflect research quality? *Journal of Applied Economics*, 13(1):1–38, 2010.

Timothy D Fry and Joan M Donohue. Outlets for operations management research: a DEA assessment of journal quality and rankings. *International Journal of Production Research*, 51(23-24):7501–7526, 2013.

Timothy D Fry and Joan M Donohue. Exploring the author affiliation index. *Scientometrics*, 98(3):1647–1667, 2014.

Eugene Garfield. Citation analysis as a tool in journal evaluation. *Science*, 178:471–479, 1972.

Janet Geary, Liz Marriott, and Michael Rowlinson. Journal rankings in business and management and the 2001 research assessment exercise in the UK. *British Journal of Management*, 15(2):95–141, 2004.

Wolfgang Glänzel and Henk F Moed. Journal impact measures in bibliometric research. *Scientometrics*, 53(2):171–193, 2002.

Wolfgang Glänzel and Henk F Moed. Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, 96(1):381–394, 2013.

CH Goh, CW Holsapple, LE Johnson, and J Tanner. An empirical assessment of influences on POM research. *Omega*, 24(3):337–345, 1996.

Borja González-Pereira, Vicente P Guerrero-Bote, and Félix Moya-Anegón. A new approach to the metric of journals scientific prestige: The SJR indicator. *Journal of informetrics*, 4(3):379–391, 2010.

George Emm Halkos and Nickolaos G Tzeremes. Measuring economic journals citation efficiency: a data envelopment analysis approach. *Scientometrics*, 88(3):979–1001, 2011.

Anne-Wil Harzing. Journal Quality List, 52nd Edition, 11 February 2014. URL http://www.harzing.com.

Clyde W Holsapple and Anita Lee-Post. Behavior-based analysis of knowledge dissemination channels in operations management. *Omega*, 38(3):167–178, 2010.

Kurt Hornik, Rainer Jankowitsch, Manuel Lingo, Stefan Pichler, and Gerhard Winkler. Validation of credit rating systems using multi-rater information. *Journal of Credit Risk*, 3(4):1–27, 2007.

John G Kemeny and LJ Snell. *Mathematical Models in the Social Sciences*. MIT Press, Cambridge, 1962.

Christian List. Social Choice Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford, USA, Winter 2013 edition, 2013.

Sebastián Lozano and Jose L Salmerón. Data envelopment analysis of OR/MS journals. *Scientometrics*, 64(2):133–150, 2005.

Guido Masarotto and Cristiano Varin. The ranking lasso and its application to sport tournaments. *The Annals of Applied Statistics*, 6(4):1949–1970, 2012.

Jack R Meredith, Michelle D Steward, and Bruce R Lewis. Knowledge dissemination in operations management: Published perceptions versus academic reality. *Omega*, 39(4):435–446, 2011.

John Mingers and Anne-Wil Harzing. Ranking journals in business and management: a statistical analysis of the Harzing data set. *European Journal of Information Systems*, 16(4):303–316, 2007.

F.H. Moed. The source normalized impact per paper is a valid and sophisticated indicator of journal citation impact. *JASIST*, 62:211–213, 2011.

Josephine E Olson. Top-25-business-school professors rate journals in operations management and related fields. *Interfaces*, 35(4):323–338, 2005.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: bringing order to the Web., 1999.

Charles G Petersen, Gerald R Aase, and Daniel R Heiser. Journal ranking analyses of operations management research. *International Journal of Operations & Production Management*, 31(4):405–422, 2011.

Gabriel Pinski and Francis Narin. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, 12(5):297–312, 1976.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.R-project.org/.

Alexander Serenko and Nick Bontis. Global ranking of knowledge management and intellectual capital academic journals: 2013 update. *Journal of Knowledge Management*, 17(2):307–326, 2013.

Alexander Serenko and Michael Dohan. Comparing the expert survey and citation impact journal ranking methods: Example from the field of Artificial Intelligence. *Journal of Informetrics*, 5(4):629–648, 2011.

Stefan Theußl, Thomas Reutterer, and Kurt Hornik. How to derive consensus among various marketing journal rankings? *Journal of Business Research*, 67(5):998–1006, 2014.

Thomson Reuters. The 2014 Journal Citation Reports® (JCR), Social sciences edition, 2014.

Ryan Joseph Tibshirani. *The solution path of the generalized lasso*. Stanford University, 2011.

Heinz Tüselmann, Rudolf R Sinkovics, and Grigory Pishchulov. Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis. *Omega*, 51:11–23, 2015.

R.J. Vokurka. The relative importance of journals used in operations management research: a citation analysis. *Journal of Operations Management*, 14(4):345–355, 1996.

