CESifo Working Papers

www.cesifo.org/wp

Who is the 'Journal Grand Master'?

A new ranking based on the Elo rating system

Robert Lehmann
Klaus Wohlrabe

CESIFO WORKING PAPER NO. 6259
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
DECEMBER 2016

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 6259*

Who is the 'Journal Grand Master'?
A new ranking based on the Elo rating system

Abstract

In this paper we transfer the Elo rating system, which is widely accepted in chess, sports and other disciplines, to rank scientific journals. The advantage of the Elo system is the explicit consideration of the factor time or the history of a journal's performance. Most other rankings that are commonly applied neglect this fact. The Elo ranking approach can easily be applied to any metric, published on a regular basis, to rank journals. We illustrate the approach using the SNIP indicator based on citation data from Scopus. Our balanced panel consists of 7,748 journals from many scientific fields for the period from 1999 to 2015. We show that the Elo approach produces a similar but not identical ranking compared to other rankings based on the SNIP. Especially the rank order for rather 'middle-class' journals can tremendously change.

JEL-Codes: A120, A140.

Keywords: Elo rating system, journal rankings, SNIP.

Robert Lehmann
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
lehmann@ifo.de

*Klaus Wohlrabe**
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
wohlrabe@ifo.de

*corresponding author

December 9, 2016

1 Introduction

Measuring the 'quality' of scientific publishing has always been an important aspect for researchers, institutions, politics and the public. Next to financial incentives for the publisher, publications in high-quality journals are necessary prerequisites for future job market signals of the scientists. What defines a journal as 'high-quality' mainly depends on the classification or ranking scheme that is applied. The question on how such a classification scheme should look like has entailed a heated debate in general, which is especially pronounced in several scientific disciplines such as economics. In this article we do not want to comment the 'right' or the 'wrong' of existing rankings, but rather adopt a system that was originally developed for chess: the Elo rating system.

One of the main criticisms which can be raised when it comes to rank journals is the time invariance of the classification scheme. Generally, many journal metrics are reported with respect to a given year. The prestige of a journal can be negatively affected in a given year if the corresponding metric significantly drops although in the years before the performance was very good. This shortcoming becomes irrelevant with the Elo rating system, since it explicitly incorporates the trajectory of the journal's performance.[1]

The rationale of the Elo rating system is the following. Each journal has an Elo number which is based on its impact. Every year, the journals compete with each other and earn Elo points which are based on the expected values for a win or a loss. After this competition, the Elo number is adjusted according to the result. In the upcoming years, the journals compete with each other based on the last available Elo number. Thus, the complete time path of the journal's performance is relevant for the latest competition and therefore the latest ranking. The aim of this article is by no means an examination of the ranking's properties, but rather to present a new approach that is subsequently compared to rather standard rankings based on, for example, the SNIP (source normalized impact per paper) indicator or the tournament method. In the end we ask, whether the inclusion of the trajectory changes the ranking of journals.

Based on a balanced set of 7,748 journals from all possible scientific fields for the period from 1999 to 2015, we can state that the time line of a journal's performance is very important for the most recent ranking. Our Elo approach produces a similar but by no means identical ranking compared to the Tournament Method, the average SNIP between 1999 to 2015 or the latest SNIP in 2015. With our approach, the top journals remain top-ranked. However, there are substantial differences observable for rather 'middle-class' journals and not only for the top 30. We also show that a 'bad' year in terms of the SNIP does not necessarily lead to a large drop in the ranking position. The Elo rating system seems to be a promising alternative to rank scientific journals compared to existing ones. A further advantage is the

[1] The factor time is indirectly incorporated for many metrics by considering different citation windows or evaluation period of articles, e.g. for two- or five year impact factor. However, these metrics are published on a yearly basis.

application of the Elo ranking system to any journal metric, like the Journal Impact Factor or citation counts, that is published on a regular basis.

The paper is organized as follows. In Section 2 we elaborate on the data and the Elo ranking system. Section 3 presents and discusses the results. The last section offers some conclusions.

