CrossMark

ORIGINAL RESEARCH

Robustness of personal rankings: the Handelsblatt example

Daniela Lorenz[1] · **Andreas Löffler**[1]

Abstract   In the last years, *Handelsblatt* has published several rankings of business economists from German, Swiss and Austrian research institutions based on their journal publication output. These rankings have a strong influence on the academic profession. We scrutinize the *Handelsblatt* methodology by examining the effect the rankings' underlying algorithms and assumptions have on the scores and ranks of individual researchers. In doing so, we clarify how robust the result is with respect to these internal parameters. Since the parameters used by *Handelsblatt* are not scientifically substantiated but defined ad hoc, this question is of great importance. For each parameter variation, we provide several robustness measures for both the *Handelsblatt* life's work ranking and the *Handelsblatt* recent research performance ranking. E.g., if one applies a weighting scheme that lays more emphasis on first tier journal publications such that the weight of a particular category is always double of the weight of the next lower category, rank correlations based on all researchers in both personal rankings exceed 80 %. However, if one solely considers the top 25 performing researchers rank correlations fall below 50 and 20 % of researchers even drop out of this top group. Further research as well as the discussion in the academic community should clarify whether these correlations verify the robustness of the ranking or manifest the opposite.

Keywords   Research evaluation · Personal ranking · Handelsblatt ranking

✉   Andreas Löffler
     al@wacc.de

     Daniela Lorenz
     dl@wacc.de

[1]   Institut für Bank- und Finanzwirtschaft, Freie Universität Berlin, Boltzmannstr. 20, 14195 Berlin, Germany

🖄 Springer

1 Introduction

In the last years, *Handelsblatt* has published several rankings of business economists from German, Swiss and Austrian research institutions based on their research performance. The first ranking was launched in 2009. In September 2012 and December 2014, two more recent versions followed. While *Handelsblatt* also produces rankings of entire research institutions (Ranking "Top-25 Departments"), its focus is placed on rankings of individual researchers (e.g., "Top-250 Life's Work Ranking", "Top-100 Recent Research Performance"). These rankings have gained great importance in the academic profession; however, they are still heavily discussed (see, e.g., Kieser and Osterloh 2012).

To produce individual rankings, *Handelsblatt* solely considers journal publication output and uses several journal rankings that classify journals according to their (perceived) quality. Using this journal classification and "certain algorithms", each scholar was assigned a score to reflect their publication success. This score was then used to sort the scholars. Because the resulting rankings have such a sustained influence on both the public perception of business administration and the impression researchers have of their peers, it is appropriate and necessary to critically scrutinize the underlying methodology. To this end, we focus on the *Handelsblatt* life's work ranking as well as the *Handelsblatt* recent research performance ranking (which is solely based on the journal publications in the last 5 years) and examine the effect the mentioned "certain algorithms" have on the ranking according to which the researchers' journal publications are translated into scores. *Handelsblatt* itself states:

> " There are several ways to make research performance comparable, each of them with their special strengths and weaknesses. For instance, when selecting and weighting journals one can reasonably justify different decisions. In individual cases this can have a significant effect on a researcher's rank. Also, in case of co-authorship the way the scores are split up among authors can influence the rank as well as whether a researcher's life's work, recent publication output or annual productivity is used as ranking criterion."(translated from Müller and Storbeck 2009).

Against this background, there is good reason to analyze the robustness of the *Handelblatt* rankings. The objective of this study is thus to measure the extent to which both personal rankings change when the methods and algorithms used by *Handelsblatt* (referred to in this context as "internal parameters") are modified. In particular, our aim is to clarify how robust the result is with respect to variations in these internal parameters. What happens if single parameters are changed slightly? Does this displace the rank of only a few researchers, or does it lead to completely different results? Also we analyze which internal parameter exhibits the strongest effect on the ranking and, given the fact that the top 10, 25 or 100 researchers receive more public attention, if there would still be the same scholars in these top performing groups. Since the parameters used by *Handelsblatt* are not scientifically substantiated but defined ad hoc, this question is of great importance. If the

Handelsblatt ranking is shown to depend heavily on the parameters, this may indicate a lack of validity of the underlying methodology. Our dataset dates from January 2013 and our robustness analyses are thus based on the 2012 ranking. A more recent ranking dates from December 2014 which, however, we did not investigate because the same methodology was used (for details see section 3) and results are assumed to be very similar.

The paper is structured as follows. The next section reviews the literature on performance measurement. Section 3 describes the dataset underlying the *Handelsblatt* personal rankings. In section 4, we examine whether they are robust to changes in the internal parameters. Section 8 provides implications and limitations of our study.

2 Indicators of research performance and previous studies on the *Handelsblatt* ranking

It is beyond dispute that the performance of individual researchers is multidimensional in nature (e.g., Hussain 2011; Bornmann and Marx 2013). Besides research, acquisition of external funds, academic self-administration, active participation in professional associations and communities as well as teaching and supporting young scholars play an important role in academia. Therefore, plenty of indicators have been proposed in the literature that measure, compare and rank individuals' performance. With regard to research output, one generally makes use of scientometrics (see, e.g., Bornmann and Marx 2013; Kreiman and Maunsell 2011; Froghi et al. 2012), i.e., metrics that capture research performance based on librarian/bibliographic resources. According to Fig. 1, such quantitative evaluation instruments can be classified into the following categories:

Productivity indicators are a measure of the researcher's publication output, e.g., total number of publications or average number of publications per year. Most productivity indicators additionally take into account journal quality weights since publication quantity does not equate with publication quality (see, e.g., Rost and Frey 2011). A feature common to these indicators is that they assume important research results are published and that publications are the only visible research output. When calculating such metrics several decision criteria have to be set, such as choice of considered publication outlets and type of publication, choice of period of time, handling of co-authorships and the determination of appropriate journal quality weights. Because of this discretionary or even arbitrary nature, productivity
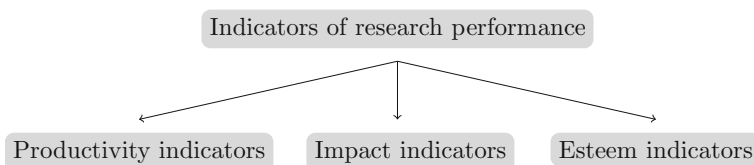

Fig. 1 Classification of research performance indicators

indicators have often been challenged and criticized (Adler and Harzing 2009). Particularly, the determination of journal quality weights is an often discussed issue. The commonly used proxies for the quality of a journal—Journal Impact Factors (JIF) (Garfield 2006) and Normalized Journal Position (NJP) (Bornmann and Marx 2013; Costas et al. 2010)—are based on a journal's citation frequency. However, this frequency can also be determined by non-scientific reasons (Bornmann and Daniel 2008; Judge et al. 2007; Jones et al. 1996) such as the reputation of a journal or its relationship to the authors and might create potential for manipulation or misuse (Archambault and Larivière 2009). Another possible way to determine accurate journal weights is to rely on researchers' assessment of a journal's quality. The German Academic Association for Business Research (VHB) uses this approach in their journal ranking JOURQUAL (Hennig-Thurau et al. 2004).

The second category of research performance measures consists of impact indicators such as the sum of citations a researcher receives, or their average number of citations per publication. Unlike productivity indicators, they attempt to capture the response to individual publications. However, it is well known that these metrics are strongly time- and field-specific, e.g., due to the different numbers of journals indexed or different citation practices across scientific disciplines (see Abramo et al. 2011). Therefore, it seems appropriate to standardize citation frequency with respect to field and time (Bornmann and Marx 2013). In particular, Leydesdorff et al. (2011) suggest using percentile ranks that rate each paper in terms of its percentile in the citation distribution. Besides non-scientific motivations for citing a paper, citation analyses are criticized for leaving room for discretionary decisions, e.g., how to deal with self-citations or what the appropriate length of the considered citation window is (Abramo et al. 2011). Recently, additional indicators have been proposed to combine productivity and impact and thus increase assessment accuracy. Most notably, Hirsch (2005) introduces the h-index that reflects the number of the researcher's publications that have been cited in other papers at least h times. An overview and critical evaluation of the h-index and its variants are provided by Froghi et al. (2012).

