

UC Santa Barbara

Ted Bergstrom Papers

Title

How often do economists self-archive?

Permalink

<https://escholarship.org/uc/item/69f4b8vz>

Authors

Bergstrom, Ted C
Lavaty, Rosemarie

Publication Date

2007-02-08

How often do economists self-archive?

Theodore C. Bergstrom and Rosemarie Lavaty*

Economics Department
University of California Santa Barbara

February 8, 2007

1 Introduction

Most scholars are eager to have their work read and even more eager to have it cited. Labor economists find that academic salaries are significantly influenced by citations. Regression analysis by Hamermesh, Johnson, and Weisbrod [8] showed the number of citations to economists' work have a significant positive effect on their salaries. Diamond [4] studied the effects of citations on academic salaries of a broad range of academics. Diamond states that "A robust finding of all studies is that citations are a positive and significant determinant of earnings over almost all of the observed range of citation levels." Diamond estimated the marginal effect of a citation on annual salary to range upward from \$50 per year. A more recent study of the determinants of economists' salaries by Baser and Pema [1] also finds a significant positive effect of citations on salaries, with an additional citation being worth about \$40 per year.

Even wealthy research universities can not afford to subscribe to all academic journals. Many smaller institutions and institutions in less wealthy countries subscribe only to a handful of relatively inexpensive top-quality journals. Scholars in government or the private sector often have no access to institutional subscriptions. Therefore an author who wishes to be read

*The authors are grateful to Anna Li and Trevor Crowe for capable research assistance and to Peter Suber for encouragement and sage advice.

has a strong incentive to make his work more widely available by posting it on the internet.

Lawrence [9] found computer science articles that are freely available online to be significantly more often cited than those that are not. Hajjem, Harnad and Gingras [7] conducted a large-scale examination of citation data in 10 disciplines over a period of 12 years. They found that in most disciplines, articles that are freely available on the internet (open access), receive at least 50 percent more citations than those that are not. In June 2004, the *Proceedings of the National Academy of Sciences* announced that authors could pay \$1000 for their articles to become immediately open access. Between June and August of that year, *PNAS* published 1492 articles, of which 212 were made open access by the authors' payments. Eysenbach [5] compared the number of citations to the open access and non-open access articles published in this period over the 10-16 month interval from their publication to October 2005. He found that the open access articles were cited significantly earlier and more often than those that were not open access.¹

For authors who have access to an online departmental or university archive or who have their own websites, the cost of posting their own papers is only a small amount of effort.² There are almost no legal obstacles to posting a final copy of one's own research article on the author's web page or in a university archive. More than two-thirds of all all journals allow authors to post a copy of the final draft of their own paper.³ Most of the remaining journals permit authors to post an earlier version of the paper that does not include all of the final copy-editing changes. Thus 94 percent of all academic journals permit authors to post essentially final versions of

¹Eysenbach found that some of the *PNAS* articles of each type were also self-archived. He found that articles that were made freely available either by self-archiving or by *PNAS* open access were significantly more cited than those that were not.

²Economics authors whose own institution lacks a working paper archive are able to post their papers at the Munich Personal RePEc archive <http://mpra.ub.uni-muenchen.de/>.

³These include the large commercial publishers, Elsevier, Springer-Kluwer, and Wiley, who do not permit posting of the publisher-prepared pdf file, but do permit the author to post a self-prepared copy of the final draft.

their own articles.⁴

Given the substantial benefits and low cost of posting their own papers, we would expect to find that most authors, and especially most economists would self-archive copies of their published papers. This paper is a report on the self-archiving practices of authors of papers in a sample of economics and political science journals.

2 Methods

We selected the two most recent (as of August, 2006) issues of a sample of 33 economics journals, selected to offer a variety of prestige-levels and field specialties. The tables of contents of these issues contained a total of 703 articles. For each article, we used Google to search the web for a freely-downloadable copy of an article with the same title and the same author(s). We also recorded whether the posted copy was either stated to be a final version of the paper or a copy of the publisher-prepared page prints or final PDF file.

We also investigated the effectiveness of two alternative search engines: Google Scholar, and OAIster,⁵ by seeking for the same titles with these tools. We also checked whether articles could be obtained through the Social Science Research Network (SSRN) and Research Papers in Economics (RePEc).