2 Data and Methodology

2.1 Data

One aim of this paper is to present the new ranking approach for a wide range of journals from different scientific fields. Therefore, we need high quality and notably comparable data. Such high-quality data are available from Scopus at `http://www.journalmetrics.com`. The data, as of June 2016, are available for the period ranging from 1999 to 2015 and comprises 35,414 journals in total.

A main challenge is the comparability of journals across different disciplines. To this end, we use the SNIP (source normalized impact per publication) indicator (Moed, 2010; Waltman *et al.*, 2013). The strength of the SNIP lies in its normalization of citations in order to make scientific fields comparable. It especially pays attention to different citation practices within and between subjects. According to Moed (2010), the SNIP is basically the ratio of the so called raw impact per paper (RIP) and the Relative Database Citation Potential (RDCP) in the journal's sub-field. Whereas the RIP is defined as the number of citations in year t for papers published in the journal in the time span $t-3$ to $t-1$, the RDCP explicitly uses the distribution of citations. For each journal in the list, one can calculate its database citation potential (DCP). Repeating this step for each journal, results in a distribution of DCPs for the whole data set. In order to gain the RDCP, each journal's DCP is divided by the median DCP of the whole distribution.

For the majority of the journals in the original data set, many entries for the SNIP are missing or zero. Since we want to have a fair comparison with our new ranking, we decided to balance the panel. We end up with 7,748 journals for which a SNIP greater than zero is available for all the years from 1999 to 2015.[2]

2.2 The Elo rating system

Fundamentals Originally developed to rate chess players, the Elo rating system is nowadays adopted by many other sports such as table tennis (see, for example, Glickman, 1995) or used to, for example, rank evolutionary algorithms (Veček *et al.*, 2014). The eponym for

[2] We are aware of the fact that the SNIP can essentially be zero. Nevertheless we dropped these cases as a draw in the Elo system between two journals would result in a positive score. We think that does not reflect the spirit of the Elo rating system.

this rating system is Arpad Emrick Elo, who was an American physician and statistician. His main objective was to develop a rating system for the United States Chess Federation (USCF) that has a statistical foundation. Later on, the rating system was also adopted by the Fédération Internationale des Échecs (FIDE), the world chess federation.

The two main steps of the ranking comprise (i) calculating the expected score and (ii) updating the 'players' rating (see here and henceforth Glickman and Jones, 1999). Additionally, we refer to Elo (1978) for a very detailed description. Since the inherent strength of a player is unknown to outsiders, one has to approximate it by a rating. Thus, the match outcome between two players A and B can be approximated with the following formula:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}} \,. \tag{1}$$

E_A is the expected score for player A to win the game, based on the unknown strengths for both players (R_A and R_B). To illustrate the expected score for player A, we use the example by Glickman and Jones (1999). Imagine a game between two players with strengths $R_A = 1,500$ and $R_B = 1,700$, respectively. The expected long-run score of player A is $E_A = 0.24$. Thus, based on these hypothetical strengths, player A is expected to win the game or gain a draw in 24 of 100 cases. The opposite is true for player B, since his expected score is $E_B = 0.76$. As mentioned, these figures are long-run scores. However, a game score can only take three possible values: 1 for a victory of player A, 1/2 if the game ended in a draw or 0 if player A loses the match. Since the strengths of both players are unknown, they are replaced by their estimates, the so called Elo number or Elo rating (for player A it is R_A).

The second step comprises the update of a player's strength. This is done by the following equation, again from player A's perspective:

$$R_{A,t+1} = R_{A,t} + k(S_A - E_A) \,. \tag{2}$$

The new Elo rating of player A ($R_{A,t+1}$) is based on his or her old rating ($R_{A,t}$) plus the difference from the game score S_A and the expected long-run score E_A, which is weighted by the factor k to allow how fast a rating can evolve. In chess, this factor is either based on the number of games played, the age of the player or the strength. Suppose that the Elo ratings of two players are $R_{A,t} = 1,500$ and $R_{B,t} = 1,700$ before they play a match. We set the adjustment parameter $k = 32$, which is mainly used in chess for weaker players. Three possible match outcomes can emerge and thus resulting ratings:

- A **wins:** $R_{A,t} = 1,500$, $S_{A,t} = 1$, $E_{A,t} = 0.24$, $R_{A,t+1} = 1,524$, $R_{B,t+1} = 1,676$,
- **Draw:** $R_{A,t} = 1,500$, $S_{A,t} = 0.5$, $E_{A,t} = 0.24$, $R_{A,t+1} = 1,508$, $R_{B,t+1} = 1,692$,
- A **loses:** $R_{A,t} = 1,500$, $S_{A,t} = 0$, $E_{A,t} = 0.24$, $R_{A,t+1} = 1,492$, $R_{B,t+1} = 1,708$.

