



# Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis

Heinz Tüselmann<sup>a,1</sup>, Rudolf R. Sinkovics<sup>b,c,\*</sup>, Grigory Pishchulov<sup>d,2</sup>

<sup>a</sup> Centre for International Business and Innovation (CIBI), Manchester Metropolitan University Business School, All Saints Campus, Oxford Road, Manchester M15 6BH, UK

<sup>b</sup> Centre for Comparative & International Business Research (CIBER), The University of Manchester, Manchester Business School, Booth Street West, Manchester M15 6PB, UK

<sup>c</sup> Lappeenranta University of Technology, Skinnarilankatu 34, PL 20, 53851 Lappeenranta, Finland

<sup>d</sup> Faculty of Business, Economics and Social Sciences, TU Dortmund University, Martin-Schmeißer-Weg 12, 44227 Dortmund, Germany

## ARTICLE INFO

### Article history:

Received 11 July 2013

Accepted 5 August 2014

Processed by Podinovski

Available online 27 August 2014

### Keywords:

Citation indices

Journal rankings

Journal lists

Research assessment

Data envelopment analysis

## ABSTRACT

The question of how to assess research outputs published in journals is now a global concern for academics. Numerous journal ratings and rankings exist, some featuring perceptual and peer-review-based journal ranks, some focusing on objective information related to citations, some using a combination of the two. This research consolidates existing journal rankings into an up-to-date and comprehensive list. Existing approaches to determining journal rankings are significantly advanced with the application of a new classification approach, ‘random forests’, and data envelopment analysis. As a result, a fresh look at a publication’s place in the global research community is offered. While our approach is applicable to all management and business journals, we specifically exemplify the relative position of ‘operations research, management science, production and operations management’ journals within the broader management field, as well as within their own subject domain.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction and objectives

The ranking of academic journals is a highly contentious element of research assessment, and thus a widely debated foundation stone for the ranking of individual research outputs and university rankings [1,2]. As it affects people’s careers and aspirations, the issue is one of perennial topicality and debate. Findings are repeatedly challenged as lists arguably bear non-intended consequences, skew scholarship and foster academic monoculturalism [3], and the methodologies underpinning the various approaches are contested as they are open to non-intended use [4,5]. Within business and management, in recent years we have witnessed an increasing proliferation of rankings, listings and productivity indicators, drawing the attention of a

wide range of academic disciplines, including accounting, economics, finance, international business and marketing [6], of associations such as the Association of Business Schools (ABS and the Association to Advance Collegiate Schools of Business (AACSB), among others, but also that of dominant industry players such as Thomson Reuters’ Web of Science, Elsevier’s Scopus, and Google Scholar. These various parties are distinguished by unique interests. The commercial providers have started to monetize a rapidly expanding and lucrative global intelligence information business by building on the academic ‘gift economy’ [7] – collecting institutional profile information and then selling it back to the institutions for strategic-planning purposes [8]. However, the aim of this paper is not to go into aspects of ‘use and abuse’ or epistemological positions regarding journal rankings [2,4]. Instead, given their broad adoption in today’s academic practice, we address some distinct methodological shortcomings of the previous attempts to rank journals and contribute to the development of a more suitable methodology, which in turn, can be used to gauge the relative standing of individual journals more realistically.

There are three conventional ways of assessing journal quality: (i) subjective (perceptual), (ii) objective (citation-based) and (iii) a combination thereof (hybrid). All three conventional ways feature well-known methodological limitations [9–11]. Recently, a fourth

\* Corresponding author. Tel.: +44 161 305 8980.

E-mail addresses: [h.tuselmann@mmu.ac.uk](mailto:h.tuselmann@mmu.ac.uk) (H. Tüselmann),

[Rudolf.Sinkovics@manchester.ac.uk](mailto:Rudolf.Sinkovics@manchester.ac.uk) (R.R. Sinkovics),

[grigory.pishchulov@tu-dortmund.de](mailto:grigory.pishchulov@tu-dortmund.de) (G. Pishchulov).

URLs: <http://www.business.mmu.ac.uk/cibi> (H. Tüselmann),

<http://www.manchester.ac.uk/research/rudolf.sinkovics> (R.R. Sinkovics),

<http://www.wiso.tu-dortmund.de/scm/> (G. Pishchulov).

<sup>1</sup> Tel.: +44 161 247 390.

<sup>2</sup> Tel.: +49 231 755 3234.

approach has gained momentum – the ‘meta’-ranking approach – which, like the hybrid approach, is intended to provide a balanced view by delivering a composite journal ranking (cf. [12,13]). In contrast to the hybrid studies, which usually combine a few rankings or ratings and often involve the hand-collection of perceptual data, meta-analyses typically rely on a comprehensive selection of existing, in many cases reputable, rankings or ratings, and aim to deliver a reproducible outcome (cf. Table 1). As outlined, the existence of journal rankings is often – justifiably – contested on philosophical grounds, and there is the fundamental question whether possible distortions in terms of scholarship and unintended consequences of ranking exercises (see e.g. [2]) may offset the advantages of increased manageability of scholarly outputs. Indeed, the emergence of meta-rankings can be seen as a result of the sheer volume and range of diverse lists that are – counter to the original motivation for developing them, which was to improve academic resource ‘management’ – proving to be unmanageable outside their respective academic institutions and often include different selections of journals. Within the academic community there seems to be agreement that if rankings are being used, the agenda should be the pursuit of a rigorous and objective perspective, based on state-of-the-art methodologies, free of individual stakeholder interests in this contentious area.

However, despite the advances made by meta-studies, a number of shortcomings remain. These include (i) arbitrary inclusion or datedness of journal lists; (ii) over-reliance on citation data; (iii) limited coverage in terms of disciplinary focus, number of journals and number of lists included; (iv) inadequate treatment of missing data and unsophisticated imputation methods; (v) treatment of ordinal rank data as metric; (vi) choice of ranking categories.

In the present study, we elaborate an approach that addresses these shortcomings while combining the strong features of existing studies, extending these and adding novel features. Therefore, we substantiate the methodological underpinnings to the current debate on journal rankings. We (i) extend recent work and offer an aggregate journal ranking based on a comprehensive number of journals, (ii) cover a significant number of disciplines within business and management, and (iii) deploy a unique methodological approach and integrate subjective and objective rankings with a focus on systematism and the production of comprehensive journal rankings. Specifically, this is the first meta-ranking to feature both the random forests framework (a non-parametric state-of-the-art predictive learning method) for missing data imputation and data envelopment analysis (DEA) (an established non-parametric approach to performance evaluation of peer entities) for the aggregation of rankings. This paper is decidedly focused on the methodological advancement of existing journal rankings. Thus, our final aggregate journal ranking outcomes (see Table 5) can be seen as a frame of reference for a substantive discussion and objectification of journal rankings, which is otherwise rather politicized.

The paper is organized as follows. The next section provides a critical review of objective, subjective and hybrid approaches to journal ranking and rating. Following this, Section 3 provides an overview of the major meta-ranking studies. Subsequently, in Sections 4–6, we present our novel meta-approach to journal ranking and rating, discuss its specific methodological advancements and apply it to our data set of journal rankings and ratings. This involves dealing with issues of database compilation, data missingness and imputation methods, classification trees, random forests and the subjection of the data to DEA (for full modeling and computational details, please refer to our working paper [78]). Section 7 concludes with a discussion of main results of our study and their implications.

With particular emphasis on operations research, management science, production and operations management (OR/MS/POM), we apply the method to ascertain the relative positions of journals

within the broader business and management discipline, as well as the relative position within the OR/MS/POM field.

## 2. Review of objective, subjective and hybrid approaches to journal ranking and rating

With regard to *objective* ranking, issues arise around the analysis of citation data. The Impact Factor delivered by the Journal Citation Reports [14] – defined as the number of cites received in the given year by an average article published in the given journal within a pre-defined number of preceding years – is the most widely accepted citation-based measure for “significance and performance of scientific journals”. It is widely acknowledged for its comprehensibility, robustness and availability [15]. Yet, it has received a considerable amount of criticism in the literature, connected to the accuracy problem in collecting citation data, undifferentiated treatment of citations, biases due to different maturing of published work across different journals, inaccurate definition of citable work and differing citation habits across different sub-disciplines. Further criticism includes biasedness towards journals with lengthy articles [15], see also [16]; and a selective disciplinary and geographical coverage [17,18]. Some of these deficits have recently been addressed by introducing a newer, prestige-oriented metric called Eigenfactor Score [19] which augments the Journal Citation Reports, and the emergence of Scopus – a citation database by Elsevier which offers a broader journal coverage together with new citation indices Source-Normalized Impact per Paper (SNIP) and SCImago Journal Rank (SJR). These aim to account for discipline-related citation habits and the prestige of the citing journals, respectively [20,21]. Yet, and despite these advancements, extensive discussions of the underlying methodological issues raise concern of the sole reliance on citation-based analysis in journal ranking exercises. This is because important work may be considered as “common knowledge” and is sometimes left uncited – with acknowledgment given to other work or citation counts frequently representing simply fashion and herding within the academic community which implicates that citing does not necessarily imply influence [9,22,23]. There are also problems of selective citations and the opportunity for self- and mutual citations, a poor association between the quality of a journal and that of individual articles in it, as well as possible subjectivity which can be pertinent to the analysis based on the objective citation data [5,24,25]. Regardless of these shortcomings, the citation impact factor remains an important indicator in the academic community to assess journal quality.