In addition to productivity and impact indicators, esteem indicators as a surrogate for quality have been proposed in the literature. E.g., Albers (2011) suggests considering honorary doctorates, too, when assessing the researchers' overall achievements. Rost and Frey (2011) focus on membership of the editorial boards of professional journals. They provide evidence that a ranking based on this esteem indicator randomly correlates with citation and publication rankings. However, Backes-Gellner (2011) cast doubt on this esteem indicator for various reasons, one of which being that unlike productivity indicators, it leaves more room for manipulation and discriminates against younger scholars.

To summarize, all research performance indicators have their advantages and disadvantages, and no standardized approach has emerged. All indicators have been discussed controversially by critics questioning the validity of these indicators and of any rankings of individuals, faculties or universities that are based on these indicators. Also, the methodology used by *Handelsblatt*, which is based solely on productivity indicators, has already been analyzed and frequently criticized in several studies. A broad critique of the *Handelsblatt* ranking even led to an online

appeal to boycott the ranking (see Kieser and Osterloh 2012; Dilger 2013). Criticism was levelled at several aspects including the inference from journal quality to paper quality, lack of neutrality with respect to different fields of specialization, and the setting of false incentives that adversely affect science and society. For a rebuttal of these criticisms, see Storbeck (2012). It was also criticized that *Handelsblatt* focuses on journal-based publications only. Note, however, that *Handelsblatt* does not claim to measure the quality of researchers in general. Concerning the 2007 *Handelsblatt* economists' ranking, Hofmeister and Ursprung (2008) complain about an incomplete weighting of the extent of research results, the distorting co-author weighting, and an overly restrictive journal selection. More recent *Handelsblatt* rankings have at least addressed the last two concerns in that they use a less distorting co-author weighting and a broadly expanded journal selection. Voeth et al. (2011) analyze the journal rankings underlying the *Handelsblatt* ranking with respect to more impact-oriented quality. For marketing, they show that journal rankings are only weakly aligned with the bibliometric impact of the journals. At the center of criticism of Müller (2010) stands the *Handelsblatt* business economists' ranking which he compares to a citation-based researchers' ranking. His comparison reveals considerable discrepancies between both rankings. Accordingly, he concludes that citations cannot be used to predict the rank assigned by *Handelsblatt*. Individual rankings should thus be interpreted with caution.

The study most similar to ours is that of Krapf (2011). *Krapf* uses alternative journal rankings (e.g., Jourqual 2, impact factors, journal-rating of the Wirtschaftsuniversität Wien) and compares the resulting department and life's work rankings to the corresponding *Handelsblatt* rankings by providing correlation measures. His findings suggest that department rankings are more or less robust to different journal classifications. By contrast, individual rankings react much more sensitive to a change in the underlying journal ranking. This remains true independently of whether the economists' or business economists' ranking of *Handelsblatt* is considered. In contrast to *Krapf*, we did neither analyze department rankings nor rankings of economists but limit our robustness analyses to rankings of individual business economists that are in return investigated in more depth. In particular, our study focuses on both the *Handelsblatt* life's work ranking and the *Handelsblatt* recent research performance ranking, and differs from his analysis in several ways.

Firstly, we not only investigate the impact of variations in journal weights but also in other internal parameters that have not been examined before such the way scores are split among co-authors and the weighting of different types of journal publications. Secondly, our approach also differs with respect to variations in journal weights. When assigning weights to each journal publication *Handelsblatt* had to face two problems: the allocation of journals to several quality categories (which journals belong to the best, second best, third best etc., quality level?) and the assignment of appropriate weights to each of these categories (which weight is assigned to the best, second best, third best etc. category?). *Krapf* concentrates on the former aspect by investigating ranking differences that arise when different journal rankings are used that classify journals differently into quality categories. By contrast, we focus on the latter aspect by still relying on the *Handelsblatt*

allocation but changing the weights that *Handelsblatt* assigned ad hoc to each quality category. This enables us to isolate the effect of the *Handelsblatt* weighting scheme. Finally, we contribute to prior research by providing more meaningful statistical measures to analyze the robustness of the rankings. Particularly, we additionally use rank correlations that do not assume equivalence of rank differences (see Sect. 5 for details) as well as median rank changes and also provide statistics for the most interesting group of top performers (top 10, 25, 100 and 250) on how many researchers drop out of the respective top group when internal parameters vary.

3 Data summary

The data used by *Handelsblatt* were collected by Thurgauer Wirtschaftsinstitut (TWI) at the University of Konstanz and is currently managed by the KOF Swiss Economics Institute (ETH Zürich).[1] TWI records all publications of the individual researchers who are able to constantly view, update and correct their data via the online portal *Forschungsmonitoring*. Our dataset dates from January 2013 and does thus essentially match the data underlying the 2012 ranking.

Originally, the database contained 3,016 researchers. 493 were excluded from our dataset since they have not published yet. Around 15 % of these researchers without any publications are professors. Thus, the database on which our analysis is based contains 2,523 researchers with at least one publication each. As shown in the distribution of academic or job titles in Table 1, only 62 % of them hold a professorship. This category includes junior professors, assistant professors, associate professors, honorary professors, irregular ("außerplanmäßiger") professors as well as full professors at universities of applied sciences ("Fachhochschulen") or universities.

Besides names and academic degrees, the data also contain researchers' dates of birth, their university or research institute affiliation, field of specialization and information on their journal publications. In total, the database covers 47,998 journal publications. However, several are co-authored so duplicate entries cannot be ruled out. For each publication, the number of co-authors, type of publication, year of publication and the weight assigned by *Handelsblatt* to the journal in question are provided. Note that we do not have information on the particular journal of publication. Concerning the assignment of weights *Handelsblatt* considered all journals that are ranked by JOURQUAL 2.1 of the VHB, journals that belong to social science citation index (SSCI) or science citation index (SCI) as well as journals that are mentioned in the 2011 version of the journal ranking of the Erasmus Research Institute of Management (EJL) (see Schläpfer and Storbeck 2012). In addition, all economics journals were added to the database as long as they were assigned a weight of 0.1 or higher in the *Handelsblatt* economists' ranking. This procedure results in a total of 950 journals that were considered by *Handelsblatt* in 2012 [meanwhile the number has increased to more than 1000

Table 1 Descriptive analysis: academic degrees

Highest academic degree or job title	Relative frequency (%)
University degree	6.5
Doctoral degree	24.0
Lecturer	1.5
Professor	62.0
n.s.	6.0
Sum	100.0

This table reports the relative frequencies of academic or job titles of all 2,523 researchers with non-zero publications considered by *Handelsblatt*

Table 2 Journal weighting

Quality level	HB1	HB2	HB3	HB4	HB5	HB6	HB7	HB8
Weight	1.0	0.7	0.5	0.4	0.3	0.2	0.1	0.0

This table reports the weights that were assigned by *Handelblatt* to the journal quality levels HB1 to HB8. HB1 stands for the category of highest journal quality

journals and the so-called Journal to Field Impact Scores (JFIS) are additionally used that account for different publication and citation practices across disciplines in business administration, see Gygli et al. (2014)] and categorized into eight quality levels. An overview of the journals' classification to these quality levels can be found in Schläpfer and Storbeck (2012). In a second step, *Handelsblatt* then assigned weights to each quality level that are shown in Table 2.

When calculating the individual scores, *Handelsblatt* also took into account the type of journal publication by assigning only half points for comments and zero points for editorials, corrections and book reviews. Moreover, the scores were split among the authors. In analogy to the *Handelsblatt* economists' ranking, the score is divided by the number of authors n. By contrast, the prior ranking of business economists in 2009 used the allocation formula $2/(n+1)$ to split up the score. However, this approach has been criticized because it incentivizes the excessive use of co-authorships. As a consequence, the formula $1/n$ was introduced in the 2012 version of the ranking. The influence of this approach is also addressed in the next section. The individual score of each researcher was then determined according to

$$\text{Score}_i = \sum_{j=1}^{\text{Pub}_i} \text{Weight(Journal)}_{ij} \cdot \text{Weight(Type of Publication)}_{ij} \cdot \frac{1}{n_{ij}}.$$

While index i denotes a particular researcher, index j stands for a particular journal publication. The summation runs over a researcher's $j = 1, \ldots, Pub_i$ journal publications. "Pub_i" denotes the total number of journal publications of researcher i as far as the life's work ranking is concerned and the total number of journal publications in the last five years as far as the recent research performance ranking is concerned.