For each journal in our sample, we recorded the impact factor, the average influence per article,⁶ the profit status of the publisher, the institutional subscription price per article, and an index of the total number of library subscriptions to this journal.⁷

⁴These statistics can be found at <http://romeo.eprints.org/stats.php> and are compiled by the Sherpa/Romeo research project, maintained by a consortium of English universities.

⁵OAIster <http://oaister.umdl.umich.edu/o/oaister/> is a search engine developed by the University of Michigan Digital Library Production Service. It is designed to find scholarly materials posted on institutional archives in OAI compliant form.

⁶Citation influence per article is taken from the website www.eigenfactor.org. This measure is similar to the ISI's impact factor, but weights citations recursively by the citation weights of the citing journals, using a method similar to that of Google's PageRank. This measure counts cites that take place within five years of publication, while the impact factor counts only those appearing within two years of publication.

⁷We index subscriptions by the number of current subscriptions held in 2004 by libraries

3 Results

3.1 Google search

Table 1 reports the results of a Google search for online copies of each article in our sample, ranked in order of their average “influence per article”. Influence per article is calculated as the average number of citations per article in the first five years after publication with citations weighted recursively by influence per article of the citing journal. The column “Pct Posted” records the percentage of all research articles in each journal for which Google found a freely available online copy of a paper with the same title and same author(s).

that participate in the Online Computer Library Center (OCLC). The OCLC listing is far from a complete catalog of worldwide subscriptions. It catalogs about 20 percent of worldwide holdings.

Table 1: Self-Archiving for Economics Journals

Journal Name	Pct Posted	Cit Infl per Art	Price per Art (\$)	OCLC Subs
Quarterly J Ec	100	13.3	7.8	595
J Polit Ec	100	8.5	6.7	611
Econometrica	100	6.5	8.1	326
J Finance	96.6	6.4	3.8	689
Rev Ec Studies	100	5.9	8.0	285
J Financial Ec	100	5.8	30.9	197
Am. Ec Review	93.5	5.0	2.1	949
J Monetary Ec	100	3.9	28.5	153
J Labor Ec	75.0	3.7	8.7	253
The Ec Journal	72.7	3.3	6.9	429
J International Ec	80.0	3.3	22.7	174
J. Econometrics	90.0	2.9	33.3	112
Eur Econ Rev	75.0	2.4	24.0	100
J. Public Ec	91.9	2.1	18.6	122
J Ec Theory	89.5	2.1	27.4	138
J Develop Ec	76.2	1.5	26.7	132
J Env Ec & Mgmt	60.0	1.3	15.9	171
Economica	87.5	1.1	8.8	283
Res & Energy Ec	45.5	1.1	34.9	27
Ec Inquiry	51.9	0.9	5.4	290
Health Econ	14.3	0.8	13.8	49
Int J Ind Org	62.5	0.8	18.9	65
Int J Game Thy	75.0	0.8	16.2	54
Kyklos	71.4	0.8	20.0	154
Public Choice	65.4	0.7	16.7	144
J of Math Ec	62.5	0.6	38.0	66
Theory & Decision	37.5	0.6	36.2	73
Ecological Ec	32.1	0.5	17.2	45
J of Ec Educ	23.8	0.5	4.4	347
J Inst and Th Ec	47.1	0.4	8.5	38
Rev Indust Org	46.2	0.4	22.5	62
Applied Ec	5.0	0.3	28.4	122
Pub Finance Rev	50.0	0.2	20.0	220

*Titles of non-profit journals are printed in **bold**.

Evidently self-archiving has become the norm among authors who publish in top economics journals. Table 1 shows that freely available versions of about 90 percent of the articles in the top fifteen listed journals can be found by Google-searching the title and author. The table also indicates that the self-archiving norm is less strong among those who publish in the less influential journals. Freely available versions of about 50 percent of the articles in the eighteen lower-ranked journals in our sample could be found on the internet.

We used regression methods to investigate factors that might influence the propensity of authors to self-archive. We included independent variables that measured characteristics of the journal in which an article appeared, average citation influence per article, institutional subscription price per article, and estimated number of institutional subscriptions. We also included characteristics of the item, such as whether it was a “nonstandard article” such as a book review or editorial. We included the article’s number of authors, whether at least one of the authors was at a U.S. “research-intensive” institution (as rated by the Carnegie commission), whether at least one author was at a U.S. institution and whether at least one author was in an economics department.