*Subjective*, or perceptual, rankings are developed via opinion surveys among the experts within an institution, a society, or a research network and may be motivated by the needs to elaborate a basis for institutional decision making and evaluation purposes as well as to provide guidance within particular disciplines [1,26,27]. For these reasons, a variety of rankings exist which are tailored to the needs of a particular institution or a discipline [10,26–28]. Generally, perceptual rankings alleviate the problems pertinent to citation data, and explicitly capture the *perceived* quality of journals [5,29]. On the other hand, they are prone to biasedness in the experts’ judgments – due to the institutional focus or self-identification with particular journals [11,26]. Furthermore, the coverage of perceptual lists is often restricted to a particular discipline or by institutional preferences [26].

Due to the shortcomings of the above two approaches, the *hybrid* lists – which in some way combine subjective and/or objective data – have gained attention in the literature (e.g. [13,29,30]). Indeed, pooling data that originates from different sources helps to produce a more balanced view and is seen as a desired approach [13,27,31]. However, hybrid ranking lists typically have a particular disciplinary or geographical focus; they

**Table 1**  
Overview of journal meta-ranking studies and selected hybrid ranking studies.

	Disciplinary focus	No. of journals	No. of rankings used	Rankings used	Age of rankings	Treatment of missing data	Imputation method	Rating/ranking by	Outcome: rating or ranking (scale type, feasible range, meaning)
<b>Meta-Rankings</b>									
Bancroft et al. [34]	Business policy /strategy	25	5	P	1987–1994	Imputation	Maximum likelihood	Mean rank	Rating (interval 1–4, 4= outstanding as a publication outlet), ranking (ordinal 1–23, 1= top rating)
Benati and Stefani [10]	Mathematics & economics	138	7	JQL, OI	2002–2005	Imputation	Separate category	Cluster analysis	Classification (nominal, 4 classes)
Theußl et al. [33]	Marketing	62	12	JQL	2001–2009	Ignoring		IP	Ranking (ordinal 1–5 <sup>†</sup> , 1= top quality)
Cook et al. [12]	Accounting	140	26	P, JQL, OI, C, U	2002–2007	Ignoring		IP	Ranking (ordinal 1–33 <sup>†</sup> , 1= top quality)
Mingers and Harzing [1]	Business, management and related disciplines	834	10	JQL, C	1994–2005	Partial imputation	Chained regression	Cluster analysis	Ranking (ordinal 1–4, 4= top quality)
Halkos and Tzeremes [22]	Business, management and related disciplines	229	8	C, JQL, OI	2009	Does not apply		DEA	Rating (ratio 0–1, 1= maximum performance in terms of citedness), ranking (ordinal A–D, A= top tier rating)
Franke and Schreier [31]	Tech. & Innovation/ Entrepreneurship	43	37	C, JQL, P, U, OI	1989–2004	Ignoring		Scoring	Rating (ratio 0–10, 10= max. quality), ranking (ordinal 1–39, 1= top rating; ordinal A–D, A= top tier rating)
Rainer and Miller [35]	Management Information Systems	50	9	P, C	1991–2001	Ignoring		Scoring	Rating (ratio 0–1, 0= maximum quality), ranking (ordinal 1–47, 1= top quality rating)
Steward and Lewis [27]	Marketing	100	11	P, C, U	1993–2006	Ignoring		Scoring	Rating (ratio 0–1, 0= maximum quality), ranking (ordinal 1–49 <sup>†</sup> , 1= top quality rating)
<b>Hybrid rankings (selected)</b>									
Zhou et al. [30]	Cross-disciplinary (Hong Kong RAE)	285*	3	C, OI, OS	1996–2000	Does not apply		Fuzzy inference	Ranking (ordinal A–C, A= top quality)
Morris et al. [26]	Business, management and related disciplines	1039**	9	OI, C, OS	2003–2008	Ignoring		Modal score & Delphi	Ranking (ordinal 0*–4*, 4*= top quality)
Crookes et al. [76]	Nursing & midwifery	144	3	OS, C	2006–2007	Does not apply		Scoring	Rating (interval 0–100, 100= maximum quality), ranking (ordinal 1–4, 1= top tier rating)
DuBois and Reeb [77]	International business	30	5	C, OS	1995–1998	Does not apply		Scoring	Ranking (ordinal 1–23, 1= top quality)
Bauerly and Johnson [29]	Marketing (mainly US background)	252	1	U, OS	2001	Does not apply		Does not apply	Rating (ratio 1–1434, citations in doctoral program syllabi), ranking (ordinal 1–34 <sup>†</sup> , 1= top rating)
Kao et al. [13]	Management (Taiwanese journals)	46	5	C, OI, OS	2003–2005	Does not apply		DEA & scoring	Rating (ratio 0–1, 1= maximum quality), ranking (ordinal 1–46, 1= top rating; ordinal A–E, A= top tier rating)

P – Perceptual rankings published in academia.  
 OS – Opinion survey as source of perceptual data.  
 JQL – Cross-disciplinary rankings present in the JQL.  
 OI – Other institutional cross-disciplinary rankings.  
 C – Citation data or citation-based rankings.  
 U – Rankings featuring other usage data (e.g. download counts, citations in syllabi etc.).  
 The symbols are in descending order of the respective rankings' share in the data set.  
 IP – Integer programming.  
 DEA – Data envelopment analysis.

\* Ranking of just a single journal within a single discipline with 285 journals is provided as an illustration.

\*\* As of 21 April 2009.

† As for the number of journals in the reported ranking.

usually combine a few rankings or ratings and involve hand-collection of perceptual data, and, with a few exceptions, use unsophisticated and less principled techniques for data aggregation (cf. [1]).

Because objective, subjective and hybrid approaches have attracted the above criticisms, the meta-approach to journal ranking and rating has recently received a substantial development, being intended to overcome the drawbacks of the hybrid approaches by relying on a comprehensive selection of existing, in many cases reputable, rankings or ratings, and aiming to deliver a reproducible outcome.

### 3. Overview of journal meta-rankings and ratings

Table 1 offers a compilation of the main journal meta-ranking studies. As can be seen, most of these studies focus on particular sub-disciplines, with the exception of Mingers and Harzing [1] and Halkos and Tzeremes [22] who take a cross-disciplinary approach. The journal coverage ranges from 25 to 229, with the exception of Mingers and Harzing [1] who cover over 800 journals. In terms of rankings used, most of the studies draw on a combination of subjective and objective rankings. Two thirds of the meta-rankings are based on journal rankings contained in Harzing's broadly accepted Journal Quality List (JQL) [32].

The number of underlying rankings is often 10 or less. There is quite a spread in terms of the recentness of the rankings, with only two studies covering recent years. As for data missingness, which arises because of selective coverage of journals, either this is not addressed, or it is not dealt with properly in these meta-rankings (see Section 5.1). For Theußl et al. [33] and Cook et al. [12], data missingness is not an issue. They effectively adopt the perspective that only the observed rank data can determine the ultimate ranking. There are a few, varied, attempts to impute missing data: for example, Bancroft et al. [34] employ a maximum likelihood approach, while Mingers and Harzing [1] implement a form of chained regression.

As for the aggregation method for rating/ranking journals, the main approaches used are scoring methods, cluster analysis and consensus ranking via integer programming, with only one study, that of Halkos and Tzeremes [22], featuring the state-of-the-art DEA. While scoring is attractive due to its simplicity, it is rather subjective in its application. Cluster analysis offers a more advanced approach, but usually only delivers a limited set of categories. DEA, in contrast, is a methodologically profound and objective approach that helps to reduce manipulation, over-interpretation and bias. The integer programming approach deployed by Theußl et al. [33] and Cook et al. [12] is very effective at producing a consensus ranking, yet it works within the confines of treating missing data as non-existent. Further, it cannot deliver an interval or ratio scale outcome.

### 4. Compiling a database for journal meta-ranking

In view of the limitations and shortcomings of meta-rankings described above, we proceed to develop a comprehensive journal database, which will subsequently be subjected to our rating and ranking exercise.

The primary databases are the journal quality ranking lists contained in the 49th edition of Harzing's Journal Quality List (JQL49) [32] and the Thomson Reuters Journal Citation Reports ([14], various years):

- The ranking lists contained in JQL49 are dated in the range from 2001 to 2013.

- To reflect an up-to-date, rather than historical, journal status, we select the 10 most recent ranking lists (out of the 22 contained in the JQL49 database), covering a 6-year time span (2008–2012).<sup>3</sup>
- We update and correct a number of the journal lists in JQL49 based on information in the most recent publicly available editions of the respective ranking lists.<sup>4</sup>
- In order to capture a comprehensive quantity of journals, all journals listed in JQL49, a total of 939 journals, are considered. This provides a broad and cross-disciplinary coverage.

The 10 ranking lists selected for aggregation by means of DEA (Section 6) are labeled 'target lists', as shown in Table 2. In an additional step, these rankings are further augmented by including 2011 Impact Factor data from the Journal Citation Reports [14].<sup>5</sup> Thus, we use 11 rankings in total.

