Journal of Informetrics 10 (2016) 1143–1152

Contents lists available at ScienceDirect

Journal of Informetrics

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

Towards field-adjusted production: Estimating research productivity from a zero-truncated distribution

Timo Koski[a], Erik Sandström[b], Ulf Sandström[c,*]

[a] KTH Royal Institute of Technology, Dept Mathematics, SE-100 44 Stockholm, Sweden
[b] Doktorsvägen 43, SE-123 52, Stockholm, Sweden
[c] KTH Royal Institute of Technology, Dept Indek, SE-100 44 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 7 April 2016
Received in revised form
15 September 2016
Accepted 15 September 2016
Available online 20 October 2016

Keywords:
Research productivity
Waring distribution
Field adjusted production
Size-dependent indicators

ABSTRACT

Measures of research productivity (e.g. peer reviewed papers per researcher) is a fundamental part of bibliometric studies, but is often restricted by the properties of the data available. This paper addresses that fundamental issue and presents a detailed method for estimation of productivity (peer reviewed papers per researcher) based on data available in bibliographic databases (e.g. Web of Science and Scopus). The method can, for example, be used to estimate average productivity in different fields, and such field reference values can be used to produce field adjusted production values. Being able to produce such field adjusted production values could dramatically increase the relevance of bibliometric rankings and other bibliometric performance indicators. The results indicate that the estimations are reasonably stable given a sufficiently large data set.

1. Introduction

Whilst methods for citation analysis have developed significantly during the latest twenty years, the same cannot be said regarding methods for publication productivity analysis. “Research productivity”, “Scientific productivity” and “Publication productivity” are frequently used keywords in about one and a half thousand Web of Science-articles over the years, but a closer look reveals little of methodological development with regard to the measurement of “productivity” and few attempts to explicitly contribute to such a development (for an exception, see papers by Abramo & D’Angelo, 2014, 2016).

This paper will address a fundamental issue in the empirical study of scientific productivity, i.e. the calculation of the average number of peer reviewed papers[1] published by researchers in a given time period. This task, which at first sight seems quite simple, is often restricted by the properties of the data available. Publication databases, such as Web of Science and Scopus, only contain information on actual (publishing) authors within a given time period, not the full population of publishing and non-publishing “authors”. Hence, a paper frequency distribution based on such data will be zero-truncated; the zero-class (number of non-publishing, or potential, “authors”) will be missing.

* Corresponding author.
E-mail addresses: tjtkoski@kth.se (T. Koski), erik.sandstrom@zoho.com (E. Sandström), ulf.sandstrom@indek.kth.se (U. Sandström).

[1] In this paper, the Web of Science is used as to provide details of “peer reviewed papers”. It should be noted that all publications databases have coverage issues. Therefore, estimated in this paper are papers recorded in Web of Science (Arts & Humanities Citation Index, Social Science Citation Index, and Science Citation Index) as document types “Article”, “Letter”, or “Review” in the selected time period. For a discussion regarding coverage issues in Web of Science, see Mongeon and Paul-Hus (2016).

http://dx.doi.org/10.1016/j.joi.2016.09.002
1751-1577/

For example, if the productivity of two countries, e.g. China and Canada, are to be investigated, one might categorize the publishing authors by name (unique authors) and divide the total number of papers by the number of actual authors. However, this calculation will not produce a trustworthy measure of productivity since the proportion of non-publishing (potential) authors might differ between the two countries. A non-biased measure requires knowledge of the full population, including the number of non-publishing authors (i.e. the zero-class of the paper frequency distribution).

2. Actual authors and potential authors

"Potential authors" is a concept used by the Budapest group to ex ante denote the total population of researchers that could publish papers, or to ex post denote researchers that could have, but have not, been publishing within a given time period (Braun, Glänzel, & Schubert, 2001; Telcs & Schubert, 1986). The (ex post) potential authors include active researchers that have been publishing before or after the given time period but for different reasons have not been publishing during the given time period.

Productivity comparisons between different categories, e.g. field, gender or country requires knowledge about the total population of authors, potential as well as actual authors. In a field with low paper productivity, e.g. the social sciences, the proportion of zero-class (the potential authors) will be high in relation to the total population (actual and potential authors). In contrast, in fields with high productivity, such as the medical and natural sciences, the proportion of potential authors will be low. Comparisons based solely on actual authors will thus be misrepresentative.

Actual authors can be extracted from publication databases such as Web of Science and Scopus. Potential authors will, however, not be included in these. An estimate of the number of potential authors would either require detailed data concerning the entire researcher population or the use of statistical estimates, such as the one presented in this paper. A successful method for estimating the number of potential authors in a productivity distribution (i.e. the zero-class of the paper frequency distribution) would enable the creation of more advanced productivity measures.

The objective of this paper is to contribute to the discussion on productivity from a scientometric perspective. If it is possible to estimate the zero-class of a truncated paper frequency distribution, then it would, in principle, be possible to create advanced (e.g. field normalized) productivity indicators (Sandström & Sandström, 2008).