Table 2: Effects on Probability of Self-Archiving

Variable	Coeff.	*Std Err.	Sig Lev
Nonstandard Article	-0.453	.077	.000
Citation Infl Per Art	.053	.019	.007
Au at US Res Inst	.086	.044	.061
Au at US Nonres Inst	-.126	.086	.150
Subscrip Price Per Art	.006	.004	.168
Author in Econ Dept	.059	.045	.204
Number Subscriptions	.0002	.0002	.343
Number of Authors	.002	.023	.9420

*Standard Errors calculated with clustering at journal level.

Not surprisingly, “nonstandard articles” such as editorials and book re-

views are less likely to be posted than research articles. Aside from this, the strongest predictor of whether an article will be posted is whether the measure “citation influence per article” of the prestige of the journal in which it is published.

Papers with at least one author employed by a U.S. university designated by the Carnegie commission as “research intensive” were more likely to be posted. Papers by U.S. authors not associated with a research-intensive university are less likely to be posted than papers by other by non-US authors or by authors at research intensive universities. Papers with at least one author employed in a university economics department are slightly more likely to be posted.

Authors of papers in journals with high subscription prices and relatively few subscribers would seem to have a stronger incentive to post their papers. We see from Table 2 that the coefficient on subscription price per article is positive as expected, but only marginally statistically significant. The coefficient on “number of subscriptions” is not negative as expected, but positive though not statistically significant. The number of authors of a paper apparently has negligible effect on the probability that it will be posted.

3.2 Exact copies of final versions?

In cases where we could not find a free self-archived copy of an article with the exact title, we searched for an article with a similar title that appeared to be an earlier draft of the same paper by the same author(s). Including such papers increases the overall percentage of available papers by about five percent. For each article, we also looked for clear evidence that the online copy was a final draft. About 14 percent of the papers in our sample were recorded as definite final drafts by our criteria. About 10 percent had freely available posted copies of the publisher-prepared pdf.

3.3 Other search engines

We compared the performance of three search engines, Google, Google Scholar, and OAIster in finding free online copies of published articles, and

we also examined the breadth of coverage of web archives SSRN and RePEc. Searching with Google Scholar, we found about 10 percent fewer articles than searches using Google. Searches with OAIster found only about 25 percent of articles in top journals. Free copies of about 26 percent of the articles could be found through SSRN and 27 percent through RePEc.

3.4 Comparison to Political Science

Political scientists are apparently less likely than economists to post their articles. We conducted a Google search for free online versions of the 302 regular articles that appeared in the most recent two issues of 16 political science journals. The results of this search are reported in Table 3. Free online versions of just over 30 percent of the articles in this sample were found. A noteworthy exception is the *American Journal of Political Science*, one of the most prestigious journals in political science, for which 94.3 percent of articles were posted.

4 Related Studies

Hajjem, Harnad and Gingras [7] conducted an ingenious large-scale study of self-archiving, using a web-crawling robot to seek open access versions of 1,307,038 articles written between 1992 and 2003 and covered by the ISI's Science and Social Science Citation Indices. They estimated the proportions of self-archived articles in each of 10 disciplines: administration, economics, education, business, psychology, health, political science, sociology, biology, and law. They found the fraction of openly accessible articles in these 10 disciplines to range from 5-16 percent. They report (in table 2a) that 13.5 percent of all economics articles and 5 percent of all political science articles are freely available online. This contrasts notably with our sample, where about 70 percent of the economics articles and 30 percent of those in political science are freely accessible. A likely explanation for most of this difference is that Hajjem et al report proportions of open access articles in a population consisting of articles published in the last twelve years, while our samples consist only of journals published in late 2006. It is also possible that the

Table 3: Self-archiving in Political Science

Journal Title	Pct Posted	Cit. Influence per Article
Am Pol Sci Review	35.0	6.9
Am J Pol Sci	94.3	4.1
J Conflict Resolut	40.0	2.6
J Democracy	44.0	1.7
J Peace Research	21.4	1.5
Eur J Pol Research	26.7	1.1
Pol Sci Quart	20.0	1.1
Party Politics	36.4	1.1
Pol. Geography	8.3	1.0
Electoral Studies	61.1	0.8
Pol Theory	0.0	0.7
The Pol Quarterly	8.3	0.5
Austral J Pol Sci	8.7	0.5
Gov & Opposition	23.1	0.5
Can J Pol Sci	18.8	0.4
J Develop Areas	13.6	N.A.

automated web search method used by Hajjem et al found fewer of the existing posted versions than our Google search.