As one can see, player A's rating either increases by winning the game or by gaining a draw since the expected long-run score of player A lies below the score for a draw ($0.24 < 0.50$). In the next match, the expected score is calculated based on the new Elo ratings. For the mathematics of such pairwise comparisons, for which the Elo rating system is a special case, we refer to Joe (1991).

Application to rank journals After the discussion of the fundamentals, it is the aim in the following to present how we apply the Elo rating system to rank journals. Therefore, we need to introduce parameter values: $R_{i,0}$, S_i and k. Each journal i is treated as a single 'player' at any point t in time. As for each sports or any other competition, the score S_i can take three values: 1 if journal i has a higher SNIP in year t compared to journal j, 0.5 if they equal each other and 0 in the case of $\text{SNIP}_{i,t} < \text{SNIP}_{j,t}$. We set the adjustment parameter to $k = 1$ in order to apply the same 'catch-up speed' for each journal from each scientific category. The main reason for this parameter value is the usage of the SNIP. Since this indicator is comparable between scientific categories as well as sub-categories of a single profession, we do not need to control for different citation patterns or anything similar that makes categories not comparable. The last parameter value we have to choose is the initial Elo number of each journal ($R_{i,0}$). It becomes immediately obvious that this number cannot be estimated from the data, thus, we decided to attribute each journal the same initial number: $R_{i,0} = 10,000$. Our resulting ranking is, however, independent from this initial value as we treat the time before 1999 as non-existing and let the journals be established in this year. Afterwards, the Elo numbers develop from this constant starting value. Choosing a different initial value does not influence the ranking that results at the end of our data set. However, it should be a sufficiently large number to avoid negative Elo ratings.

Applying our notation to Equation (1) and (2), the expected long-run score of journal i to beat journal j and the corresponding update of journal i's Elo number transform into:

$$E_{i,t} = \frac{1}{1 + 10^{(R_{j,t} - R_{i,t})/400}} \, , \tag{3}$$

$$R_{i,t} = R_{i,t-1} + (S_{i,t} - E_{i,t}) \, . \tag{4}$$

Since our data set comprises 7,748 journals, we have to calculate 7,747 pairwise comparisons for each journal and each year. So the natural question to raise is: How does the Elo number develop between these pairwise comparisons? The answer can also be found in the chess system. The Elo rating is adjusted only once a year, after a journal has 'played' against all the other journals. Thus, the final Elo rating of a journal at the end of year t is: $R_{i,t} = R_{i,t-1} + \sum_{j \neq i} (S_{i,t} - E_{i,t})$. Each pairwise result is summed up and added to the previous Elo number at the end of all comparisons. Based on the Elo ratings in 2015 ($R_{i,2015}$), we calculate the overall ranking of all journals. At this point, our main contribution of the paper sets in: the Elo rating in 2015 incorporates the complete trajectory or history of the journal's performance and thus produces a more realistic ranking.

2.3 An alternative: the Tournament Method

An alternative approach using pairwise comparisons is the so called tournament method, which was introduced by Kóczy and Strobel (2010) with an application to economics journals. In the 'tournament' the journals compete in 'citation games' against each other. Thus, the ranking is based on cross-comparisons of citations between the journals.

In terms of our notation, the score $\sigma_{i,t}$ of journal i for a given year t is simply the share of games it wins against competitors or matches that end in a draw:

$$\sigma_{i,t} = \frac{|\{j \in J, SNIP_{i,t} > SNIP_{j,t}\}| + \frac{1}{2}|\{j \in J, SNIP_{i,t} = SNIP_{j,t} > 0\}|}{|\{j \in J, SNIP_{i,t} + SNIP_{j,t} > 0\}|} \, . \tag{5}$$

A victory of journal i is defined as $SNIP_{i,t} > SNIP_{j,t}$. The main difference to the Elo rating system is that it the relative position of a journal does not matter. A win of a 'bad' journal against a 'good' gives the same score as a win in a 'less good' journal.

