Journal of Informetrics 6 (2012) 389–402

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# Ranking economists on the basis of many indicators: An alternative approach using RePEc data☆

Christian Seiler[1], Klaus Wohlrabe*

*Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany*

ARTICLE INFO

*Article history:*
Received 28 October 2011
Received in revised form 19 January 2012
Accepted 23 January 2012

*JEL classification:*
A12
A14

*Keywords:*
RePEc
Ranking aggregation
Principal components analysis
Economics profession

ABSTRACT

In economics the Research Papers in Economics (RePEc) network has become an essential source for the gathering and the spread of both existing and new economic research. Furthermore, it is currently the largest bibliometric database in economic sciences containing 33 different indicators for more than 30,000 economists. Based on this bibliographic information RePEc calculates well-known rankings for authors and academic institutions. We provide some cautionary remarks concerning the interpretation of some provided bibliometric measures in RePEc. Moreover, we show how individual and aggregated rankings can be biased due to the employed ranking methodology. In order to select key indicators describing and assessing research performance of scientist, we propose to apply principal component analysis in this data-rich environment. This approach allows us to assign weights to each indicator prior to aggregation. We illustrate the approach by providing a new overall ranking of economists based on RePEc data.



## 1. Introduction

The assessment of research performance in economic sciences has been of long interest to the profession.[2] Rankings can be regarded as a concrete realization of an unobserved process representing scientific output, achievements, performance, or merits.[3] They may play an important role both for the ranked scientists and decision makers. In the latter case e.g. for promotions, tenure decisions, or funding. A generally accepted academic ranking approach would be desirous but has not been achieved yet. Each specific ranking has its pros and cons. Furthermore, desired rankings cannot be calculated as data gathering is often prohibitive. Consider for instance the collection of data for all potential dimensions of scientific work: research (works, citations, weighting), teaching, press relations, acquisition of grants, supervision of students, among others. Therefore, many existing rankings are solely based on one or two bibliometric indicators. These are often quality weighted counts of citations or published work. Even in case of a large available database several questions remain open. Do all bibliometric indicators measure the same desired unobserved process: the research performance? Are there some key indicators? This calls for an appropriate selection or aggregation procedure which assigns specific weights according to its importance. In the literature there are different ranking aggregation approaches, notably some weighted means based on (standardized) bibliometric indicators. Either the indicators are aggregated first and than ranked, or vice versa. But how

☆ We thank Christian Zimmermann for providing us with the RePEc data. Furthermore, we thank two anonymous referees for helpful comments.

* Tel.: +49 0 89 9224 1229.
*E-mail addresses:* seiler@ifo.de (C. Seiler), wohlrabe@ifo.de (K. Wohlrabe).
[1] Tel.: +49 0 89 9224 1248.

[2] Early U.S. studies for economics institutions and/or departments can found in Graves, Marchand, and Thompson (1982), Hirsch, Austin, Brooks, and Moore (1984), or Dusansky and Vernon (1998).

[3] In the following we use these terms interchangeably.

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2012.01.007

to choose these weights remains largely unexplained. As there is no natural benchmark at which these indicators can be evaluated, this remains a difficult task. This paper deals both with the aspect of many available bibliometric measures and the selection of key indicators as a basis for an aggregated ranking.

In economics, Research Papers in Economics (RePEc; http://www.repec.org) has become an essential source for the spread of knowledge and ranking of individual authors and academic institutions. RePEc is based on the 'active participation principle', i.e. that authors, institutions and publishers have to register and to provide information to the network. This approach has the main advantage that a clear assignment of works and citations to authors and articles respectively is possible.[4] Indeed, the RePEc story has become a success, with more than 29,000 registered authors and 11,000 institutions in economic sciences worldwide as of August 2011. This success is based on a considerable amount of time spent by numerous volunteers who set up and maintain the web page. RePEc provides 33 quantitative bibliometric measures both for authors and institutions. Among these measures are, for example, number of published works, number of citations, h-index, and number of downloads. Thus, RePEc can be considered as currently one of the largest bibliometric database in the field of economics. Based on these measures, RePEc calculates corresponding rankings for both authors and institutions. Finally, several aggregated rankings are provided. Although the RePEc network considers its rankings as 'experimental' (see Zimmermann, 2007), they provide more and more a comprehensive overview of the competitive situation in the economic sciences.

We provide some cautionary remarks for each subcategory and demonstrate how the results can be biased in some cases. Furthermore, we illustrate how inconsistencies between the worldwide ranking and regional rankings arise due to the ranking methodology employed by RePEc. Doing this, we complement Zimmermann (2007) who notes that there are some limitations in RePEc. Additionally, we add the standardization approach (Vinkler, 2006) as an robust alternative to aggregate all rankings provided by RePEc.

The second, and more important, contribution of this paper is to answer the question how to extract key indicators of research performance. We suggest to apply a principal component analysis (PCA) to the data. Although this method has been used in the literature before but rather to classify determinants of research productivity,[5] PCA allows us to extract common components which, in the best case, explain most of the variance common to all included indicators. For each component, factor loadings can be calculated, which can be interpreted as weights, i.e. indicators with the high factor loadings are more important to explain the underlying latent process.

We illustrate the PCA with 27 bibliometric indicators for a sample of about 29,000 registered authors (economists) in RePEc from July 2011. We find that the first factor explains almost 90% of the variance in all indicators. Four of the included indicators have by far the highest loadings, i.e. the highest impact, for explaining scientific achievement in economics: number of journal pages and number of citations both weighted each by a simple impact factor and authors. We provide a worldwide ranking of economists based on the PCA. Our results are similar to the ones provided by RePEc, especially for the top economists. Nevertheless individual results may differ substantially.

The paper is structured as follows: Section 2 gives a literature overview of existing rankings of economists. Section 3 provides an overview of the RePEc database and gives some cautionary remarks concerning the rankings. We compare the existing rankings in RePEc with the robust standardization approach in Section 4. The alternative ranking aggregation approach is described and illustrated in Section 5. Finally, we conclude.

## 2. Existing rankings for economists

The assessment and ranking of research has a long tradition. The focus has been primarily on the assessment of universities, departments or research institutes.[6] In Table 1 we report existing studies for economists. It catches the eye that many of these rankings are based on counting published research or citations. The research output was often weighted by some quality measures, mostly impact factors. In a different approach rankings are based on citation counts or variations of the h-index. In addition to Ursprung and Zimmer (2006), in Germany the Handelsblatt ranking of German speaking economists and economic departments has gained a lot of attention. This ranking is also based on counting weighted research output, see Hofmeister and Ursprung (2008) for details. A second important point can be read from Table 1. Although some papers provide several rankings no aggregated ranking is provided. Furthermore, besides Baltagi (1999, 2003), who studies foremost econometricians, and Tol (2009) which ranks economists by the h-index, all rankings have a country-specific focus. The article by Coupé (2003) is one of the most comprehensive studies but is more then 10 years old and considers only data from 1990 to 2000. Thus, RePEc provides constantly updated rankings (single and aggregated) for institutions and authors. Furthermore, these rankings are available not only worldwide, but for different geographic regions and countries.

[4] For instance, Google Scholar as a source for citation analysis potentially suffers from the problem of clear identification of citations, which can lead to overestimation of citations, see Harzing and van der Wal (2009).

[5] See for instance Docampo (2011), Ramesh Babu and Singh (1998), or Ortega, Lopez-Romero, and Fernandez (2011).

[6] In addition to the U.S. studies mentioned in footnote 2, ranking results for European departments are stated in Kalaitzidakis, Mamuneas, and Stengos (2003), Portes (1987), or Combes and Linnemer (2003) among others. Besides the United States and Europe there exists a larger literature on German (speaking) authors and institutions. See Ketzler and Zimmermann (2009) for a recent example.

**Table 1**
Rankings for economists.

| Study | Ranking approach(es) | Aggregated ranking |
|---|---|---|
| Baltagi (1999) | (Quality adjusted) standardized pages | No |
| Baltagi (2003) | (Quality adjusted) standardized pages | No |
| Ben-David (2010) | Number of citations | No |
| Coupé (2003) | Quality weighted publications and citations | No |
| Dolado, Garcí a-Romero, and Zamarro (2003) | Number of citations, quality weighted publication counts | No |
| Henrekson and Waldenström (2011) | Citation measures | No |
| Medoff (1989) | Citation measures | No |
| Ruane and Tol (2008) | Publications, citations, variants of the h-index | No |
| Ruane and Tol (2009) | Successive h-indices | No |
| Sinha and Macri (2004) | Publication counts with different quality measures | No |
| Tol (2009) | Citations, variants of h-index | No |
| Ursprung and Zimmer (2006) | Citations, variants of h-index | No |

## 3. RePEc data and ranking system

### 3.1. The database

Based on all available bibliographic information within the network, RePEc calculates every month 33 different bibliometric indicators for registered authors and institutions. Table 2 provides an overview of these measures. There are five main categories: number of (published) works, citations, citation indices, citing authors, journal pages, and RePEc access statistics. Each of these main categories can be combined with different weighting schemes: simple or recursive impact factors, number of authors and combination of them. For the category *distinct number of works* different version of a paper are counted only once. Published work is only counted if, first the publisher provides the meta data to RePEc and second, the author assigns this work to his/her account. Currently there are more 1300 journals and almost 3000 working paper series listed in RePEc and the list is constantly expanding. To the best of our knowledge no major journal or working paper series is missing in RePEc. The indicators are not publicly available on the web page, RePEc only reports the bibliometric scores for the top 5% listed authors for each category. Therefore, only for authors belonging to the top 5% list in each category a complete record can be established. RePEc provides all scores with its corresponding worldwide rank for each author every month via email. Table 2 reveals that there is a focus on citations both directly and indirectly. In 14 out of 33 rankings citations are count with quality and time adjustments. The indirect channel is the different impact factors.