Table 3 (see p. 8) summarizes the data and displays information about the researchers' scores which form the basis of the *Handelsblatt* ranking. It provides

Table 3 Descriptive analysis: publication performance

Performance indicator	All 2,523 researchers			Top 10			Top 25			Top 100			Top 250		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Total score	1.45	0.63	2.27	19.20	17.49	5.78	14.86	12.89	5.10	9.70	8.57	4.04	6.80	5.73	3.53
Total score in last 5 years	0.64	0.25	0.99	5.50	5.50	2.40	4.09	3.54	2.34	3.13	2.58	2.13	2.40	2.09	1.78
Average score per year	0.12	0.07	0.14	0.74	0.66	0.30	0.56	0.49	0.27	0.46	0.41	0.23	0.36	0.32	0.20
Average score per publ.	0.09	0.07	0.09	0.17	0.16	0.06	0.18	0.17	0.08	0.16	0.15	0.08	0.15	0.13	0.09
Total no. of publications	19.02	9.00	29.59	122.20	136.00	47.68	101.40	87.00	54.88	75.17	58.00	50.13	61.25	50.00	45.85
Total no. of publ. in last 5 years	6.90	4.00	10.88	31.70	35.00	11.66	26.76	22.00	17.58	22.39	16.50	18.50	19.51	14.00	18.48
Average no. of publ. per year	1.41	1.00	1.48	4.61	4.94	1.86	3.74	3.62	2.13	3.36	2.90	1.92	3.03	2.45	2.08
Thereof in HB1	0.02	0.00	0.10	0.24	0.07	0.35	0.19	0.06	0.28	0.11	0.04	0.19	0.07	0.00	0.15
Thereof in HB2	0.08	0.00	0.20	0.57	0.57	0.30	0.42	0.34	0.28	0.40	0.27	0.45	0.29	0.19	0.35
Thereof in HB3	0.13	0.00	0.25	0.92	0.67	0.91	0.64	0.48	0.66	0.52	0.38	0.50	0.41	0.32	0.42
Thereof in HB4	0.07	0.00	0.15	0.34	0.28	0.25	0.23	0.22	0.22	0.23	0.15	0.21	0.20	0.13	0.20
Thereof in HB5	0.07	0.00	0.17	0.33	0.23	0.32	0.23	0.14	0.25	0.17	0.10	0.19	0.17	0.11	0.20
Thereof in HB6	0.21	0.08	0.36	0.54	0.39	0.43	0.64	0.43	0.73	0.57	0.41	0.63	0.52	0.35	0.59
Thereof in HB7	0.19	0.06	0.38	0.52	0.48	0.37	0.40	0.27	0.40	0.33	0.23	0.43	0.38	0.22	0.73
Thereof in HB8	0.62	0.33	1.01	1.15	0.97	0.81	0.99	0.78	0.96	1.03	0.66	1.10	0.98	0.60	1.10
No. of authors per Publ.	2.29	2.00	1.54	2.34	2.00	1.41	2.21	2.00	1.43	2.25	2.00	1.43	2.34	2.00	1.75
Average % of full articles	0.96	1.00	0.12	0.96	0.97	0.04	0.92	0.94	0.07	0.94	0.96	0.09	0.94	0.96	0.08

This table reports mean, median and standard deviation (SD) of several performance indicators for all 2,523 researchers with non-zero publications in the database, for the top 10, 25, 100 as well as top 250 researchers in the *Handelsblatt* life's work ranking. The total score and number of publications in the last 5 years only take journal publications into account that were published between 2008 and 2012. In case of per-year-indicators, the individual number of years is measured as difference between 2012 and the year of each researcher's first publication. The table also reports descriptive statistics about the number of authors per publication as well as the percentage of full articles averaged over all researchers. The eight quality categories to which journals are assigned by *Handelsblatt* are denoted by HB1 (highest quality) to HB8 (lowest quality). In contrast to *Handelsblatt* we did not exclude boykotters

statistics for all 2,523 researchers as well as for the top performing groups in the life's work ranking. According to the first row, researchers achieved a mean total score of 1.45 when averaged over all 2,523 researchers whereas the mean total score amounts to 19.20 (14.86) among the top 10 (top 25) researchers. However, the distribution of scores is not symmetric. The median value shows that 50 % of all researchers (of the top 10 researchers) only obtained 0.63 (17.49) points. The table also provides information on the total score obtained within the last 5 years (from 2008 to 2012) and the average score per year. In line with Kreiman and Maunsell (2011), we define the corresponding number of years as the difference between 2012 and the year of each researcher's first publication. E.g., among the top 25 researchers, 4.09 points were obtained on average in the last 5 years which strongly exceeds five times the mean score per year that amounts to 0.56 points only. This indicates that publication intensity has increased over time. The score per publication averaged over 2,523 researchers is 0.09 which, however, does not provide information about the quantity of research output. Quantity is measured by the number of journal publications, which is 19.02 in total, thereof 6.90 obtained within the last 5 years, and 1.41 per year as far as all researchers are concerned and 122.2, 31.7 and 4.61, respectively, as far as the top 10 performers are concerned. On average, a researcher achieved 0.02 publications per year in a journal of highest quality (HB1) and 0.62 publications per year in a journal of lowest quality (HB8) according to the *Handelsblatt* classification. By contrast, top 10 researchers publish an average of 0.24 publications per year in HB1 and 1.15 publications per year in HB8. Furthermore, Table 3 shows that an average of 2.29 researchers produced a journal publication, only very few of which were comments or book reviews.

To provide descriptive statistics of the performance of various disciplines in business administration in the *Handelsblatt* life's work ranking, we grouped researchers according to their fields of specialization. However, the allocation by *Handelsblatt* is associated with an enormously high error rate and is incomplete. In particular, more than 600 researchers were not allocated to any field of specialization. Therefore, we rely on the researchers' self-assignment which follows from their affiliation with one or more of the 16 scientific sections of the VHB. 964 of the 2,523 business economists belong to at least one section. We manually assigned another 1,529 researchers to one or more fields of specialization in business administration by comparing their websites with the allocation made by *Handelsblatt*. This approach proved successful, with only 30 researchers who could not be allocated to a discipline. Table 4 (see p. 10) shows the absolute and relative frequencies of each specialization's affiliates. The sum of all affiliates is 3,390 rather than 2,493, which implies that some researchers are assigned to more than one specialization. Moreover, Table 4 describes the disciplinary composition of top performing groups in the *Handelsblatt* life's work ranking. It reads as follows: Out of the 25 best researchers according to *Handelsblatt* 11.1 % have specialized in banking and finance whereas not a single top 25 researcher has specialized in business taxation. This reveals that the disciplines are not equally represented in the top performing groups. Differences across disciplines also become obvious when

Table 4 Descriptive analysis: researchers' specialization

Field of specialization	Frequency		Composition top		Quality level HB	
	Absolute	Relative (%)	Top 25 (%)	Top 250 (%)	Mean	SD
Banking/finance	430	12.7	11.1	11.5	5.8	1.8
Marketing	386	11.4	9.3	7.4	6.4	1.4
Organization	336	9.9	9.3	9.9	6.2	1.4
Accounting/auditing	309	9.1	0.0	4.1	7.0	1.1
Technology, innovation and entrepreneurship	286	8.4	13.0	9.2	6.1	1.4
Business information technology	242	7.1	11.1	8.3	6.6	1.1
International management	222	6.5	3.7	6.0	6.1	1.4
Human resource management	206	6.1	5.6	6.2	6.4	1.2
Operations research	159	4.7	11.1	10.1	5.5	1.4
Logistics	153	4.5	11.1	7.6	6.0	1.5
Business taxation	147	4.3	0.0	0.9	7.4	0.9
Philosophy of science, ethics in business economics	129	3.8	3.7	3.7	6.4	1.3
Production management	129	3.8	11.1	9.2	6.3	1.3
Sustainability management	107	3.2	0.0	2.1	6.5	1.3
Public business administration	88	2.6	0.0	1.8	6.8	1.1
University management	61	1.8	0.0	1.8	6.7	1.1
All researchers	3.390	100	100	100	6.2	1.5