Kóczy and Strobel (2010) propose to account for the ranking's time line by applying a geometric decay function to calculate the total score of journal i:

$$S_{i,T} = \frac{1-\delta}{1-\delta^T} \sum_{t=1}^{T} \delta^{T-t} \sigma_{i,t} \, . \tag{6}$$

To be in line with Kóczy and Strobel (2010), we choose $\delta = 0.5$ in our application.

3 Results

In the following, we present our results in three steps. First, we show and discuss the top 30 ranked journals, followed by a summary of 'Winners' and 'Losers' in a second step. Finally, this section closes with some statements on how the different rankings (Elo, Tournament and SNIP) are connected.

Let us start with the presentation of the top journals. Table 1 shows the top 30 ranked journals based on the Elo rating system.[3] For reasons of comparison, we also include the ranks resulting from the Tournament Method, a ranking based on the average SNIP for the years 1999 to 2015 and the ranking of the latest available SNIP for 2015. The top 3 journals are: *CA - A Cancer Journal for Clinicians*, *Reviews of Modern Physics* and *Annual Review of Fluid Mechanics.* Whereas the former two journals are also among the top 3 by applying a different methodology, the *Annual Review of Fluid Mechanics* is just ranked on ninth place. The top 10 are dominated by journals from natural sciences. The first journal from

[3]The full ranking is available from the authors upon request.

Table 1: Top 30 ranked journals

Journal	Elo	Tournament Method	Average SNIP (1999-2015)	SNIP 2015
CA - A Cancer Journal for Clinicians	1	1	1	1
Reviews of Modern Physics	2	2	3	2
Annual Review of Fluid Mechanics	3	9	9	9
New England Journal of Medicine	4	4	2	4
Progress in Energy and Combustion Science	5	7	12	11
Physics Reports	6	8	27	10
Chemical Reviews	7	11	4	7
Progress in Materials Science	8	3	86	5
Progress in Polymer Science	9	12	28	16
Lancet	10	5	22	3
Advances in Physics	11	10	17	8
ACM Computing Surveys	12	13	24	19
Journal of Economic Literature	13	17	7	21
Clinical Microbiology Reviews	14	24	14	24
Physiological Reviews	15	15	6	13
JAMA - Journal of the American Medical Association	16	16	8	14
Annual Review of Psychology	17	6	11	6
IEEE Transactions on Pattern Analysis and Machine Intelligence	18	19	19	17
Annual Review of Immunology	19	18	5	15
Nature	20	21	13	18
Psychological Bulletin	21	28	26	29
Annual Review of Plant Biology	22	40	20	38
Science	23	22	15	25
Annual Review of Materials Research	24	31	60	41
Annual Review of Astronomy and Astrophysics	25	14	25	22
Surface Science Reports	26	23	56	114
International Materials Reviews	27	42	54	34
Quarterly Journal of Economics	28	30	22	27
Annual Review of Biochemistry	29	45	10	33
Materials Science and Engineering: R: Reports	30	27	29	12

Note: The journals are ordered according to the Elo ranking. *Source:* Data are taken from Scopus and are available at `http://www.journalmetrics.com`.

social sciences, the *Journal of Economic Literature*, is ranked 13. From Table 1 we can also state that many different (sub)disciplines are part of the top 30. For instance, astronomy, economics, health sciences and physics are on the list.

For most of the top journals, the position across different rankings show rather similar results. However, the different methodologies produce results that are by no means identical. For example, the journals *Progress in Material Sciences* and *Surface Science Report* show a large variation in their rankings. The most impressive example is *Progress in Material Sciences*. Its rank is either 8 based on our Elo system or 3 to 5 by applying the Tournament

Method or the SNIP of 2015. However, by using the average SNIP for the years from 1999 to 2015, its rank drops down to 86. Our main criticism deals with the missing consideration of the time line or the history of a journal in most of the common rankings. The evidence from Table 1 strengthens our hypothesis that timely variation is very important for the ranking outcome. Even for the top 30 journals in our data set, we observe a certain degree of ranking heterogeneity.