### 3.2. Some cautionary remarks

#### 3.2.1. Citations and impact factors

As noted above, citations and impact factors play a central role in RePEc, as in the assessment of science in general. They allow to differ between journals with respect to their importance, prestige and their position in the journal system. RePEc

**Table 2**
Bibliometric measures in RePEc.

| | Without any further weighting | Simple impact factor | Recursive impact factor | Number of authors | Number of authors + simple impact factor | Number of authors + recursive impact factor |
|---|---|---|---|---|---|---|
| Works | | | | | | |
| Overall | × | | | | | |
| Distinct | × | × | × | × | × | × |
| Citations | | | | | | |
| Overall | × | × | × | × | × | × |
| Discounted by citation year | × | × | × | × | × | × |
| Citing authors | | | | | | |
| Overall | × | | | | | |
| Weighted by authors rank | × | | | | | |
| Journal pages | × | × | × | × | × | × |
| Access via RePEc | | | | | | |
| Abstract views | × | | | × | | |
| Downloads | × | | | × | | |
| Indices | | | | | | |
| h-Index | × | | | | | |
| Wu-Index[a] | × | | | | | |

[a] Only for authors.

started to extract citations in 2003. It is aimed to gather all citations from listed works. Given the large number of registered series it is, besides the standard *Web of Science* (WoS), a further tool for citation tracking in economic sciences. Recently, *Scopus* and *Google Scholar* have been emerged as serious competitors.[7] How the RePEc citation database compares to others is an open question for future research.

RePEc has two main sources for extracting citations: First, it reads out all publicly available documents within the network. Due to missing (open) access to the article or technical problems it is not always possible to extract all citations. Second, archive maintainers may provide meta information on citations for their journals. Currently more than 1 million items are listed in RePEc where the majority allocates to working papers and journal articles. In contrast, only about 300,000 items have been processed by RePEc. Therefore, it is obvious there are still many missing citations. It is important to note that both the citing and the cited work have to be listed in RePEc. Assuming that almost all important series are indexed in RePEc and citations of articles outside of economics are rather minor, we assume that this fact does not introduce any large bias.

As noted in Table 1, some rankings are based on weights with various impact factors. The most well-known yearly impact factors are provided by WoS from Thomson Scientific in its Journal Citation Report (JCR). Although they are criticized for a number of reasons, see Glänzel and Moed (2002) for an overview, they still provide a glimpse of the quality of a journal. Focusing on the economic sciences, the JCR impact factors have two major drawbacks: First, the average time for a journal article from publication to peak in citations is not always two years. Furthermore, the publication process in economics is rather slow compared to natural sciences, see Ellison (2002), which leads to the fact that the impact factors are rather small. Second, the impact factors from JCR is restricted to a specific journal list. The subsection 'economics' lists only 304 journals for the JCR 2010. Thus, many citations from other economic journals are potentially missing.[8] RePEc accounts for these two issues: First, citations of articles from the whole journal history available in the network are included. Second, RePEc considers citations from all indexed series. Based on this, impact factors for all listed series are available (journals, working papers and book series). Although impact factors in RePEc are also restricted to citations from listed series, this list is much larger compared to the economics subcategory in the JCR. Currently more than 1300 journals (including some statistics and mathematics journals) are listed in RePEc. Another difference between the standard and the RePEc impact factor is the exclusion of 'self-citations' to prevent 'self-inflation'. Finally, the JCR impact factors are only updated once a year, whereas in RePEc updates on a regular basis. In addition to the standard impact factor, RePEc provides also a recursive impact factor. It gives citations from journals with higher impact larger weights than citations from low-impact journals. In economics this method goes back to Liebowitz and Palmer (1984).

Besides the number of citations, the impact factor is influenced by the number of published articles in the respective series. Concerning this point we have to note that different journals provide different records to RePEc. For example, for the *Journal of Political Economy* (JPE) almost the whole journal history is listed, starting in 1896 comprising currently more than 5300 items (August 2011). In contrast the *Quarterly Journal of Economics* (QJE) provides articles from volume 83 in 1969 on. As of August 2011 more than 2000 articles are listed in RePEc. Thus, it may not be surprising that the impact factor for the QJE is higher than for the JPE as can be seen in Table 3. In this table we compare the JCR 2- and 5-year impact factor with the corresponding RePEc ones.[9] We took the 304 journals from the economics subsection in the JCR. In the last row we document the Pearson and the Spearman correlation coefficient relative to the 2-year impact factor. First, it can be noted that the majority of impact factors in RePEc are large in values compared to the one obtained by JCR. One explanation is the inclusion of citations from different sources, such as working papers and more economic journals compared to SSCI mentioned above. Second, the 2- and 5-year impact factor is similar both in absolute terms as well as ranking positions with a relatively large correlation. Looking at the RePEc impact factors one can see that the relative ranking substantially differs compared to WoS. It will be interesting to see how these rankings compare as the citation record in RePEc improves in the future. We leave this for future research.

### 3.2.2. Access statistics

Zimmermann (2007) notes that access statistics of articles indicate attractiveness of past and current research. This leads to the assumption that the higher the number of abstract views and downloads the higher is the possible impact on current research and public discussions. We have three notes on this. First, the number of real downloads of journals is highly sceptically, because the access to downloads for majority of journals is restricted.[10] But there exists a download button that does not refer directly to the PDF document (as it is recommended by RePEc) but to web page of the publisher where the abstract is listed. In almost all cases one has to pay for a download of a specific article. A possible solution is that these kinds of pseudo-downloads should not be counted. Or, the publisher provides information about actually carried out downloads. The provided ranking on access statistics may be misleading for another reason. The researcher is free to choose

[7] See Mingers and Lipitakis (2010), Norris and Oppenheim (2007) or Neuhaus and Daniel (2008) for comparisons.

[8] See Nederhof (2006) for the issue of coverage in the Social Science Citation Index (SSCI), which contains the economics category as a subgroup, for the social sciences.

[9] Not for all journals a corresponding RePEc impact factor is available, as these journals are not listed yet.

[10] As of August 2011, e.g., for the American Economic Review (ranked first in the download ranking) RePEc counted 8046 downloads. PDF-Files are only available via payments.

**Table 3**
Comparison of impact factors.

| | Thomson Scientific | | | | RePEc | | | |
|---|---|---|---|---|---|---|---|---|
| | 2Y IF | Rank | 5Y IF | Rank | IF | Rank | RIF | Rank |
| Journal of Economic Literature | 7.432 | 1 | 8.076 | 1 | 31.011 | 2 | 2.095 | 6 |
| Quarterly Journal of Economics | 5.940 | 2 | 8.053 | 2 | 33.491 | 1 | 3.581 | 1 |
| Technological and Economic Development of Economy | 5.605 | 3 | | | | | | |
| Review of Financial Studies | 4.602 | 4 | 5.016 | 9 | 10.706 | 19 | 1.109 | 17 |
| Journal of Finance | 4.151 | 5 | 6.529 | 4 | 11.345 | 16 | 1.162 | 16 |
| Journal of Political Economy | 4.065 | 6 | 6.896 | 3 | 17.591 | 8 | 2.122 | 5 |
| Journal of Business Economics and Management | 3.866 | 7 | | | | | | |
| Journal of Financial Economics | 3.810 | 8 | 5.631 | 6 | 16.064 | 12 | 1.556 | 10 |
| Brookings Papers on Economic Activity | 3.783 | 9 | 3.364 | 16 | 18.061 | 9 | 2.157 | 4 |
| Journal of Economic Perspectives | 3.702 | 10 | 5.958 | 5 | 17.737 | 7 | 1.386 | 12 |
| Journal of Economic Geography | 3.662 | 11 | 4.487 | 10 | 2.893 | 80 | 0.134 | 90 |
| Pharmacoeconomics | 3.440 | 12 | 3.122 | 21 | 0.030 | 245 | 0.002 | 238 |
| Econometrica | 3.185 | 13 | 5.330 | 7 | 28.485 | 3 | 2.430 | 2 |
| American Economic Review | 3.150 | 14 | 4.278 | 12 | 15.866 | 13 | 1.637 | 9 |
| Review of Economic Studies | 3.113 | 15 | 4.300 | 11 | 18.534 | 6 | 1.850 | 8 |
| Economic Geography | 3.028 | 16 | 3.195 | 19 | 0.131 | 234 | 0.006 | 219 |
| Journal of Environmental Economics and Management | 2.989 | 17 | 3.029 | 22 | 6.809 | 34 | 0.208 | 71 |
| Journal of Urban Economics | 2.892 | 18 | 2.607 | 32 | 4.716 | 55 | 0.297 | 55 |
| Review of Economics and Statistics | 2.883 | 19 | 4.163 | 13 | 9.537 | 24 | 0.811 | 22 |
| Journal of Accounting and Economics | 2.817 | 20 | 5.268 | 8 | 4.077 | 59 | 0.244 | 64 |
| Review of Environmental Economics and Policy | 2.781 | 21 | 3.146 | 20 | 1.163 | 158 | 0.070 | 133 |
| Ecological Economics | 2.754 | 22 | 3.232 | 18 | 1.238 | 154 | 0.019 | 193 |
| Journal of Banking and Finance | 2.731 | 23 | 2.528 | 33 | 3.708 | 61 | 0.223 | 68 |
| Journal of Economic Growth | 2.458 | 24 | 3.467 | 15 | 28.111 | 4 | 2.231 | 3 |
| Energy Economics | 2.449 | 25 | 2.903 | 25 | 1.314 | 149 | 0.022 | 187 |
| Economics and Human Biology | 2.438 | 26 | | | 1.059 | 165 | 0.064 | 139 |
| Value in Health | 2.342 | 27 | 2.992 | 23 | | | | |
| Economic Journal | 2.271 | 28 | 2.710 | 30 | 10.798 | 18 | 0.789 | 24 |
| Journal of Policy Analysis and Management | 2.246 | 29 | 2.326 | 41 | 0.563 | 200 | 0.056 | 147 |
| Journal of Labor Economics | 2.244 | 30 | 3.708 | 14 | 16.078 | 11 | 1.501 | 11 |
| Correlation | | | 0.955[a] | 0.961[b] | 0.709[a] | 0.530[b] | 0.685[a] | 0.488[b] |

*Notes*: Correlations are calculated for the whole sample of 304 journals. SSCI impact factors are for 2010. 2- and 5-year impact factor includes citations for articles from the two and five preceding years respectively. RePEc impact factors were retrieved in August 2011, considers all available citations irrespective of a given period.