Hitherto, the most interesting discussion on publication productivity has been given within the framework of frequency distributions (Braun, Glänzel, & Schubert, 1990), which is a core element in bibliometric theory. The Waring distribution is a statistical distribution used for describing publication productivity processes. The distribution was originally introduced by Simon (1955) as a generalization of the Yule distribution and further analysed by Irwin in (1963), who gave the distribution its current name.

In this paper, we aim to further explore the Waring distribution as a method to estimate the zero-class of a truncated paper frequency distribution. We note that there are certainly other possible methods and that only the Waring method has been reviewed as part of this paper. We welcome further research focused on comparison with the results of alternative methods.

3. The waring distribution

The Waring distribution can be derived and justified by many means. We shall here present an argument related to a simple probabilistic picture of publishing. Let us suppose that an author during a certain period keeps on submitting new papers until rejected for the first time. Let us suppose that the probability of rejection of any paper is equal to p, $0 < p < 1$. Let us also suppose that the rejection or acceptance of a paper does not depend on the rejection or acceptance of any other paper.

Then it follows that the probability of publishing exactly k papers is the geometric distribution

$$Pr(k\,published\,papers|p) = (1-p)kp, (k) = 0, 1, \ldots \tag{1}$$

Let us next suppose that p itself is a random variable. This could be an expression of uncertainty about the actual value of p in the Bayesian sense, where the uncertainty is expressed as a prior probability density $f(p)$. Or, we could think that the author has been allotted p that has been drawn from probability density $f(p)$. Now we can compute a new probability of k published papers by averaging, i.e.,

$$Pr\,(k\;\;published\;\;papers) = \int_0^1 Pr\,(k\;\;published\;papers|\;\;p) f\,(p)\,dp = \int_0^1 (1-p)^k pf\,(p)\,dp. \tag{2}$$

Let us take $f(p)$ as the Beta density with parameters $\rho > 0$ and $\alpha > 0$, i.e,

$$fp\frac{\Gamma\alpha+\rho}{\Gamma\rho\Gamma\alpha}p^{\rho-1}1-p^{\alpha-1} \tag{3}$$

where $\Gamma(z)$ is the Euler Gamma function. The (hyper)parameters ρ and α can be thought of as kind of virtual frequencies of refusal and acceptance, respectively. Then a computation using the properties of the Beta integral entails that

$$Pr\ (k\ \ published\ \ \ papers) = \rho \frac{\Gamma(\rho+\alpha)\Gamma(k+\alpha)}{\Gamma(\rho+\alpha+k+1)\Gamma(\alpha)}. \tag{4}$$

Let $\alpha_{(r)}$ be the ascending factorial

$$\alpha_{(r)} = \alpha \cdot (\alpha+1) \cdot \ldots \cdot (\alpha+r-1) \tag{5}$$

It holds that

$$\alpha_{(r)} = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)},$$

by the well-known recursion formula of the Gamma function, i.e., $\Gamma(z+1) = z\Gamma(z)$. Then we get

$$Pr\ (k\ \ Published\ \ \ papers) = \rho \cdot \frac{\alpha_{(k)}}{(\rho+\alpha)_{(k+1)}}, \quad k = 0, 1, 2, \ldots, \tag{6}$$

which is the familiar standard expression of the probability mass function of the Waring distribution (Irwin, 1963). The derivation above explains why the Waring distribution is also known as the Beta-geometric distribution (Bernardo & Smith, 1993).

The geometric distribution above is a special case ($n = 1$) of the Negative Binomial distribution whose parameters are n, and p. This would be the distribution of the number of papers published until n of them have been refused (n is fixed in advance, say in the beginning of the period). If we average the probability mass function of the Negative Binomial distribution over p with a Beta density as above, we would obtain a so-called generalized Waring distribution as $Pr(k\ \ published\ \ papers)$ which is further discussed in Irwin (1963).

Of course, $Pr(k\ \ published\ \ papers)$ as given above can be shown to be the equilibrium distribution of the immigration-birth-emigration process, introduced by Glänzel and Schubert in (1984) for modelling of scientific productivity. This model was later applied to the citation process in Glänzel and Schubert (1995). In the setting of Glänzel and Schubert (1984) the parameters are coefficients in the linear rates that are proportional to the frequency of authors with published papers. This means, in plain words, that success breeds success, or, that there is a cumulative advantage in higher levels of productivity. The parameter ρ is proportional to the total number of authors and is the rate of external source emitting new authors.

Then it can be shown that the expectation μ of the distribution is

$$\mu = \frac{\alpha}{\rho - 1} \quad if \rho > 1. \tag{7}$$

The condition $\rho > 1$ is equivalent to $\sigma > \beta$, which means that the rate of infusion of new authors is higher than the rate of transfer of authors to higher publication numbers.

