

Agricultural & Applied Economics Association

Citations and Individuals: First Authorship across the Alphabet

Author(s): Bruce A. McCarl

Source: *Review of Agricultural Economics*, Vol. 15, No. 2 (May, 1993), pp. 307-312

Published by: Oxford University Press on behalf of Agricultural & Applied Economics Association

Stable URL: <https://www.jstor.org/stable/1349450>

Accessed: 03-05-2021 17:27 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Agricultural & Applied Economics Association, Oxford University Press are collaborating with JSTOR to digitize, preserve and extend access to *Review of Agricultural Economics*

CITATIONS AND INDIVIDUALS: FIRST AUTHORSHIP ACROSS THE ALPHABET

Bruce A. McCarl

The presence of a bias in citation counts drawn from sources with first-author citations only are investigated. The study finds systematic bias against authors whose last names begin with letters that occur later in the alphabet, with authors receiving approximately one-half percent less first author citations of the total pieces per letter the later their names are in the alphabet.

Given these biases, counts should certainly be used with care. Perhaps there are ways of improving them by either maintaining professionally-based citation counts or by lobbying sources, such as the Social Science Citation Index, to include all authors in their counts.

In recent years, citations have become one measure of the worth of professional publications, individuals, and departments. In particular: (1) the criteria for nomination for the American Agricultural Economics Association Publication of Enduring Quality award encourages citation counts; (2) the faculty promotion process frequently generates evaluation letters and/or candidate dossiers that contain citation count information; (3) authors such as Beilock and Polopolus and earlier Beilock, Polopolus, and Correal have used citation counts in generating information on department rankings; and (4) authors such as Medoff have used citation counts to create ranked lists of professionals. There have also been mailings to department heads of citation counts for their faculty and former students.¹

Most of the citation count data supporting the efforts mentioned above are based on the Social Sciences Citation Index wherein only the first author's name is associated with the citation (Beilock and Polopolus maintained a system

with all authors; however, they have discontinued this effort). Furthermore, the Social Sciences Citation Index regularly publishes such citations which is, and will in all likelihood be, the dominant source of citations data as entry of all citations is a very laborious task.²

The utilization of citation counts restricted to first authors raises questions of potential bias. The hypothesis is raised and tested here that first-author citations are biased against those whose last names appear later in the alphabet.³ Two anecdotal pieces of evidence give rise to this hypothesis. First, the reviewers of Medoff's paper noted his top 100 list was biased toward names early in the alphabet (although Medoff argues there is not systematic bias). Second, it appears that many Agricultural Economists deviate from alphabetic ordering only when there is a great difference in contribution. Obviously, under the latter circumstances, first-author citations could cause an alphabetically-based bias.⁴

Data Set

This study examines the above hypothesis using data on all citations from the American, North Central,⁵ Northeastern, Southern, and Western Journals of Agricultural Economics for the years 1988-1990. Information on 15,225 citations was entered into a data base program.

²The SSCI enters over 1,500 journals including two agricultural economics journals (*AJAE*, *WJAE*) and is predominantly intended as a bibliographic source.

³In the context of departmental rankings, Beilock and Polopolus show that reliance on single author citations changes rankings.

⁴Many other biases are potentially present in these data as a measure of the "worth" of an individual. For example, the target audience who would consider worth as largely reflected by agricultural economics literature citations is limited and not totally appropriate for a large number of agricultural economists.

⁵The *North Central Journal of Agricultural Economics* was subsequently renamed *Review of Agricultural Economics*.

B.A. McCarl is a Professor in the Agricultural Economics Department, Texas A&M University.

Technical paper TA31058 of the Texas Agricultural Experiment Station.

¹The mailing is done by Dale Adams at Ohio State who maintains a citation data base.

Of these references, 1,432 did not contain author attribution (being associated with the USDA, the World Bank, newspapers, etc.), 6,843 were associated with single authors, 4,516 with two authors, 1,711 with three, 452 with four, 138 with five, 35 with six, 24 with seven, five with more than seven authors, and finally, 45 articles were cited using *et al.* Thus, the average was 1.74 authors per piece. These data involve citations to the work of 8,043 individuals with 23,984 total authors mentioned including both agricultural economists and others. The distribution of first-author citations to individuals ranged from a maximum of 77 cites to a minimum of 0. Approximately 35.5 percent of the people in the sample had 0 first-author citations since they only appeared as second or later author in the sample data. No corrections were made for self citations. The distribution of the total number of cites to any author, regardless of whether they were first or later, ranged from 114 to 1.⁶ A summary of the data by number of citations to both first and any author appears in Table 1.