[a] Pearson correlation coefficient.

[b] Spearman rank correlation coefficient, both with respect to the 2-year impact factor.

the download directly from the publisher's web page. To give an example: The most downloaded paper from the IZA web page (http://www.iza.org) in December 2010 is by Gonzalez-Navarro and Quintana-Domeque (2010) with 1390 downloads. From the RePEc page this working paper was downloaded only 6 times in December. Thus, attractiveness of current research does not have to be signalled via the RePEc network. Finally, it is doubtful whether abstract views signal scientific quality. The latter one is only revealed *after* the abstract view, or better after the download and the reading of the article. Abstract views can, e.g., easily inflated by fancy titles. Thus, individual quality may not be reflected by abstract views or downloads. Rather aggregated trends are revealed.

### 3.2.3. *Inconsistencies in regional rankings*

Within the various rankings of individual authors there exists the phenomenon that, e.g. clearly U.S.-based authors, appear in rankings of different regions or countries. This is due to the fact that the overall author score is distributed to the different listed affiliations of each author. To give an example, Harald Uhlig is based at the economics department in Chicago. Because of his German Bundesbank affiliation he appears in German ranking at 48th position (August 2011). Many European rankings are 'contaminated' by non-European researchers (mostly Americans), due to the European based networks: CEPR (United Kingdom), CESifo and IZA (both Germany). In August 2011 43 authors of the top 25% economists in Germany appear in that ranking only due to their IZA or the CESifo affiliation. Among them are well-known economists as Joshua Angrist (MIT), Eric Hanushek (Standford) or Orley Ashenfelter (Princton). A possible solution to this problem could be that registered authors are only ranked in one country with their full score, which remains the institution rankings unaffected.

The final comment concerns these inconsistencies in regional rankings. Besides the 33 different rankings, RePEc calculates an average rank score for both, authors and institutions. One main disadvantage of this score is that it can produce some inconsistencies when comparing worldwide and regional rankings. This feature arise due to the fact that rankings are calculated for each region separately. For example, the Ifo Institute for Economic Research in Munich is ranked 5th in

**Table 4**
Illustration of regional ranking inconsistencies.

| | | I | II | III | IV | V | Harmonic mean | Arithmetic mean |
|---|---|---|---|---|---|---|---|---|
| Worldwide | A | 9 | 11 | 202 | 234 | 198 | 23.1 | 130.8 |
| Ranking | B | 175 | 182 | 135 | 152 | 178 | 162.3 | 164.4 |
| Regional | A | 1 | 1 | 2 | 2 | 2 | 1.4 | 1.6 |
| Ranking | B | 2 | 2 | 1 | 1 | 1 | 1.3 | 1.4 |

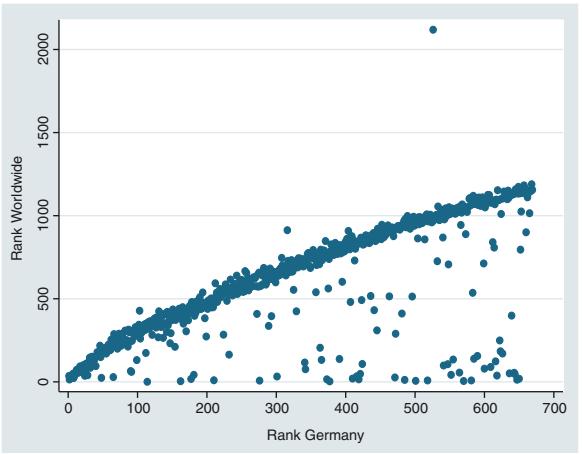


**Fig. 1.** Rank differences between regional and worldwide rankings.

the German ranking as of August 2011, but is the second best institution from Germany in the worldwide ranking. These inconsistencies arise from averaging the mean ranks instead of averaging the underlying scores. We explain this problem by a simple example: Suppose there are two authors A and B in a particular region and you have five ranking criteria I–V, see Table 4. Author A leads clearly in rankings I and II but is only slightly behind B in rankings III–V. Because of the significant lead in the first two rankings, A gets a better mean rank. If this is transferred to the regional ranking (and keeping all scores equal to the worldwide), the great lead of A has vanished. Since B is the leader in 3 out of 5 rankings, it gets a better average rank score and therefore leads the regional ranking. This phenomenon is known as *Simpsons paradox* (Simpsons, 1951).

How are the regional rankings affected by this paradox? In Fig. 1 we give an example for Germany's author ranking. In case of a consistent regional ranking that is derived from the overall worldwide ranking, we would obtain a straight line. We find large inconsistencies in the regional rankings. Many non-German based authors have a low ranking in Germany but reach a top position worldwide. This is due to fact that only a fraction of scores is attributed to Germany as their main affiliation is outside of Germany.

### 3.3. A descriptive look at the RePEc data

Before we turn to ranking calculations we provide some descriptive statistics on the bibliometric measures. We obtained a data set from RePEc containing all 33 indicators for 29,082 authors from the July 2011 ranking.[11] In Table 5 we report the mean, median, the minimum and maximum score, and the relative share of authors with a zero score. It is obvious that the scores are not comparable across categories, thus a ranking based on the simple average mean across categories would be highly distorted. For instance a score of 56 is very large in the *h-index* category but not for the *number of journal pages*. Looking at the category *number of citations* and its variation one can see that there are about 30% of all authors with no recorded citations. But it is unknown whether the authors have not been cited or the potentially existing citations have not been indexed by the network yet. The share of 19% of authors with no journal pages can by explained by the fact that the recorded items are either working papers, books, chapters or software codes.[12] Comparing the mean with the median we see that the data is highly skewed. The last column in Table 5 describes the ratio between the second largest to the largest value in each category. One can clearly see that there some categories with a large distance between the best and the second best score.

[11] The indicators are not publicly available on the web page. RePEc only reports the bibliometric scores for the top 5% listed authors for each category.
[12] The Munich RePEc personal archive (http://mpra.ub.uni-muenchen.de/) allows each author to submit a paper. This opportunity is well taken by authors who do not have access to (institutional) working paper series.