Let us now suppose that X is a random variable with the Waring distribution with parameters ρ and α, i.e.,
$\Pr(X=k)=p_k$, $k=0,1,2,\ldots$with p_k as given in (4). Then it holds that the truncated expectation of X, $E[X|X > k]$, is given as

$$E[X \mid X \geq k] = \mu + (k+1) \cdot \mu_1, \quad k = 1, 0, 1, \ldots \tag{8}$$

where $\mu (= E[X|X > -1])$ is given in (7) and $\mu_1 = \frac{\rho}{\rho-1}$. This is shown in Glänzel, Telcs, & Schubert (1984), where it is also shown that the property (8) is a characterization of the Waring distribution, i.e., the Waring distribution has this property, and if (8) is true for a distribution, then it must be a Waring distribution. There is an elegant simplified proof of these facts in Dimaki & Xekalaki, 1996).

The Waring probability mass function p_k is an example of a power law, since it holds that

$$p_k = P(X = k) \approx k^{-(1+\rho)}, \quad as \quad k \to \infty. \tag{9}$$

This means that we can call ρ the tail parameter, as it controls the tail of the distribution. A probabilistic analysis that gives (9) as a special case is specified in Chen (1980).

Power-law tails have been observed in the distributions of the sizes of incomes, cities, internet files, biological taxa, and, after the sequencing of genomes, in (size) distributions of molecular parts, see, e.g., Reed & Hughes (2002a, 2002b).

If it holds exactly that

$$p_k = c \cdot k^{-\gamma}, \tag{10}$$

where c is the normalizing constant, the probability mass function in (10) is known as Lotka's Law, and was found by Lotka (1926) as a bibliometric distribution on the number of authors making contributions. The work by Egghe (2005) gives a theoretical justification of the law, which arises in many situations even outside bibliometrics. A similar law had been earlier found by the economist V. Pareto, as a frequency of wealth as a function of income category (above a certain bottom level). In plain words this means: most success seems to migrate to those people or companies who are already popular.

The well-known birth-and-death processes are a natural source of power law distributions (Reed & Hughes 2002a). This was first realized by G.U. Yule, who established a model (a pure birth process) to explain the observed size distribution of genera with respect to the number of species (Yule, 1925). Yule obtained a special case of the following probability mass function, which was later generalized in Simon (1955).

$$p_k = \delta B\left(\delta + 1, k\right), k = 1, 2, \ldots, . \tag{11}$$

Here $\delta > 0$ is real, $B\left(\delta + 1, k\right)$ is the *Beta function*, i.e.,

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{12}$$

where $\Gamma(\cdot)$ is, for z with positive real part, the Euler gamma function.

It can be checked that if X is a random variable with the Waring distribution with parameters ρ and α, then

$$\Pr(X = k | X > 0) \to \rho B(\rho + 1, k)$$

as $\alpha \to 0$.

The publication productivity data is by its very definition zero-truncated, i.e., as there is no information of those that are not publishing (in a certain period of time). But, as is clear from (9), the Waring distribution is not hampered by the truncation, as pointed out by Braun et al. (1990). In fact the linear expression in (8) gives a way of finding by linear regression against k the estimated intercept $\hat{\mu}$ and this can be used to estimate the frequency of zero via (4) as

$$p_0 = \frac{\alpha + \hat{\mu}}{\alpha\left(\hat{\mu} + 1\right) + \hat{\mu}} \tag{13}$$

if α is known or estimated from data.

The problem of estimation of zero-frequency (under a Poisson distribution) from zero-truncated data was first considered by McKendrick (1926). McKendrick was considering a case of estimating the number of individuals in an Indian village, who were susceptible to infection but did not develop the symptoms. He developed a differential equation and solved it to get the negative binomial distribution, from which he obtained the Poisson distribution as a limiting case. In fact this is related to the Waring distribution, which can be obtained by mixing a negative binomial distribution with a beta distribution with parameters α and ρ. McKendrick developed a moment estimator to find the number of individuals susceptible to infection but did not develop the symptoms. His data contained the number of individuals that did not develop the symptoms, including thus the immune ones. The work of McKendrick and the development of it by Irwin (1959) is surveyed and put into a modern framework in Meng (1997). Burrell (2004) points out that Kendall (1960) raised the question of estimating the "potentially contributory class" (of journals), when discussing Bradford's work on journal productivity.

4. Testing the accuracy of waring based estimation

4.1. Constructing data sets for precise tests

Earlier empirical studies regarding Waring based estimation of the zero-class have been subject to an important shortcoming: lack of precise data sets. Studies have been performed using truncated distributions, without knowledge of the actual number of zero-frequencies, which has made it difficult to assess the accuracy of the results.

Precise studies of the accuracy of Waring based estimation of the zero-class (non-publishing "authors") require known paper frequency distributions that include potential authors. The construction of such paper frequency data sets is set out below, using statistics on researchers associated with two different Swedish universities.

The distributions created below are based on selections of (*ex ante*) "potential authors", i.e. categories of people that we expect to be publishing research papers. A number of these selected people will not have been publishing during the selected time period (but possibly in another time period) and will thus form the zero-class of the paper frequency distributions. It should be noted that the frequency distributions will vary depending on selected categories of people. For example, if we define potential authors as professors at the universities, the number of non-publishing "authors" will likely be small. However, as we include more categories in the potential author definition, the head of the distribution, i.e. the low producing authors, will likely become larger.

