



www.sciencemag.org/cgi/content/full/321/5887/395/DC1

Supporting Online Material for
Electronic Publication and the Narrowing of Science and Scholarship

James A. Evans

E-mail: jevans@uchicago.edu

Published 18 July 2008, *Science* **321**, 395 (2008)

DOI: 10.1126/science.1150473

This PDF file includes:

Materials and Methods
Tables S1 to S4
References

Supporting Online Material

Methods

FSO online availability provides a high resolution picture of journal availability over time. For example, linked issues of FSO show that in the second half of 1998, *Annual Review of Ecology & Systematics (ARES)* was not offered online by any major commercial source. The journal first came online in the first half of 1999 through Dialog, with issues going back to January, 1997. ProQuest and H. W. Wilson also began offering the title in the first half of 2000 from 1997. In the second half of 2000, OCLC began to offer *ARES*—the fourth commercial provider, and the journal became available for free through its own website (<http://ecolsys.annualreviews.org>). A year later, Westlaw and FirstSearch offered it, but Dialog (and Westlaw—which had simply ported Dialog’s offerings) stopped offering the most recent year. In the second half of 2002, the journal website pushed online issues back to 1996, and in the first half of 2003, back to November, 1970. In 2002 EBSCO Online—the fifth commercial provider—began offering the journal, and in the second half of 2005, pushed its offerings back through January, 1970. In the first period of 2006, EBSCO and Wilson dropped the most recent three years of the journal.

This *ARES* example shows that the FSO data begins just as library scholars argue the “evolving phase” of online journal availability began—“in the late 1990s,” a phase “marked by the availability of both print and electronic journals, with readings from print journals, electronic journals, and journal alternatives” (8). As such, it is appropriate data with which to test propositions concerning online availability on citation patterns over time.

In order to test the effect of online availability on citation depth, the following regression model was estimated, using OLS and specifying fixed effects:

$$CiteDepth_{it} = \alpha_{it} + t\delta + Pages_{it}\gamma_0 + Cites_{it}\gamma_1 + TitleAge_{it}\gamma_2 + IssueDepth_{it}\beta + v_i + \varepsilon_{it}$$

where $CiteDepth_{it}$ is the average number of years between publication of an article and year t —the year in which citations are measured—for journal or subfield i ; $IssueDepth_{it}$ is a matrix of variables indicating how many years prior to t that one, two or three commercial archives carried journal i (or any journal from subfield i) and for how many years it was available freely; $Pages_{it}$, $Cites_{it}$ and $TitleAge_{it}$ are the average number of pages, cites and age of weighted title words in citing articles; and $v_i + \varepsilon_{it}$ is the residual with v_i representing the journal or subfield-specific component that differs only between and not within journals or subfields.

To test for the effect of online availability on the distinct number of articles cited in journals and research fields, conditional, fixed-effect negative binomial models were used to account for the overdispersion of $ArticlesCited_{it}$ ($S1$, $S2$). Specifically, we estimated the following model for both journals and subfields:

$$\begin{aligned} & \Pr(ArticlesCited_{i1} = y_{i1}, \dots, ArticlesCited_{in_i} = y_{in_i} \mid \mathbf{X}_i, \sum_{t=1}^{n_i} ArticlesCited_{it} = \sum_{t=1}^{n_i} y_{it}) \\ &= \frac{\Gamma(\sum_{t=1}^{n_i} \lambda_{it}) \Gamma(\sum_{t=1}^{n_i} y_{it} + 1)}{\Gamma(\sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it})} \prod_{t=1}^{n_i} \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it}) \Gamma(y_{it} + 1)} \end{aligned}$$

where $ArticlesCited_{it}$ is the number of articles cited in the twenty years prior to citing year

t from journal or subfield i , $\sum_{t=1}^{n_i} y_{it}$ is the sum of these counts for i , Γ is the gamma

function, $\lambda_{it} = e^{(\mathbf{x}_{it}\beta + offset_{it})}$ and $\mathbf{x}_{it}\beta = \alpha_{it} + t\delta + Pages_{it}\gamma_0 + Cites_{it}\gamma_1 + IssueDepth_{it}\beta$.