**Table 5**
Descriptive statistics for different rankings in RePEc.

| Label | Category | Mean | Median | SD | Min | Max | Zero | OL |
|---|---|---|---|---|---|---|---|---|
| Nb_Works | No. of works | 22.69 | 10.00 | 36.25 | 1 | 869.00 | 0.00 | 0.69 |
| Dnb_works | Distinct No. of works | 17.01 | 8.00 | 25.40 | 0 | 774.00 | 0.00 | 0.53 |
| Sc_works | No. of distinct works, W. by simple IF | 96.99 | 16.43 | 261.44 | 0 | 5582.82 | 0.01 | 0.94 |
| WSc_works | No. of distinct works, W. by Recursive IF | 0.66 | 0.07 | 1.98 | 0 | 39.77 | 0.18 | 1.00 |
| Anb_works | No. of distinct works, W. by No. of authors | 11.13 | 5.17 | 17.87 | 0 | 387.30 | 0.00 | 0.96 |
| Asc_works | No. of distinct works, W. by No. of authors and simple IFs | 55.18 | 8.89 | 158.98 | 0 | 4620.52 | 0.02 | 0.73 |
| AWSc_works | No. of distinct works, W. by No. of authors and recursive IFs | 0.38 | 0.04 | 1.20 | 0 | 32.92 | 0.23 | 0.75 |
| Nb_cites | No. of citations | 81.55 | 6.00 | 325.28 | 0 | 11865.00 | 0.28 | 0.81 |
| D_cites | No. of citations, discounted by citation age | 19.86 | 1.79 | 71.69 | 0 | 2549.10 | 0.28 | 0.79 |
| Sc_cites | No. of citations, W. by simple IF | 425.87 | 12.18 | 2013.63 | 0 | 73245.69 | 0.28 | 0.78 |
| DSc_cites | No. of citations, W. by simple IF, discounted by citation age | 23.79 | 0.97 | 99.98 | 0 | 3637.82 | 0.28 | 0.79 |
| WSc_cites | No. of citations, W. by recursive IF | 2.92 | 0.06 | 14.38 | 0 | 512.44 | 0.36 | 0.82 |
| WDSc_cites | No. of citations, W. by recursive IF, discounted by citation age | 0.81 | 0.02 | 3.53 | 0 | 123.94 | 0.32 | 0.78 |
| ANb_cites | No. of citations, W. by No. of authors | 44.36 | 3.00 | 186.82 | 0 | 6979.16 | 0.28 | 0.91 |
| AD_cites | No. of citations, W. by No. of authors, discounted by citation age | 10.54 | 0.92 | 39.59 | 0 | 1274.93 | 0.28 | 0.95 |
| ASc_cites | No. of citations, W. by No. of authors and simple IFs | 234.05 | 5.95 | 1159.67 | 0 | 46200.35 | 0.28 | 0.91 |
| ADSc_cites | No. of citations, W. by No. of authors and simple IFs, discounted by citation age | 12.68 | 0.48 | 55.01 | 0 | 1875.77 | 0.28 | 0.90 |
| AWSc_cites | No. of citations, W. by No. of authors and recursive IFs | 1.61 | 0.03 | 8.31 | 0 | 367.97 | 0.39 | 0.82 |
| AWDSc_cites | No. of citations, W. by No. of authors and recursive IFs, discounted by citation age | 0.43 | 0.01 | 1.95 | 0 | 73.60 | 0.33 | 0.82 |
| H-Index | h-Index | 2.63 | 1.00 | 3.68 | 0 | 56.00 | 0.27 | 0.82 |
| NC_authors | No. of registered citing authors | 52.89 | 5.00 | 166.03 | 0 | 4036.00 | 0.30 | 0.99 |
| RC_authors | No. of registered citing authors, W. by rank (max 1 per author) | 40.28 | 3.71 | 127.86 | 0 | 3066.52 | 0.30 | 0.97 |
| Nb_pages | No. of journal pages | 143.86 | 55.00 | 245.63 | 0 | 4822.00 | 0.19 | 0.97 |
| Sc_pages | No. of journal pages, W. by simple IF | 782.45 | 76.87 | 2222.97 | 0 | 57360.42 | 0.20 | 0.95 |
| WSc_pages | No. of journal pages, W. by recursive IF | 5.26 | 0.25 | 16.54 | 0 | 433.81 | 0.28 | 0.93 |
| Anb_pages | No. of journal pages, W. by No. of authors | 82.87 | 31.25 | 146.04 | 0 | 3145.41 | 0.19 | 0.99 |
| ASc_pages | No. of journal pages, W. by No. of authors and simple IFs | 451.01 | 40.51 | 1340.65 | 0 | 37293.48 | 0.20 | 0.98 |
| AWSc_pages | No. of journal pages, W. by No. of authors and recursive IFs | 3.04 | 0.13 | 9.96 | 0 | 272.90 | 0.30 | 0.93 |
| Abs_views | No. of abstract views in RePEc over the past 12 months | 871.38 | 332.00 | 1840.58 | 0 | 58941.00 | 0.00 | 0.85 |
| Downloads | No. of downloads through RePEc over the past 12 months | 278.52 | 101.00 | 646.12 | 0 | 19799.00 | 0.01 | 0.88 |
| AAbs_views | No. of abstract views in RePEc over the past 12 months, W. by No. of authors | 441.77 | 169.00 | 1016.35 | 0 | 39966.00 | 0.00 | 0.89 |
| A_Downloads | No. of downloads through RePEc over the past 12 months, W. by No. of authors | 139.89 | 52.00 | 352.34 | 0 | 12136.00 | 0.02 | 0.92 |

*Notes*: This table reports the descriptive statistics for all categories for the July 2011 in RePEc. No. = number, W. = weighted, IF = impact factor, *SD* = standard deviation, *Zero* reports the percentage of authors with a score of zero. *OL* denotes the ratio of the second largest to the largest value in each category.

In Table 6 we tabulate the Pearson linear cross-correlations between all 31 bibliometric measures.[13] In contrast to Zimmermann (2007) we report the linear and not the rank correlations. All pairwise correlations are significantly different from zero at the 1% level. The average correlation is $\bar{r} = 0.797$ and varies between 0.460 and 0.999.[14] The table groups criteria in categories (number of works, citations, derived from citations, article pages, visibility on RePEc), and not surprisingly, correlations within these categories tend to be higher than with other categories.[15] Let us take a look at the details: Publishing more has a positive effect on the number of cites but this relationship is not that strong as may be expected comparing the other correlations. It can also be seen that quality weighted works have about 0.3 higher correlations with the citations measures than the (unweighted) distinct number of works. In contrast to the general expectations, a higher publication record (weighted or unweighted) is not strongly correlated with the access statistics.

## 4. Ranking calculation in RePEc

### 4.1. Aggregated rankings provided in RePEc

Based on categories in Table 1 RePEc computes an ordinal rankings for each indicator and all registered authors. In order to get an overall picture an aggregated ranking is provided. For the overall ranking the category *number of works* is omitted.[16] Furthermore, the personal best and worst ranking results are excluded. It avoids both 'one hit ranking wonder' at the top and single outliers at the bottom.

The generalized mean for $N$ different rankings $r_i$ is given by

$$M_p = \left( \frac{1}{N} \sum_{i=1}^{n} w_i r_i^p \right)^{1/p}. \tag{1}$$

In RePEc the weights $w_i$ are set to one, i.e. all rankings have the same weight. For $p = 1$ we obtain the arithmetic mean, which penalizes poor ranks, $p = -1$ results in the harmonic mean, which favors good ranks. The latter one is the standard approach as this ranking is reported on the web page.[17] To illustrate the difference between the arithmetic and harmonic mean consider Christopher Baum from Boston as an example. He is ranked 15th as of August 2011 based on the harmonic mean in the worldwide ranking. Employing the arithmetic mean for ranking aggregation his rank would be 882nd. The reason is that Christopher Baum is top ranked in the four access statics categories and *number of works* (software components) but much lower ranked in the citations categories. For $p = 0$ we obtain the geometric mean which balances both. Two further aggregation approaches, the lexicographic and the graphicolexic ordering of ranks, both rely on the ordering of the ranks, where the first rewards most extreme positive ranks and the second the other way round. See Zimmermann (2007) for details. All these aggregation approaches are provided by RePEc on its web page.

### 4.2. An alternative: Rankings based on standardized scores

The transformation of scores to an ordinal ranking in RePEc prior to aggregation has the large disadvantage that the true underlying distribution of scores is discarded, i.e. relative distance between two authors vanishes. To give an example: Peter Nijkamp is ranked first in the category *number of distinct works* with a score of 766 as of August 2011. Nicholas Cox, ranked 2nd, has a score of 411. Although Nijkamp has almost a twice as large score this advantage vanishes in the ordinal ranking. A score of 412 would be enough to end up at the same position in the aggregate ranking based on generalized means. Therefore, RePEc also offers the percentage criterion. The best score is attributed 100% and then proportionally percentages to the smaller scores. Finally, all percentages are averaged by the arithmetic mean and ranked:

$$M_{\text{percentage}} = \frac{1}{N} \sum_{j=1}^{N} \frac{S_{ij}}{\max S_j} w_i, \tag{2}$$

where $S_{ij}$ denotes the score for individual $i$ in category $j$. Although, this criterion has the advantage that it accounts for the relative distances in the underlying scores but is prone to outliers. McAllister, Narin, and Corrigan (1983) suggested to

[13] We exclude *number of works* and the *Wu-Index*.
[14] The average correlation is highly significantly different from zero using an Chi-Square test, thus the indicators are not independent.
[15] This confirms the finding of Zimmermann (2007) with a smaller database.
[16] One obvious reason is that this category can easily be inflated by publishing the same work in many working paper series.
[17] This aggregation approach is also used in the personal ranking analysis provided monthly for each registered author.