Hugh Willmott. Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organization*, 18(4):429–442, 2011.

Fang Xu, Wenbin Liu, and Ronald Rousseau. Introducing sub-impact factor (SIF-) sequences and an aggregated SIF-indicator for journal ranking. *Scientometrics*, 102 (2):1577–1593, 2015.

Zhou Xu, Brenda Cheang, Andrew Lim, and Qi Wen. Evaluating OR/MS journals via PageRank. *Interfaces*, 41(4):375–388, 2011.

H Peyton Young. Optimal ranking and choice from pairwise comparisons. *Information pooling and group decision making*, pages 113–122, 1986.

Duanning Zhou, Jian Ma, and Efraim Turban. Journal quality assessment: An integrated subjective and objective approach. *IEEE Transactions on Engineering Management*, 48(4):479–490, 2001.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.

A Journal abbreviations and ranking lists

Table 4: Journal abbreviations

Abbrev	Journal name	Abbrev	Journal name
4OR	4OR – A Quarterly Journal of Operations Research	MILOR	Military Operations Research
AOR	Annals of Operations Research	MMOR	Mathematical Methods of Operations Research
APJOR	Asia Pacific Journal of Operational Research	MOR	Mathematics of Operations Research
ASMB	Applied Stochastic Models in Business and Industry	MP	Mathematical Programming
CEJOR	Central European Journal of Operations Research	MS	Management Science
COA	Computational Optimization and Applications	MSOM	Manufacturing and Service Operations Management
COR	Computers & Operations Research	NAVRL	Naval Research Logistics
DEDS	Discrete Event Dynamic Systems	NETW	Networks
DO	Discrete Optimization	NSE	Networks and Spatial Economics
DSS	Decision Support Systems	OCAM	Optimal Control Applications and Methods
EJOR	European Journal of Operational Research	OE	Optimization and Engineering
EO	Engineering Optimization	OL	Optimization Letters
EXSA	Expert System with Applications	OMEGA	OMEGA - International Journal of Management Science
FODM	Fuzzy Optimization and Decision Making	OMS	Optimization Methods and Software
IIE	IIE Transactions (Institute of Industrial Engineers)	OPT	Optimization
IJITD	International Journal of Information Technology & Decision Making	OR	Operations Research
IJPR	International Journal of Production Research	ORL	Operations Research Letters
IMAJMM	IMA Journal Management Mathematics	ORS	OR Spectrum
INFOR	INFOR: Information Systems and Operational Research	PJO	Pacific Journal of Optimization
INFORMS	INFORMS Journal on Computing	POM	Production and Operations Management
INTER	Interfaces	PPC	Production Planning & Control
ITOR	International Transactions in Operational Research	QS	Queueing Systems
JGO	Journal of Global Optimization	RAIRO	RAIRO – Operations Research
JMS	Journal of Manufacturing Systems	SCL	Systems & Control Letters
JOM	Journal of Operations Management	TNV	Technovation
JORS	Journal of the Operational Research Society	TOP	TOP – An Official Journal of the Spanish Society of Statistics and Operations Research
JOTA	Journal of Optimization Theory & Applications	TRBM	Transportation Research Part B: Methodological
JSCH	Journal of Scheduling	TRELT	Transportation Research Part E: Logistics and Transportation Review
JSIM	Journal of Simulation	TS	Transportation Science