This table describes the allocation of researchers to their fields of specialization and provides information on the research performance across these specializations. The allocation was possible for 2,493 of the 2,523 researchers with non-zero publications considered by *Handelsblatt*. The first columns report absolute and relative frequencies of researchers in each discipline. Note that the sum of all affiliates add up to 3,390 rather than to 2,493 because some researchers have specialized in more than one field of business administration. The next two columns report the disciplinary composition of the top 25 and top 250 group according to the *Handelsblatt* life's work ranking. The last columns display mean and standard deviation (SD) of journal quality levels, ranging from 1 (highest quality) to 8 (lowest quality), per discipline. The numbers are based on 47,551 publications of 2,493 researchers in the *Handelsblatt* database that have non-zero publications and that could be allocated to the fields of business administration

looking at the *Handelsblatt* quality classification of their publications, ranging from 1 (highest quality level) to 8 (lowest quality level). Researchers specialized in business taxation published in journals with an average quality level of 7.4. No other specialization shows a lower mean or smaller standard deviation. The reason might lie in the fact that they predominantly publish in German journals that are (by intention) lower ranked than international ones. Researchers specialized in accounting or auditing published in journals with an average quality level of 7.0, so they come in second.

4 Robustness with respect to internal parameters

4.1 Robustness measures

In this section, we explore whether and if so, to what extent variations in the internal parameters change the *Handelsblatt* life's work ranking as well as the *Handelsblatt* recent research performance ranking of business economists. We begin by attempting to reconstruct both rankings on the basis of our dataset by calculating and sorting the scores according to the approach described above. In line with the approach of *Handelsblatt*, we placed researchers with identical scores at the same rank. We then compare our results with the top 250 life's work ranking of business economists and with the top 100 recent research performance ranking of business economists, respectively, that were published by *Handelsblatt* in 2012. Differences between our reconstruction and published version of the rankings arise only when a researcher refused to be ranked and named in the public rankings. At the time of publication of the ranking, the number of boycotters was 339; most of them are lower-ranked. For the purpose of our study, we include all boycotters in the data. Next, we modify the underlying score calculation in a consistent way and analyze the resulting effect on the (ordinal) ranks, (cardinal) scores and the composition of the top performing groups of researchers. More precisely, we provide the following robustness measures for each parameter variation.

Score correlation We analyze the correlation between the obtained cardinal scores before and after the parameter change measured by *Pearson*'s correlation coefficient (*Pearson*'s rho). A high coefficient indicates a strong linear relationship between the scores before and after the change in internal parameters.

Rank correlation With respect to individual ranks, we raise the question to what extent these *ordinal* data react to parameter changes. As is known from the literature, there are two such rank correlation measures: *Spearman*'s correlation coefficient and *Kendall*'s tau. To assess which of them is most appropriate for our purposes, we take a closer look at both rank correlation coefficients. Although the *Spearman* correlation coefficient is calculated like the *Pearson* correlation coefficient, it is based on ranks instead of raw scores. This approach assumes equivalence between all rank neighbors which implies that, e.g., the difference between the first and second rank is equivalent to the difference between ranks 100 and 101. The above-mentioned study Krapf (2011) uses *Spearman*'s rho. However, due to the features we just mentioned, we are not convinced that it is appropriate for our purposes. A closer look at, e.g., the first twenty researchers reveals immense score differences. The further down in the ranking, the smaller these differences. From the thirtieth rank downwards, the differences are marginal and practically negligible. Therefore, we do not use *Spearman*'s rho. Instead, *Kendall*'s tau seems suitable as a correlation measure in our study as it does not assume equivalence of rank differences (see, e.g., Cleff 2008: 118). Moreover, *Kendall*'s tau is easy to interpret. Tau takes values between $-100\,\%$ and $+100\,\%$; the actual value is determined by all rank comparisons that are possible within the sample. A value of $+100\,\%$ indicates that both rankings, those before and after the parameter change, have the exact same order. In case of a correlation amounting to $-100\,\%$ all

comparisons reverse to the opposite. A zero correlation implies that even and odd comparisons are perfectly balanced. Consequently, we use *Kendall*'s tau –instead of *Spearman*'s rho– as a correlation measure for ordinal ranks.

Percentage out Because most public attention is given to the top performing group of researchers, we complement all analyses by additionally providing statistics for the top 10, 25, 100, 250 and 500 researchers. We then investigate the percentage of researchers that drop out of the respective top group if a parameter variation takes place. The lower this percentage, the more robust is the composition of the top performing group.

Median absolute rank change Finally, we want to describe whether researchers are ranked quite differently on average or if they only slightly de- or increase in ranks if internal parameters are varied.[2] To prevent that negative and positive rank changes neutralize each other, we look at absolute rank changes only. Instead of arithmetic means, we rely on the median (absolute) change in number of ranks because the median is robust to outliers. If, e.g., within the top 10 the best ranked researcher goes down from rank 1 to 10 due to a parameter change and all nine other scores remain stable, the median absolute rank change is 1 whereas the mean absolute rank change would amount to 1.8 even though all (but one) researchers just move up by one rank.

4.2 Internal parameters

To investigate the impact of internal parameters on the rankings, we deviate from the *Handelsblatt* procedure and verify how individual scores, individual ranks as well as the composition of the top performing groups change. The internal parameters of interest are:

Co-author weight First, we deal with parameter variations that we expect to have no serious effect on the ranking. Concerning co-authorships we change the current allocation formula $\frac{1}{n}$ into $\frac{2}{n+1}$ that was formerly used by *Handelsblatt* to split up scores among n authors. Accordingly, in cases with two authors both receive 0.67 points instead of 0.5 points as in the recent ranking. The formerly used formula has been heavily criticized for creating (false) incentives for multi-coauthorship (Hofmeister and Ursprung 2008). The robustness measures associated with the alternation of this formula inform about the extent to which such multi-coauthorships have been used.

Weight of publication type Furthermore, we integrate and fully weight comments, editorials, corrections and book reviews when determining the scores. Such a weighting can be seen as problematic since they may not be classified as original research output. However, they are part of a researcher's journal publication output and overall productivity. As prior research claims that rankings should be based on multiple criteria, not only on research output (see, e.g., Albers 2011), we additionally rely on these journal publication types to measure productivity and want to examine if even small changes like this lead to serious adjustments in the ranking. By contrast, *Handelsblatt* assigned only half points for comments and zero points for editorials, corrections and book reviews.

---

[2] We thank the editor for suggesting this robustness measure.

Journal weights The selection of weights is another subjective influence on the ranking and there is no reasonable justification for the concrete values shown in Table 2. However, there does not exist an overall accepted or "theoretically correct" weighting scheme. Thus, we apply some alternative journal weighting schemes that we believe are reasonable and leave it to the readers' judgement which scheme reflects the journals' quality best. We then want to test how sensitive the ranking reacts to these changes in journal weights. Our changes in journal weights can be divided into two blocks.

First, we only change two weights at most. For instance, we raise or lower one of the values presented in Table 2 and verify the effect of this modification on the obtained score and the resulting rank. For consistency, we redefine the weights by merging two categories. That is, in one analysis we combine the *Handelsblatt* journal categories HB4, HB5 and HB6 and assign them a weight of 0.3, thereby downgrading the former category HB4 and upgrading the former category HB6. Note that *Handelsblatt* assigns zero points to publications in journals of the category HB8, meaning that of approximately 2,500 considered researchers, more than 300 received no points and had to share the last rank. Since category HB8 does not weight journals in, e.g., JOURQUAL category E, journals that play an important role in transferring research findings to professional practice are not taken into account. In light of the strong applied nature of business administration, the decision to disregard these publications as research output should be considered carefully. Therefore, we also analyze the impact on the ranking when taking into account the lowest journal category.

In a second step, we scrutinize the ranking more closely by modifying the weights of *all* journal categories instead of those of just one or two. We assess the effects that result from arranging the weights in an arithmetic or geometric progression. There are also other degrees of freedom. For instance, in the case of an arithmetic series one could assign a value of 0.7 to the highest journal category and reduce this weight downwards in steps of $c = 0.1$. In the case of geometric progressions, the growth ratio g between two journal categories could reach, e.g., 0.5 or 0.8 without one value being more appropriate than the other.