**Table 6**
Correlations between bibliometric indicators in RePEc.

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Dnb_works | 1 | 0.72 | 0.67 | 0.96 | 0.70 | 0.65 | 0.55 | 0.56 | 0.48 | 0.49 | 0.47 | 0.48 | 0.54 | 0.55 | 0.48 | 0.49 | 0.46 | 0.47 | 0.73 | 0.59 | 0.59 | 0.88 | 0.66 | 0.60 | 0.85 | 0.64 | 0.59 | 0.76 | 0.69 | 0.70 | 0.65 |
| 2 | Sc_works | 0.72 | 1 | 0.99 | 0.77 | 0.98 | 0.97 | 0.83 | 0.83 | 0.82 | 0.83 | 0.81 | 0.82 | 0.81 | 0.82 | 0.81 | 0.82 | 0.79 | 0.81 | 0.85 | 0.87 | 0.87 | 0.79 | 0.93 | 0.91 | 0.79 | 0.91 | 0.89 | 0.78 | 0.76 | 0.71 | 0.70 |
| 3 | WSc_works | 0.67 | 0.99 | 1 | 0.72 | 0.97 | 0.98 | 0.83 | 0.83 | 0.84 | 0.84 | 0.83 | 0.83 | 0.81 | 0.82 | 0.82 | 0.83 | 0.81 | 0.82 | 0.83 | 0.86 | 0.87 | 0.75 | 0.92 | 0.92 | 0.75 | 0.90 | 0.90 | 0.75 | 0.74 | 0.68 | 0.69 |
| 4 | Anb_works | 0.96 | 0.77 | 0.72 | 1 | 0.78 | 0.74 | 0.58 | 0.58 | 0.53 | 0.53 | 0.51 | 0.51 | 0.59 | 0.60 | 0.54 | 0.55 | 0.53 | 0.53 | 0.73 | 0.62 | 0.62 | 0.86 | 0.69 | 0.65 | 0.88 | 0.71 | 0.66 | 0.76 | 0.70 | 0.75 | 0.70 |
| 5 | Asc_works | 0.70 | 0.98 | 0.97 | 0.78 | 1 | 0.99 | 0.80 | 0.78 | 0.79 | 0.78 | 0.78 | 0.77 | 0.81 | 0.81 | 0.81 | 0.81 | 0.80 | 0.80 | 0.81 | 0.83 | 0.84 | 0.76 | 0.90 | 0.89 | 0.80 | 0.92 | 0.91 | 0.74 | 0.72 | 0.71 | 0.71 |
| 6 | AWSc_works | 0.65 | 0.97 | 0.98 | 0.74 | 0.99 | 1 | 0.79 | 0.78 | 0.80 | 0.79 | 0.80 | 0.79 | 0.81 | 0.81 | 0.82 | 0.82 | 0.81 | 0.81 | 0.79 | 0.83 | 0.84 | 0.73 | 0.89 | 0.89 | 0.76 | 0.92 | 0.92 | 0.71 | 0.70 | 0.69 | 0.69 |
| 7 | Nb_cites | 0.55 | 0.83 | 0.83 | 0.58 | 0.80 | 0.79 | 1 | 0.99 | 0.98 | 0.97 | 0.96 | 0.96 | 0.97 | 0.97 | 0.94 | 0.95 | 0.93 | 0.94 | 0.80 | 0.97 | 0.96 | 0.65 | 0.83 | 0.83 | 0.64 | 0.80 | 0.80 | 0.80 | 0.83 | 0.72 | 0.77 |
| 8 | D_cites | 0.56 | 0.83 | 0.83 | 0.58 | 0.78 | 0.78 | 0.99 | 1 | 0.96 | 0.98 | 0.94 | 0.97 | 0.94 | 0.97 | 0.92 | 0.95 | 0.90 | 0.93 | 0.82 | 0.97 | 0.96 | 0.66 | 0.83 | 0.82 | 0.63 | 0.79 | 0.78 | 0.81 | 0.84 | 0.72 | 0.77 |
| 9 | Sc_cites | 0.48 | 0.82 | 0.84 | 0.53 | 0.79 | 0.80 | 0.98 | 0.96 | 1 | 0.98 | 1.00 | 0.99 | 0.95 | 0.94 | 0.97 | 0.97 | 0.96 | 0.97 | 0.76 | 0.95 | 0.95 | 0.59 | 0.82 | 0.83 | 0.58 | 0.79 | 0.80 | 0.75 | 0.79 | 0.68 | 0.73 |
| 10 | DSc_cites | 0.49 | 0.83 | 0.84 | 0.53 | 0.78 | 0.79 | 0.97 | 0.98 | 0.98 | 1 | 0.98 | 1.00 | 0.93 | 0.95 | 0.94 | 0.97 | 0.93 | 0.96 | 0.78 | 0.95 | 0.95 | 0.60 | 0.82 | 0.84 | 0.58 | 0.78 | 0.79 | 0.77 | 0.81 | 0.68 | 0.73 |
| 11 | WSc_cites | 0.47 | 0.81 | 0.83 | 0.51 | 0.78 | 0.80 | 0.96 | 0.94 | 1.00 | 0.98 | 1 | 0.98 | 0.94 | 0.93 | 0.97 | 0.96 | 0.97 | 0.97 | 0.75 | 0.94 | 0.94 | 0.57 | 0.81 | 0.83 | 0.57 | 0.78 | 0.80 | 0.73 | 0.77 | 0.67 | 0.72 |
| 12 | WDSc_cites | 0.48 | 0.82 | 0.83 | 0.51 | 0.77 | 0.79 | 0.96 | 0.97 | 0.99 | 1.00 | 0.98 | 1 | 0.92 | 0.94 | 0.94 | 0.96 | 0.94 | 0.97 | 0.77 | 0.94 | 0.94 | 0.58 | 0.81 | 0.83 | 0.57 | 0.77 | 0.79 | 0.75 | 0.79 | 0.67 | 0.72 |
| 13 | ANb_cites | 0.54 | 0.81 | 0.81 | 0.59 | 0.81 | 0.81 | 0.97 | 0.94 | 0.95 | 0.93 | 0.94 | 0.92 | 1 | 0.99 | 0.98 | 0.97 | 0.96 | 0.96 | 0.77 | 0.94 | 0.94 | 0.63 | 0.81 | 0.81 | 0.65 | 0.82 | 0.82 | 0.76 | 0.80 | 0.74 | 0.80 |
| 14 | AD_cites | 0.55 | 0.82 | 0.82 | 0.60 | 0.81 | 0.81 | 0.97 | 0.97 | 0.94 | 0.95 | 0.93 | 0.94 | 0.99 | 1 | 0.96 | 0.98 | 0.95 | 0.97 | 0.80 | 0.95 | 0.95 | 0.65 | 0.82 | 0.82 | 0.66 | 0.82 | 0.82 | 0.79 | 0.82 | 0.75 | 0.80 |
| 15 | ASc_cites | 0.48 | 0.81 | 0.82 | 0.54 | 0.81 | 0.82 | 0.94 | 0.92 | 0.97 | 0.94 | 0.97 | 0.94 | 0.98 | 0.96 | 1 | 0.99 | 1.00 | 0.99 | 0.73 | 0.92 | 0.93 | 0.58 | 0.80 | 0.82 | 0.60 | 0.81 | 0.82 | 0.72 | 0.76 | 0.70 | 0.76 |
| 16 | ADSc_cites | 0.49 | 0.82 | 0.83 | 0.55 | 0.81 | 0.82 | 0.95 | 0.95 | 0.97 | 0.97 | 0.96 | 0.96 | 0.97 | 0.98 | 0.99 | 1 | 0.98 | 1.00 | 0.77 | 0.94 | 0.94 | 0.60 | 0.82 | 0.83 | 0.61 | 0.82 | 0.83 | 0.75 | 0.79 | 0.71 | 0.76 |
| 17 | AWSc_cites | 0.46 | 0.79 | 0.81 | 0.53 | 0.80 | 0.81 | 0.93 | 0.90 | 0.96 | 0.93 | 0.97 | 0.94 | 0.96 | 0.95 | 1.00 | 0.98 | 1 | 0.99 | 0.72 | 0.91 | 0.91 | 0.56 | 0.79 | 0.81 | 0.58 | 0.80 | 0.82 | 0.70 | 0.74 | 0.69 | 0.74 |
| 18 | AWDSc_cites | 0.47 | 0.81 | 0.82 | 0.53 | 0.80 | 0.81 | 0.94 | 0.93 | 0.97 | 0.96 | 0.97 | 0.97 | 0.96 | 0.97 | 0.99 | 1.00 | 0.99 | 1 | 0.75 | 0.93 | 0.93 | 0.58 | 0.81 | 0.83 | 0.59 | 0.81 | 0.82 | 0.73 | 0.77 | 0.69 | 0.75 |
| 19 | H-index | 0.73 | 0.85 | 0.83 | 0.73 | 0.81 | 0.79 | 0.80 | 0.82 | 0.76 | 0.78 | 0.75 | 0.77 | 0.77 | 0.80 | 0.73 | 0.77 | 0.72 | 0.75 | 1 | 0.86 | 0.86 | 0.80 | 0.83 | 0.81 | 0.77 | 0.80 | 0.77 | 0.77 | 0.76 | 0.68 | 0.69 |
| 20 | NC_authors | 0.59 | 0.87 | 0.86 | 0.62 | 0.83 | 0.83 | 0.97 | 0.97 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.92 | 0.94 | 0.91 | 0.93 | 0.86 | 1 | 1.00 | 0.69 | 0.86 | 0.85 | 0.67 | 0.83 | 0.83 | 0.79 | 0.82 | 0.72 | 0.76 |
| 21 | RC_authors | 0.59 | 0.87 | 0.87 | 0.62 | 0.84 | 0.84 | 0.96 | 0.96 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.93 | 0.94 | 0.91 | 0.93 | 0.86 | 1.00 | 1 | 0.68 | 0.86 | 0.86 | 0.67 | 0.84 | 0.83 | 0.79 | 0.81 | 0.72 | 0.75 |
| 22 | Nb_pages | 0.88 | 0.79 | 0.75 | 0.86 | 0.76 | 0.73 | 0.65 | 0.66 | 0.59 | 0.60 | 0.57 | 0.58 | 0.63 | 0.65 | 0.58 | 0.60 | 0.56 | 0.58 | 0.80 | 0.69 | 0.68 | 1 | 0.82 | 0.77 | 0.97 | 0.79 | 0.75 | 0.73 | 0.69 | 0.67 | 0.64 |
| 23 | Sc_pages | 0.66 | 0.93 | 0.92 | 0.69 | 0.90 | 0.89 | 0.83 | 0.83 | 0.82 | 0.82 | 0.81 | 0.81 | 0.81 | 0.82 | 0.80 | 0.82 | 0.79 | 0.81 | 0.83 | 0.86 | 0.86 | 0.82 | 1 | 0.99 | 0.81 | 0.97 | 0.96 | 0.73 | 0.73 | 0.67 | 0.69 |
| 24 | WSc_pages | 0.60 | 0.91 | 0.92 | 0.65 | 0.89 | 0.89 | 0.83 | 0.82 | 0.83 | 0.84 | 0.83 | 0.83 | 0.81 | 0.82 | 0.82 | 0.83 | 0.81 | 0.83 | 0.81 | 0.85 | 0.86 | 0.77 | 0.99 | 1 | 0.76 | 0.97 | 0.97 | 0.71 | 0.71 | 0.66 | 0.67 |
| 25 | Anb_pages | 0.85 | 0.79 | 0.75 | 0.88 | 0.80 | 0.76 | 0.64 | 0.63 | 0.58 | 0.58 | 0.57 | 0.57 | 0.65 | 0.66 | 0.60 | 0.61 | 0.58 | 0.59 | 0.77 | 0.67 | 0.67 | 0.97 | 0.81 | 0.76 | 1 | 0.83 | 0.78 | 0.70 | 0.67 | 0.69 | 0.66 |
| 26 | ASc_pages | 0.64 | 0.91 | 0.90 | 0.71 | 0.92 | 0.92 | 0.80 | 0.79 | 0.79 | 0.78 | 0.78 | 0.77 | 0.82 | 0.82 | 0.81 | 0.82 | 0.80 | 0.81 | 0.80 | 0.83 | 0.84 | 0.79 | 0.97 | 0.97 | 0.83 | 1 | 0.99 | 0.70 | 0.70 | 0.68 | 0.69 |
| 27 | AWSc_pages | 0.59 | 0.89 | 0.90 | 0.66 | 0.91 | 0.92 | 0.80 | 0.78 | 0.80 | 0.79 | 0.80 | 0.79 | 0.82 | 0.82 | 0.82 | 0.83 | 0.82 | 0.82 | 0.77 | 0.83 | 0.83 | 0.75 | 0.96 | 0.97 | 0.78 | 0.99 | 1 | 0.68 | 0.68 | 0.66 | 0.68 |
| 28 | Abs_views | 0.76 | 0.78 | 0.75 | 0.76 | 0.74 | 0.71 | 0.80 | 0.81 | 0.75 | 0.77 | 0.73 | 0.75 | 0.76 | 0.79 | 0.72 | 0.75 | 0.70 | 0.73 | 0.77 | 0.79 | 0.79 | 0.73 | 0.73 | 0.71 | 0.70 | 0.70 | 0.68 | 1 | 0.97 | 0.94 | 0.92 |
| 29 | Downloads | 0.69 | 0.76 | 0.74 | 0.70 | 0.72 | 0.70 | 0.83 | 0.84 | 0.79 | 0.81 | 0.77 | 0.79 | 0.80 | 0.82 | 0.76 | 0.79 | 0.74 | 0.77 | 0.76 | 0.82 | 0.81 | 0.69 | 0.73 | 0.71 | 0.67 | 0.70 | 0.68 | 0.97 | 1 | 0.91 | 0.94 |
| 30 | AAbs_views | 0.70 | 0.71 | 0.68 | 0.75 | 0.71 | 0.69 | 0.72 | 0.72 | 0.68 | 0.68 | 0.67 | 0.67 | 0.74 | 0.75 | 0.70 | 0.71 | 0.69 | 0.69 | 0.68 | 0.72 | 0.72 | 0.67 | 0.67 | 0.66 | 0.69 | 0.68 | 0.66 | 0.94 | 0.91 | 1 | 0.96 |
| 31 | A_Downloads | 0.65 | 0.70 | 0.69 | 0.70 | 0.71 | 0.69 | 0.77 | 0.77 | 0.73 | 0.73 | 0.72 | 0.72 | 0.80 | 0.80 | 0.76 | 0.76 | 0.74 | 0.75 | 0.69 | 0.76 | 0.75 | 0.64 | 0.69 | 0.67 | 0.66 | 0.69 | 0.68 | 0.92 | 0.94 | 0.96 | 1 |

