How much do journal titles tell us about the academic interest and relevance of economic research? An empirical analysis

Felix Schläpfer

December 2009
How much do journal titles tell us about the academic interest and relevance of economic research? An empirical analysis

December 2009

Author’s address: Felix Schläpfer
E-mail: felix.schläpfer@env.ethz.ch
How much do journal titles tell us about the academic interest and relevance of economic research? An empirical analysis

Felix Schläpfer* †

December 2009

Abstract

Unlike in other disciplines, research output in economics is commonly measured based on the journal titles in which an author has published. Here, I examine how much output measures based on journal titles tell us about the academic interest and relevance of economic papers as measured by citation frequency. Using data from the 2008 Handelsblatt ranking of economists in German speaking countries and interdisciplinary citation data from the Web of Science, I find that researcher scores based on journal titles explain only about one fourth of the variation (variance) in article citations. When the top 10 (20) percent of the researchers according to journal title scores are excluded, the percentage of explained variation in citation frequency drops to 5 (3) percent. These findings empirically confirm the hypothesis that the measures of research output in economics promote narrow and complacent work that is of interest to few, even among an academic audience. They suggest that responsible hiring committees and funding institutions should re-examine existing standards in evaluation and abandon the heavy reliance on journal titles as a measure of individual research output.

Keywords: citation index, incentives, publication, research evaluation, scientometrics

* Correspondence to: Environmental Policy and Economics, Institute for Environmental Decisions, ETH Zurich, Universitätstrasse 16, 8092 Zürich, Switzerland, Tel. +41 44 632 85 08, E-mail: felix.schlaepfer@env.ethz.ch.
† I am grateful to Silvana Wölfle for careful database research in the Web of Science and to Bruno S. Frey, Hans-Werner Sinn and Peter Zweifel for comments on an earlier draft of the manuscript.
1. Introduction

For good reasons, the use of citation frequency as a sole or main criterion for measuring researcher productivity is being increasingly criticised (e.g. Adler et al. 2008). Key issues with citation based output measures include their limited responsiveness to an article’s interest and relevance outside academia and the problem that these measures can provide incentives for unproductive herd behaviour and “citation cartels” in research. Furthermore, it is often argued that the coverage of sources by the existing citation databases is not sufficiently objective or comprehensive.

Less widely appreciated is the idea that the incentives in research can be even worse where research output is not measured based on article citations but instead relies on a mono-disciplinary expert rating of the journals in which an author has published. Such is the situation in economics where a small number of authoritative individuals define “quality” through a largely subjective weighting of journal titles and at the same time act as gate keepers who select the articles published in those same journals (Laband and Tollison 2003, Oswald 2007). In this system, the quality of an individual article is defined solely, and for all eternity, by the journal title in which it has been accepted for publication—regardless of whether the article is ever read, understood or cited by any audience.

In the economic discipline, universities and funding agencies largely base their hiring and funding decisions on this quality criterion. This is not a contentious claim but would be confirmed by almost any academic working in the discipline. Anecdotal evidence from job and funding applications and personal communications from many of colleagues support the impression that the use of journal titles is the central criterion in research evaluations.

Since citations, in contrast, play virtually no role in the evaluation of economic research output, any rational economist who wishes to survive and succeed in academia should focus on chasing journal titles – and sacrifice the objectives of relevance and interest (as measured by future citations) in his or her research. This implies the testable hypothesis that economists’ research output based on journal titles should, as a result, be only very weakly, if at all, associated with their output based on citations. If this hypothesis is confirmed, this finding would imply that the incentives introduced by these output measures do not encourage interesting and widely relevant work. Originating from a time when this problem was not apparent (and citation data were unavailable), these output measures would thus be highly undesirable from a societal and efficiency perspective.
Here, I examine this hypothesis empirically. Specifically, I investigate the relationship between a well known traditional measure of researcher output based on expert-rated journal titles – the Handelsblatt-Ranking of economists in German speaking countries‡ – with the same authors’ citation rates in Web-of-Science-listed journals. The following specific questions are addressed:

1. How much of the variation in researchers’ number of citations is explained by their score based on journal titles?
2. How strong is the association between researcher rank based on citations and researcher rank based on journal scores?