Table 5: Journal ranking lists

Abbrev	Ranking	Source	Year	Scale
IF[1]	Impact Factor	Thomson Reuters	2013	Numeric from 0.
IF5Y[1]	5-Year Impact Factor	Thomson Reuters	2013	Numeric from 0
IMMI[1]	Immediacy Index	Thomson Reuters	2013	Numeric from 0
CHL[1]	Cited Half-Life	Thomson Reuters	2013	Numeric from 0
EFS[1]	Eigenfactor® Score	University of Washington	2013	Numeric from 0 to 1
AIS[1]	Article Influence® Score	University of Washington	2013	Numeric from 0
SNIP[2]	Source Normalized Impact per Paper	Elsevier	2013	Numeric from 0
IPP[2]	Impact per Publication	Elsevier	2013	Numeric from 0
SJR[2]	SCImago Journal Rank	Elsevier	2013	Numeric from 0
Wie01	WU Wien Journal Rating	WU Vienna University of Economics and Business	2001	D < C < B < A < A+
Vhb03	VHB ranking	Association of University Professors of Business in German speaking countries	2003	E < D < C < B < A < A+
Bjm04	British RAE Rankings	Geary et al. (2004)	2004	Numeric from 1 to 7
Hkb05	HKBU ranking	Hong Kong Baptist University School of Business	2005	B- < B < B+ < A
Theo05	Theoharakis integrated ranking	Theoharakis et al. index	2005	Numeric from 1.2 to 95
Ejis07	European Journal of Information Systems ranking	Mingers and Harzing (2007)	2007	1 < 2 < 3 < 4
EjisCI07	European Journal of Information Systems ranking including citation impact factors	Mingers and Harzing (2007)	2007	1 < 2 < 3 < 4
UQ07	University of Queensland Rating	University of Queensland	2007	5 < 4 < 3 < 2 < 1
Ast08	Aston ranking	Aston Business School	2008	1 < 2 < 3 < 4
Wie08	WU Wien Journal Rating	WU Vienna University of Economics and Business	2008	A < A+
ABS10	ABS Academic Journal Quality Guide	Association of Business Schools	2010	1 < 2 < 3 < 4 < 4*
Den11	Danish Ministry Journal list	Danish Ministry	2011	1 < 2
HEC11	HEC Paris Ranking	Hautes Études Commerciales de Paris	2011	C < B < B+ < A
UQ11	Adjusted ERA Rankings List	University of Queensland	2011	4 < 3 < 2 < 1
Vhb11	VHB ranking	Association of University Professors of Business in German speaking countries	2011	E < D < C < B < A < A+
Aeres12	AERES journal list	Agence d'évaluation de la recherche et de l'enseignement supérieur	2012	C < B < A
Cra12	CRA ranking	Cranfield University School of Management	2012	1 < 2 < 3 < 4
EJL12	ERIM Journals Listing	Erasmus Research Institute of Management	2012	S < P A < P < STAR
Abdc13	ABDC Journal Rankings	Australian Business Deans Council	2013	C < B < A < A*
Cnrs13	CNRS ranking	Le Centre national de la recherche scientifique	2013	4 < 3 < 2 < 1 < 1*
Ess13	ESSEC journal ranking	ESSEC Business School Paris	2013	3 < 2 < 1 < 0 < 0+
Fneg13	FNEGE ranking	French Management Association	2013	4 < 3 < 2 < 1 < 1*

[1] based on the Web of Science® database
[2] based on the Scopus® database

Figure 1: Journal quality scores based on the whole sample of rankings – ALASSO estimators based on AIC

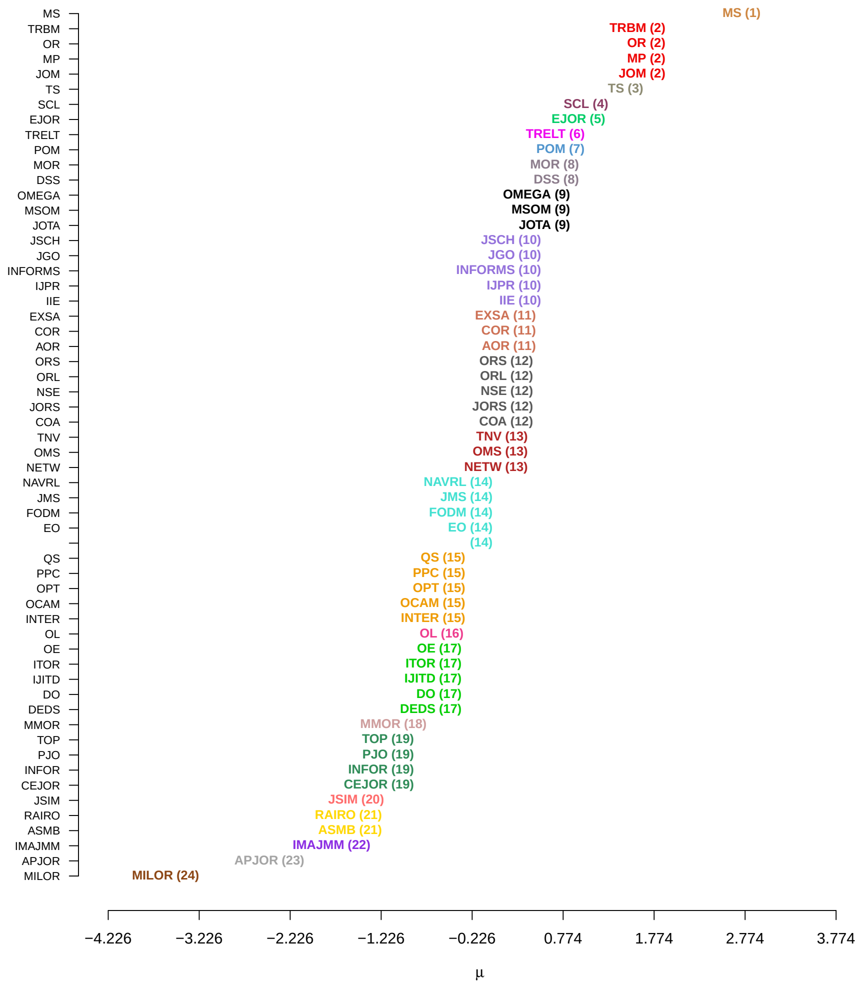

Figure 2: Percentage of times rated versus journal quality scores estimated by maximum likelihood based on the whole sample of rankings. The number of the cluster to which a journal is assigned based on AIC is given in parentheses. Journals in one cluster are marked by the same color.