Wren [14] used Google’s “application programming interface” to study the availability of open access versions of 48,516 papers published in 13 biomedical journals over the period from 1994 to 2004. To be counted as a match for a published article, Wren required that the posted article be a pdf file whose title, authors’ names, and affiliations match, and that the page length of the posted item be the same as that of the published article. He found that the likelihood of a posted copy of an article being available is highest for articles that are one year old and diminishes with age after one year. Wren also found that articles that appeared in journals with high impact factors were much more likely to be posted than less prestigious journals. Wren’s program found copies of about a 33 percent of the articles from the 6 journals in his sample with highest impact factors, and copies of

less than 10 percent of the articles published in the six journals with lowest impact factors.

In late 2004, Swan and Brown [13] surveyed 1296 scholars from a wide variety of disciplines to determine their practices and their beliefs about self-archiving their own scholarly work. They found that about half of the persons surveyed have “placed at least one article during the last three years in at least one of the three possible ways— by placing a copy of an article in an institutional (or departmental) archive, in a subject-based repository, or on a personal or institutional website.” They report that the likelihood of self-archiving is higher for more prolific authors. About 35 percent of the surveyed population were not aware of the possibility of self-archiving. About half of the respondents (47 percent) believed that they did not need the publisher’s permission to post their own papers, while 36 percent didn’t know, and 17 percent believed that they needed permission.

In their discussion, Swan and Brown note that physicists and computer scientists have a long-standing tradition of posting their papers in arXiv, an open access repository for preprints and postprints. They estimate that arXiv contains about one third of all articles written in physics and virtually all articles in the specialist areas—condensed matter, astrophysics, and high energy physics.

5 Implications

This paper’s findings indicate that economists who publish in prestigious journals respond appropriately to clear economic incentives. But as economists, we like to believe that others respond to the same incentives. We find it puzzling that authors of papers in less prestigious journals do not self-archive at a similar rate. Less prestigious journals typically have fewer institutional subscribers than the top journals.⁸ Therefore the gains in exposure from posting an open access copy would seem to be larger for authors of papers in these journals. Peter Suber [12] expressed similar puzzlement over the results found in Wren’s [14] study of biomedical journals. It is also perplex-

⁸The average number of subscriptions listed in the OCLC is about 342 for the top 15 economics journals in our sample and 130 for the bottom 18.

ing that, with the exception of those who write for the *American Journal of Political Science*, political scientists are much less likely to post their articles than economists.

The large differences in posting behavior between economists and political scientists, and that between economists at research intensive universities and those at teaching universities suggests to us that the decision about whether to self-archive is often not an informed rational decision, but a response to the information and norms supported by their peer groups. Authors are more likely to learn about the possibilities of self-archiving by observing their colleagues, than from their own investigations. Swan and Brown observed [13] that only about half of their respondents were aware that there are rarely any legal impediments to posting one’s own papers on an institutional website, either before or after publication. This is not surprising, given that many of the major publishers’ policies toward author self-posting their own work have only recently been clarified and liberalized.

We have argued that authors have an incentive to post their papers on open access sites because this is likely to increase their citations. While many studies show a strong positive relationship between citations and open access, Eysenbach [5] has suggested that some of this positive relationship is likely to be the result of “selection bias”. That is, papers that are more likely to be cited are more likely to be posted. It may be, for example, that authors are more likely to post a paper, the better they think it is. Or it may be that the most able authors tend to be located at high profile institutions that encourage self-archiving. Our regression results in Table 2 confirm that there are strong effects running in this direction. In economics, authors at research institutions and authors publishing in high impact journals are much more likely to self-archive. It is interesting to see that despite the presence of this effect, the Eysenbach study indicates that citation rates are higher for open access articles, even when all are published in the same prestigious journal.⁹

Another useful lesson from this survey is that the best way to search for

⁹This result persisted when Eysenbach used logistic regressions to control for other observable characteristics of articles and authors.

an article when you know its title and/or author seems to be to use plain Google rather than a more specialized search engine. While Google Scholar excludes non-academic articles, it evidently misses some articles stored only on authors' home pages. When one knows the title and/or author of the article, there is no need to exclude non-academic pages. If the paper is available, a simple Google search will almost always place the desired paper at or near the top of the search results. Thus when searching for an article whose title is known, there is no advantage, and a significant disadvantage, to using the more restrictive tools such as Google Scholar, OAIster, SSRN, or RePEc.