4.2. Potential authors at two Swedish universities

Employee data were obtained from two Swedish universities.[2] A selection of ex ante potential authors was made based on the status position of the employments, where professors, researchers, senior lectures and doctoral students were selected. Two selections of potential authors from the respective universities were hereby obtained, containing 729 and 949 employees.

[2] Linkoping University (employee data covering the period 2005–2007) and Swedish University of Agricultural Sciences (employee data from December 2008).

Table 1
Paper frequencies for two universities and authorship variants.

	University 1				University 2			
	All AU	Random AU	First AU	Reprint AU	All AU	Random AU	First AU	Reprint AU
0	[140]	[321]	[321]	[334]	[110]	[370]	[400]	[371]
1	81	159	144	131	172	238	229	224
2	73	95	114	97	117	136	162	130
3	77	53	73	66	102	80	74	86
4	49	32	34	36	77	48	39	57
5	44	29	20	19	59	27	18	32
6	33	12	9	7	46	18	8	10
7	39	8	5	8	40	11	10	9
8	18	5	3	9	34	8	2	9
9	25	5	1	2	25	6	2	5
10	19	3	1	5	25	2	2	2
>10	131	7	4	15	142	5	3	14

Publication data pertaining to a fixed period were downloaded for each ex ante potential author from the Web of Science and compiled into a data set, summarizing the number of papers (article, letter and review) associated with each potential author.[3] In addition, the number of "first author papers" and "reprint papers" (see below) by each author was extracted and listed in Table 1. Furthermore, a "random author" was selected from each downloaded paper by randomly selecting a single author from the author list of each paper, resulting in a selection of one author per paper. The number of randomly selected authorships of each author was added to the table of paper frequencies.

An author is considered when counting first author papers if listed first among the authors, and for a reprint paper if designated as the corresponding author. Using this paper as an example, Timo Koski would be considered when counting first author papers, Ulf Sandström would be considered when counting reprint author papers, whereas Erik Sandström would only be considered in the all author counting (and possibly the "random author" counting, in which any one of the three authors could be selected).

For each university and authorship type (all, random, first and reprint), a paper frequency distribution, i.e. the number of authors having one paper, two papers and so forth, was compiled (see Table 1).

All zero-frequencies were subsequently removed to construct zero-truncated samples on which the estimations could be performed.

4.3. Waring based estimation of the population mean

The paper frequency distributions constructed above are zero-truncated samples: the zero-class is missing (or in this case, intentionally hidden). The objective of the method presented and tested in this paper is to estimate this zero-class and in turn the mean of the non-truncated distribution (the population mean).

The method is very simple and carried out using the following steps, as set out in Telcs, Glänzel, & Schubert (1985).

1 Extraction of the stepwise left truncated sample mean: The mean of the full (zero-truncated) sample was calculated after which all one-frequencies (authors with one paper) were removed (resulting in a one-truncated sample). Subsequently, the one-truncated sample mean was calculated, the two-frequencies removed and so on. The result is a set of data points where the x-axis range from one (zero-truncated) to the maximum value of the distribution, and the y-axis present increasing values (the calculated means).
2 Fitting of the straight line: The data points were plotted and a straight line fitted through the points using weighted least square regression. Weights presented in Telcs et al. (1985) were applied. The intercept of the fitted line is an estimate of the mean of the non-truncated distribution (the population mean).

4.4. Simplified waring estimation

The estimation presented above may be further simplified and calculated only based on the sample mean, the total number of frequencies and the fraction of one-frequencies (Braun et al., 2001; Schubert & Braun 1992). Using this method, estimation of the mean of the non-truncated Waring distribution μ can be carried out using the following formula:

$$\mu = \frac{S}{S(1-f)/[x(1-f)+f(1-x)]} \tag{14}$$

[3] Linkoping University publication data covered the period 2005–2008 and for Swedish University of Agricultural Sciences the period 2005–2007. Selections of time periods for employee and publication data were due to availability of data from previous studies.

Table 2
Comparison between actual numbers and Waring estimates.

	University 1				University 2			
	All AU	Random AU	First AU	Reprint AU	All AU	Random AU	First AU	Reprint AU
Calculated	6.56	1.55	1.38	1.65	5.43	1.55	1.30	1.68
Estimated	7.26	1.75	1.39	2.04	5.22	1.51	1.17	1.63
(Simplified)	6.98	1.64	1.67	2.03	4.82	1.47	1.35	1.65

where S is the total number of papers, x is the average number of papers per author and f is the fraction of authors with exactly one paper.

4.5. Agreement between expected and estimated productivity

For each constructed zero-truncated distribution, the population means were estimated using the Waring approach. The results of the estimations and the actual population means are presented in Table 2.