This table ranks economists based on the RePEc database from July 2011 and the PCA applied to 27 bibliometric indicators. 'Top 4' represents the ranks based on the four indicators with the highest loading in Fig. 3. The last two columns reports the corresponding recalculated RePEc ranks either based on the harmonic or arithmetic mean.

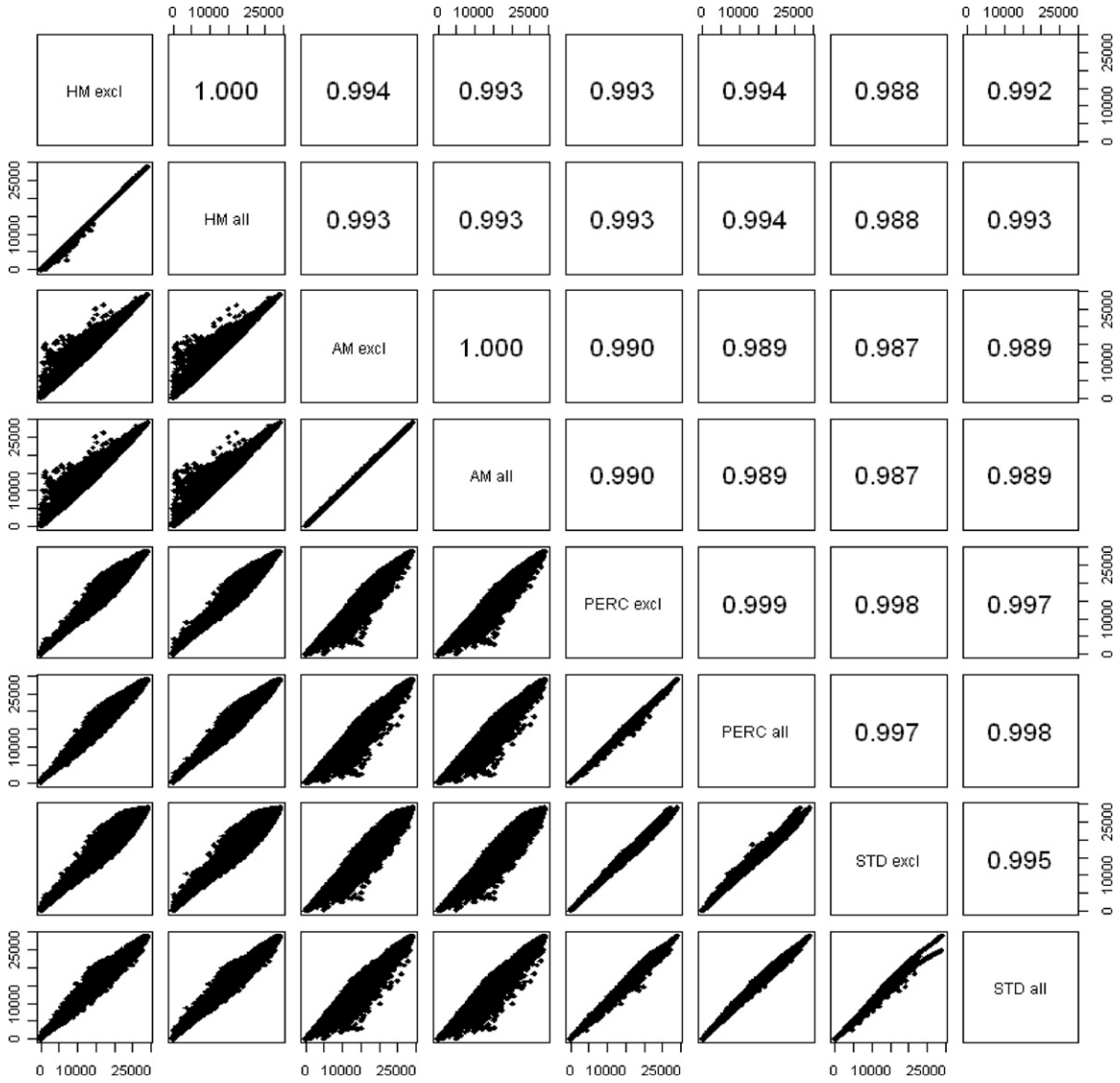


**Fig. 2.** Ranking comparison in RePEc.

standardize the underlying scores to obtain robust results. Given the mean $m(\cdot)$ and standard deviation $SD(\cdot)$ for category $j$ the aggregated ranking is given by

$$M_z = \frac{1}{N}\sum_{j=1}^{N}\frac{S_{ij} - m(S_j)}{SD_j}w_i. \tag{3}$$

These so-called $Z$-scores were also used in Vinkler (2006) for research evaluation.

### 4.3. An empirical comparison of ranking aggregation approaches

We now illustrate how the aggregated ranking results differ across the different approaches. In Fig. 2 we show the crossplots between the harmonic mean (HM), the arithmetic mean (AM), the percentage approach (PERC) and the aggregation based on standardized scores (STD) which are given in Eqs. (1), (2), and (3) respectively. For all four approaches we both include all 31 measures (all) and 31 excluding the personal best and worst ranking (excl). In addition to the graphs we tabulate the Spearman rank correlations. It is obvious that the results differ only marginally. The correlations are very close to one. Nevertheless the change in ranks can be substantial for authors across methods for authors. Comparing the harmonic and arithmetic mean, which could be considered as the first 'natural' alternative for ranking aggregation like it is done for school marks, these differences are highly skewed to the right. There are larger losses than gains for individual authors. The