The argumentation on the time line of journals is underpinned by taking a closer look on the 'Winners' and 'Losers' for the period from 1999 to 2015. Table 2 shows the difference of the rank a journal gets attributed for 2015 and its rank for the year 1999. Both ranks are based on the Elo system. The upper part of the table shows the ten journals with the largest increase in the rank, thus, these journals are 'Winners'. The 'Losers' are displayed in the lower part of Table 2. The largest increase by 7,296 places gains the journal *Transportation Research Part E: Logistics and Transportation Review*. *Food Technology* is the journal with the largest decrease in its rank, namely 7,055 places.

Table 2: Top 10 Winner and Loser

Journal	Ranking Change 1999 to 2015
Transportation Research Part E: Logistics and Transportation Review	7,296
Annual Reviews in Control	7,205
Critical Care	6,872
Materials and Design	6,742
Canadian Journal of Electrical and Computer Engineering	6,705
Alternative Medicine Review	6,667
Journal of Iron and Steel Research International	6,516
Progress in Planning	6,477
Engineering Failure Analysis	6,417
Internet and Higher Education	6,382
Food Technology	-7,055
Tribology and Lubrication Technology	-6,927
Zeitschrift für Wirtschaftsgeographie	-6,774
Advances in Inorganic Chemistry	-6,764
Advances in Chemical Physics	-6,680
Historia Mathematica	-6,625
Yearbook of Physical Anthropology	-6,394
AgBioForum	-6,261
Bioremediation Journal	-6,208
Journal of Family Practice	-6,155

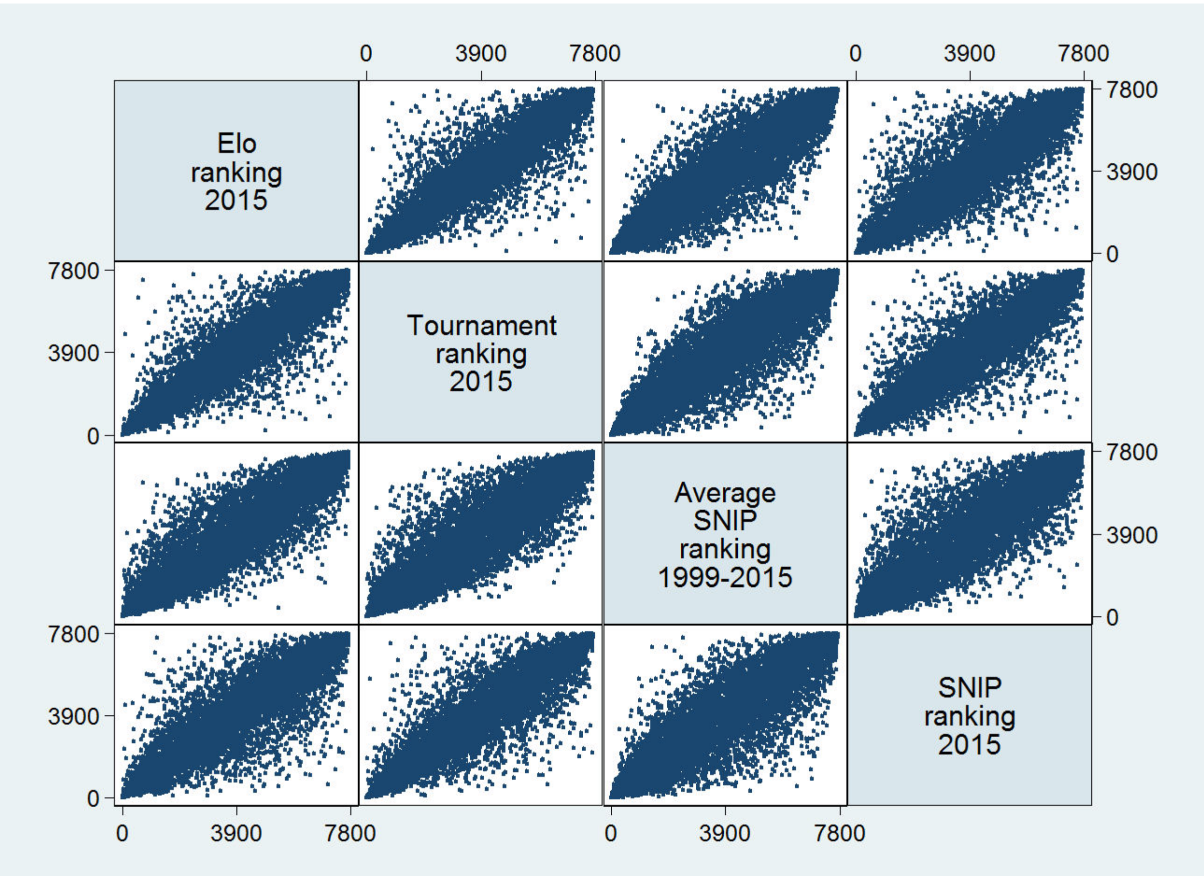
Source: Data are taken from Scopus and are available at `http://www.journalmetrics.com`.