The estimated values are for the most part very good. In many cases the estimates are within a 10 % margin from the expected values. Only in a few cases the estimates are considerably far from the expected values (>20 %). The Simplified Waring estimation performs on par with the full version.

4.6. Testing the reliability of the estimate

The results above indicate that Waring based estimation of the zero-class in general produce good results. A concern is, however, that the estimates will be very sensitive to small variations in the sample, i.e. that the reliability of the estimates will be low. That is certainly the case in the estimates provided above since they are based on relatively small samples (~500–1000 authors). But if we do not need to know the zero-class, larger samples can easily be created. In the following section, the reliability of the estimates is studied using larger samples.

4.7. Field delimited nordic papers

The potential author data sets outlined above provide a full population of paper frequencies, including the zero-class. This provides for detailed comparisons of expected and estimated population means. However, the data sets are rather small and the estimates can therefore be expected to be unstable. To study the confidence interval of estimates on a larger sample, further data sets has been compiled.

From the Web of Science, papers with Nordic (excluding Iceland) addresses published between 2003 and 2006 were downloaded. A selection of authors was obtained by extracting "first authors" (see above) from the downloaded papers and connecting these to the designed addresses. Authors with non-Nordic (or Icelandic) addresses were removed. The restriction to first authors was necessary since other authors, in the case of collaborative papers, could not be associated with specific addresses.[4]

The names of the selected set of author fractions were manually adjusted to distinguish between homonyms and to harmonize author fractions relating to the same person.[5] The number of first authorships of each distinctive author represented in the data were extracted and compiled into a table.

Each paper was designated to one of seven "fields" based on the classification of Web of Science subject categories proposed by Zhang, Glänzel, & Liang (2009). Following this, each distinctive author was designated to the field where the author had most papers. In cases where the number of papers was equal for two fields, one field was randomly selected.

For each field, a paper frequency distribution, i.e. the number of authors having one paper, two papers and so forth, was compiled (Table 3). The number of potential authors (authors with zero papers) in the selected time period is not known. Hence, the distributions are zero-truncated.

An indication of the "goodness-of-fit" of the Waring distribution is given by the correlation coefficient of the regression line through the truncated means (Telcs et al., 1985).[6] Five of the seven distributions above yield correlation coefficients above 0.99 and the other two yield values of 0.988 (Computer...) and 0.975 (Sociology...). This indicates a good fit.

[4] Complete author-address information only exists in Web of Science after 2008. An alternative solution is to use reprint authors (which reveal similar results).

[5] To simplify the process, an automatic disambiguation procedure was carried out before the manual adjustment. As automatic disambiguation procedures continue to develop (see e.g. Caron & van Eck, 2014; Gurney, Horlings, &, van den Besselaar, 2012; Soler, 2007), it will be possible to use such procedures without manual adjustment.

[6] The standard chi-squared test is not considered suitable for testing the goodness of fit to a Waring distribution, see Telcs et al. (1985).

Table 3
Paper frequency distributions for seven fields of science.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	1759	2770	3761	2257	6452	1548	2121
2	759	1255	1611	827	2835	647	659
3	385	627	961	368	1593	387	285
4	221	327	516	211	903	172	126
5	112	183	333	108	500	98	52
6	79	107	179	45	327	58	28
7	53	57	149	29	259	40	16
8	24	44	97	17	138	29	7
9	25	23	76	19	113	28	6
10	27	14	44	12	85	12	5
>10	45	37	224	26	352	46	13

[1] Agriculture & Food Science; [2] Biology, Environmental Science & Geography; [3] Chemistry, Physics & Engineering; [4] Computer Science & Mathematics; [5] Life Science & Medical Science; [6] Psychology & Education; [7] Sociology, Economics & Political Science.

4.8. Bootstrap

In several simple applications of statistical methods to data, the uncertainty of an estimate may be assessed by a mathematical calculation based on an assumed probability model for the available data. In more complex situations the mathematical analysis may be time consuming and difficult, and may require additional simplifications that make the results unreliable. Bootstrap is a well-established method for obtaining reliable standard errors and confidence intervals for estimates of interest (Davison & Hinkley, 1997). The main idea is to resample from the original data and thus create replicate data sets from which the variability of the estimate can be inferred without analytic calculation.

With regard to the Waring model of scientific productivity we want basically to estimate the parameter μ, which is the mean of the non-truncated Waring distribution using the linear representation (13) from the above, i.e.,

$$y_s = \mu + s \cdot \mu_1 + e_s, s = 0, 1, \ldots, s_{max}$$

where y_s is the left truncated sample mean i.e., an estimate of $E[X|X \geq s]$, s_{max} is the maximum value of the publications in data and e_s are respective random deviations (or residuals) of y_s from the 'true' regression line. Then by fitting of a straight line by weighted least squares, as described above, we may estimate the intercept μ and the regression coefficientμ_1. The question is to obtain a figure of the uncertainty of the estimateμ. There seems to be no immediate analytic procedure for assessment of this uncertainty, as, in particular, we should perhaps not assume the homoscedasticity and independence of the residuals.