Can subscription-based journals survive if self-archiving becomes much more widespread? Even if versions of all published articles could be found on open archives, there is value and convenience in being able to download an official final version of an article directly from the publisher's website. The evidence from physics suggests that reasonably priced subscription-based journals will remain economically viable. Swan and Brown [13] report that about one third of all articles in physics are available in the open access depository arXiv, while in the specialties, condensed matter, astrophysics, and high energy physics, virtually all articles are deposited in ArXiv. Swan and Brown asked the two main non-profit physics publishers, the American Physical Society and the Institute of Physics Publishing, about their experiences over the 14 years of ArXiv's existence. The societies responded that they could not identify loss of subscriptions due to ArXiv. They found that subscription trends for the areas that are most thoroughly represented in ArXiv were no different from those in other areas of physics. The experience from physics suggests that research libraries are not likely to abandon their subscriptions to reasonably priced journals of high quality, even if self-archiving becomes almost universal.

Although ArXiv seems to have had little effect on the major non-profit publishers, available evidence suggests that it has influenced the pattern of pricing across other journals in the discipline. According to data taken from Bergstrom and McAfee's journal pricing database www.journalprices.com, about 90% of the citations to articles in physics come from journals whose

index of value per subscription dollar is no more than 1.5 times the median price of non-profit physics journals. By contrast, in engineering, economics, and political science, the fractions of citations found in journals priced below 1.5 times the median price of non-profits are respectively, 68%, 60% and 57%.

Some economists and librarians [3], [2], [10], [6], [11] have noted with alarm that commercial publishers have taken advantage of libraries' relatively inelastic demand for academic journals and the difficult coordination of new entrants, by setting prices that are much higher than average cost. As more content becomes available in open access archives, publishers are faced with greater availability of close substitutes for their products and library demand for journals is likely to become more price-elastic. The increased price-responsiveness means that profit-maximizing prices will fall. As a result, it can be hoped that commercial publishers will no longer be able to charge subscription prices greatly in excess of average cost. Thus the benefits of self-archiving to the academic community are twofold. There is the direct effect of making a greater portion of the body of research available to scholars everywhere and the secondary effect of reducing the prices charged by publishers who exploit their monopoly power.

References

- [1] Onur Baser and Elda Pema. The return of publications for economics faculty. *Economics Bulletin*, 1(1):1–13, August 2003.
- [2] Carl Bergstrom and Theodore Bergstrom. The costs and benefits of library site licenses to academic journals. *Proceedings of the National Academy of Sciences*, 101(3):897–902, January 2004.
- [3] Theodore Bergstrom. Free labor for costly journals? *Journal of Economic Perspectives*, 15(4):183–198, 2001.
- [4] Arthur Diamond. What is a citation worth? *Journal of Human Resources*, 21(2):200–215, 1986.
- [5] Gunther Eysenbach. Citation advantage of open access articles. *Public Library of Science: Biology*, 4(5), May 2006.

- [6] Kenneth Frazier. The librarian's dilemma: Contemplating the costs of the "Big Deal". *D-Lib Magazine*, 7(3), March 2001.
- [7] Chawki Hajjem, Stevan Harnad, and Yves Gingras. Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Eng. Bull.*, 28(4):39–46, 2005.
- [8] Daniel Hamermesh, George Johnson, and Burton Weisbrod. Scholarship, citations, and salary: Economic rewards in economics. *Southern Economic Journal*, 49(2), 1982.
- [9] Steven Lawrence. Online or invisible? *Nature*, 411(6837):521, 2001.
- [10] Mark McCabe. Journal prices and mergers: A portfolio approach. *American Economic Review*, 92(1):259–269, March 2002.
- [11] Mark McCabe, Aviv Nevo, and Daniel Rubinfeld. Academic journal pricing and the demand of libraries. *American Economic Review*, 95(2):447–452, May 2005.
- [12] Peter Suber. Open access, impact, and demand. *BMJ*, 330:1097–1098, 14 May 2005.
- [13] Alma Swan and Sheridan Brown. Open access self-archiving: an author study. Technical report, Joint Information Systems Committee JISC, Southampton, UK, 2005.
- [14] Jonathan D. Wren. Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, 330:1128, 2005.