Most of the journal quality lists rank the journals on an ordinal scale, using differing numbers of scale gradations (ranks) and their designations. Thus, we relabeled the ranks in each of the lists as 1, 2, etc., from highest to lowest. The length of the original scale is maintained in all lists. This overcomes the problems related to adjusting original scale lengths to a common scale length, and the resulting subjectivity/arbitrariness [31].

In addition, all journals with an Impact Factor are ranked and divided into quintiles, with 1 denoting the top quintile, 5 the lowest quintile, and a value of 6 being assigned to journals that are not indexed in the 2011 Journal Citation Reports. This procedure helps to alleviate several of the well-known shortcomings of using the Thomson Reuters metric score in analyses (cf. [18]), as well as the problems with conventional normalization procedures [26].

### 5. Resolving the data missingness problem in journal rankings and ratings

#### 5.1. Data missingness and imputation approaches

A significant problem pertinent to journal meta-ranking approaches is the considerable amount of missing data. In our database of 939 journals, target ranking lists from 1 to 10 (see Table 2) contain 4770 entries out of the 9390 possible. This corresponds to an overall missingness rate of nearly 50%. The pattern of missingness varies across journals, and coverage rates range from approximately 28% to 88% across lists. As can be seen from Fig. 1, three strategies for dealing with data missingness can be identified in the existing journal ranking studies:

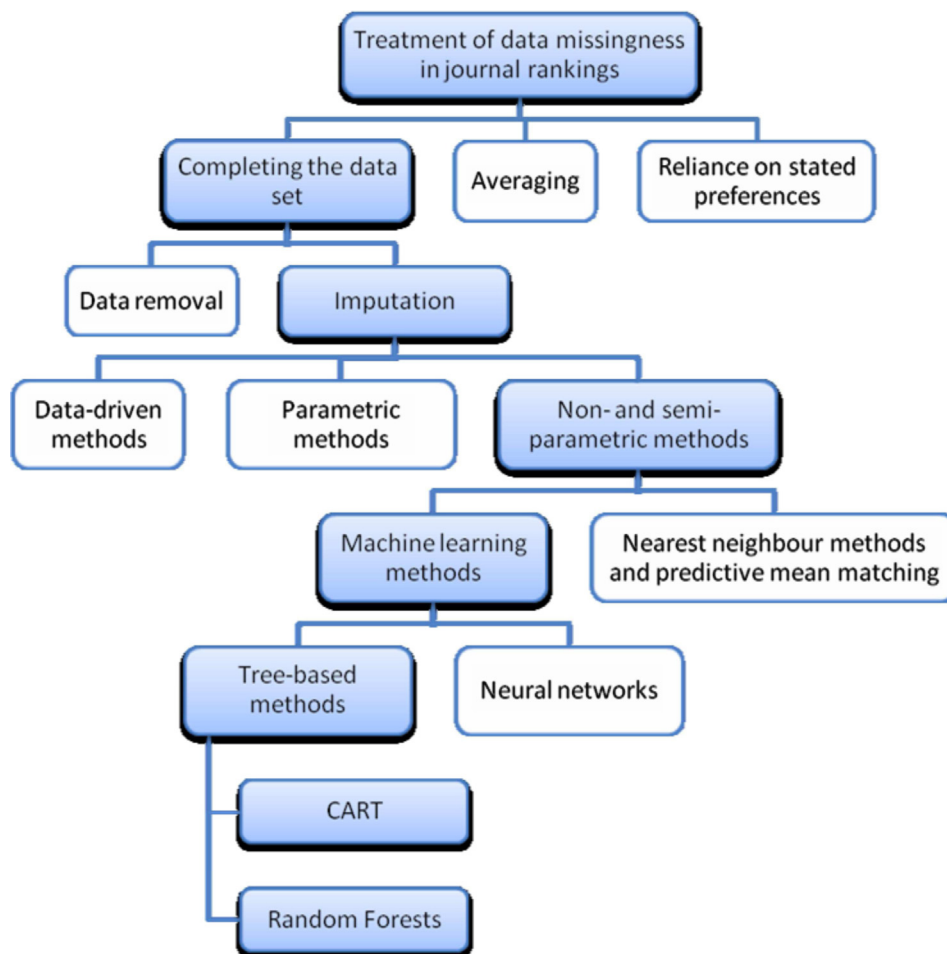
<sup>3</sup> The Wirtschaftsuniversität Wien Journal Rating 2008 (WIE 2008) list was excluded because it now publishes only its A+ and A ratings and no longer its B, C and D ratings. If we had included WIE 2008, journals that it ranked below an A would have been wrongly recorded as missing cases. We also excluded Den 2011 (Danish Ministry Journal List) because it has only two categories: top journal and others, it thus lacks differentiation. We further excluded FNEGE (Foundation National pour l'Enseignement de la Gestion des Entreprises) 2011 because it merely replicates the CNRS (Centre National de la Recherche Scientifique) 2011 ratings for management and business journals. Finally, we excluded AERES (Agence d'évaluation de la recherche et de l'enseignement supérieur) 2012 because it mainly maps CNRS 2011 ratings to a scale with fewer gradations and does not substantially add to the existing data.

<sup>4</sup> We have in particular made corrections in the ranking lists ABS 2010, CNRS 2011, UQ 2011 and HEC 2011. These and other adjustments of the JQL can be obtained from the authors on request.

<sup>5</sup> We use the two-year average of the 2011 Impact Factor [14]. An alternative would have been to use the five-year average. However, for a number of journals, no five-year average data exists. If we had used a five-year average, these journals would have received a non-entry, despite being included in the citation list. The same rationale applies to the exclusion of alternative measures such as the article influence score.

**Table 2**  
Target lists.

No.	Title	Year	Abbreviation
1	Aston	2008	Ast 2008
2	Australian Business Deans Council Journal Ranking List	2010	ABDC 2010
3	Association of Business Schools Academic Journal Quality Guide	2010	ABS 2010
4	Centre National de la Recherche Scientifique	2011	CNRS 2011
5	Hautes Études Commerciales de Paris Ranking List	2011	HEC 2011
6	University of Queensland Adjusted ERA Ranking List	2011	UQ 2011
7	Association of Professors of Business in German-speaking countries	2011	VHB 2011
8	Cranfield University School of Management	2012	Cra 2012
9	ERASMUS Research Institute of Management Journal Listing	2012	EJL 2012
10	ESSEC Business School Paris	2013	ESS 2013
11	Impact Factor from the Thomson Reuters Journal Citation Reports	2011	Thomson Reuters 2012

**Fig. 1.** Approaches to treatment of missing data in journal rankings and methods of completing the data set.

- (1) *Completing the data set*. This can be achieved either by the *removal* of records with missing data – which would however lead to an undesirable loss of information – or by *imputation*. The latter involves replacing missing entries with artificially generated values (see below).
- (2) *Averaging*. For example, Rainer and Miller [35] calculate the average score from the ranks that are available. However, this may lead to biased results (see [34]). The same criticism applies to the work of Franke and Schreier [31] and Steward and Lewis [27] who use a form of weighting to replace the missing data.
- (3) *Reliance on stated preferences*. Cook et al. [12] and Theußl et al. [33] employ an integer programming method that seeks to find the consensus ranking that exhibits the least total deviation from the underlying rankings. They thus neither extrapolate nor disregard existing data. Instead, their approach relies purely on the pairwise preference relations between the journals, effectively stated by the underlying ranking lists. Despite its advantages, we do not use the approach in this paper, instead favoring imputation for the following reasons: firstly, as Mingers and Harzing [1] point out, lists can be biased in their selective coverage and imputation reduces this bias.

Tse [36] supports this, referring to humans' limited information-processing capability (cf. [12]). Secondly, imputation enables us to extend lists while retaining their original spirit [37]. Therefore, this paper considers imputation to be the most viable strategy for dealing with missingness in journal lists.