Negative binomial models were compared with Poisson models to assess whether the dispersion parameter α was significantly greater than 0—the special case in which the negative binomial is equivalent with the Poisson model. The substantial difference (e.g.,

$\chi^2 = 3.9 \times 10^5$, $p < .001$ for the journal-level model predicted by any online provision)

suggests that a Poisson model fits the data poorly and the negative binomial specification is appropriate. Negative binomial models with varying dispersion were also tested against ones with pooled estimates and found significantly different (e.g., $\bar{\chi}^2 = 3.7 \times 10^5$, $p < .001$ for the journal-level model with any online provision).

In order to capture the degree to which citations are concentrated within particular cited articles, we estimate the following model:

$$\begin{aligned} \text{ArticleCiteConcentration}_{it} = \\ \alpha_{it} + t\delta + \text{Pages}_{it}\gamma_0 + \text{Cites}_{it}\gamma_1 + \text{IssueDepth}_{it}\beta + v_i + \varepsilon_{it} \end{aligned} \quad (3)$$

where *ArticleCiteConcentration_{it}* is a Herfindahl index capturing the concentration of citations within particular cited articles in the twenty years prior to citing year *t* from journal or subfield *i*.

Tables S1-S4

Table S1 lists descriptive statistics for the variables used in each model of the study. The mean of Citation Depth is 5.6 years for journals and subfields, suggesting that across fields, published work cites articles an average of 5.6 years old. An average of 632 and 25,241 articles are cited from each journal and each subfield each year, with maxima of 2,159 and 35,847, respectively. Not surprisingly, the distribution of citations across journals is more uneven than that across subfields. The mean of article citation in journals and subfields is .088 and .008, both with a much larger standard deviation. Citation concentration to journals within subfields is much higher (.337) and more stable. Citing articles contain are, on average, 12.5 pages in length with 42 references. Citing articles also use overwhelmingly conventional title words—words that have long

appeared in the scientific literature—with an average weighted age of 19 years using a 20 year window. Other independent variables in the analysis underscore the patterns highlighted by Figure 1 in the paper—many more journals are available online through commercial sources than freely, and commercially sources do not all pick up journals at the same time. One will get it, followed gradually by others.

To detail the pattern of online influence across subjects, the 311 fine-grained CI subject codes were consolidated into 14 broad areas: 1) Interdisciplinary Sciences, 2) Biology, 3) Medicine, 4) Environmental & Earth Sciences, 5) Agriculture, 6) Chemistry, 7) Physical Sciences, 8) Engineering, 9) Mathematics, 10) Computer Science & Information Technology, 11) Social Sciences, 12) Business & Management, 13) Law, and 14) Humanities. Tables S2 through S4 illustrate how consistent the effect of years of online availability are on citation depth, distinct articles cited, and article citation concentration across the scientific and scholarly subfields. Journal-level analyses estimates are presented, which are consistent with those at the level of detailed research subfield. All models account for citation year, and are consistent with model presented in the paper which account for age of title words and page and reference numbers from citing articles.

Table S2 shows that the OLS coefficient estimate for any online availability is negative in every subfield and statistically significant ($p < .01$) in all but the computer science & information technology subfield. Commercial and free availability exhibit slightly greater variation, but underlie the consistency of the pattern. R^2 for the models range from .493 for engineering (739 journals) to .722 for business & management (477 journals). ML incidence rate ratio estimates from negative binomial models predicting

the number of distinct article citations in journals with years of online availability are listed in Table S3. These rate ratios all indicate a negative relationship (an incidence ratio less than one) between online availability and distinct articles cited. The largest effect appears in the fields of biology, medicine and agriculture whose incidence ratio estimates go down to .884 (agriculture), indicating a nearly 12% drop (e.g., $(.88-1.0)*100$) in the number of distinct articles cited for each additional year of online availability. In every field, the influence of free electronic availability is stronger than that of commercial access, going down to .815 in the area of medicine.

Table S4 lists estimates for the journal-level models of citation concentration within articles. These effects reveal a wider variance across subfields. In agriculture, there is no effect and in mathematics and engineering, years of free availability appear to exert a significant negative ($.002, p<.05$) influence on citation concentration. In the remaining 11 fields, however, the effect of years online is associated with a rise in the article citation concentration. R^2 fit statistics for the models range from .042 (law) to .130 (social science), suggesting poorer overall fit than in the citation depth models. Collectively these tables show the general consistency of these trends across areas of science and scholarship, and their greatest prevalence in the life sciences.