A bootstrap technique for this is to resample, say B times, the authorship data (described above) thus creating a replicate authorship data. For each replicate data set one calculates the regression line obtaining $\mu_1 \ldots .\mu_2$ from which one can calculate the empirical distribution of the estimate of the intercept its bootstrap mean, bootstrap standard deviation and find fractals to compute a bootstrap confidence interval for the intercept.

4.9. Confidence intervals of estimated productivity

The bootstrap method was applied to calculate confidence intervals for estimates of papers per researcher in each field set out in Table 2. The estimations were carried out using the simplified Waring estimation (14). The results are presented in Fig. 1.

The stability of the estimates would be acceptable for a wide range of applications. Except for "Sociology, Economics & Political Science", the confidence intervals are in the range of ±4–8%. Moreover, the differences between the fields are most often significant. The results further indicate that Waring based estimation of productivity (papers per researcher) can be used for field comparisons, although some additional improvements (e.g. larger data sets and more homogenous field delineations) would certainly be desirable.

5. Some application areas for the new methodology: productivity, relevant weighting, and performance based funding

5.1. Productivity

There are several applications areas of the outlined methodology, but the first and foremost would be to handle the productivity problem discussed above. The methodology can be used to produce field normalized production values per macro field, which could be larger or smaller areas depending on the area of application. Sandström & Sandström (2007a, 2007b, 2008, 2009) introduced the Waring method for such an exercise based on the Nordic (excluding Iceland) countries when they calculated Field Adjusted Production (FAP) values for Swedish (and Australian) universities.

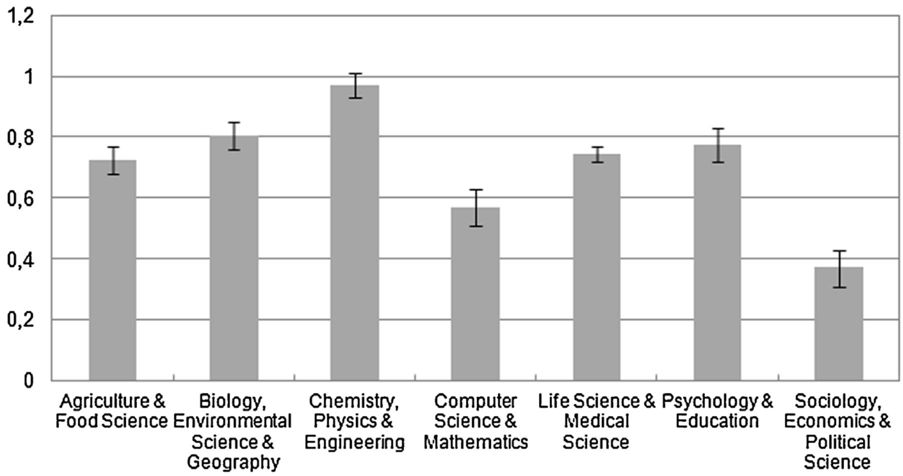

Fig. 1. Estimated Papers per Researcher with Confidence Intervals for Seven Fields of Science.

Most academic studies have kept away from institutional or organisational productivity comparisons because of the methodological problems involved. However, several of these problems are discussed and accounted for in papers by Abramo and D'Angelo (2014) and Abramo, D'Angelo, & Di Costa (2011). The Italian group states that although it might be tempting to "simply divide aggregate output values by aggregate input values", such a procedure would be fundamentally and seriously flawed. The reason is quite obvious: a varying intensity of papers and a varying specialization over disciplines between institutions (and countries).

Abramo and his colleagues contend that most of the country studies using bibliometric data fail in this respect. It is well known that medical researchers tend to produce more, often shorter papers where methodology and prior knowledge is codified in citations; and engineering scientists produce less frequently and have fewer cross-references (Narin & Hamilton, 1996). These field differences affect both citation rates and mean number of papers per author, and the differences are to some extent explained by shifting coverage of fields in the Web of Science database. Not even methodologically sophisticated scholars as Leydesdorff and Wagner (2009), take the different specialization patterns into consideration when they employ an input-output analysis on the basis of cost per publication and country.

New rankings of universities, as the Shanghai ranking, do not take these differences into account. Also the Leiden ranking (CWTS, 2011–2015) fails on this specific point. Through multiplication of relative citations by number of papers, universities active in the medical areas receive higher rankings, given the same citation performance, due to more frequent publishing.

The attempt to solve the methodological issues of counting of papers with a normalisation procedure – Field Adjusted Production (FAP) – includes several crucial steps.

One first step is to delineate a geographical area in which the relations of production within the academic sector are about the same. That could be West-Europe, Great Britain, Central Europe or Latin Europe. Reference values for all Swedish, Finnish, Danish and Norwegian universities were used in the first FAP application that was used by the Swedish government for redistribution of floor funding to universities.

Field delineation is an important issue. For citations the Web of Science subject categories are used, but these 250 categories create too small samples when Nordic (excluding Iceland) authors are used to create productivity data. There are several alternative ways of producing macro classes (e.g. Thomson ESI field categories).