The specific weighting schemes that we tested in our analysis are displayed in Table 5. Note that the choice of one specific weighting schemes directly defines whether quality or quantity is rewarded better in the ranking. Researchers who want to obtain high scores in the *Handelsblatt* personal rankings have to decide whether to spend their time producing many papers of low quality or a smaller number of top papers. In particular, not the absolute but relative weights determine the quality-quantity-trade-off and are thus also reported in Table 5. The higher the relative weight of first tier publications, the stronger emphasis is placed on quality over quantity. With a relative weight of 50 % for publications with highest quality (HB1), the geometric weighting scheme with $g = 0.5$ thus creates the strongest incentive to publish in first tier journals, i.e., the quality of one category is always double of the quality of the next lower category. Such a weighting scheme would thus punish researchers that (almost) never publish in the highest journal category. By contrast, out of the various weighting schemes tested in our analysis the arithmetic progression with $HB1 = 1$ and $c = 0.1$ lays least emphasis on HB1 publications but puts more equal weights to all journals instead. In between ranges,

Table 5 Variations in weighting scheme

None	HB1	HB2	HB3	HB4	HB5	HB6	HB7	HB8	Sum
According to HB	1	0.7	0.5	0.4	0.3	0.2	0.1	0	3.2
	31 %	22 %	16 %	13 %	9 %	6 %	3 %	0 %	
Combining HB									
HB1 = HB2 = 0.7	0.7	0.7	0.5	0.4	0.3	0.2	0.1	0	2.9
	24 %	24 %	17 %	14 %	10 %	7 %	3 %	0 %	
HB2 = HB3 = 0.6	1	0.6	0.6	0.4	0.3	0.2	0.1	0	3.2
	31 %	19 %	19 %	13 %	9 %	6 %	3 %	0 %	
HB2 = HB3 = 0.7	1	0.7	0.7	0.4	0.3	0.2	0.1	0	3.4
	29 %	21 %	21 %	12 %	9 %	6 %	3 %	0 %	
HB1 = HB2 = HB3 = 0.7	0.7	0.7	0.7	0.4	0.3	0.2	0.1	0	3.1
	23 %	23 %	23 %	13 %	10 %	6 %	3 %	0 %	
HB4 = HB5 = HB6 = 0.3	1	0.7	0.5	0.3	0.3	0.3	0.1	0	3.2
	31 %	22 %	16 %	9 %	9 %	9 %	3 %	0 %	
Weighting of lowest HB									
HB8 = 0.05	1	0.7	0.5	0.4	0.3	0.2	0.1	0.05	3.25
	31 %	22 %	15 %	12 %	9 %	6 %	3 %	2 %	
HB7 = HB8 = 0.05	1	0.7	0.5	0.4	0.3	0.2	0.05	0.05	3.2
	31 %	22 %	16 %	13 %	9 %	6 %	2 %	2 %	
Arithmetic progression									
HB1 = 0.7, $c = 0.1$	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	2.8
	25 %	21 %	18 %	14 %	11 %	7 %	4 %	0 %	
HB1 = 0.8, $c = 0.1$	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	3.6
	22 %	19 %	17 %	14 %	11 %	8 %	6 %	3 %	
HB1 = 1, $c = 0.1$	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	5.2
	19 %	17 %	15 %	13 %	12 %	10 %	8 %	6 %	
Geometric progression									
HB1 = 1, $g = 0.5$	1	0.5	0.25	0.125	0.063	0.031	0.016	0.008	1.992
	50 %	25 %	13 %	6 %	3 %	2 %	1 %	0 %	
HB1 = 1, $g = 0.6$	1	0.6	0.36	0.216	0.130	0.078	0.047	0.028	2.458
	41 %	24 %	15 %	9 %	5 %	3 %	2 %	1 %	
HB1 = 1, $g = 0.7$	1	0.7	0.49	0.343	0.240	0.168	0.118	0.082	3.141
	32 %	22 %	16 %	11 %	8 %	5 %	4 %	3 %	
HB1 = 1, $g = 0.8$	1	0.8	0.64	0.512	0.410	0.328	0.262	0.210	4.161
	24 %	19 %	15 %	12 %	10 %	8 %	6 %	5 %	

This table reports the various weighting schemes that are tested. The upper row displays the absolute weights assigned to each quality level (HB). The lower row presents the relative weight which corresponds to the relative importance of each category. The eight quality categories to which journals are assigned by *Handelsblatt* are denoted by HB1 (highest quality) to HB8 (lowest quality)

the weighting scheme was applied by *Handelsblatt*. To illustrate the effect of these three weighting schemes, consider the following example (see Table 6) of authors having different publication structures.

The geometric scheme presented in the first row gives an advantage to quality as author 1 would receive a higher score than authors 2 and 3. The opposite is true for the arithmetic weighting presented in the second row where quantity is rewarded better. By contrast, *Handelsblatt* would position author 1 and author 2 at the same rank. When comparing the scores of authors with different publication structures, it becomes obvious that the score reacts stronger to changes in the weighting, the more an author chooses quantity over quality. Thus, a larger number of low quality publications (which is also prevalent in our data as Table 3 shows) leads to leverage effects.

4.3 Results

The robustness measures that result from a comparison between the *Handelsblatt* ranking and its modifications due to alternations of the aforementioned internal parameters are reported in Table 7 for the life's work ranking and in Table 8 for the recent research performance ranking. Because there are no generally accepted guidelines or thresholds when robustness measures show stability, we leave it to the reader to judge whether or not reported numbers show that the ranking is robust.

Co-author weight Based on all 2,523 researchers with non-zero publications and independently from whether all or only their most recent publications are considered the alternation of co-author weights to the formerly used version $2/(n+1)$ results in rank and score correlations that exceed 96 % (see first row of Tables 7 and 8). This indicates that less than 2 % of rank comparisons reverse to the opposite. However, the modification of co-author weights more strongly impacts the ranks of top performers: Rank correlations no longer exceed 85 % and even drop to 59.1 % in the top 25 list of the recent research performance ranking. Two researchers drop out of the top 10 list.

Weight of publication type With respect to including all types of journal publications the *Handelsblatt* life's work ranking (recent research ranking) shows rank and score correlations that almost all exceed 85 % (90 %) in all (sub)samples (see second row of Table 7 and 8, respectively); only as far as the top 25 researchers are concerned Kendall's rank correlation falls to 68 % (83.6 %). Thus, the vast majority of comparisons remain stable which is consistent with what we expected as only 4.8 % of all publications are comments, editorials, corrections or book reviews. The median absolute rank change based on all 2,523 researchers indicates that 50 % of them de- or increase their position in the life's work ranking (recent research ranking) by 17 (10) ranks at most.

Journal weights In both personal rankings, combining journal quality categories yields score correlations that almost all amount to more than 95 % independently

Table 6 Example: effect of weighting schemes on ranking

Weighting scheme	Score of author 1 with one HB1	Score of author 2 with two HB3	Score of author 3 with four HB5
Geometric Progression (HB1 = 1, $g = 0.5$)	1	0.5	0.25
Arithmetic Progression (HB1=1, $c = 0.1$)	1	1.6	2.4
Handelsblatt weights (see Table 2)	1	1	1.2