Analysis of the Alphabetical Question

Table 2 gives summary data on how the number of authors, the number of first-author citations, the number of total citations, and the percentage of first-author citations relative to total citations varies across the alphabetic position of the first letter of an author's last name (the forecast column will be discussed later). These data show that bias is potentially present. Note that the percentage data seems to show that people whose last names begin with letters in the first part of the alphabet have first-author citations more frequently than those whose last names begin with letters in the latter half of the alphabet. For example, below the letter "I" there are no alphabetic groups with more than 60 percent of their total citations being first-author citations; whereas, above "I," 60 percent occurs in five of the nine cases. Similarly, frequencies below 50 percent only occur in the latter part of the alphabet.

Regression was utilized to see if there was statistical evidence of interdependence between

name and the proportion of first-author citations to total citations. The number of first-author citations by an individual was regressed on both the number of total citations and the number of total citations times a variable (hereafter called NAME) indicating the alphabetic position of the person's last name. The NAME variable took on the value of zero for someone whose last name began with an "A," one for "B," two for "C," etc. down to 25 for "Z." This regression was done using both ordinary least squares and Tobit estimation (to account for the censoring at zero since 35.5 percent of the authors in the data set had zero first-author citations). The parameter on the name variable estimates the proportion of first-authored citations for one whose name begins with an "A." The parameter on the NAME*TOTAL CITES variable gives the reduction in proportion of first-author cites per letter later in the alphabet where the name falls.

The regression results (Table 3) show little difference between the estimation methods and reveal the equation to explain about 76 percent of the variation while containing highly significant coefficients. These results not only show that on average someone whose name begins with "A" will have between 60 and 62 percent of their total citations as first-author citations, but show that this percentage falls between 0.52 percent and 0.58 percent the later the author's last name falls in the alphabet. An evaluation of the percentage of total citations expected to be first-author citations as forecast by the OLS equation appears as the last column in Table 2. These projections show, for example, that someone whose last name begins with "Z" has an 11.9 percent lower chance of having one of their citations be a first-author citation than does someone whose last name begins with "A." This would seem to indicate that all other things being equal, there are biases against people whose last names begin with letters in the latter part of the alphabet when using first-author citation counts only.

Why the Bias?

The results discussed in the above section may not only be due to alphabetization, but they could also reflect other forces. Two

⁶No attempt was made to remove auto citations.

Table 1. Frequency of Citations by Author Position

Number of Citations to an Author's Work	Number of Authors with This Number of Citations when Citations Consider:	
	First Author Only	Any Listed Author
0	2,853	0
1	3,153	4,797
2	825	1,261
3-5	718	1,116
6-10	292	480
11-15	93	167
16-20	36	72
21-25	26	40
26-30	18	29
31-35	8	19
36-40	6	23
41-45	1	6
46-50	6	5
51-55	5	7
56-60	2	4
61-65	0	1
66-70	0	2
71-75	0	0
76-80	1	4
81-85	0	2
86-90	0	0
91-95	0	2
96-100	0	1
101-105	0	1
106-110	0	0
111-115	0	1
TOTAL	8,043	8,043

Table 2. Summary Citation Data by First Letter of Last Name

First Letter of Last Name	Number of Authors Cited	Number of First Author Citations	Number of Any Listed Author Citations	Percentage First Over Any Listed	Regression Forecast
A	263	550	778	70.7%	61.9%
B	752	1,596	2,481	64.3%	61.4%
C	511	952	1,549	61.5%	60.9%
D	382	608	1,065	57.1%	60.4%
E	160	256	446	57.4%	59.8%
F	298	523	833	62.8%	59.3%
G	424	702	1,238	56.7%	58.8%
H	638	1,105	1,985	55.7%	58.3%
I	45	60	92	65.2%	57.8%
J	187	445	768	57.9%	57.3%
K	389	675	1,128	59.8%	55.7%
L	407	751	1,376	54.6%	56.2%
M	664	1,055	1,946	54.2%	55.7%
N	169	267	449	59.5%	55.2%
O	122	146	265	55.1%	54.7%
P	370	633	1,100	57.5%	54.2%
Q	18	21	44	47.7%	53.6%
R	423	675	1,279	52.8%	53.1%
S	854	1,343	2,522	53.3%	52.6%
T	279	482	837	57.6%	52.1%
U	25	28	66	42.4%	51.6%
V	117	148	263	56.3%	51.1%
W	442	604	1,113	54.3%	50.6%
X	0	0	0	--	50.0%
Y	46	71	155	45.8%	49.5%
Z	58	97	206	47.1%	49.0%
TOTAL	8,043	13,793	23,984	57.5%	--

Table 3. Regression Relating Number of First Citations to Total Citations and the NAME Variable Times Total Citations

	OLS	Tobit
Total Cities	0.618 (111.15) ^a	0.605 (62.13)
Total * NAME	-0.00508 (-11.02)	-0.00577 (-11.433)
R ²	0.761	0.756

^at and asymptotic t ratios are given in parentheses.

investigations were undertaken to examine alternate explanations. The first involved productivity as it is associated with the NAME variable. Simply put, there could be systematic differences in the total amount of citations across people whose last names begin with letters occurring in the latter part of the alphabet, i.e., they may not be as highly cited as those whose names begin with letters occurring in the earlier part of the alphabet. However, regression of total citations on a constant and the last name variable basically showed no effect explaining less than 0.03 percent of the total variation in the data with an insignificant coefficient on the NAME variable.