We find that these relationships are very weak, confirming the hypothesis that the traditional measurement of research output in economics produces problematic incentives and should for this reason be abandoned by universities and funding institutions as a sole or major indicator of research quality.

The remainder of the paper is organized as follows. The next section introduces the relevant journal weighting schemes and discusses the incentives for research that may result from these measures. Section 3 presents the data used in the empirical analysis. Section 4 contains the results, followed by conclusions.

2. Journal weighting schemes and resulting incentives

In economics, a variety of journal weighting schemes have been proposed (see e.g. Kodrzycki and Yu 2006). In Europe, the two most established weighting schemes are probably the ‘Tinbergen list’ and the weighting scheme developed by Combes and Linnemer (2003). Each of these is based on a subjective weighting of journals by selected experts. To provide a flavour of the considerations that went into these schemes, and to document their limited scope, these two schemes are in the following described in the words of the authors.

The Tinbergen list classifies journals as “generally accepted top-level journals” (AA), “very good journals covering economics in general and the top journals in each field” (A), and “good journals for all research fields within the Tinbergen Institute”. The fields covered by the Tinbergen Institute cover “economics, econometrics, finance, operations research,

‡ Starting in 2010, the Handelsblatt-Ranking will be jointly published by the Newspaper Handelsblatt, the German Verein für Socialpolitik and the Konjunkturforschungsstelle (KOF) of the Federal Institute of Technology ETH, Switzerland. The ranking will thus be officially endorsed by an academic association and a public research institution.
marketing and accounting”. On its website, which also displays the resulting journal weights, the Tinbergen Institute describes the classification as follows:


The weighting scheme by Combes and Linnemer (2003), was developed for the European Economic Association (EEA) for a ranking of the leading European economics departments. The weighting is described as follows (Combes and Linnemer 2003, p. 4):

We built an original journal weighting scheme denoted CL that weights all EconLit journals from 1 to 1/12. After a long and repetitive procedure which started in 1998 [...] we divided the EconLit journals in six groups. First, top journals are significantly differentiated from other ones with a weight equal to 1. A weight of 8/12 only is given to the next 16 journals. Then, a series of 39 journals are weighted 6/12, 68 journals 4/12, 138 journals 2/12 and all remaining journals 1/12 [...].

Our choices, which could be discussed endlessly as more than 800 journals are considered, tried to be consistent with citation/impact indicators when they are available. We do not think, however, that these have to be followed blindly. Independently of the journal average quality, the number of citations can vary from one field to the other and from a young journal to an older one. To counter this kind of effects, in any case, we tried to put at least 6/12 to any journal which is a leader in its field. Conversely, we did not put 8/12 or more to a journal too specialized. We do not believe that our scheme is perfect but the centre rankings proved to be very robust to moderate changes in weights even if such changes could be important at the individual level.

As suggested earlier, measures based on expert-rated journal titles are not only a highly noisy signal of quality but may also provide a number of problematic incentives in publication beyond those of typically criticized with citation based measures. At least four potentially problematic incentives can be identified (Table 1).

First, journal ratings may be a very noisy signal of the quality of individual articles. Previous research has demonstrated a substantial overlap of article citation rates among journals with higher and lower reputation (Laband and Tollison 2003, Starbuck 1005, Oswald 2007). Measurement based on journal titles thus provides a strong incentive for wasteful investments in repeated submissions to top journals with high rejection rates.

Second, it is well known that papers can be purposefully written to please the editors and reviewers of the journal to which they are submitted. Articles are more likely to be

---

accepted in highly ranked economics journals if they support (and slightly extend) rather than contradict received wisdom (Frey 2003). Hence, research evaluation based on journal titles may discourage innovative and challenging work that may, for this very reason, be disadvantaged in highly ranked journals (but would nevertheless be successful in term of future citations if published in a ‘lower’ journal).