The last step we want to undertake is a formal statement on the relation between the different rankings. Therefore, we first calculate Spearman Rank Correlations for ranking-pairs. The outcome is shown in Table 3. We find the highest rank correlation of 0.915 between our Elo ranking and the Tournament Method. By comparing our ranking to the latest available SNIP, the correlation drops to 0.887. Thus, the rankings are by no means identical. This findings also supports our main criticism that the complete history of journal has to taken into account for a ranking. As one can see: the lowest rank correlation can be found for the pair 'Average SNIP - SNIP 2015'. Therefore, the performance of a journal over time has to be taken into account. Figure 1 shows the relationships between the rankings in a graphical way. As suggested by the correlations, the rankings show a distinct linear relationship. However, we also observe a large mass of journals in the middle for which the methodologies deliver different ranking signals.

Table 3: Spearman rank correlation between the different rankings

	Elo	Tournament Method	Average SNIP (1999-2015)	SNIP 2015
Elo	1.000			
Tournament Method	0.915	1.000		
Average SNIP (1999-2015)	0.896	0.888	1.000	
SNIP 2015	0.887	0.908	0.870	1.000

Figure 1: Cross-plot between the different rankings



4 Conclusion

Most of the commonly applied rankings for scientific journals mainly neglect the time line of a journal's performance. This paper explicitly accounts for this shortcoming by transferring a concept that is widely accepted in chess, sports and other disciplines to the field of publishing: the Elo rating system. The data set on which we base our analysis comprises 7,748 journals from all possible scientific categories for the period from 1999 to 2015. In order to make the journals comparable, we use the source normalized impact per publication (SNIP) index. It turns out that the time line is very important for the ranking since the Elo rating system produces similar but by no means identical rankings compared to outcomes based on either the Tournament Method, the average SNIP for 1999 to 2015 or the latest SNIP from 2015. Since the Elo ranking is very easy to compute and widely accepted in other fields, it seems a promising alternative to already existing ranking approaches.

References

Elo, A. E. (1978). *The Rating of Chessplayers, Past & Present*. Arco Publishing, New York NY.

Glickman, M. E. (1995). A Comprehensive Guide to Chess Ratings. *American Chess Journal*, **3**, 59–102.

— and Jones, A. C. (1999). Rating the chess rating system. *Chance*, **12** (2), 21–28.

Joe, H. (1991). Rating systems based on paired comparison models. *Statistics & Probability Letters*, **11** (4), 343–347.

Kóczy, L. A. and Strobel, M. (2010). *The World Cup of Economics Journals: A Ranking by a Tournament Method*. IEHAS Discussion Papers No. MT-DP – 2010/18.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, **4** (3), 265–277.

Veček, N., Črepinšek, M., Mernik, M. and Hrnčič, D. (2014). A Comparison between Different Chess Rating Systems for Ranking Evolutionary Algorithms. In *Proceedings of the 2014 Federated Conference on Computer Science and Information System (FedCSIS)*, pp. 511–518.

Waltman, L., van Eck, N. J., van Leeuwen, T. N. and Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, **7** (2), 272–285.