Table 7 Robustness of the *Handelsblatt* life's work ranking

Parameter change	All 2,523 researchers				Top 10				Top 25			
	τ	ρ	% out	med |Δ|	τ	ρ	% out	med |Δ|	τ	ρ	% out	med |Δ|
Co-authorship 2/n + 1	0.962	0.994	n.a.	28	0.822	0.922	20	1	0.753	0.952	16	2
Full inclusion of all publication types	0.973	0.976	n.a.	17	0.911	0.993	30	2	0.680	0.903	8	3
Combining HB												
HB1 = HB2 = 0.7	0.992	0.998	n.a.	7	0.911	0.983	0	0	0.873	0.987	4	1
HB2 = HB3 = 0.6	0.982	0.999	n.a.	11	0.867	0.991	10	0	0.920	0.995	8	1
HB2 = HB3 = 0.7	0.973	0.997	n.a.	23	0.778	0.984	10	1	0.887	0.988	8	1
HB1 = HB2 = HB3 = 0.7	0.967	0.995	n.a.	24	0.644	0.948	10	1	0.787	0.964	12	2
HB4 = HB5 = HB6 = 0.3	0.955	0.995	n.a.	35	0.911	0.995	10	1	0.800	0.982	8	1
Weighting of lowest HB												
HB8 = 0.05	0.858	0.968	n.a.	104	0.956	0.998	10	1	0.860	0.991	24	4
HB7 = HB8 = 0.05	0.857	0.970	n.a.	101	0.956	0.998	10	0	0.893	0.994	24	2
Arithmetic progression												
HB1= 0.7, c = 0.1	0.985	0.998	n.a.	15	0.867	0.978	0	1	0.840	0.982	8	2
HB1= 0.8, c=0.1	0.782	0.905	n.a.	157	0.867	0.975	60	5	0.627	0.920	36	11
HB1 = 1, c = 0.1	0.677	0.758	n.a.	240	0.822	0.965	80	16	0.500	0.821	60	23
Geometric progression												
HB1 = 1, g = 0.5	0.836	0.947	n.a.	128	0.733	0.603	30	3	0.487	0.716	20	5
HB1 = 1, g = 0.6	0.849	0.973	n.a.	94	0.867	0.865	10	1	0.673	0.905	8	2
HB1 = 1, g = 0.7	0.795	0.92	n.a.	151	0.956	0.986	40	4	0.807	0.972	32	8
HB1 = 1, g = 0.8	0.702	0.792	n.a.	230	0.822	0.981	70	11	0.587	0.877	60	19
Average	0.884	0.949	n.a.	85	0.856	0.948	25	3	0.748	0.934	21	5

Table 7 continued

Parameter change	Top 100				Top 250				Top 500			
	τ	ρ	% out	med |Δ|	τ	ρ	% out	med |Δ|	τ	ρ	% out	med |Δ|
Co-authorship 2/n+1	0.770	0.968	9	9	0.822	0.979	5.6	13	0.850	0.985	4.6	22
Full inclusion of all publication types	0.861	0.874	10	7	0.883	0.922	6.4	12	0.904	0.946	3.6	19
Combining HB												
HB1 = HB2 = 0.7	0.904	0.991	5	2	0.937	0.994	0.8	5	0.957	0.996	3.0	7
HB2 = HB3 = 0.6	0.892	0.995	3	3	0.923	0.996	2.8	5	0.924	0.997	2.2	9
HB2 = HB3 = 0.7	0.870	0.991	6	3	0.893	0.993	4.8	8	0.895	0.994	3.2	15
HB1 = HB2 = HB3 = 0.7	0.820	0.977	9	5	0.855	0.984	5.2	10	0.873	0.988	4.2	18
HB4 = HB5 = HB6 = 0.3	0.824	0.984	9	6	0.844	0.985	4.8	12	0.865	0.989	5.6	20
Weighting of lowest HB												
HB8 = 0.05	0.817	0.967	12	9	0.831	0.965	11.2	17	0.832	0.970	8.6	34
HB7 = HB8 = 0.05	0.836	0.971	11	8	0.842	0.969	8.8	16	0.844	0.974	8.2	29
Arithmetic progression												
HB1 = 0.7, c = 0.1	0.869	0.989	6	3	0.902	0.992	2.0	7	0.927	0.995	3.6	11
HB1 = 0.8, c = 0.1	0.593	0.856	28	19	0.635	0.859	21.6	39	0.665	0.880	14.8	70
HB1 = 1, c = 0.1	0.445	0.664	45	47	0.479	0.665	35.2	79	0.518	0.703	26.2	124
Geometric progression												
HB1 = 1, g = 0.5	0.538	0.827	18	14	0.612	0.879	17.6	35	0.620	0.905	14.2	58
HB1 = 1, g = 0.6	0.695	0.939	10	8	0.755	0.956	10.0	18	0.767	0.966	8.2	34
HB1 = 1, g = 0.7	0.747	0.912	21	15	0.764	0.912	18.0	29	0.756	0.922	13	56
HB1 = 1, g = 0.8	0.525	0.726	36	40	0.552	0.728	32.4	69	0.577	0.759	23.6	112
Average	0.750	0.914	15	12	0.783	0.924	11.7	23	0.798	0.936	9.2	40

This table contains several robustness measures for the *Handelsblatt* life's work ranking of all 2,523 researchers with non-zero publications as well as of the top 10, 25, 100, 250 and 500 performing researchers. τ stands for Kendall's rank correlation coefficient; ρ for Pearson's correlation of scores. % out describes the percentage of researchers that drop out of the respective top group. med \|Δ\| is the median (absolute) rank change. HB stands for the quality category to which a journal is assigned by *Handelsblatt* ranging from HB1 (highest quality) to HB8 (lowest quality)

Table 8 Robustness of the *Handelsblatt* recent research performance ranking

Parameter change	All 2,523 researchers				Top 10				Top 25			
	τ	ρ	% out	Med |Δ|	τ	ρ	% out	Med |Δ|	τ	ρ	% out	Med |Δ|
Co-Authorship 2/n+1	0.969	0.992	n.a.	20	0.733	0.845	20	2	0.591	0.896	12	4
Full inclusion of all publication types	0.986	0.979	n.a.	10	1.000	0.988	20	1	0.836	0.962	8	2
Combining HB												
HB1 = HB2 = 0.7	0.992	0.998	n.a.	3	0.644	0.952	0	2	0.893	0.975	4	0
HB2 = HB3 = 0.6	0.981	0.997	n.a.	6	0.778	0.955	10	1	0.805	0.968	8	1
HB2 = HB3 = 0.7	0.975	0.996	n.a.	13	0.778	0.942	10	1	0.751	0.953	8	2
HB1 = HB2 = HB3 = 0.7	0.968	0.993	n.a.	15	0.556	0.884	10	1	0.705	0.912	8	2
HB4 = HB5 = HB6 = 0.3	0.963	0.995	n.a.	23	0.956	0.995	10	0	0.892	0.995	12	1
Weighting of lowest HB												
HB8 = 0.05	0.910	0.988	n.a.	86	1.000	0.998	10	0	0.912	0.997	4	1
HB7 = HB8 = 0.05	0.903	0.988	n.a.	77	0.867	0.994	20	1	0.912	0.995	4	1
Arithmetic progression												
HB1 = 0.7, c = 0.1	0.986	0.997	n.a.	11	0.733	0.931	10	1	0.838	0.963	4	1
HB1 = 0.8, c = 0.1	0.850	0.948	n.a.	128	0.689	0.843	30	2	0.778	0.912	20	4
HB1 = 1, c = 0.1	0.767	0.840	n.a.	150	0.378	0.669	70	7	0.624	0.812	52	17
Geometric progression												
HB1 = 1, g = 0.5	0.859	0.946	n.a.	128	0.644	0.671	40	3	0.364	0.701	20	5
HB1 = 1, g = 0.6	0.881	0.981	n.a.	113	0.689	0.857	30	2	0.538	0.884	12	3
HB1 = 1, g = 0.7	0.863	0.966	n.a.	123	0.867	0.988	20	0	0.851	0.989	12	1
HB1 = 1, g = 0.8	0.792	0.878	n.a.	133	0.644	0.878	40	5	0.725	0.914	40	10
Average	0.915	0.968	n.a.	65	0.747	0.899	22	2	0.751	0.927	14	3