Third, the journal ratings, and hence evaluations based on them, reward and thus promote technicality and (unwarranted) mathematical sophistication, as many economists use and perceive these to as signals of high-quality research.** This is directly reflected in the weighting schemes. For instance, the journal *Ecological Economics* places little emphasis on technicality and disciplinary sophistication. Although it ranks among the top economics journals in terms of its *Web-of-Science* impact factor, it receives the lowest possible score in major disciplinary rankings (including the *Handelsblatt* ranking described below).

Finally, since articles published in journals of related disciplines are not ‘counted’, the economic output measures provide powerful and virtually irresistible incentives for a purely disciplinary orientation in publishing. (As a matter of fact, the authors of some economic weighting schemes seem to have quite willingly cultivated a narrow disciplinary orientation.††)

In sum, output measures based on journal titles have the potential to provide incentives for economic research that is technically sophisticated, supportive of prior work and close to the core of the discipline – at a possible expense of relevance, interest, originality and innovation.

3. Data and measures

3.1 Output measure based on journal titles

On 22 September 2008, the German newspaper *Handelsblatt* published the 2008 edition of its ranking of the “top 200” active economists at universities in the German speaking countries

** In discussing the origins of the recent economic crisis, Paul Krugman (2009) argues that “the economics profession went astray because economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth.”

†† Liebowitz and Palmer (1984) who analyze the influence of economic journals state: “[E]conomists, being a rather narrow-minded and self-centered group, are probably more concerned with a journal’s impact on the economic profession [than on other disciplines].”
Germany, Austria and Switzerland. Furthermore, the newspaper published a list ranking the “top 100” researchers based on articles published in 2004 through 2008.

The *Handelsblatt* ranking is based on research output in 220 journals in the fields of economics and statistics. In weighting the journals, the *Handelsblatt* borrowed from two established European journal rankings, the ‘Tinbergen list’ and the weighting scheme developed by Combes and Linnemer (2003) (see section 2). Five top journals obtain the value 1, the remaining journals obtain values between 0.67 and 0.2 points. A list with the journals and corresponding points is available on the internet‡‡. Short research notes and comments are counted half. Book reviews, replies and corrections are not counted.

The journal points received by the author of an article is calculated by the formula \(2p/(n+1)\), where \(p\) is the point value of the journal and \(n\) is the number of authors of the article. Points are summed over articles to obtain a researcher’s total score. The dataset relies on researcher-reported article lists. The article lists of all researchers (with journal points of each article) are published in the internet, such that misrepresentation is unlikely.

### 3.2 Output measure based on citations

For all authors listed in those *Handelsblatt* ranking lists, citation data were researched from the *Web of Science* database using the “search” and “cited reference search” tools. Where an author’s work could not be unambiguously identified by his or her last name and initial(s), we used his or her publication list to exclude articles from other authors. In the cases where we could proceed with the efficient “cited reference search” tool (since no or only few papers from other authors with identical name and initials had to be manually excluded), the citations numbers include also citations to articles that are not themselves listed in the *Web of Science*. In the remaining cases, where we used the “search” tool, the citations include only those to articles that are themselves referenced in the *Web of Science*. Due to resource constraints, and since the citations to non-*Web of Science*-listed publications are only a small fraction of the total, we chose to tolerate this (unsystematic) measurement error

‡‡ [http://www.handelsblatt.com/_t=dgtool,id=15,obj=1;singleclip](http://www.handelsblatt.com/_t=dgtool,id=15,obj=1;singleclip) (accessed 18 November, 2009).
4. Results

4.1. Relationship between scores based on journals titles and number of citations

Figure 1a shows scatter plots of the “top 200” researchers’ number of citations in 2008 and their lifetime scores (all articles) based on journal titles as published in the *Handelsblatt* ranking. Visual inspection suggests that the correlation between the two measures is relatively weak. Many of the “top” researchers in the *Handelsblatt* ranking are relatively rarely cited. On the other hand, there are some apparent “second rate” researchers whose work obtains relatively low scores but apparently makes a substantive impact on the literature.