Tables

Table S1. Variables in Analysis

Variable	Mean	St.dev.	Min	Max
<i>Journal Analysis</i>				
Citation Depth ^a	5.633	3.380	0	20
Articles Cited ^b	632.125	2159.431	0	59891
Article Citation Concentration ^c	0.088	0.195	.00009	1
Citation year ^a	1991.905	10.246	1966	2006
Years online	1.836	5.217	0	61
Years online in one pay source	1.738	4.974	0	61
Years online in two pay sources	.849	2.741	0	36
Years online in three pay sources	.528	2.199	0	36
Years online free	.529	2.503	0	61
Avg. pages in citing articles	12.543	9.405	0	147
Avg. references in citing articles	42.162	19.990	1	236
Avg. title age in citing articles	19.046	1.253	0	20
<i>Subfield Analysis</i>				
Citation Depth ^d	5.622	2.393	0	20
Articles Cited ^e	25241.32	35847.93	0	452800
Journals Cited ^f	22.246	21.542	0	160
Article Citation Concentration ^g	.008	.0536	.00002	1
Journal Citation Concentration ^h	.337	.254	.021	1
Citation year ^d	1986.96	11.488	1966	2005
Years online	3.715	9.656	0	61
Years online in one pay source	3.478	8.960	0	61
Years online in two pay sources	1.928	4.705	0	36
Years online in three pay sources	1.590	4.194	0	36
Years online free	1.678	6.395	0	61

^a8250 Journals, 171715 Journal-Cite Years; ^b5515 Journals, 119901 Journal-Cite Years

^c8253 Journals, 195225 Journal-Cite Years; ^d311 Subfields, 10718 Subfield-Cite Years

^e306 Subfields, 10887 Subfield-Cite Years; ^f298 Subfields; 8567 Subfield-Cite Years

^g308 Subfields, 10500 Subfield-Cite Years; ^h308 Subfields, 10308 Subfield-Cite Years

Table S2. OLS Coefficient Estimates for Models of Average Citation Depth in Journals:
by Broad Subject

Variables	Years online	Years online, pay	Years online, free	R ²
Multidisciplinary Sciences: 82 journals; 2123 citing years				
Model 1	-0.026 (0.007)**			0.537
2		-0.014 (0.008) [†]		0.535
3			-0.058 (0.011)**	0.540
4		-0.006 (0.008)	-0.056 (0.011)**	0.540
Biology: 1432 journals; 31187 citing years				
	-0.009 (0.002)**			0.526
		-0.001 (0.002)		0.525
			-0.029 (0.004)**	0.526
		0.004 (0.002) [†]	-0.031 (0.004)**	0.526
Medicine: 2008 journals; 42577 citing years				
	-0.018 (0.002)**			0.518
		-0.015 (0.002)**		0.517
			-0.028 (0.004)**	0.517
		-0.012 (0.002)**	-0.021 (0.004)**	0.518
Environmental & Earth Sciences: 533 journals; 12841 citing years				
	-0.026 (0.003)**			0.572
		-0.024 (0.003)**		0.572
			-0.055 (0.007)**	0.572
		-0.020 (0.004)**	-0.047 (0.007)**	0.573
Agriculture: 189 journals; 4604 citing years				
	-0.024 (0.007)**			0.549
		-0.013 (0.008) [†]		0.548
			-0.047 (0.012)**	0.549
		-0.005 (0.008)	-0.045 (0.012)**	0.549
Chemistry: 393 journals; 8924 citing years				
	-0.026 (0.004)**			0.453
		-0.022 (0.004)**		0.452
			-0.039 (0.008)**	0.451
		-0.020 (0.004)**	-0.033 (0.008)**	0.453
Physical Sciences: 570 journals; 12841 citing years				
	-0.023 (0.003)**			0.515
		-0.015 (0.003)**		0.514
			-0.044 (0.006)**	0.515
		-0.013 (0.003)**	-0.042 (0.006)**	0.516
Engineering: 739 journals; 17003 citing years				
	-0.014 (0.003)**			0.493
		-0.013 (0.003)**		0.493
			-0.043 (0.008)**	0.493
		-0.011 (0.003)**	-0.038 (0.008)**	0.494
Mathematics: 333 journals; 7397 citing years				