In line with Farhangfar et al. [38] and Gheyas and Smith [39], three approaches to missing data imputation can generally be identified (see Fig. 1):

- (1) *Data-driven imputation methods* [38]. Missing items are replaced with artificial values, for example the mean, median or mode of the respective variable, or with a random draw from the observed values [39,40]. However, these methods distort the association between variables [40]. In the context of the journal ranking problem, the approach would lead to the distortion of the aggregate ranks of individual journals. While this is partly overcome by Benati and Stefani [10], who associate missing rank data with a separate category, their approach is not tailored to offer a rank ordering of journals.
- (2) *Parametric imputation methods*. These methods assume an explicit data model, such as the *regression imputation* [40,41] or the *maximum likelihood* approach featured by the *expectation–maximization (EM) algorithm* [42]. The *multiple imputation methodology* (see e.g. [40,43]) represents a further advancement but its reliance on the assumed data model can lead to incorrect inferences (e.g. [41,44]) and it should be used with caution [43,45]. With regard to journal rankings, Bancroft et al. [34] use the maximum likelihood approach to arrive at rank estimates for 25 journals related to business policy/strategic management research, previously ranked in a longitudinal study with censoring. Mingers and Harzing [1] use a form of chained regression imputation to estimate missing ranks for a restricted subset of journals drawn from seven ranking lists of the JQL (17th ed.). Similarly, Schulze et al. [37] carry out repeated imputation through a sequential univariate regression and a single imputation through a sequential multivariate regression, while utilizing a number of additional ranking lists as predictor variables (yet they deal only with imputation and do not attempt to derive an aggregate rating or ranking). However, parametric methods have received criticism regarding potential model misspecification and validity concerns [39,41,46].
- (3) For the purposes of our study, we pursue the branch of *non- and semi-parametric imputation methods*, as these do not (or do not fully) rely on a data model [39,41]. A major advancement within this branch is the group of *machine learning* approaches [46], which we draw on for our study (see e.g. [39,47])<sup>6</sup>. In particular, the work by Twala et al. [48] demonstrates the competitiveness of tree-based methods compared to parametric imputation methods in terms of predictive accuracy, see also Hapfelmeier et al. [49]. More specifically, we utilize the *random forests* method [50] which represents a recent and remarkable advancement in non-parametric classification and regression. This method employs an ensemble of classification or regression trees (see Section 5.2) for predicting the response variable as a committee, while the process of constructing the individual trees in the ensemble involves randomness. This approach results in a superior prediction accuracy that compares favorably or competitively 'to the best statistical and machine learning methods' [51–53]. At the same time, the random forests method is deemed more

versatile than the conventional statistical methods and can flexibly accommodate a wide range of prediction problems – even those that are 'nonlinear and involve complex interactions' [53], while being acknowledged, among others, for robustness and ease of training as compared to other machine learning methods [52,53].

## 5.2. Classification trees and random forests and their application

*Classification and regression trees (CART)* represents a well-established and widely used non-parametric predictive learning method [46,52], which has been developed with a strong emphasis on the possibility of missing data among the variables. It seeks to determine the association between the response and predictor variables via recursive, data-driven, partitioning of the predictor space and exhibits a degree of accuracy comparable to the best of the classical statistical methods [54], while producing highly interpretable models and exhibiting other strong advantages [52]. Breiman [50] has advanced CART to produce the random forests framework which effectively reduces variability of individual tree predictions by de-correlating and aggregating them across a tree ensemble, offering as a result a remarkably high prediction accuracy and a number of other advantages [52,53]. Random forests are particularly easy to train, basically requiring to fine tune a few parameters only.

Drawing on the random forest framework, we proceed towards imputing the missing data in each of the target journal-ranking lists<sup>7</sup>. Imputation in each individual list is based on predictor variables which are comprised of: (i) journals' subject areas as per JQL49; (ii) the remaining target lists<sup>8</sup>, (iii) other journal ranking lists included in JQL49, and (iv) Citation Impact Factors from the Journal Citation Reports (see Table 2 and Table 3). Specifically, we utilize ranking lists from 2001 onwards (see Table 3). Although these are older than the cut-off date for the target lists, and are therefore based on more historical data, their inclusion is warranted to improve imputation accuracy.<sup>9</sup>

The first step in the application of random forests is to (i) *pre-impute missing entries* in each single predictor.<sup>10</sup> This task is necessary as the predictor variables themselves have missing values. While random forests have a built-in mechanism for this step, we use CART to accomplish this task.<sup>11</sup> The second step involves (ii) *checking the imputation accuracy* in the target lists using cross-validation (see e.g. [52]). We find differences in the accuracy of the imputations for different ranking lists. For instance, missing values for Ast 2008 are found to be more difficult to predict than missing values in other lists. Additionally, we perform numeric experiments using different settings for CART and

<sup>7</sup> Table 2 exhibits 11 target lists. Imputation has to be carried out in 10 of these.

<sup>8</sup> As indicated in Section 4, target ranking list no. 11 is based on 2011 Impact Factor data [14] and features an ordinal rank scale with a few gradations for the purposes of aggregate ranking. When acting as a predictor variable for missing data imputation, this ranking list however maintains original ratio scale data of the 2011 Impact Factor if the latter is available, and indicates a missing value otherwise.

<sup>9</sup> Although VHB 2011 and UQ 2011 are included in the primary list, VHB 2003 and UQ 2007 are also used for imputation purposes because they use different methodologies, scoring systems or ranking procedures from the newer versions of the lists [for details and a discussion of the VHB and UQ lists, see [32]].

<sup>10</sup> All necessary computations have been conducted in R software environment (version 3.0.0). We have used CART implementation delivered by the R package *rpart* (version 4.1-1) and the implementation of the random forest method delivered by the R package *randomForest* (version 4.6-7).

<sup>11</sup> This approach had to be adopted because *randomForest* package (see footnote 10) does not allow for missing data when predicting an unknown response. Handling such situations is however an inherent feature of CART, thus the said approach has been adopted (see e.g. Hastie et al. [52], p. 333). After pre-imputing the missing values in the predictor variables, we add one dummy variable per each such predictor to indicate whether the respective predictor value is original or has been pre-imputed.

<sup>6</sup> Gheyas and Smith [39] provide an overview of imputation approaches, and in particular those featuring neural networks. However, we do not consider this group of methods in our study, preferring instead a methodology which is more straightforward in its application.

**Table 3**  
Additional lists used for imputation purposes.

No.	Title	Year	Abbreviation
1	Wirtschaftsuniversität Wien Journal Rating	2001	WIE 2001
2	Association of Professors of Business in German-speaking countries	2003	VHB 2003
3	British Journal of Management (Business & Management RAE rankings)	2001	BJM 2004
4	Theoharakis et al.	2005	Theo 2005
5	Hong Kong Baptist University School of Business	2005	HKB 2005
6	European Journal of Information Systems	2007	EJIS 2007
7	European Journal of Information Systems (including citation impact factors)	2007	EJIS-CI
8	University of Queensland Journal Rating	2007	UQ 2007
9–14	Impact Factor from the Thomson Reuters Journal Citation Reports	2005 to 2010	Thomson Reuters 2006–2011

random forests to determine the optimal parameter settings for the imputation engine. The third step is (iii) *the actual imputation of missing data* in the target lists. Having regard for misprediction rates in all of the target lists in step *ii*, we find that it would be inappropriate to stick to the point estimates of missing rank data; instead, the uncertainty involved must be reflected in rank predictions. We therefore adopt, similarly to Zhou et al. [30], a fuzzy rank approach – by letting each journal belong to two or more different ranks within the same ranking list, while the respective degrees of rank membership are required to sum up to unity (e.g. in ABS 2010, journal X is 60% associated with rank ‘1’ and 40% with rank ‘2’). A particular advantage of this approach is that our aggregate ranking method (see Section 6 below) accommodates fuzzy rank membership in a natural way.

Notably, random forests have a built-in mechanism for estimating individual rank probabilities when making a prediction. We accordingly adopt these probabilities as the respective degrees of rank membership predicted for the given journal in the given ranking list. Random forests exhibited a superior performance in producing such estimates [55]; however, that performance can be further improved by means of *calibration* techniques. For this purpose we have employed the calibration method suggested by Boström [56] and similarly used the *Brier score* (mean squared deviation of the predicted rank probabilities from the true ones) as performance measure, while the calibration data set has been comprised of all test data samples which had been formed in the course of cross-validations conducted in step *ii*. In our experience, calibration has yielded only a marginal improvement of the *Brier score*, which is in line with Niculescu-Mizil and Caruana [55]. By completing this step we have produced a comprehensive and complete data set, which is then subjected to DEA.

## 6. Rating and ranking journals by DEA

DEA [57] represents an established management science approach to multi-attribute rating of peer entities [58–60], in our case journals. A typical DEA setup involves measuring the efficiency of a number of peer entities called *decision-making units*, or DMUs (e.g. universities) that have a number of common inputs (e.g. budgets, number of staff) and outputs (e.g. research outputs, teaching quality). These inputs and outputs constitute the basis for evaluating the efficiency of the DMUs. There are no *a priori* weights attached to the inputs and outputs. Instead, DEA offers each DMU an opportunity to cross-evaluate and apply input and output weights that most favorably express its own efficiency. Essentially, DEA determines ‘frontiers rather than central tendencies’ in the data [58], [61]. As a non-parametric method, it requires no *a priori* assumptions on the interaction between the variables in the data set [58].

Conventionally, the DEA methodology is applied to metric data, but it has been extended to cover a variety of settings with ordinal

rank data (see [62] for a recent discussion). Cook et al. (see e.g. [63]) further addressed settings with a differentiated treatment of individual rankings – an approach that particularly suits the aggregate journal rating purposes. Against this background, DEA treats rank positions in individual ranking lists as outputs of the DMUs (i.e., journals) while assuming away any variable inputs. It then allows each journal to attach weights to the individual rank positions in each target ranking list. These rank weights should represent the respective journal in the best possible light or, more specifically, provide it with the maximum possible weighted average rank, representing the journal's self-rating of its own performance. Furthermore, the weights chosen by the journal also determine performance ratings of all other journals from its perspective. Thus, by choosing its own rank weights, each journal explicitly evaluates itself vis-à-vis all other journals. In this way, a cross-evaluation matrix is obtained, from which the ultimate ratings of the individual journals can be derived [64,65].