Table 8 continued

Parameter change	Top 100				Top 250				Top 500			
	τ	ρ	% out	Med |Δ|	τ	ρ	% out	Med |Δ|	τ	ρ	% out	Med |Δ|
Co-authorship 2/n+1	0.761	0.954	8	7	0.796	0.970	8.4	16	0.829	0.978	5.0	27
Full inclusion of all publication types	0.936	0.986	9	5	0.911	0.985	5.2	12	0.938	0.938	2.4	12
Combining HB												
HB1 = HB2 = 0.7	0.901	0.990	5	3	0.925	0.993	3.2	7	0.942	0.995	2.8	10
HB2 = HB3 = 0.6	0.878	0.987	8	4	0.882	0.990	4.8	8	0.905	0.993	2.8	13
HB2 = HB3 = 0.7	0.848	0.981	9	5	0.858	0.986	6.4	13	0.876	0.989	4.0	21
HB1 = HB2 = HB3 = 0.7	0.792	0.964	12	7	0.821	0.975	8.0	15	0.843	0.981	5.2	23
HB4 = HB5 = HB6 = 0.3	0.866	0.988	8	4	0.835	0.988	6.8	11	0.867	0.989	5.0	17
Weighting of lowest HB												
HB8 = 0.05	0.909	0.996	10	3	0.868	0.980	7.2	11	0.904	0.986	4.6	18
HB7 = HB8 = 0.05	0.884	0.993	7	2	0.878	0.982	7.2	7	0.903	0.987	3.8	15
Arithmetic progression												
HB1= 0.7, c = 0.1	0.858	0.985	11	4	0.872	0.989	4.4	9	0.907	0.992	3.4	16
HB1 = 0.8, c = 0.1	0.716	0.917	23	14	0.639	0.875	15.6	33	0.724	0.911	11.0	49
HB1 = 1, c = 0.1	0.595	0.777	42	34	0.490	0.655	26.8	74	0.584	0.744	18.8	99
Geometric progression												
HB1 = 1, g = 0.5	0.499	0.820	22	17	0.560	0.860	20.4	39	0.593	0.887	16.0	66
HB1 = 1, g = 0.6	0.662	0.940	13	10	0.712	0.951	12.4	24	0.758	0.963	7.4	39
HB1 = 1, g = 0.7	0.844	0.986	16	7	0.784	0.942	10.8	21	0.839	0.960	9.4	32
HB1 = 1, g = 0.8	0.676	0.871	32	25	0.581	0.747	22.8	56	0.664	0.818	16.2	79
Average	0.789	0.946	15	9	0.776	0.929	10.7	22	0.817	0.944	7.4	34

This table contains several robustness measures for the *Handelsblatt* recent research performance ranking of all 2,523 researchers with non-zero publications as well as of the top 10, top 25, top 100, top 250 and top 500 performing researchers. τ stands for Kendall's rank correlation coefficient; ρ for Pearson's correlation of scores. % out describes the percentage of researchers that drop out of the respective top group. med \|Δ\| is the median (absolute) rank change. HB stands for the quality category to which a journal is assigned by *Handelsblatt* ranging from HB1 (highest quality) to HB8 (lowest quality)

from whether all or only the top researchers are considered. Rank correlations take values that are on average 11 % points lower and reach a minimum of 0.556 as far as the impact of merging the first three journal quality levels on the top 10 list of the recent research performance ranking is concerned. In comparison to other parameter variations, the percentage of researchers that drop out of the top groups (which does not exceed 12 %) as well as the median change in ranks is rather moderate.

When taking into account the lowest journal category, the recent research performance shows greater stability than the life's work ranking in almost all subsamples. This might indicate a change in the researchers' publication strategies across time because weighting of low tier publications causes less variation in the ranking if researchers' lifetime publications instead of only their most recent publications are taken into account. If one, e.g., assigns the lowest journal a value of only 0.05 (instead of 0) the rank correlation is still above 90 % in the recent research performance ranking of all 2,523 researchers but drops to approximately 86 % if researchers' entire publication records are considered. This means that almost 7 % of comparisons between specific researchers reverse. Note that *Handelsblatt* also considers researchers with zero points, meaning that of 2,523 considered researchers, 303 received no points and had to share the last rank. One could argue that the majority of these researchers would obtain positive scores when weighting the lowest journal category and that the changes in the life work's ranking are mainly explained by the resulting new ranks. To verify this, we additionally calculate correlation measures based on only those researchers that received positive scores in the *Handelsblatt* life work's ranking, i.e., we exclude the 303 researchers that shared the last rank. The results are provided in Table 9 in the Appendix and show that correlation coefficients remain stable, indicating that the researchers with a zero score are not causal for the decline in rank correlation.

Finally, we more generally varied journal weights by arranging them in arithmetic and geometric progressions. For the life's work ranking (recent research performance ranking), Table 7 (Table 8) reveals that the rank correlation decreases to only 67.7 % (76.7 %) if all 2,523 researchers are taken into account and more equal weights are assigned to all eight journal quality levels, i.e., an arithmetic progression with HB1 $= 1$ and $c = 0.1$. Corresponding rank correlations in the top 250 rankings even drop below 50 % which indicates that more than 25 % of rank comparisons reverse. If, in contrast, one better rewards quality by attaching greatest relative importance on publications in first tier journals (i.e., geometric progression with HB1 $= 1$ and $g = 0.5$) rank correlations based on all 2,523 researchers exceed 80 % but again take numbers below 50 % in some subsamples. Also, the composition of the top groups and the median rank change show that both *Handelsblatt* personal rankings react more sensitively to these systematic modifications of journal weights than to the journal weight variations described before. One possible explanation is that a researcher's number of journal publications per quality level is correlated between adjacent quality levels. Therefore, weight alterations that simultaneously affect the weights of the highest as well as lowest categories should have the largest impact on the rankings.

5 Conclusion

Handelsblatt has ranked more than 2,500 business economists according to their life's work (based on their total journal publication output) as well as according to their recent research performance (based on their journal publications within the last 5 years). Our aim is to analyze the impact of the underlying internal parameters on these personal rankings. In particular, we investigate to what extent they are robust with respect to changes in internal parameters, such as the splitting of scores in case of co-authorships, inclusion of all types of journal publications, and journal weighting.

We find that there are differences in journal publication intensity between researchers that persist even when internal parameters are varied. However, individual performance evaluations based on a researcher's specific rank can strongly depend on the internal parameters of the ranking. Specifically, our findings suggest that the *Handelsblatt* rankings of business economists tend to be more robust with respect to the allocation of scores among co-authors and full integration of all types of journal publication (i.e., book reviews, comments, corrections and editorials besides full original articles) than to systematic changes in journal weights. The underlying weighting scheme directly determines whether quality or quantity is rewarded better in the rankings. In particular, *Handelsblatt* applies a weighting scheme such that, e.g., the quality of the second best journal category is considered 70 % of the quality of first tier journal publications. Our study investigates the effect of several alternative weighting schemes on both *Handelsblatt* personal rankings. If one applies a scheme that lays more emphasis on quality, e.g., such that the weight of a particular category is always double of the weight of the next lower category, rank correlation based on all researchers amounts to 83.6 % in the life's work ranking (85.9 % in the recent research performance ranking) which indicates that about 8 % (7 %) of rank comparisons reverse to the opposite. However, if, e.g., solely the top 25 performing researchers are considered the corresponding rank correlation fall below 50 %, and 20 % of researchers even drop out of the top 25 list. In general, our analyses show that rank and score correlations tend to be lower if only the group of top performers (e.g., top 10, 25, 100, 250, 500) are taken into account. Future research as well as the (ongoing) discussion in the academic community will ascertain whether or not such numbers allow the conclusion that the ranking is robust.

In general, when comparing researchers' publication output in journals, e.g., in the course of filling a vacancy, quantitative measures cannot replace an evaluation of the papers' actual content. In this respect, our result is conform to Krapf (2011) who has already claimed that the actual rank assigned by *Handelsblatt* is not suitable (and not intended to be suitable) for evaluating individual researchers. However, "managing by numbers" is still common, and Germany's academic community is not the only one to have raised concerns about the danger of a mechanistic use of ratings when comparing applicants' profiles [for the UK see Hussain (2011)]. If one nevertheless relies on "managing by numbers" researchers', metric score is a more robust estimate of their publication output than their rank in the *Handelsblatt* rankings; the comparison of both correlation coefficients used in this study shows that in most cases, correlation of scores takes higher values than correlation of ranks.

This study has a number of limitations. Firstly, we have not addressed all the problems that are generally associated with rankings. For example, it is possible that the correlation between reviewers' evaluations is rather low, and that the quality classification of a journal can be manipulated (e.g., by coalitions). Moreover, another restriction of the underlying data is that it neither contains information on the number of pages per publication nor on other types of publications such as books or articles in outlets other than journals. Therefore, an analysis of researcher's overall publication record and its impact is not possible.

Appendix

See Table 9.