	-0.027 (0.004)**				0.598
		-0.025 (0.004)**			0.597
			-0.041 (0.009)**		0.597
		-0.022 (0.005)**	-0.033 (0.009)**		0.598
Computer Science & Information Technology: 416 journals; 8389 citing years					
	-0.001 (0.004)				0.610
		0.010 (0.004)*			0.611
			-0.042 (0.007)**		0.612
		0.015 (0.004)**	-0.046 (0.007)**		0.613
Social Sciences: 1314 journals; 31064 citing years					
	-0.052 (0.002)**				0.666
		-0.052 (0.002)**			0.666
			-0.075 (0.005)**		0.663
		-0.045 (0.002)**	-0.044 (0.005)**		0.667
Business & Management: 477 journals; 10204 citing years					
	-0.041 (0.003)**				0.721
		-0.042 (0.003)**			0.722
			-0.025 (0.007)**		0.715
		-0.043 (0.003)**	0.006 (0.007)		0.722
Law: 155 journals; 4336 citing years					
	-0.063 (0.005)**				0.633
		-0.064 (0.005)**			0.633
			-0.071 (0.017)**		0.620
		-0.062 (0.005)**	-0.043 (0.016)**		0.633
Humanities: 586 journals; 14534 citing years					
	-0.119 (0.004)**				0.678
		-0.122 (0.004)**			0.677
			-0.127 (0.008)**		0.664
		-0.114 (0.004)**	-0.086 (0.008)**		0.680

* Coefficient for citation year from .184 to .34; Constant from -671.544 to -360.875; all $p < .001$.

Standard errors in parentheses

† significant at 10%; * significant at 5%; ** significant at 1%

Table S3. ML Estimated Incidence Ratios for Models of Distinct Articles Cited in Journals:
by Coarse Subject

Variables	Years Online	Years Online, Pay	Years online, free
Multidisciplinary Sciences: 54 journals; 1516 citing years			
Model 1	.920 (0.006)**		
2		.913 (0.007)**	
3			.890 (0.011)**
4		.942 (0.006)**	.918 (0.011)**
Biology: 908 journals; 21056 citing years			
	.886 (0.002)**		
		.881 (0.002)**	
			.836 (0.003)**
		.935 (0.002)**	.873 (0.004)**
Medicine: 1386 journals; 31001 citing years			
	.893 (0.002)**		
		.845 (0.002)**	
			.815 (0.003)**
		.947 (0.002)**	.846 (0.003)**
Environmental & Earth Sciences: 304 journals; 7829 citing years			
	.907 (0.003)**		
		.905 (0.003)**	
			.851 (0.006)**
		.943 (0.003)**	.879 (0.006)**
Agriculture: 114 journals; 2927 citing years			
	.884 (0.006)**		
		.880 (0.006)**	
			.831 (0.009)**
		.946 (0.006)**	.862 (0.010)**
Chemistry: 215 journals; 4967 citing years			
	.920 (0.005)**		
		.910 (0.005)**	
			.843 (0.008)**
		.968 (0.004)**	.863 (0.008)**
Physical Sciences: 318 journals; 7349 citing years			
	.911 (0.003)**		
		.919 (0.004)**	
			.855 (0.006)**
		.954 (0.003)**	.874 (0.006)**
Engineering: 439 journals; 10903 citing years			
	.929 (0.002)**		
		.935 (0.003)**	
			.852 (0.005)**
		.965 (0.002)**	.868 (0.005)**
Mathematics: 229 journals; 5337 citing years			

	.908 (0.004)**		
		.923 (0.004)**	
			.836 (0.007)**
		.962 (0.003)**	.850 (0.007)**
Computer Science & Information Technology: 272 journals; 5620 citing years			
	.937 (0.003)**		
		.938 (0.003)**	
			.887 (0.005)**
Social Sciences: 1051 journals; 26412 citing years		.965 (0.003)**	.902 (0.005)**
	.924 (0.001)**		
		.926 (0.001)**	
			.826 (0.003)**
Business & Management: 386 journals; 8818 citing years		.952 (0.001)**	.847 (0.003)**
	.956 (0.002)**		
		.957 (0.002)**	
			.856 (0.005)**
Law: 141 journals; 4112 citing years		.972 (0.001)**	.868 (0.005)**
	.951 (0.002)**		
		.950 (0.002)**	
			.824 (0.013)**
Humanities: 491 journals; 13308 citing years		.955 (0.002)**	.839 (0.012)**
	.902 (0.002)**		
		.899 (0.002)**	
			.833 (0.006)**
		.919 (0.002)**	.864 (0.005)**

Citeyear coefficient estimates ranged from .023 to .052; Constant estimates ranged from -102.7 to -44.424, all $p < .001$.