Due to the DEA's advantage of avoiding *a priori* assumptions and subjective bias, we adopt the above approach to derive an aggregate journal rating and ranking. To this end, we employ the DEA framework for aggregation of ordinal preferences by Green et al. [64] while further extending it to include a rank discrimination threshold in line with Noguchi et al. [65] and a differentiated treatment of individual rankings as in Cook et al. [63]. In addition to that, we enforce convexity constraints on the rank weights in line with Hashimoto [66]. Further, we use the aggressive form of cross-evaluation [64] to give each journal the opportunity to appear most strongly against its peers, and derive the ultimate journal ratings from the cross-evaluation matrix using the arithmetic means so that all journals have an equal say in determining the final result. Appendix A describes our modeling approach in detail.

As explained in Section 4, we subject 11 target ranking lists to the above aggregation procedure, while the missing rank data has to be imputed in these lists by means of the random forests method as per Section 5.2, and supplied to DEA in the form of fuzzy membership degrees to which the respective journal is associated with the individual ranks of the respective ranking list. This represents a distinctive feature of our model as compared to DEA approaches to ordinal rank data [62–66]. Random forests method can impute the fuzzy rank membership in a natural way and ordinal DEA can also accommodate fuzzy rank data. Thus, these two approaches are complementary to each other for the purposes of aggregate journal rating.

Before proceeding with DEA, we exclude from the final list of journals used in this study those journals with ranks available for less than 25% of the 11 target lists (see Table 2). This reduces the list from 939 to 786 journals, representing around 84% of all journals in JQL49. This approach is taken because the ranks that are available for sparsely ranked journals may not be representative enough, and it also ensures that the imputations are ‘pluralistic’ enough rather than being based on just one or two

rankings. Our conservative choice of this lower limit of 25% for the number of original rankings per journal is in line with previous related studies, such as Cook et al. [12] and Theußl et al. [33].<sup>12</sup>

A particular problem in attaching weights to the individual journal ranks in our DEA exercise is the arbitrary choice of a rank discrimination threshold to separate the weights of any two consecutive ranks (see [62,64,65,67]). If the threshold is virtually '0', this leads to the undesirable suggestion that there may be no difference between any pair of journal ranks. If the threshold value is set to the maximum, this infringes on the spirit of DEA, since it largely restricts the freedom of choice in determining the rank weights [64]. We resolve this dilemma by setting up the process so that journals settle on an intermediate value of the threshold via Nash bargaining [68].<sup>13</sup> Accordingly, we find the compromise value of the threshold to be 31.3% of the maximal possible value. We then use DEA to rate the journals, producing in effect rating scores in the range from 0.55705 to 1, which yield 729 unique ranks, with 786 tied ranks. Tables 4 and 5 offer a selection of the results.

We also conduct a series of tests to address the sensitivity of the final rating to the choice of the rank discrimination threshold. We find that the results differ across the entire range of feasible threshold values – with Pearson correlations among the corresponding ratings ranging from 80.4% to 100% and Spearman rank correlations from 79.7% to 100%. At the same time, the rating remains robust in the proximity of the selected threshold value; neither of the above two correlation measures falls below 99.97% within the range of  $\pm 10\%$  around the selected threshold value. The final rating exhibits a Pearson correlation of 88.2% and a Spearman rank correlation of 89.9% with the rating produced by means of the Borda count – a points-based system that specifies equidistant weights for the individual ranks in each of the ranking lists.

## 7. Conclusion and implications

The debates over the use and abuse of journal rankings are heated and have recently heightened in their intensity. Much of the effort in the scholarly exchange regarding these rankings is concerned with the construction and publication of list data. However, fundamental issues related to epistemological positions and their implications for scholarly exchange and the scientific production system [71] are still to be resolved [72]. This paper empathizes with these concerns and criticisms in relation to issues such as the homogenization of research cultures, the reduction of pluralism, the skewness of scholarship and the polarization and entrenchment of orthodoxies [2], to mention just a few. Notwithstanding the importance of the wider and philosophical discourse, the main contribution of this paper is a methodological one, driving the advancement of journal rankings. Our position is that,

<sup>12</sup> We found that the final results remain robust when this lower limit is set to a higher value, e.g. 35%.

<sup>13</sup> To be specific, we consider a bargaining problem with  $n=786$  players [68] where journals are acting as players. The utility that a journal attaches to a particular threshold value is taken to be its own standing in the DEA rating that arises under this threshold value. A journal's standing is defined as the difference between this journal's rating score and the average one across the list, normalized to account for the length of the rating scale. The analytic form of each journal's utility function is obtained by fitting a cubic polynomial to 10 equally spaced data points computed for each journal within the feasible range of the threshold. Further, instead of using the disagreement point in the sense of the original Nash bargaining problem, we refer to the minimum utility point [69] – where, accordingly, a journal's minimum utility is its lowest possible standing throughout the entire feasible range of the threshold. The bargaining solution is then determined as the threshold value that maximizes the Nash product over the entire feasible range (see also [70]).

if journal rankings are here to stay, we better pursue a rigorous perspective based on state-of-the-art methodologies that transcend the individual stakeholder interests in this contested field.

With this paper we provide a meta-ranking that overcomes some of the specific shortcomings of the existing meta-rankings in terms of the construction of the underlying database, the treatment of missing data and the ranking approach. To the best of our knowledge, this is the first study to go beyond previous ranking snapshots, and to uniquely feature a combined application of the random forest framework and DEA, two established non-parametric methods, in the construction of the aggregate list. This makes our study wholly non-parametric and therefore free from subjective a priori assumptions about the interaction between the various ranking and rating data included in the study. In this process, we ensure that we retain the strong features of existing and relevant methods, extend them and add novel features (such as fuzzy rank membership and rank discrimination via Nash bargaining) so as to arrive at a 'state-of-the-art' meta-ranking. Confidence in our findings is established through a series of extensive robustness checks, and reliability and cross-validation procedures. However, despite the recency of our methodological approach, future work may still direct its attention towards some possible extensions. For example, it could be explored whether a form of 'discounting' or weighting should be introduced for the imputed journal ranks, due to their omission from the original ranking studies. In our research, they are treated on an equal basis to the existing ranks.

Table 4 offers a selection of the final aggregate journal ranks. We deliberately refrain from making any judgement as to the quality of various ranks, or the 'star-rating' of certain journals, as is frequently found in other ranking lists. We simply provide a rank-ordering of the journals along with their numerical ratings, leaving stakeholder or user groups to arrive at their own subjective judgments regarding the cut-off points for quality grades. There are also a number of useful applications of this list. It allows for the relative standing of a particular journal to be ascertained vis-à-vis all other journals, as well as within its own subject area.

Based on our meta-ranking, Table 5 highlights the ranking order of journals within the OR/MS/POM domain. *Management Science*, *Journal of Operations Management* and *Operations Research* occupy the top-three positions. This discipline is represented over proportionally well when looking at the top 50% of all the journals within the business and management area. Overall, the journals in this subject area perform well vis-à-vis other disciplines that are included in our meta-ranking. Table 5 also offers a look at the relative position of journal outlets from a disciplinary perspective. OR/MS/POM journals account for about 10% of all the 786 journals in the final list of our meta-analysis. They account for around 10% of the top 5% of the whole journal list (*Management Science*, *Journal of Operations Management* and *Operations Research*, *Journal of the Royal Statistical Society: Series B*) and for around 12% of the top 10% and top 20%. Journals up to a tied rank of 19 in Table 5, including *Decision Sciences*, *Risk Analysis*, *European Journal of Operational Research* and *Omega* fall within the top quartile of their own subject discipline and within the top 20% of all management and business journals (see Tables 4 and 5). On the other hand, OR/MS/POM journals account for only around 7% of both, the lower third and lower quartile of journals in our meta-ranking list.

Besides, our meta-ranking may also serve as a reference point onto which the grade and/or star-rating of a particular journal or the population of journals in other lists (e.g. ABS, VHB, Cranfield) can be mapped (see Table 5). This allows pinpointing whether there is congruence between the journal grading of those lists and the results of our meta-approach.