Table 9 Correlation coefficients based on different sample sizes of the *Handelsblatt* life's work ranking

Parameter change	Based on 2,523 researchers		Based on 2,220 researchers	
	Kendall's τ	*Pearson*'s ρ	*Kendall*'s τ	*Pearson*'s ρ
None				
According to HB	1.000	1.000	1.000	1.000
Combining HB				
HB1 = HB2 = 0.7	0.992	0.998	0.989	0.998
HB2 = HB3 = 0.6	0.982	0.999	0.976	0.999
HB2 = HB3 = 0.7	0.973	0.997	0.965	0.997
HB1 = HB2 = HB3 = 0.7	0.967	0.995	0.957	0.995
HB4 = HB5 = HB6 = 0.3	0.955	0.995	0.943	0.994
Weighting of lowest HB				
HB8 = 0.05	0.858	0.968	0.873	0.968
HB7 = HB8 = 0.05	0.857	0.970	0.875	0.970
Arithmetic progression				
HB1 = 0.7, c = 0.1	0.985	0.998	0.980	0.998
HB1 = 0.8, c = 0.1	0.782	0.905	0.788	0.903
HB1=1, c = 0.1	0.677	0.758	0.686	0.754
Geometric progression				
HB1 = 1, g = 0.5	0.836	0.947	0.827	0.945
HB1 = 1, g = 0.6	0.849	0.973	0.861	0.973
HB1 = 1, g = 0.7	0.795	0.920	0.813	0.920
HB1 = 1, g = 0.8	0.702	0.792	0.715	0.790

This table contains score correlations according to *Pearson*'s ρ and rank correlations (*Kendall*'s τ) between the ranking before and after varying the journal weights. The first two columns are based on all 2,523 researchers with non-zero publications that were considered by the *Handelsblatt* life's work ranking. In contrast, the last columns show correlation measures based on 2,220 researchers with non-zero publications who obtained positive scores in this ranking. Thus, 303 researchers sharing the last rank (non-zero publications but zero score) are excluded

References

Abramo, Giovanni, Tindaro Cicero, and Ciriaco Andrea D'Angelo. 2011. Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics* 5: 659–667.

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

Albers, Sönke. 2011. Esteem indicators: membership in editorial boards or honorary doctorates, discussion of quantititive and qualitative rankings of scholars by Rost and Frey. *Schmalenbach Business Review* 63: 92–98.

Archambault, Éric, and Vincent Larivière. 2009. History of the journal impact factor: contingencies and consequences. *Scientometrics* 79: 635–649.

Backes-Gellner, Uschi. 2011. Rankings upon rankings—and no end in sight, discussion of quantititive and qualitative rankings of scholars by Rost and Frey. *Schmalenbach Business Review* 63: 99–108.

Bornmann, Lutz, and Hans-Dieter Daniel. 2008. What do citation counts measure? citing behavior. *Journal of Documentation* 64: 45–80.

Bornmann, Lutz, and Werner Marx. 2013. Vorschläge für Standards zur Anwendung der Szientometrie bei der Evaluation von einzelnen Wissenschaftler(inne)n im Bereich der Naturwissenschaften. *Zeitschrift für Evaluation* 12: 103–127.

Cleff, Thomas. 2008. *Deskriptive Statistik und moderne Datenanalyse*. Wiesbaden: Gabler.

Clermont, Marcel, and Christian Schmitz. 2008. Erfassung betriebswirtschaftlich relevanter Zeitschriften in den ISI-Datenbanken sowie der Scopus-Datenbank. *Zeitschrift für Betriebswirtschaft* 78: 987–1010.

Costas, Rodrigo, Thed N. van Leeuwen, and Maria Bordons. 2010. A bibliometric classificatory approach for the study and assessment of research performance at the individual level: the effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology* 61: 1564–1581.

Dilger, Alexander. 2013. Soll man das Handesblatt-ranking BWL boykottieren? *Beiträge zur Hochschulforschung* 35: 100–119.

Froghi, Saied, Kamran Ahmed, Adam Finch, John M. Fitzpatrick, Mohammed S. Khan, and Prokar Dasgupta. 2012. Indicators for research performance evaluation: an overview. *BJU International* 109: 321–324.

Garfield, Eugene. 2006. The history and meaning of the journal impact factor. *The Journal of the American Medical Association* 295: 90–93.

Gygli, Savina, Hans Christian Müller-Dröge, and Jörg Schläpfer. 2014. Handelsblatt BWL-ranking 2014: Methodik und Zeitschriftengewichte, available online: http://htmldb-hosting.net/pls/htmldb/FMONITORING.download_my_file?p_file=921. Accessed 07 July 2015.

Hennig-Thurau, Thorsten, Gianfranco Walsh, and Ulf Schrader. 2004. VHB-JOURQUAL: Ein ranking von betriebswirtschaftlich relevanten Zeitschriften auf der Grundlage von Expertenurteilen. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 56: 520–545.

Hirsch, Jorge E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102: 16569–16572.

Hofmeister, Robert, and Heinrich W. Ursprung. 2008. Das Handelsblatt-Ökonomen Ranking 2007: Eine kritische Beurteilung. *Perspektiven der Wirtschaftspolitik* 9: 254–266.

Hussain, Simon. 2011. Food for thought on the ABS academic journal quality guide. *Accounting Education* 20: 545–559.

Jones, Michael J., Tony Brinn, and Maurice Pendlebury. 1996. Journal evaluation methodologies: a balanced response. *Omega* 24: 607–612.

Judge, Timothy A., Daniel M. Cable, Amy E. Colbert, and Sara L. Rynes. 2007. What causes a management article to be cited - article, author, or journal? *Academy of Management Journal* 50: 491–506.

Kieser, Alfred, and Margit Osterloh. 2012. Warum wir aus dem Handelsblatt BWL-Ranking ausgestiegen sind—ein offener Brief an das Handelsblatt, unterzeichnet von Professoren der Betriebswirtschaftslehre, (September 7th, 2012, 11 o'clock, 309 Signers), available online: http://handelsblattranking.wordpress.com/2012/08/29/handelsblatt-ranking/. Accessed 07 July 2015.

Krapf, Matthias. 2011. Research evaluation and journal quality weights: much ado about nothing. *Zeitschrift für Betriebswirtschaft* 81: 5–27.

Kreiman, Gabriel, and John H.R. Maunsell. 2011. Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience* 5: 1–6.

Leydesdorff, Loet, Lutz Bornmann, Rüdiger Mutz, and Tobias Opthof. 2011. Turning the tables on citation analysis one more time: principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology* 62: 1370–1381.

Müller, Harry. 2010. Wie valide ist das Handelsblatt-BWL-Ranking? - Zeitschriften- und zitationsbasierte Personenrankings im Vergleich. *Betriebswirtschaftliche Forschung und Praxis* 62: 150–164.

Müller, Anja, and Olaf Storbeck. 2009. BWL-ranking: Methodik und Interpretation, available online: http://www.handelsblatt.com/politik/oekonomie/bwl-ranking/bwl-ranking-methodik-und-interpretation/3180850.html. Accessed 07 July 2015.

Raffournier, Bernard, and Alain Schatt. 2010. Is European accounting research fairly reflected in academic journals? *An Investigation of Possible Non-mainstream and Language Barrier Biases, European Accounting Review* 19: 161–190.

Rost, Katja, and Bruno S. Frey. 2011. Quantitative and qualitative rankings of scholars. *Schmalenbach Business Review* 63: 63–91.

Schläpfer, Jörg, and Olaf Storbeck. 2012. BWL-ranking 2012: Methodik und Zeitschriftenliste, available online: http://www.handelsblatt.com/politik/konjunktur/bwl-ranking/-bwl-ranking-2012-bwl-ranking-2012-methodik-und-zeitschriftenliste/v_detail_tab_comments/6758368.html. Accessed 07 July 2015.

Storbeck, Olaf. 2012. Falsche Anreize schaden der Wissenschaft, available online: http://www.handelsblatt.com/politik/oekonomie/bwl-ranking/dokumentation-falsche-anreize-schaden-der-wissenschaft-seite-all/7114064-all.html. Accessed 07 July 2015.

Voeth, Markus, Uta Herbst, and Jeanette Loos. 2011. Bibliometrische analyse der Zeitschriftenrankings VHB-JOURQUAL 2.1 und Handelsblatt-Zeitschriftenranking BWL. *Die Betriebswirtschaft* 71: 439–458.