The methodology described in Section 4 above was used to establish reference values based on all Swedish, Finnish, Danish and Norwegian universities. By dividing each paper fraction[7] of a university (or country, research group, etc) with the relevant reference value for that paper, we obtain the relative quantity of production performed by the university. This indicator is called the "Field Adjusted Production (FAP)". Then, simply by multiplying the production values by some citation values, e.g. field-normalized citation scores, it is possible to establish a combined value incorporating production and "quality". Such multiplication should be carried out on the paper (fraction) level to ensure that each paper is given a relevant weight also in the "quality" indicator (see comments on citation averages below).

The advantage of this method is that units are made comparable although they have their main activities in separate fields of science.

[7] This could be an author fraction (e.g. 0.25 if there are four authors of the paper) or an organisation fraction (e.g. 0.5 if there are two organisation addresses listed on the paper. Note that all paper fractions can be used, even if "random author" is used to make up the reference value. The important aspect is that each paper is only counted to an aggregated sum of one (1), since each paper is only counted once in the making of the reference values.

5.2. More relevant citation averages

The methodology described above also includes a new view on weighted citation averages that makes it possible to account for and give reasonable weight to those fields that have an average paper production which is considerably lower than other areas.

Traditionally, when calculating a (field normalized) citation score using fractional counting, each paper is given an aggregate weight of one (1), which weight is split between authors or organisations depending on level of fractionalisation. Hence, when calculating the (field normalized) citation score of a university, the university's papers in social science has been given the same weight per paper as the university's papers in chemistry. However, since the paper productivity is very different in social science and chemistry (as apparent from the results presented above), such calculation would give too much weight to the chemistry papers and the citation score of the university would be unjustifiably based on the score of the chemistry papers. Note that this applies even though the citation values are properly field normalized and proper author/organisation fractionalisation is used.

Therefore, the traditional (field normalized) citation scores, such as the scores in the Leiden ranking, are misrepresentative when applied to units (e.g. universities) that include several fields of science: they overestimate the weight of research fields with high paper production and underestimate the weight of research fields with low paper production.

This discussion is remotely related to the discussion on fractional counting methods. In their article Waltman and van Eck (2015) state that when field normalisation is applied it has to be accompanied by a proper method for fractional counting. Although the authors account for most of the problems with counting they miss the mentioned important aspect. Even if we count only fractions of papers the production of article shares is higher in natural and medical science and compared to humanities and soft social sciences. How to handle that problem? Waring weighted citations averages makes organisational units comparable as this procedure gives humanities and soft social sciences a weight that puts the activities in these areas in line with its financial and scientific importance for the respective universities.

5.3. A new ranking of universities

The range of methodologies applied here can also be the rational for a ranking and measurement of all universities within a country if there is interest for applying a performance based research funding model.

The range of methodologies applied here has also been the ground for the so called Swedish Model (as an alternative to the Norwegian Model for publication counting). This model has been used for a performance based research funding of direct resources (institutional funding or floor funding) from the State to universities and university colleges since 2009 and is in 2016 still in operation, but under discussion partly due to resistance from the soft sciences, partly due to the complexity of the method which seems to be hard do understand for politicians as well as for those academics involved in the policy discussion (c.f. Hicks, 2012; OECD, 2010).

6. Conclusions and areas for further research

The results presented in this paper indicate that Waring based estimation of the zero-class of a zero-truncated paper frequency distribution (and estimates of publications per researcher based thereon) bear real significance and that the estimates are reasonably stable (given a sufficiently large data set). We are convinced that the ability to perform this type of estimates will be the key to more advanced measures of productivity (e.g. field normalized). Consequently, it is our intention to inspire further methodological development on estimates based on publication frequencies. We welcome suggestions of how to improve the Waring model as well as suggestions of alternative methods for the estimation of mathematical probability.

Authors contribution

Conceived and designed the analysis: Timo Koski, Erik Sandström and Ulf Sandström.
Collected the data: Ulf Sandström.
Contributed data or analysis tools: Erik Sandström.
Performed the analysis (performed the statistical analysis): Erik Sandström.
Wrote the paper (concerns the mathematical parts of the paper): Timo Koski, Erik Sandström and Ulf Sandström.

Acknowledgements

A first version of this paper was presented at the ISSI conference in Durban 2011 (Koski, Sandström, & Sandström, 2011). Valuable financial support was given by the Riksbankens Jubileumsfond (P12-1302:1) which made the finalization of this paper possible.

References

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*, 1129–1144.

Abramo, G., & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, *10*(2), 646–651.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). Research productivity: Are higher academic ranks more productive than lower ones? *Scientometrics*, *88*(3), 915–928.

Bernardo, J. M., & Smith, A. F. M. (1993). *Bayesian theory*. Chichester: John Wiley and Sons.

Braun, T., Glänzel, W., & Schubert, A. (1990). Publication productivity: From frequency distributions to scientometric indicators. *Journal of Information Science*, *16*, 37––44.