Since we have deliberately refrained from attaching grade categories to our journal rankings, the interpretation of such a



**Table 4**  
Aggregate journal ranks: selected results (N=786).

Journal	Subject area	Rating	Ranking	Tied rank
Academy of Management Review	General Management & Strategy	1	1	1
Administrative Science Quarterly	General Management & Strategy	1	1	1
Journal of Finance	Finance & Accounting	1	1	1
Journal of Marketing	Marketing	1	1	1
Quarterly Journal of Economics	Economics	0.99853	2	5
Journal of Political Economy	Economics	0.99832	3	6
Econometrica	Economics	0.99701	4	7
American Economic Review (The)	Economics	0.99411	5	8
Accounting Review (The)	Finance & Accounting	0.98425	6	9
MIS Quarterly	MIS, KM	0.98425	6	9
Strategic Management Journal	General Management & Strategy	0.98425	6	9
Academy of Management Journal	General Management & Strategy	0.98163	7	12
Information Systems Research	MIS, KM	0.98163	7	12
Journal of Consumer Research	Marketing	0.98163	7	12
Journal of Financial Economics	Finance & Accounting	0.98163	7	12
Marketing Science	Marketing	0.98163	7	12
Review of Financial Studies	Finance & Accounting	0.98163	7	12
Journal of Economic Literature	Economics	0.98101	8	18
Journal of Applied Psychology	Psychology	0.97685	9	19
Accounting, Organizations and Society	Finance & Accounting	0.96588	10	20
Journal of Accounting & Economics	Finance & Accounting	0.96588	10	20
Journal of Accounting Research	Finance & Accounting	0.96588	10	20
Organization Science	OS/OB, HRM/IR	0.96588	10	20
American Journal of Sociology	Sociology	0.96549	11	24
Annual Review of Psychology	Psychology	0.95627	12	25
Management Science	OR, MS, POM	0.95627	12	25
Journal of Marketing Research	Marketing	0.95361	13	27
American Sociological Review	Sociology	0.95354	14	28
American Political Science Review	Public Sector Management	0.95274	15	29
Journal of International Business Studies	International Business	0.95187	16	30
Organizational Behavior and Human Decision Processes	OS/OB, HRM/IR	0.95187	16	30
Journal of Operations Management	OR, MS, POM	0.94574	17	32
American Journal of Public Health	Economics	0.94227	18	33
Review of Economic Studies	Economics	0.93991	19	34
Journal of Economic Perspectives	Economics	0.9382	20	35
Operations Research	OR, MS, POM	0.9379	21	36
Journal of the American Statistical Association	Economics	0.93394	22	37
Organization Studies	OS/OB, HRM/IR	0.93087	23	38
American Psychologist	Psychology	0.93029	24	39
Journal of the Royal Statistical Society, Series B	OR, MS, POM	0.92495	25	40
Research Policy	Economics	0.92386	26	41
Journal of Financial & Quantitative Analysis	Finance & Accounting	0.92215	27	42
Annals of Statistics	OR, MS, POM	0.91961	28	43
Journal of Management Studies	General Management & Strategy	0.91688	29	44
Journal of Retailing	Marketing	0.91512	30	45
Journal of Personality & Social Psychology	Psychology	0.91436	31	46
Review of Economics & Statistics	Economics	0.91409	32	47
Annual Review of Sociology	Sociology	0.90852	33	48
Journal of Development Economics	Economics	0.90058	34	49
Journal of Monetary Economics	Economics; Finance & Accounting	0.89924	35	50
<b>First ranked journals within subject areas, outside top 50</b>				
Business History	Business History	0.6963	214	126
Journal of Communication	Communication	0.7647	111	126
Journal of Business Venturing	Entrepreneurship	0.88189	45	60
Journal of Product Innovation Management	Innovation	0.80578	83	98
Annals of Tourism Research	Tourism	0.83548	65	78

Abbreviations: OR, MS, POM=Operations Research, Management Science, Production & Operations Management; MIS, KM=Management Information Systems, Knowledge Management; OS/OB, HRM/IR=Organization Behavior/Studies, Human Resource Management/Industrial Relations [32].

**Table 5**  
Ranking position of journals within the subject area Operations Research, Management Science, Production & Operations Management (OR/MS/POM).

Journal	Rating	Ranking	Tied rank	Rank within subject area	ABS 2010
Management Science	0.95627	12	25	1	4
Journal of Operations Management	0.94574	17	32	2	4
Operations Research	0.9379	21	36	3	4
Journal of the Royal Statistical Society, Series B	0.92495	25	40	4	4
Annals of Statistics	0.91961	28	43	5	
Transportation Research Part B: Methodological	0.88095	46	61	6	4
Decision Sciences	0.83024	68	83	7	3
Transportation Research Part A: Policy & Practice	0.81931	72	87	8	3
Risk Analysis	0.77628	102	117	9	4

Table 5 (continued)

Journal	Rating	Ranking	Tied rank	Rank within subject area	ABS 2010
Mathematical Programming	0.77052	106	121	10	3
Annals of Probability	0.76941	107	122	11	
European Journal of Operational Research	0.76367	116	131	12	3
IEEE Transactions on Intelligent Transportation Systems	0.76316	117	132	13	
Journal of the Royal Statistical Society, Series A	0.75692	123	138	14	3
SIAM Journal on Control & Optimization	0.74447	134	149	15	
Transportation Science	0.74422	136	151	16	3
IEEE Transactions on Engineering Management	0.73753	137	152	17	3
OMEGA – International Journal of Management Science	0.73751	138	153	18	3
Production and Operations Management	0.73667	139	155	19	3
Biometrika	0.73	150	166	20	
Mathematics of Operations Research	0.7298	152	168	21	3
International Journal of Production Research	0.72265	162	178	22	3
International Journal of Operations & Production Management	0.71567	176	194	23	3
Journal of Business Logistics	0.70445	197	215	24	2
Manufacturing and Service Operations Management	0.70421	200	218	25	3
Transportation Research Part C: Emerging Technologies	0.70373	201	219	26	
Journal of the Operational Research Society	0.69812	211	229	27	3
Transportation Research Part E: Logistics	0.69586	216	236	28	3
Journal of Scheduling	0.69383	222	242	29	3
International Journal of Production Economics	0.69378	223	243	30	3
Journal of Optimization Theory & Applications	0.69018	238	259	31	
Journal of Transport Geography	0.68921	242	263	32	2
Transportation Research Part D: Transport & Environment	0.68609	253	277	33	2
Reliability Engineering & System Safety	0.68578	254	278	34	3
Journal of Supply Chain Management	0.68438	260	284	35	1
Computers & Operations Research	0.68327	263	287	36	2
Service Industries Journal	0.68288	265	289	37	2
Supply Chain Management: An International Journal	0.67861	282	307	38	3
OR Spectrum	0.67193	300	329	39	2
Advances in Applied Probability	0.66825	319	349	40	
Operations Research Letters	0.666	324	355	41	2
Journal of Productivity Analysis	0.66502	327	358	42	3
Naval Research Logistics	0.66481	329	360	43	3
INFORMS Journal on Computing	0.66231	342	376	44	3
International Journal of Human-Computer Studies	0.66154	344	378	45	3
Annals of Operations Research	0.65699	359	393	46	2
Applied Statistics: Journal of the Royal Statistical Society Series C	0.6569	360	396	47	
Transportation	0.65667	361	397	48	2
Theory and Decision	0.64877	391	430	49	2
American Statistician	0.64865	393	432	50	
Production Planning & Control	0.64473	417	458	51	3
Interfaces	0.6436	425	468	52	2
Journal of Combinatorial Optimization	0.64295	429	473	53	1
Transport Reviews	0.64287	432	476	54	2
Research Technology Management	0.64176	435	479	55	
Queueing Systems	0.641	440	484	56	
International Journal of Project Management	0.63986	445	490	57	2
Journal of Multivariate Analysis	0.63637	462	507	58	
Computers & Industrial Engineering	0.63589	464	509	59	2
International Journal of Physical Distribution & Logistics Management	0.63518	470	515	60	2
Journal of Manufacturing Systems	0.62404	527	579	61	
Mathematical Methods of Operations Research	0.62394	529	581	62	
Quality & Quantity	0.62121	543	596	63	
Journal of Purchasing and Supply Management	0.62039	546	599	64	2
Journal of Service Management	0.61963	549	602	65	2
Industrial Management and Data Systems	0.61871	553	606	66	1
International Journal of Flexible Manufacturing	0.61455	574	630	67	2
International Journal of Logistics: Research and Applications	0.61056	596	652	68	2
International Journal of Logistics Management	0.60808	609	665	69	2
Journal of Manufacturing Technology Management	0.60211	636	692	70	2
Quality Management Journal	0.60151	638	694	71	
International Transactions in Operational Research	0.59995	642	698	72	2
Journal of Multi-Criteria Decision Analysis	0.59936	648	704	73	
Business Process Management Journal	0.59595	658	714	74	1
International Journal of Quality & Reliability Management	0.5937	669	725	75	2
Total Quality Management & Business Excellence	0.59064	674	730	76	2
International Journal of Manufacturing Technology & Management	0.5893	677	733	77	
Benchmarking: An International Journal	0.57895	704	761	78	1
Knowledge and Process Management	0.57747	707	764	79	1

Note: The ABS ranking scale has four quality ratings ranging from 4 to 1. The 4 category comprises journals that publish the most original and best executed research, the 3 category journals that publish original and well executed research papers and are highly regarded, the 2 category journals that publish original research of acceptable standards and the 1 category journals that publish research of recognized standards. For a full specification of the journal quality grades, see ABS – Academic Journal Quality Guide, Version 4, 2010.

comparison lies in the eye of the beholder. However, if we were to find gross discrepancies between our meta-ranking and other journal ranking lists, this would not be easy to argue away. Instead, it may serve as an invitation to the authors of the journal list in question to revisit their assessment and ameliorate such discrepancies. While our list is certainly not a panacea, we introduce a ‘dose of objectivity’ into some of the issues picked up in the wider debates on journal rankings, such as vested interests, gamesmanship and politicking. To this end, we hope to contribute to shifting the discussion back towards the essence of scholarly endeavors, namely the development of interesting and relevant contributions.