To quantify how much of the variation in citations can be explained by the *Handelsblatt* scores, the citation measure was regressed on the journal score to obtain the $R^2$ values (explained variation in the linear model) and the respective $P$-values (from testing the restriction that the coefficient on “journal score” is zero). The data were not log-transformed due to the occurrence of zero values in the citation variable. (The original data are available from the author on request.)

Table 2 (upper part) summarizes these results. In the linear model, journal score explains 27 percent of the variation in citation rates. However, much of this explanatory power is due to a small number of top researchers. When the top 10 percent of researchers (based on *Handelsblatt* scores) are omitted, the percentage of explained variation drops to 5 percent. When the top 20 percent are omitted, the proportion of explained variation drops to 3 percent. These results show that the association between the *Handelsblatt* ranking and citation frequency is very weak. Evaluating research output based on the *Handelsblatt* ranking implies that relevance and innovation as measured by citations is almost completely discounted.

One might object that lifetime journal scores may be only weakly correlated with citations, since older articles might not be cited any more in 2008 (although articles citing those articles might still be). To entertain this possibility, I also examined the relationship between journal scores accumulated in 2004 through 2008 and citations to articles published in 2004 through 2008 (Figure 1b). The finding is similar. The percentages of explained variation are even somewhat smaller, ranging from 11 percent for the full sample to 2 percent when the top 20 percent of researchers are omitted (Table 2, lower part).
4.2. Relationship between rank based on journals score and rank based on citations

How do these numbers translate into researcher ranks? For illustrative purposes, I also plotted rank based on journal scores (Handelsblatt rank) against rank based on citations (Figure 2)§§. The Handelsblatt ranks based on lifetime journal score and 2008 citations to all articles are shown in Figure 2a. The data for the journal score of articles published in 2004 through 2008 and the citations in 2008 to articles published in 2004 through 2008 are shown in Figure 2b. Again, the scatter plots shows that the Handelsblatt ranks are a very poor predictor of researcher ranks based on number of citations.

5. Conclusion

Article citations – the standard measure of research output in many disciplines – are increasingly questioned as a sole indicator of research quality. The present study argues that journal titles – the standard measure of research quality in economics – are even more problematic. The present empirical analysis shows that there is a lack of consistent relationship between the standard measure of research output and citations rates. This finding supports the hypothesis that the standard measure of research output in economics provides incentives for unfruitful research that is not (even) interesting and relevant to an academic audience.

By their very nature, the standard measures based on journal titles not only fail to adequately measure the quality of an individual researcher’s work. More importantly, the measure introduces powerful incentives against interesting and relevant work. Since these standard measures do not reflect (multi-disciplinary) citation rates, they provide incentives for investing in unfruitful technical sophistication, academic followership, and disciplinary isolation which are rewarded by that standard. The financial crisis has most impressively demonstrated the failure of the economic discipline to respond to real world issues (e.g. Krugman 2009). Responsible universities and funding institutions are advised to re-consider existing practices in research evaluation and abandon the heavy reliance on journal titles in the evaluation of economic research.

§§ Where group of two or more researchers had received an identical number of citations, the ranks within the group were assigned based on journal scores. (Alternatively, one could assign equal ranks to these individuals without changing the overall pattern of the results.)
References


Table 1. Problematic characteristics of output measures, resulting incentives, and relevance of the problem (+) for measures based on journal ratings and article citations.

<table>
<thead>
<tr>
<th>Problematic characteristic</th>
<th>Resulting incentives</th>
<th>Journal ratings</th>
<th>Article citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blind to relevance outside academia</td>
<td>Incentive for irrelevant “ivory tower” research</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Reward for “fashionable” topics (in review stage or after publication)</td>
<td>Incentive for publication of superficially interesting results, herd behaviour</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Measurement at group (journal) rather than individual (article) level</td>
<td>Incentive for wasteful investments in repeated submissions to top journals with high rejection rates</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Quality appraisal only during review process</td>
<td>Incentive for “pleasing” rather than innovative research</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Inappropriate reward for technicality/sophistication</td>
<td>Incentive for wasteful investments in technical sophistication</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Exclusion of work in other disciplines</td>
<td>Disincentive for collaboration with other disciplines</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Proportion of variation in citations explained by journal scores ($R^2$-values and $P$-values from linear regression models).