Table S4. OLS Coefficient Estimates for Models of Citation Concentration in Journals:
by Broad Subject

Variables	Years Online	Years Online, Pay	Years online, free	R ²
Multidisciplinary Sciences: 54 journals; 1393 citing years				
Model 1	0.003 (0.001)**			0.070
2		0.003 (0.001)**		0.072
3			0.001 (0.001)	0.061
4		0.003 (0.001)**	0.000 (0.001)	0.072
Biology: 908 journals; 19772 citing years				
	0.001 (0.000)**			0.075
		0.001 (0.000)**		0.075
			0.001 (0.000)*	0.075
		0.001 (0.000)**	0.000 (0.000)	0.075
Medicine: 1385 journals; 29001 citing years				
	0.001 (0.000)**			0.077
		0.001 (0.000)**		0.077
			0.001 (0.000)*	0.076
		0.001 (0.000)**	0.000 (0.000)	0.077
Environmental & Earth Sciences: 304 journals; 7206 citing years				
	0.002 (0.000)**			0.093
		0.002 (0.000)**		0.092
			0.001 (0.001)*	0.090
		0.002 (0.001)**	0.000 (0.001)	0.092
Agriculture: 114 journals; 2751 citing years				
	0.001 (0.001)			0.053
		0.001 (0.001)		0.053
			0.000 (0.001)	0.053
		0.001 (0.001)	-0.001 (0.001)	0.053
Chemistry: 215 journals; 4579 citing years				
	0.001 (0.001)*			0.058
		0.000 (0.001)		0.057
			0.001 (0.001)	0.058
		0.000 (0.001)	0.001 (0.001)	0.058
Physical Sciences: 318 journals; 6741 citing years				
	0.001 (0.000)**			0.050
		0.001 (0.001)*		0.049
			0.000 (0.001)	0.049
		0.001 (0.001)*	0.000 (0.001)	0.049
Engineering: 439 journals; 9551 citing years				
	0.000 (0.001)			0.084
		0.000 (0.001)		0.084
			-0.002 (0.001)*	0.084
		0.001 (0.001)	-0.002 (0.001)*	0.084
Mathematics: 229 journals; 4971 citing years				

	0.000 (0.000)				0.083
		0.000 (0.001)			0.083
			-0.002 (0.001)*		0.083
			-0.002 (0.001)*		0.084
Computer Science & Information Technology: 272 journals; 5059 citing years	0.003 (0.001)**				0.056
		0.003 (0.001)**			0.056
			-0.002 (0.001)†		0.053
		0.004 (0.001)**	-0.003 (0.001)**		0.058
Social Sciences: 1051 journals; 24600 citing years	0.003 (0.000)**				0.130
		0.003 (0.000)**			0.130
			0.001 (0.000)*		0.120
		0.004 (0.000)**	-0.001 (0.000)**		0.130
Business & Management: 386 journals; 8040 citing years	0.002 (0.000)**				0.093
		0.002 (0.000)**			0.093
			-0.001 (0.001)		0.087
		0.002 (0.000)**	-0.003 (0.001)**		0.094
Law: 141 journals; 3861 citing years	0.003 (0.000)**				0.059
		0.003 (0.000)**			0.060
			-0.001 (0.001)		0.042
		0.003 (0.000)**	-0.002 (0.001)†		0.060
Humanities: 490 journals; 12114 citing years	0.006 (0.000)**				0.137
		0.006 (0.000)**			0.137
			0.002 (0.001)**		0.122
		0.006 (0.000)**	0 (0.001)		0.137

Coefficient estimate for citation year ranged from -.009 to -.003; Constant estimates ranged from 5.08 to 18.061, all $p < .001$

Standard errors in parentheses

† significant at 10%; * significant at 5%; ** significant at 1%

Additional References

S1. J. Hausman, B. H. Hall, Z. Griliches. *Econometrica* **52**, 909-938 (1984).

S2. B. H. Hall, Z. Griliches, J. A. Hausman. *International Economic Review* **27**, 265-284 (1986).