**Acknowledgment**

We are grateful for comments received from participants of the session “Learning: Methods and Algorithms II” at the 26th EURO–INFORMS Conference on Operational Research in July 2013. We are also grateful to Editor-in-Chief Professor Ben Lev and three anonymous reviewers for constructive and highly insightful comments on the paper throughout the review process. Financial support from the Economic and Social Research Council (ESRC), UK, who funded part of Rudolf Sinkovics’ time [Grant number RES-075-25-0028], is gratefully acknowledged. We also appreciate financial support from the Leverhulme Trust who funded the work of Grigory Pishchulov on this paper during his Leverhulme Overseas Fellowship at Manchester Metropolitan University.

**Appendix A. DEA model for aggregate journal rating**

Let  $T = \{1, \dots, 11\}$  comprise the target ranking lists as per Table 2, and let  $R_\ell$  denote the number of rank gradations in the list  $\ell \in T$ . As explained in Section 6, we exclude from the aggregate rating those journals which are ranked only sparsely across the target lists  $T$ . Let  $J \subset \{1, \dots, 939\}$  be the remaining set of journals ( $|J| = 786$ ), and let  $\hat{f}_{jke}$  represent the degree of membership of journal  $j$  to rank  $k$  in the list  $\ell$  ( $j \in J, k \in \{1, \dots, R_\ell\}, \ell \in T$ ). Note that  $\hat{f}_{jke} = 1$  if journal  $j$  belongs originally to rank  $k$  in the list  $\ell$ ; if journal  $j$  is not scoring in the list  $\ell$  then  $\hat{f}_{jke}$  is determined for  $k = 1, \dots, R_\ell$  by the imputation procedure described in Section 5.2. It holds  $\sum_{k=1}^{R_\ell} \hat{f}_{jke} = 1$  for each  $j \in J$  and  $\ell \in T$ .

We employ the ordinal DEA framework proposed by Green et al. [64] to derive an aggregate rating of journals in  $J$  as follows. In the spirit of DEA [58], each journal  $j \in J$  is being given the opportunity to determine rank weights  $w_{ke}$  ( $k \in \{1, \dots, R_\ell\}, \ell \in T$ ) that would maximize its own rating defined in terms of the weighted average rank:

$$\theta_{jj} := \max_{\{w_{ke}\}} \sum_{\ell \in T} \sum_{k=1}^{R_\ell} w_{ke} \hat{f}_{jke} \tag{A.1}$$

subject to:

$$\sum_{\ell \in T} \sum_{k=1}^{R_\ell} w_{ke} \hat{f}_{ike} \leq 1 \quad \forall i \in J \tag{A.2}$$

$$w_{ke} - w_{k+1,\ell} \geq w_{k+1,\ell} - w_{k+2,\ell} \quad \forall \ell \in T, k = 1, \dots, R_\ell - 2 \tag{A.3}$$

$$w_{\ell,R_\ell-1} - w_{\ell,R_\ell} \geq \varepsilon \quad \forall \ell \in T \tag{A.4}$$

$$w_{\ell,R_\ell} \geq \varepsilon \quad \forall \ell \in T \tag{A.5}$$

Rank weight  $w_{ke}$  can be interpreted as the ‘worth of being ranked in  $[k]$ th place’ in the ranking list  $\ell$  [63]. Constraints (A.2) represent the usual DEA constraints expressing in their left-hand

sides the respective rating score of each individual journal in  $J$  under the rank weights chosen by the given journal  $j$  and therefore requiring that none of the journals can attain a score higher than 1. Constraints (A.3)–(A.5) are the weak convexity constraints imposed on the rank weights in each ranking list  $\ell \in T$  which essentially require that the difference between two consecutive ranks expressed in terms of their weights is at least as large as the difference between two respectively lower consecutive ranks. Including convexity constraints in the DEA model of Green et al. [64] has been suggested by Noguchi et al. [65]; however, their variant of convexity constraints received criticism [62] which we share and therefore include convexity constraints (A.3)–(A.5) in the weak form as advocated by Hashimoto [66]. These constraints also ensure that rank weights are nonnegative and non-decreasing from the lowest rank ( $R_\ell$ ) to the highest (1) in each ranking list  $\ell \in T$ . The nonnegative constant  $\varepsilon$  represents the *rank discrimination threshold* which particularly determines the minimum amount by which the weights of any two consecutive ranks have to differ. Note that (A.1)–(A.5) represent a DEA model with  $|J|$  DMUs, each having  $\sum_{\ell \in T} R_\ell$  outputs and one constant input equal to unity, and *assurance region* (cf. [66, 73,74]).

Note that model (A.1)–(A.5) represents a further departure from the approach adopted in [64–66] in the following two important aspects. Firstly, following Cook et al. [63], we keep rank memberships differentiated by ranking list  $\ell \in T$  – what is explained by different number of rank gradations in different ranking lists and different meaning attached to them. Journals are therefore allowed to differentiate between ranking lists by assigning weights to the rank positions in each list separately. As a result, (A.1)–(A.5) comprise a linear optimization problem with 51 variables and 837 constraints. Note that the individual ranking lists represent *ranking criteria* in the terminology of Cook et al. [63]. Their DEA model further introduces additional constraints on the magnitude of rank weights to discriminate between them across the ranking criteria of different priority. We deliberately assume all target ranking lists to be of equal importance for the purposes of aggregate journal rating and therefore forego these constraints. The model by Cook et al. [63] also allows to distinguish between the criteria of different degree of *clearness* – so that a clearer ranking list would have a stronger rank discrimination in (A.3)–(A.5) than a less clear one. We however assume all target ranking lists to be of equal degree of clearness and apply therefore the same rank discrimination threshold to all  $\ell \in T$ . The second distinctive feature of our model consists in introducing fuzzy rank memberships to accommodate the uncertainty associated with imputations of missing rank data.<sup>14</sup>

Solving (A.1)–(A.5) for each  $j \in J$  and applying an aggressive form of cross-evaluation [64, eq. (9)], we obtain rank weights  $\{w_{ke}^{(j)} \mid k = 1, \dots, R_\ell, \ell \in T\}$  most favorable to the respective journal  $j \in J$ . The left-hand sides of constraints (A.2), when being evaluated under rank weights  $\{w_{ke}^{(j)}\}$ , accordingly express *peer-ratings*  $\{\theta_{ji}\}$  of journals  $i \in J$  by journal  $j$  for  $i \neq j$  and its *self-rating*  $\theta_{jj}$  for  $i = j$  – equal to the optimal objective value in (A.1). The aggregate rating of journal  $i \in J$  is finally obtained as  $A_i = (1/|J|) \cdot \sum_{j \in J} \theta_{ji}$ .

<sup>14</sup> Note further that the DEA approaches developed in [64–66] are intended to produce a rank order of the candidates in a preferential election, what assumes that each voter ranks only a predefined number of the candidates most preferred by him or her, but not necessarily all of them – so that the input data may only partially reflect the voters’ preferences. This can lead to instability of the resulting ranking with respect to the perturbations in the voters’ preferences towards lowly ranked candidates. See [75] for an approach addressing that problem, and the critical review in ([62, Section 5]). In our setting, however, the input data represents complete preferences of the voters (i.e., journal rankings) with respect to the candidates (i.e., journals), what counters the said instability.