Subsequently, the presence of authorship alphabetization was examined by looking at the relative frequency of alphabetical ordering. This was done by counting the frequency with which data appeared in alphabetical order for two and three authored pieces. Namely, if a random, uniform process determined authorship order, one would expect alphabetic ordering for one-half of the cited pieces in the two author cases and one-sixth of the cited pieces in the three author cases (since there are six permutations of three names). Data pertinent to these cases are given in Table 4. There were 4,516 citations with two authors, and analysis showed that about 63 percent were in alphabetical order. The binomial probability of this percentage or more alphabetic occurrences out of a sample of 4,520 is very close to zero. Simultaneously, the three author case shows that about 32 percent of the data were in alphabetical order while 68 percent were not. Here alphabetization appears about twice as frequently as a naive forecast would predict based on a uniform distribution of the six possible author orderings. Again, the probability of this percentage or more alphabetic draws is basically zero if authorship is uniformly distributed. These data contain strong evidence that systematic alphabetization is occurring biasing the first-author citation count against those whose names occur later in the alphabet.

Citation Counts and Individuals

Another way to look at the consequences of the above biases is to examine how they affect

individual rankings. For example, consider a comparison of citation rankings by total and first citation count. Five of those with the 20 most total citations regardless of authorship placement are not in the top 20 of the first-author citation list. One of those individuals has one first-author citation which places him/her below the top 3,000 people in total citations while two individuals have two first-author citations placing them basically below the top 2,000. However, when considering total citations, these individuals' counts rise to 68, 58, and 58 in total placing them in thirteenth, sixteenth, and seventeenth places out of the 8,043 cited authors. Yet another individual rises from 31 first-author citations up to 101, becoming the second most cited person.

Conclusions

Citation data shows the presence of a systematic bias against people with names beginning with letters that occur later in the alphabet when citation rankings are based on the first author only counting procedure as supported by the Social Science Citations Index. This bias should be recognized and citation counts based only on first-author citations should not be used to rank individual productivity. Certainly when authors names appear on a piece, it is a reflection of their contribution and should not be ignored. The question then is, should we try to do better? Perhaps Agricultural Economics Associations should either maintain their own lists (which is becoming more feasible as scanners become cheaper, but is still very expensive), or should lobby the Social Sciences Citations Index suppliers to include all authors in their citation database.

The bias also has implications for other uses such as departmental rankings. For example, if departmental composition is biased toward people whose names begin with letters that occur later in the alphabet, then there would be ranking bias (Beilock and Polopolus indicate ranking biases exist in departmental rankings due to first versus multiple author citation counts, but do not explore the causes of this bias). McCarl shows that such bias is present in a companion paper to this one.

Table 4. Occurrence of Permutations of Alphabetic Ordering

Name Ordering Sequences	Number of Citations	
	Observed Count	Naively Predicted
Citations Authored by Two Authors by Name Ordering Sequence		
Alphabetic Order Maintained	2,868	2,258
Alphabetic Order Violated	<u>1,648</u>	2,258
TOTAL	4,516	

Author Ordering Sequence ^a	Number of Citations	
	Observed Count	Naively Predicted
Citations Authored by Three Authors by Name Ordering Sequence		
abc	557	285
acb	224	285
bac	266	285
bca	220	285
cab	232	285
cba	<u>212</u>	285
TOTAL	1,711	

^aIn the table, "a" stands for the first author in alphabetical order, "b" the second, and "c" the third. Potential orderings of the author's names then range from abc which is strictly alphabetical to cba which is strictly non-alphabetical.

References

- Beilock, R.P. and L.C. Polopolus. "Ranking of Agricultural Economics Departments: Influence of Regional Journals, Joint Authorship, and Self-Citations." *American Journal of Agricultural Economics* 70(1988):403-9.
- Beilock, R.P., L.C. Polopolus, and M. Correal. "Ranking of Agricultural Economics Departments by Citations." *American Journal of Agricultural Economics* 68(1986):595-604.
- McCarl, B.A. "Economics Journal Citations and Individuals: Bias When Relying on First Authorship." Unpublished paper, Texas A&M University, 1992.
- Medoff, M.H. "The Ranking of Economists." *Journal of Economic Education* 20(1989):405-15.