<table>
<thead>
<tr>
<th>Sample</th>
<th>$n$</th>
<th>$R^2$</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All articles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample</td>
<td>200</td>
<td>0.27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Top 10 percent of scores omitted</td>
<td>180</td>
<td>0.05</td>
<td>0.002</td>
</tr>
<tr>
<td>Top 20 percent of scores omitted</td>
<td>160</td>
<td>0.03</td>
<td>0.039</td>
</tr>
<tr>
<td>Articles in 2004-2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample</td>
<td>100</td>
<td>0.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Top 10 percent of scores omitted</td>
<td>90</td>
<td>0.07</td>
<td>0.014</td>
</tr>
<tr>
<td>Top 20 percent of scores omitted</td>
<td>80</td>
<td>0.02</td>
<td>0.163</td>
</tr>
</tbody>
</table>
**Figure legends**

**Figure 1.** Relationship between journal score and citations in 2008: a) total journal score and citations to all articles; b) journal score in 2004-2008 and citations to articles published in 2004-2008.

**Figure 2.** Relationship between researcher rank based on journal titles and researcher rank based on citations: a) ranks based on total journal score and citations to all articles; b) ranks based on journal score in 2004-2008 and citations to articles published in 2004-2008.
a) All articles

b) Articles published 2004-2008

Note: 3 observations are out of range: (16/299), (22/422), (35/312)
a) All articles

b) Articles published 2004-2008
<table>
<thead>
<tr>
<th>Working Papers of the Socioeconomic Institute at the University of Zurich</th>
</tr>
</thead>
<tbody>
<tr>
<td>0916 Fine Tuning of Health Insurance Regulation: Unhealthy Consequences for an Individual Insurer, Johannes Schoder, Michèle Sennhauser, Peter Zweifel, August 2009, 18 p.</td>
</tr>
<tr>
<td>0911 Physician dispensing and the choice between generic and brand-name drugs — Do margins affect choice?, Maurus Rischatsch, Maria Trottmann, July 2009, 15 p.</td>
</tr>
<tr>
<td>0905 Patents versus Subsidies — A Laboratory Experiment, Donja Darai, Jens Großer, Nadja Trhal, March 2009, 59 p.</td>
</tr>
<tr>
<td>0904 Simple tests for exogeneity of a binary explanatory variable in count data regression models, Kevin E. Staub, February 2009, 30 p.</td>
</tr>
<tr>
<td>0815 Delay and Deservingness after Winning the Lottery, Andrew J. Oswald, Rainer Winkelmann, December 2008, 29 p.</td>
</tr>
</tbody>
</table>
0810 Why does the amount of income redistribution differ between United States and Europe? The Janus face of Switzerland; Sule Akkoyunlu, Ilja Neustadt, Peter Zweifel, September 2008, 32 p.
0718 Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.
0716 The relation between competition and innovation – Why is it such a mess? Armin Schmutzler, November 2007, 26 p.
0714 Competition and Innovation: An Experimental Investigation, Dario Sacco, October 2007, 36p.
0713 Hedonic Adaptation to Living Standards and the Hidden Cost of Parental Income, Stefan Boes, Kevin Staub, Rainer Winkelmann, October 2007, 18p.
0711 Self-Reinforcing Market Dominance, Daniel Halbheer, Ernst Fehr, Lorenz Goette, Armin Schmutzler, August 2007, 34p.
0710 The Role of Landscape Amenities in Regional Development: A Survey of Migration, Regional Economic and Hedonic Pricing Studies, Fabian Waltert, Felix Schläpfer, August 2007, 34p.
0707 I’m not fat, just too short for my weight – Family Child Care and Obesity in Germany, Philippe Mahler, May 2007, 27p.