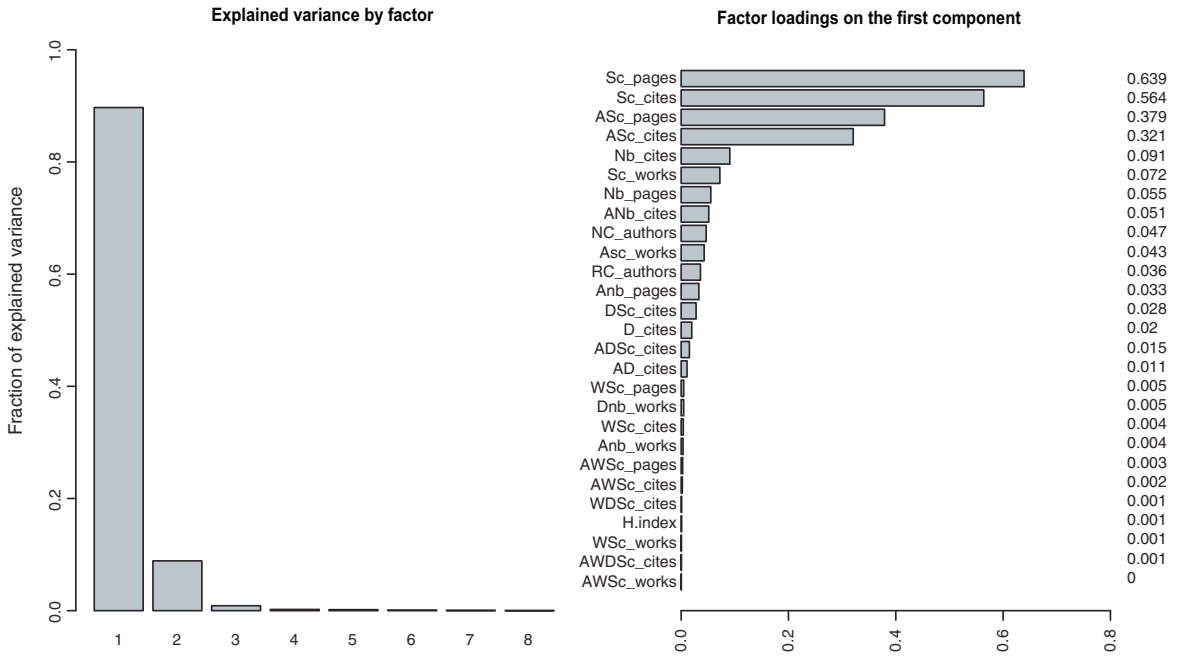


**Fig. 3.** Explained variance and factor loadings from the PCA.

standard deviation of the differences is approximately 930 ranks. The maximum loss of ranking positions of 13,867 is for an author which is highly ranked in the access statistics but very low in the other categories. The maximum gain is 1211 positions. The exclusion of outliers has a small effect on individual results. Although there are some outliers (maximum loss is 4300), a standard deviation of 140 ranks indicates that individual authors are rarely affected by the exclusion their best and worst ranking. How does the standardization approach affects the overall ranking compared to the percentage criterion? The standard deviation is about 500 positions and the maximum position change is 2186, i.e. for individual authors there can be substantial rank differences between these two approaches.

## 5. An alternative approach

We already stated that 33 bibliometric measures to assess scientific achievement are available in RePEc. For all indicators you can find pros and cons. For example: Is the unweighted publication record a correct measure? Which impact factor is the right one? Is discounting a good idea, as science can be regarded as time-independent? Do abstract views reflect any quality? This list can be expanded in several ways. In the previous sections we saw that these measures are very similar. But do they measure all by what is understood by research performance? Are there some key indicators? It is obvious that we cannot set up an objective list from a theoretical point of view that represents all aspects. As such a list of indicators is unknown it would be nevertheless desirable to have a 'shortlist' with the key indicators. The 33 indicators in RePEc cannot be considered to be short. The presented aggregation approaches in RePEc all assume equal weighting. For parsimony we need an approach to select the relevant indicators from it. Vinkler (2006) calls for an appropriate weighting scheme prior to aggregation. But how to choose these weights? Unfortunately there is no benchmark at which all rankings can be evaluated.

Therefore, we propose to define research performance as a latent process. Each of our 33 indicators can be regarded as an observed representation of this process. To extract the main variables, we run a principle component analysis to extract the most important components. Although this method has been used the literature before but rather to classify determinants of research productivity. See for instance Costas and Bordons (2007), Docampo (2011), Franceschet (2009), Ramesh Babu and Singh (1998), or Ortega et al. (2011). We propose to use the factor loadings as the basis for constructing weights for each available indicator.

We apply PCA to our data set of about 29,000 economists from all over the world. We started with the full set of 31 indicators.[18] We obtained a dominant factor which explains more than 80% of the variance. But the category *abstract views* showed a high loading on this first factor. As we outlined in Section 3.2.2 we do not think that this category refers to quality, but rather to actual trends. Therefore, we decided to leave out all four indicators from the access statistics categories. Thus, the following results are based on the remaining 27 indicators.

[18] We excluded *Number of Works* and the *Wu-Index*.

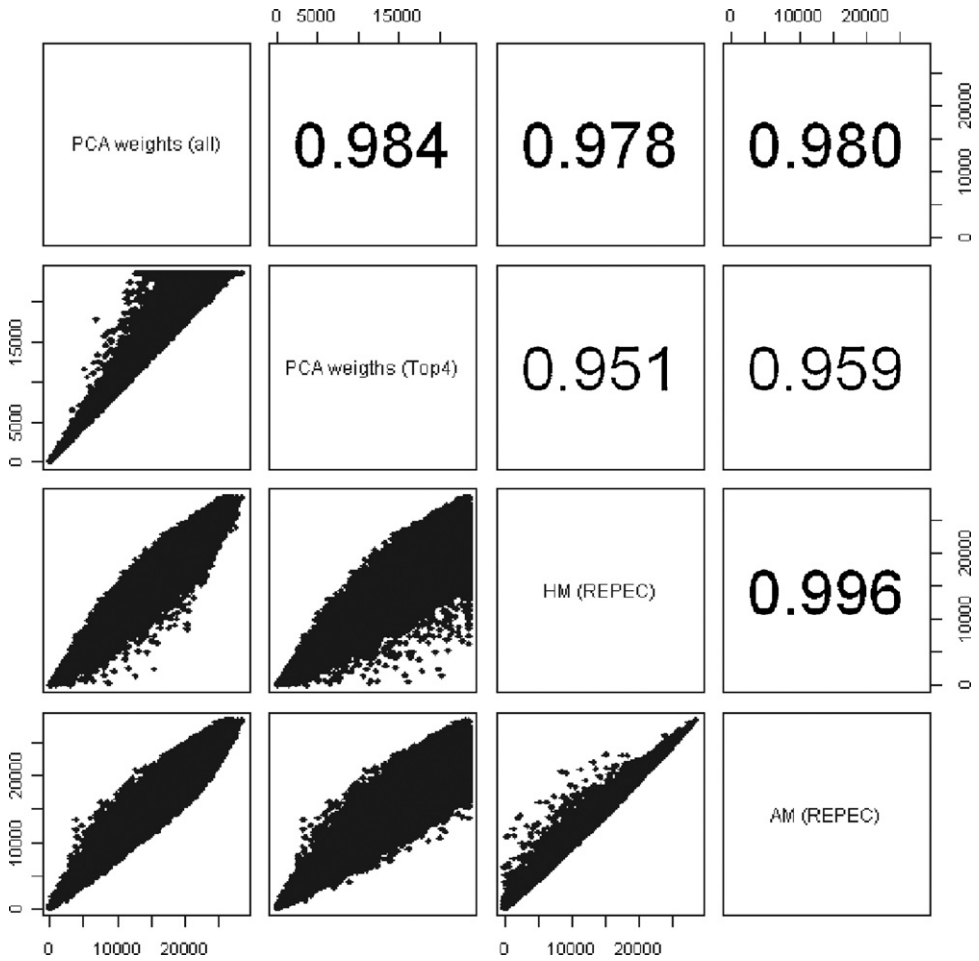


**Fig. 4.** Ranking comparison between PCA and RePEc rankings.

In the left panel of Fig. 3 we show the fraction of the explained variance of the first eight factors. The first factor explains almost 90% of the variance of all series included. The second one only explains about 9% and the other ones are negligible. Thus, we decide to focus on the first factor.[19] The right panel plots the factor loadings for the first factor. There are four indicators that stand out: *number of journal pages weighted by simple impact factor*, *number of cites weighted by simple impact factor*, *number of journal pages weighted by simple impact factor and number of authors*, and *number of cites weighted by simple impact factor and number of authors*. We have two groups which have the highest impact on the latent factor: number of journal pages and number of cites. Whereas the latter one is an expected result, the former one may come as a surprise. This can be explained by the fact that the number of journal articles is included in the *number of distinct works*, which also includes working paper, books and chapters. Thus, journal pages can be seen as a proxy for journal articles. The two groups can also explain the low loading of the famous h-index, which combines quantity and quality in one measure.[20] Another explanation might be that there is only little variation in this indicator. Our results confirm the previous ranking approaches given in the literature section. Either citations or quality weighted output are taken to rank economists.

Based on our results we calculate two new aggregated rankings. First we take all indicators into account, except access statistics. Second, we focus on the four main indicators found by the PCA. For both approaches we take the standardized scores, weight them by the obtained factor loadings and finally take the average.[21] In Table 7 we show the top 50 economists worldwide based on all 27 included weighted standardized bibliometric indicators. Furthermore, we report the rank based the top 4 indicators and the recalculated RePEc rankings. Comparing the full PCA results with the Top 4 we have the same ranking for the first 13 economists. The other ranks in this table are very similar. This is confirmed by looking at the crossplots in Fig. 4. Larger individual differences emerge at the lower end of the ranking. This effect is most pronounced for authors

[19] The loadings on the second factor show a similar ranking as the for the first one.
[20] For a literature review of the h-index and its variants see Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009).
[21] We rescale the loadings such that the weights add up to one.