Braun, T., Glänzel, W., & Schubert, A. (2001). Publication and co-operation patterns of the authors of neuroscience journals. *Scientometrics*, *51*(3), 499–510.

Burrell, Q. L. (2004). Sample-size dependence or time-dependence of statistical measures in informetrics? *Journal of the American Society for Information Science and Technology*, *55*(2), 183–184.

CWTS (2011–2015): Leiden ranking <http://www.leidenranking.com/> last accessed 2016-02-17.

Caron, E., & van Eck, N. J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. In *Proceedings of the science and technology indicators conference* (pp. 79–86).

Chen, W.-C. (1980). On the weak form of Zipf's law. *Journal of Applied Probability*, *17*, 611–622.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Dimaki, C., & Xekalaki, E. (1996). Towards a unification of certain characterizations by conditional expectations. *Annals of the Institute of Statistical Mathematics*, *48*, 157–168.

Egghe, L. (2005). The power of power laws and an interpretation of Lotkaian informetric systems as self-similar fractals. *Journal of the American Society for Information Science and Technology*, *56*(7), 669–675.

Glänzel, W., & Schubert, A. (1984). A dynamic look at a class of skew distributions. A model with scientometric applications. *Scientometrics*, *6*, 149–167.

Glänzel, W., & Schubert, A. (1995). Predictive aspects of a stochastic model for citation processes. *Information Processing & Management*, *3*, 169–180.

Glänzel, W., Telcs, A., & Schubert, A. (1984). Characterization by truncated moments and its application to Pearson type systems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *66*, 173–183.

Gurney, T., Horlings, E., & van den Besselaar, P. (2012). Author disambiguation using multi-aspect similarity indicators. *Scientometrics*, *91*(2), 435–449.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*, 251–261.

Irwin, J. O. (1959). On the estimation of the mean of a Poisson distribution from a sample with the zero class missing. *Biometrics*, *15*, 324–326.

Irwin, J. O. (1963). The place of mathematics in medical and biological statistics. *Journal of the Royal Statistical Society. Series A (General)*, *126*, 1–45.

Kendall, M. G. (1960). The bibliography of operational research. *Operational Research Quarterly*, *11*, 31–36.

Koski, T., Sandström, E., & Sandström, U. (2011). Estimating research productivity from a zero-truncated distribution. *Paper to the 2011 ISSI conference in durban*,

Leydesdorff, L., & Wagner, C. (2009). Macro-level indicators of the relations between research funding and research output. *Journal of Informetrics*, *3*, 353–362.

McKendrick, A. G. M. (1926). Applications to mathematics in medical problems. *Proceedings of edinburgh mathematical society*, *44*, 98–130.

Meng, X. L. (1997). The EM algorithm and medical studies: A historical link. *Statistical Methods in Medical Research*, *6*, 3–23.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228.

Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, *36*(3), 293–310.

OECD (2010): Performance-based funding of public research in tertiary education institutions. Report.

Reed, W. J., & Hughes, B. D. (2002a). On the size distribution of live genera. *Journal of Theoretical Biology*, *217*, 125–135.

Reed, W. J., & Hughes, B. D. (2002b). From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, *66*, 67–103.

Sandström, E., Sandström, U. (2007): Modell för beräkning av direktanslag till svenska lärosäten baserad på forskningsproduktion och citeringsgrad. Bilaga 8 till SOU 2007:81 *Resurser för kvalitet*, pp. 389–443. [In Swedish].

Sandström, U., Sandström, E., (2007): A metric for academic performance applied to Australian Universities 2001–2004. E-prints in Library and Information Science <http://eprints.rclis.org>.

Sandström, U., Sandström, E. (2008): Resources for citations (in Swedish: Resurser för citeringar). Stockholm : National Board for Higher Education, Report 2008:18R.

Sandström, U., & Sandström, E. (2009). The field factor: Towards a metric for academic institutions. *Research Evaluation*, *18*(3), 243–250.

Schubert, A., & Braun, T. (1992). Three scientometric etudes on developing countries as a tribute to Michael Moravcsik. *Scientometrics*, *23*(1), 3–19.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425–440.

Soler, J. M. (2007). Separating the articles of authors with the same name. *Scientometrics*, *72*(2), 281–290.

Telcs, A., & Schubert, A. (1986). Publication Potential: An indicator of scientific strength for cross-national comparison. *Scientometrics*, *9*(5–6), 231–238.

Telcs, A., Glänzel, W., & Schubert, A. (1985). Characterization and statistical test using truncated expectations for a class of skew distributions. *Mathematical Social Sciences*, *10*(2), 169–178.

Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indictors and the choice of an appropriate counting method. *Journal of Informetrics*, *9*, 872–894.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philosophical Transactions Royal Society B*, *213*, 431–444.

Zhang, L., Glänzel, W., & Liang, L. M. (2009). Tracing the role of individual journals in a cross-citation network based on different indicators. *Scientometrics*, *81*(3), 821–838.