## References

- [1] Mingers J, Harzing A-W. Ranking journals in business and management: a statistical analysis of the Harzing data set. *European Journal of Information Systems* 2007;16(4):303–16.
- [2] Willmott H. Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organization* 2011;18(4):429–42.
- [3] Adler NJ, Harzing A-W. When knowledge wins: transcending the sense and nonsense of academic rankings. *Academy of Management Learning and Education* 2009;8(1):72–95.
- [4] Rowlinson M, Harvey C, Kelly A, Morris H. The use and abuse of journal quality lists. *Organization* 2011;18(4):443–6.
- [5] Hult GTM, Reimann M, Schilke O. Worldwide faculty perceptions of marketing journals: rankings, trends, comparisons, and segmentations. *Global EDGE Business Review* 2009;3(3):1–23.
- [6] Bruton GD, Lau C-M. Asian management research: status today and future outlook. *Journal of Management Studies* 2008;45(3):636–59.
- [7] Bollen J, Van de Sompel H, Hagberg A, Chute R. A principal component analysis of 39 scientific impact measures. *PLoS One* 2009;4(6):e6022.
- [8] Hazelkorn E. Rankings and the reshaping of higher education: the battle for world-class excellence. Houndmills, Basingstoke, UK: Palgrave Macmillan; 2011.
- [9] Frey BS, Rost K. Do rankings reflect research quality? *Journal of Applied Economics* 2010;13(1):1–38.
- [10] Benati S, Stefani S. The academic journal ranking problem: a fuzzy-clustering approach. *Journal of Classification* 2011;28(1):7–20.
- [11] Fam K-S, Shukla P, Sinha A, Luk C-L, Parackal M, Chai JCY. Rankings in the eyes of the beholder: a vox populi approach to academic journal ranking. *Asian Journal of Business Research* 2011;1(1):1–17.
- [12] Cook WD, Raviv TAL, Richardson AJ. Aggregating incomplete lists of journal rankings: an application to academic accounting journals. *Accounting Perspectives* 2010;9(3):217–35.
- [13] Kao C, Lin H-W, Chung S-L, Tsai W-C, Chiou J-S, Chen Y-L, et al. Ranking Taiwanese management journals: a case study. *Scientometrics* 2008;76(1):95–115.
- [14] Thomson Reuters. Journal citation reports. Philadelphia: Thomson Reuters; 2011; 2012.
- [15] Glänzel W, Moed HF. Journal impact measures in bibliometric research. *Scientometrics* 2002;53(2):171–93.
- [16] Leydesdorff L. Caveats for the use of citation indicators in research and journal evaluations. *Journal of the American Society for Information Science and Technology* 2008;59(2):278–87.
- [17] Bordons M, Fernández MT, Gómez I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics* 2002;53(2):195–206.
- [18] Mahdi S, D'Este P, Neely A. Citation counts: Are they good predictors of RAE scores? A bibliometric analysis of RAE 2001 Bedford: Cranfield: University School of Management & Advanced Institute of Management (AIM); 2008.
- [19] Bergstrom CT, West JD, Wiseman MA. The Eigenfactor metrics. *Journal of Neuroscience* 2008;28(45):11433–4.
- [20] González-Pereira B, Guerrero-Boteb VP, Moya-Anegón FI. The SJR Indicator: A New Indicator of Journals' Scientific Prestige 2009. arXiv:0912.4141v1.
- [21] Moed HF. Measuring contextual citation impact of scientific journals. *Journal of Informetrics* 2010;4(3):265–77.
- [22] Halkos GE, Tzeremes NG. Measuring economic journals' citation efficiency: a data envelopment analysis approach. *Scientometrics* 2011;88(3):979–1001.
- [23] Jones MJ, Brinn T, Pendlebury M. Journal evaluation methodologies: a balanced response. *Omega* 1996;24(5):607–12.
- [24] Baum JAC. Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies. *Organization* 2011;18(4):449–66.
- [25] Albers S. Misleading rankings of research in business. *German Economic Review* 2009;10(3):352–63.
- [26] Morris H, Harvey C, Kelly A. Journal rankings and the ABS journal quality guide. *Management Decision* 2009;47(9):1441–51.
- [27] Steward MD, Lewis BR. A comprehensive analysis of marketing journal rankings. *Journal of Marketing Education* 2010;32(1):75–92.
- [28] Meredith JR, Steward MD, Lewis BR. Knowledge dissemination in operations management: published perceptions versus academic reality. *Omega* 2011;39(4):435–46.
- [29] Bauerly RJ, Johnson DT. An evaluation of journals used in doctoral marketing programs. *Journal of the Academy of Marketing Science* 2005;33(3):313–29.
- [30] Zhou D, Ma J, Turban E. Journal quality assessment: an integrated subjective and objective approach. *IEEE Transactions on Engineering Management* 2001;48(4):479–90.
- [31] Franke N, Schreier M. A meta-ranking of technology and innovation management/entrepreneurship journals. *Die Betriebswirtschaft* 2008;2008(2):185–216.
- [32] Harzing A-W. Journal Quality List. 49th ed. (<http://www.harzing.com>) [08.06.2013].
- [33] Theußl S, Reutterer T, Hornik K. How to derive consensus among various marketing journal rankings? *Journal of Business Research* 2014;67(5):998–1006.
- [34] Bancroft DRE, Gopinath C, Kovács ÁM, Rejtő LDK. A new methodology for aggregating tables: summarizing journal quality data. *Journal of Business Venturing* 1999;14(3):311–9.
- [35] Rainer RK, Miller MD. Examining differences across journal rankings. *Communications of the ACM* 2005;48(2):91–4.
- [36] Tse ACB. Using mathematical programming to solve large ranking problems. *Journal of the Operational Research Society* 2001;52(10):1144–50.
- [37] Schulze GG, Warning S, Wiermann C. Zeitschriftenrankings für die Wirtschaftswissenschaften – Konstruktion eines umfassenden Metaindexes. *Perspektiven der Wirtschaftspolitik* 2008;9(3):286–305.
- [38] Farhangfar A, Kurgan LA, Pedrycz W. A novel framework for imputation of missing values in databases. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* 2007;37(5):692–709.
- [39] Gheyas IA, Smith LS. A neural network-based framework for the reconstruction of incomplete data sets. *Neurocomputing* 2010;73(16–18):3039–65.
- [40] Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods* 2002;7(2):147–77.
- [41] Durrant GB. Imputation methods for handling item-nonresponse in practice: methodological issues and recent debates. *International Journal of Social Research Methodology* 2009;12(4):293–304.
- [42] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 1977;39(1):1–38.
- [43] van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* 2007;16(3):219–42.
- [44] Nielsen SF. Proper and improper multiple imputation. *International Statistical Review* 2003;71(3):593–607.
- [45] Paul C, Mason WM, McCaffrey D, Fox SA. A cautionary case study of approaches to the treatment of missing data. *Statistical Methods and Applications* 2008;17(3):351–72.
- [46] Breiman L. Statistical modeling: the two cultures. *Statistical Science* 2001;16(3):199–231.
- [47] Farhangfar A, Kurgan L, Dy J. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition* 2008;41(12):3692–705.
- [48] Twala BETH, Jones MC, Hand DJ. Good methods for coping with missing data in decision trees. *Pattern Recognition Letters* 2008;29(7):950–6.
- [49] Hapfelmeyer A, Hothorn T, Ulm K. Recursive partitioning on incomplete data using surrogate decisions and multiple imputation. *Computational Statistics and Data Analysis* 2012;56(6):1552–65.
- [50] Breiman L. Random forests. *Machine Learning* 2001;45(1):5–32.
- [51] Biau G, Devroye L, Lugosi G. Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research* 2008;9:2015–33.
- [52] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. 2nd ed. New York: Springer; 2009.
- [53] Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods* 2009;14(4):323–48.
- [54] Lim T-S, Loh W-Y, Shih Y-S. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning* 2000;40(3):203–29.
- [55] Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: *Proceedings of the 22nd International Conference on Machine Learning, ACM; 2005. p. 625–32.*
- [56] Boström H. Calibrating random forests. In: *Proceedings of the 2008 IEEE Seventh International Conference on Machine Learning and Applications; 2008. p. 121–6.*
- [57] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978;2(6):429–44.
- [58] Cooper WW, Seiford LM, Tone K. Data envelopment analysis. 2nd ed. New York: Springer; 2007.
- [59] Liu JS, Lu LYY, Lu W-M, Lin BJJ. A survey of DEA applications. *Omega* 2013;41(5):893–902.
- [60] Liu JS, Lu LYY, Lu W-M, Lin BJJ. Data envelopment analysis 1978–2010: a citation-based literature survey. *Omega* 2013;41(1):3–15.
- [61] Cook WD, Tone K, Zhu J. Data envelopment analysis: prior to choosing a model. *Omega* 2014;44(0):1–4.
- [62] Llamazares B, Peña T. Preference aggregation and DEA: an analysis of the methods proposed to discriminate efficient candidates. *European Journal of Operational Research* 2009;197(2):714–21.
- [63] Cook WD, Doyle J, Green R, Kress M. Multiple criteria modeling and ordinal data: evaluation in terms of subsets of criteria. *European Journal of Operational Research* 1997;98(3):602–9.
- [64] Green RH, Doyle JR, Cook WD. Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research* 1996;90(3):461–72.
- [65] Noguchi H, Ogawa M, Ishii H. The appropriate total ranking method using DEA for multiple categorized purposes. *Journal of Computational and Applied Mathematics* 2002;146(1):155–66.
- [66] Hashimoto A. A ranked voting system using a DEA/AR exclusion model: a note. *European Journal of Operational Research* 1997;97(3):600–4.
- [67] Park KS, Jeong I. How to treat strict preference information in multicriteria decision analysis. *Journal of the Operational Research Society* 2011;62(10):1771–83.
- [68] Osborne MJ, Rubinstein A. Bargaining and markets. London: Academic Press; 1990.
- [69] Diskin A, Felsenthal DS. Individual rationality and bargaining. *Public Choice* 2007;133(1–2):25–9.
- [70] Wang Y-M, Chin K-S. Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics* 2010;128(1):332–8.

- [71] Whitley R. Changing governance and authority relations in the public sciences. *Minerva* 2011;49(4):359–85.
- [72] Clark T, Wright M. Reviewing journal rankings and revisiting peer reviews: editorial perspectives. *Journal of Management Studies* 2007;44(4):612–21.
- [73] Lovell CAK, Pastor JT. Radial DEA models without inputs or without outputs. *European Journal of Operational Research* 1999;118(1):46–51.
- [74] Liu WB, Zhang DQ, Meng W, Li XX, Xu F. A study of DEA models without explicit inputs. *Omega* 2011;39(5):472–80.
- [75] Foroughi AA, Tamiz M. An effective total ranking model for a ranked voting system. *Omega* 2005;33(6):491–6.
- [76] Crookes PA, Reis SL, Jones SC. The development of a ranking tool for refereed journals in which nursing and midwifery researchers publish their work. *Nurse Education Today* 2010;30(5):420–7.
- [77] DuBois FL, Reeb D. Ranking the international business journals. *Journal of International Business Studies* 2000;31(4):689–704.
- [78] Pishchulov G, Tüselmann H, Sinkovics RR. Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis. Working paper, TU Dortmund University; June 2014. <http://ssrn.com/abstract=2492472>.