**Table 7**
World ranking for economists – top 50.

| | Rank PCA All indicators | Rank PCA Top 4 | Rank RePEc Harmonic mean | Rank RePEc Arithmetic mean |
|---|---|---|---|---|
| Andrei Shleifer | 1 | 1 | 1 | 6 |
| James J. Heckman | 2 | 2 | 4 | 2 |
| Joseph E. Stiglitz | 3 | 3 | 2 | 1 |
| Robert J. Barro | 4 | 4 | 3 | 4 |
| Daron Acemoglu | 5 | 5 | 8 | 3 |
| Robert E. Lucas Jr. | 6 | 6 | 5 | 54 |
| Jean Tirole | 7 | 7 | 9 | 5 |
| Peter C.B. Phillips | 8 | 8 | 6 | 17 |
| Olivier Blanchard | 9 | 9 | 14 | 7 |
| Lawrence H. Summers | 10 | 10 | 17 | 8 |
| John Y. Campbell | 11 | 11 | 18 | 13 |
| Martin S. Feldstein | 12 | 12 | 7 | 14 |
| Edward C. Prescott | 13 | 13 | 11 | 26 |
| Kenneth S Rogoff | 14 | 17 | 15 | 18 |
| David E. Card | 15 | 15 | 30 | 9 |
| Thomas J. Sargent | 16 | 14 | 25 | 10 |
| Robert Ernest Hall | 17 | 16 | 20 | 15 |
| Elhanan Helpman | 18 | 18 | 23 | 12 |
| Mark L. Gertler | 19 | 20 | 13 | 59 |
| Maurice Obstfeld | 20 | 19 | 27 | 11 |
| N. Gregory Mankiw | 21 | 23 | 21 | 28 |
| Paul R. Krugman | 22 | 24 | 19 | 22 |
| Alan B. Krueger | 23 | 21 | 43 | 23 |
| Michael Woodford | 24 | 25 | 26 | 19 |
| Lars E. O. Svensson | 25 | 26 | 22 | 16 |
| Robert W. Vishny | 26 | 28 | 24 | 167 |
| Ben S. Bernanke | 27 | 27 | 29 | 20 |
| Donald W. K. Andrews | 28 | 22 | 39 | 44 |
| Alberto Alesina | 29 | 30 | 31 | 29 |
| Robert G. King | 30 | 29 | 47 | 53 |
| Gary S. Becker | 31 | 32 | 32 | 49 |
| Richard Blundell | 32 | 34 | 35 | 21 |
| Ross Levine | 33 | 33 | 36 | 32 |
| James H. Stock | 34 | 35 | 28 | 72 |
| Lawrence F. Katz | 35 | 31 | 51 | 57 |
| James Poterba | 36 | 37 | 37 | 24 |
| Lawrence J. Christiano | 37 | 40 | 40 | 30 |
| Martin Eichenbaum | 38 | 39 | 48 | 70 |
| Raghuram G. Rajan | 39 | 42 | 53 | 47 |
| Angus S. Deaton | 40 | 44 | 54 | 27 |
| Lars Peter Hansen | 41 | 36 | 67 | 86 |
| Edward Ludwig Glaeser | 42 | 50 | 45 | 39 |
| Jean-Jacques Laffont | 43 | 43 | 38 | 51 |
| Peter A. Diamond | 44 | 45 | 61 | 35 |
| George A. Akerlof | 45 | 41 | 66 | 93 |
| Jordi Gali | 46 | 49 | 41 | 99 |
| Robert J. Gordon | 47 | 38 | 49 | 82 |
| Christopher Sims | 48 | 46 | 55 | 55 |
| Robert F. Engle | 49 | 55 | 56 | 33 |
| John B. Taylor | 50 | 51 | 46 | 52 |

This table ranks economists based on the RePEc database from July 2011 and the PCA applied to 27 bibliometric indicators. 'Top 4' represents the ranks based on the four indicators with the highest loading in Fig. 3. The last two columns reports the corresponding recalculated RePEc ranks either based on the harmonic or arithmetic mean.

who do not have any citation record yet. Similar results we obtain for the PCA vs. RePEc comparison. But larger changes in relative positions are now observed in the range around the median.

## 6. Conclusion

In economics, Research Papers in Economics (RePEc; http://www.repec.org) has become an essential source both for the spread of knowledge and ranking of individual authors and academic institutions. With 33 bibliometric measures it is currently one of the most comprehensive databases in the field of economics. In this paper, we provide some cautionary remarks concerning the interpretation of rankings provided by the RePEc network. Distortions of rankings can be due to missing citations, calculation of impact factors, or 'unreal' access statistics. Furthermore, we provide evidence how inconsistencies between worldwide and regional rankings may arise.

Given this large database we ask how to select the most important indicators describing the scientific achievements of an economist. This selection can be used for a new worldwide ranking. We propose to use the principal component analysis to derive weights for each ranking. In our example of more than 29,000 economists from all over the world we find that the first component explains almost 90% of the variance common to all included 27 indicators. Furthermore, we identify two groups of indicators that are most important: number of journal pages as a proxy for journal articles and the number of citations. Both are weighted with simple impact factors and number of authors. This confirms the recent ranking approaches in the literature for economists, which use variations of them, in particular, the h-index is a combination of these but a relatively rough measure.

## References


Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289.

Baltagi, B. (1999). Applied econometrics rankings: 1989–1995. *Journal of Applied Econometrics*, *14*(4), 423–441.

Baltagi, B. (2003). Worldwide institutional and individual rankings in econometrics over the period 1989–1999: An update. *Econometric Theory*, *19*(01), 165–224.

Ben-David, D. (2010). Ranking Israel's economists. *Scientometrics*, *82*(2), 351–364.

Combes, P., & Linnemer, L. (2003). Where are the economists who publish? Publication concentration and rankings in Europe based on cumulative publications. *Journal of the European Economic Association*, *1*(6), 1250–1308.

Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, *1*(3), 193–203.

Coupé, T. (2003). Revealed performances: Worldwide rankings of economists and economics departments, 1990–2000. *Journal of the European Economic Association*, *1*(6), 1309–1345.

Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of university systems. *Scientometrics*, *86*(1), 77–92.

Dolado, J., Garcí a-Romero, A., & Zamarro, G. (2003). Publishing performance in economics: Spanish rankings (1990–1999). *Spanish Economic Review*, *5*(2), 85–100.

Dusansky, R., & Vernon, C. (1998). Rankings of US economics departments. *The Journal of Economic Perspectives*, *12*(1), 157–170.

Ellison, G. (2002). The slowdown of the economics publishing process. *Journal of Political Economy*, *110*(5), 947–993.

Franceschet, M. (2009). A cluster analysis of scholar and journal bibliometric indicators. *Journal of the American Society for Information Science and Technology*, *60*(10), 1950–1964.

Glänzel, W., & Moed, H. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193.

Gonzalez-Navarro, M., Quintana-Domeque, C. (2010). Urban infrastructure and economic development: Experimental evidence from street pavement, IZA Discussion Papers 5346, Institute for the Study of Labor (IZA).

Graves, P., Marchand, J., & Thompson, R. (1982). Economics departmental rankings: Research incentives, constraints, and efficiency. *The American Economic Review*, *72*(5), 1131–1141.

Harzing, A., & van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, *60*(1), 41–46.

Henrekson, M., & Waldenström, D. (2011). How should research performance be measured? A study of Swedisch economists. *The Manchester School*, *79*(6), 1139–1156.

Hirsch, B., Austin, R., Brooks, J., & Moore, J. (1984). Economics departmental rankings: Comment. *The American Economic Review*, *74*(4), 822–826.

Hofmeister, R., & Ursprung, H. (2008). Das Handelsblatt Ökonomen-Ranking 2007: Eine kritische Beurteilung. *Perspektiven der Wirtschaftspolitik*, *9*(3), 254–266.

Kalaitzidakis, P., Mamuneas, T., & Stengos, T. (2003). Rankings of academic journals and institutions in economics. *Journal of the European Economic Association*, *1*(6), 1346–1366.

Ketzler, R., & Zimmermann, K. (2009). Publications: German economic research institutes on track. *Scientometrics*, *80*(1), 231–252.

Liebowitz, S., & Palmer, J. (1984). Assessing the relative impacts of economics journals. *Journal of Economic Literature*, *22*(1), 77–88.

McAllister, P., Narin, F., & Corrigan, J. (1983). Programmatic evaluation and comparison based on standardized citation scores. *IEEE Transactions on Engineering Management*, *30*, 205–211.

Medoff, M. (1989). The ranking of economists. *Journal of Economic Education*, *20*(4), 405–415.

Mingers, J., & Lipitakis, E. (2010). Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics*, *85*, 613–625.

Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81–100.

Neuhaus, C., & Daniel, H. (2008). Data sources for performing citation analysis: An overview. *Journal of Documentation*, *64*(2), 193–210.

Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the web of science for coverage of the social sciences' literature. *Journal of Informetrics*, *1*(2), 161–169.

Ortega, J., Lopez-Romero, E., & Fernandez, I. (2011). Multivariate approach to classify research institutes according to their outputs: The case of the CSIC's institutes. *Journal of Informetrics*, *5*(3), 323–332.

Portes, R. (1987). Economics in Europe. *European Economic Review*, *31*(6), 1329–1340.

Ramesh Babu, A., & Singh, Y. (1998). Determinants of research productivity. *Scientometrics*, *43*(3), 309–329.

Ruane, F., & Tol, R. (2008). Rational (successive) h-indices: An application to economics in the Republic of Ireland. *Scientometrics*, *75*(2), 395–405.

Ruane, F., & Tol, R. (2009). A Hirsch measure for the quality of research supervision, and an illustration with trade economists. *Scientometrics*, *80*(3), 613–624.

Simpsons, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, *13*(2), 238–241.

Sinha, D., & Macri, J. (2004). Rankings of economists in teaching economics departments in Australia 1988–2000. *Economics Bulletin*, *1*(4), 1–19.

Tol, R. (2009). The h-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics*, *80*(2), 317–324.

Ursprung, H., & Zimmer, M. (2006). Who is the "Platz-Hirsch" of the German economics profession? A citation analysis. *Jahrbücher für Nationalökonomie und Statistik*, *227*, 187–202.

Vinkler, P. (2006). Composite scientometric indicators for evaluating publications of research institutes. *Scientometrics*, *68*(3), 629–642.

Zimmermann, C. (2007). Academic rankings with RePEc, Working papers 2007–36, University of Connecticut, Department of Economics.