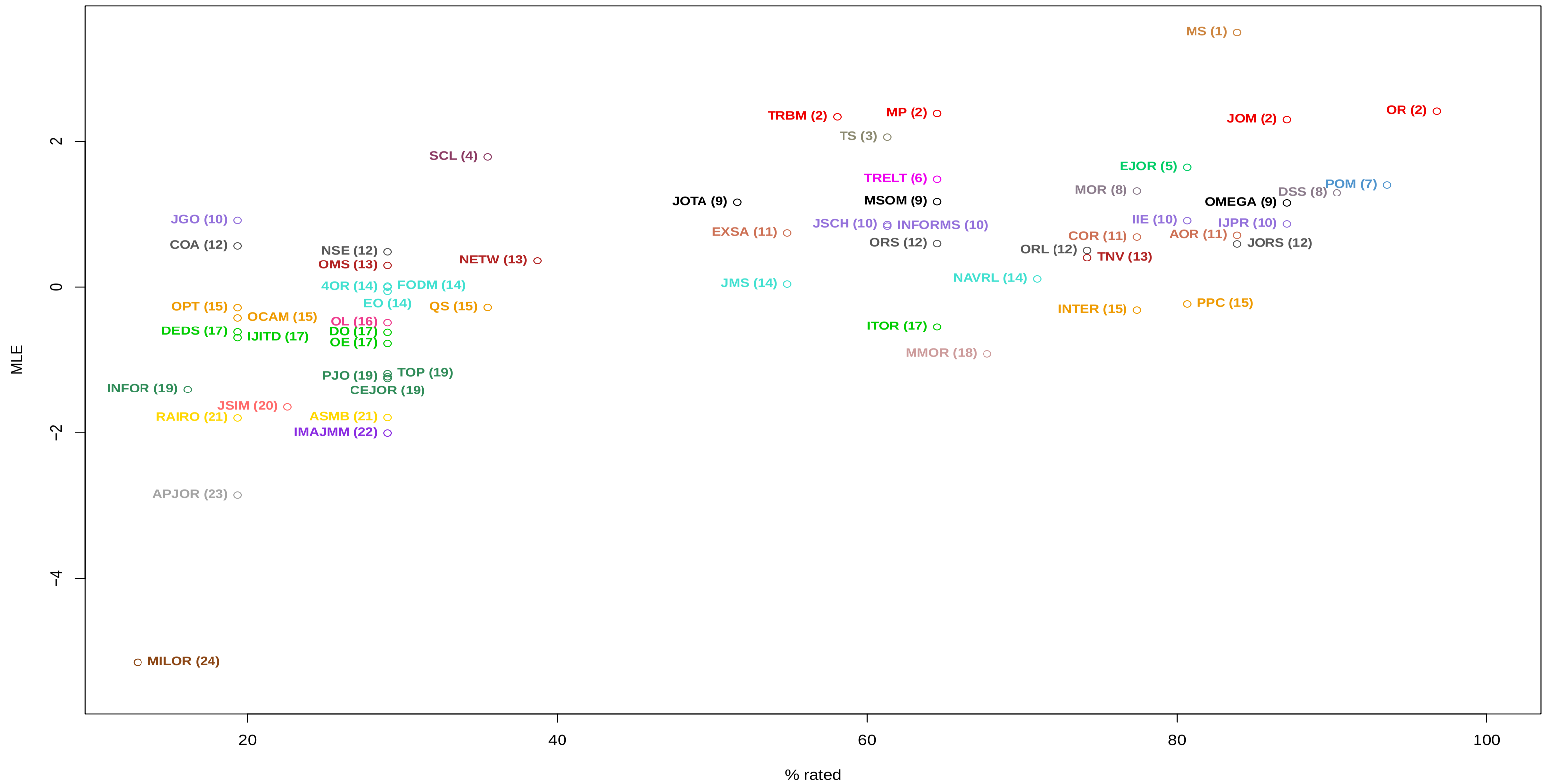

Figure 3: Journal quality scores based on the 2013 rankings– ALASSO estimators based on AIC