PA

Are Papers Written by Women Authors Cited Less Frequently?

Justin Esarey[1] and Kristin Bryant[2]

[1] Department of Politics and International Affairs, Wake Forest University, Winston-Salem, MC 27109, USA.
Email: justin@justinesarey.com
[2] Department of Political Science, Rice University, Houston, TX 77005, USA. Email: kab18@rice.edu

Abstract

Dion, Sumner, and Mitchell (2018) find that a published article is more likely to cite at least one female-authored paper if that article is itself authored by women. To complement their work, we study the number of times that an article in their data set is cited given that it has at least one female author. We find that articles with at least one female author are cited no more or less often than male-authored articles once we control for the publishing journal and the number of authors. The importance of controlling for author count in our model suggests that spurious correlation and/or self-citation might explain at least some of the gender differences found by Dion, Sumner, and Mitchell (2018).

Keywords: linear regression, observational studies, bibliometrics

Dion, Sumner, and Mitchell (2018) demonstrate that a published article is more likely to cite a female-authored paper if that article is itself authored by women.[1] They also show that articles in *Political Analysis* are substantially less likely to cite female-authored papers when compared to articles in the discipline's pre-eminent general interest journal, *American Political Science Review*.[2] This difference is not peculiar to political science as a discipline: articles in *Econometrica* are also less likely than articles in *American Political Science Review* to cite a female-authored paper.[3]

Some of the potential explanations for these results are extremely concerning from the perspective of achieving gender equality in the political methodology community. For example, it may be that women are not as well-networked in the profession as male counterparts, and thus have greater difficulty gaining attention for their work. Or, perhaps authors simply take the work of women authors less seriously. These possibilities represent what Dion, Sumner, and Mitchell (2018, p. 3) call the "Matilda effect." Either possibility is worrisome.

To complement the findings of Dion, Sumner, and Mitchell (2018), we study the difference in citation rates between male- and female-authored papers. That is, we are interested in the following:

a. (citation count|at least one female author) − (citation count|only male author(s)), and
b. Pr(paper cited|at least one female author) − Pr(paper cited|only male author(s)).

By contrast, Dion, Sumner, and Mitchell (2018) study the probability that an article's reference list contains at least one paper authored only by women as a function of the gender composition

1 This is the result in Table 3 of Dion, Sumner, and Mitchell (2018). In Appendix Table A8, they also show that a published article is more likely to cite a paper with at least one female author (including mixed-gender author teams) if that article is itself authored by at least one woman.

2 This is the result in Table 3; Appendix Table A8 shows that articles in *Political Analysis* are also less likely than articles in *American Political Science Review* to cite papers with at least one female author (including mixed-gender author teams).

3 *Econometrica* articles are also less likely than those in *American Political Science Review* to cite a paper with at least one woman author (including mixed-gender author teams).

of the citing article's authors.[4] We argue that our dependent variables speak more directly to diversity in the methods community because citation counts are an increasingly important metric for assessing scholarly impact, for example, when evaluating a tenure case (Fowler and Aksnes 2007; Fischer, Ritchie, and Hanspach 2012; Cameron, White, and Gray 2016; Marini 2017). If women systematically receive fewer citations to their papers than men, this could directly harm these women's career prospects and pose a barrier to diversifying the methods field. For the same reason, we choose to code our key independent variable ("female author") such that both women-only author teams and mixed-gender author teams are classified as 1, while male-only author teams are classified as 0. Our choice focuses on citations to all female authors, not just those who write alone or with other women, and increases the proportion of "treated" cases in the data set (see Table 1 in Dion, Sumner, and Mitchell (2018)).

We collected citation data from Crossref (www.crossref.org) for all the publications in the Dion, Sumner, and Mitchell (2018) data set using the `rcrossref` package of Chamberlain *et al.* (2017).[5] The resulting citation count is our dependent variable. We were able to collect citation data for 1907 of the 1939 publications in the data set.[6] Papers in this data set had been cited between 0 and 984 times, with a mean of 28.92 citations per paper. As reported by Dion, Sumner, and Mitchell (2018), the data set contains all articles published between the years 2007 and 2016 in five prominent journals, including three methodological journals (*Political Analysis*, *Econometrica*, and *Sociological Methods and Research*), a general interest political science journal (*American Political Science Review*), and a political science journal focusing on gender (*Politics and Gender*).

A regression analysis of the citation data is shown in Table 1. Models 1 and 2 use ordinary least squares (OLS) regression to predict the number of citations using whether at least one author is a female, the publishing journal, and the number of authors (model 2 only).[7] Model 3 uses OLS to predict whether the paper has been cited more than five times (1 if so, 0 if not);[8] this model tries to directly assess Pr(paper cited) using five citations as a cutoff for "not being cited."

The combined evidence of Table 1 indicates that female authors are *not* at a systematic disadvantage when it comes to being cited, holding the publishing journal and the number of authors constant. Model 1 in Table 1 seems to indicate that articles with female authors receive about five *more* citations on average than those with only male authors, ceteris paribus. This relationship is sensible: for nearly any fixed probability that a woman is a paper author, a greater number of co-authors implies a higher probability that at least one author is female.[9] At the same time, previous research has found that the number of co-authors on a paper is associated with a greater number of citations to that paper—a pattern that holds across multiple fields (Wuchty, Jones, and Uzzi 2007; Franceschet and Costantini 2010; Abramo and D'Angelo 2015). This increase in citations may be due in part to the fact that as the number of authors increases, so does the network of scholars who might know and cite those authors (Leimu and Koricheva 2005, p. 30). We thus have reason to suspect spurious correlation between our "female author" variable and the number of citations; the potential back-door pathway leads through the number of authors of the paper. Once author count is included as a variable in the regression (in Model 2), this relationship becomes statistically indistinguishable from zero and substantively small. The findings of Model 3 (using a binary dependent variable for citations) are substantively similar to those of Model 2.

---

4   Dion, Sumner, and Mitchell (2018) also include mixed-gender teams in the coding of their dependent variable, but only in a supplementary analysis; see Appendix Table A8 in their paper.
5   The citation data was collected on April 5, 2018, and citation counts therefore are current as of this date.
6   One article was missing author gender information, and therefore the total sample size for our analysis is 1906.
7   The number of authors varies between 1 and 5; articles with more than five authors are classified as having five authors. We also only use gender information from the first five authors in an article. These choices are consistent with the coding in the original data set, which only codes the gender of the first five authors.
8   612 articles in the data set have been cited five times or fewer.
9   If the probability of any author being a woman is p, then the probability that at least one author on a paper is a man is $1 - (1 - p)^n$, where n is the number of authors. This quantity is strictly increasing in n for $p \in (0, 1)$.

Table 1. Regression analysis of citation counts for published papers in the Dion, Sumner, and Mitchell (2018) data set.

	Dependent Variable:		
	Citation Count		Citations > 5
	(1)	(2)	(3)
Female author	5.370	1.313	0.014
	(2.958)	(2.990)	(0.022)
Author count		9.049	0.067
		(1.376)	(0.010)
APSR	69.692	55.206	0.839
	(5.074)	(5.480)	(0.041)
Econometrica	71.008	52.957	0.834
	(4.917)	(5.584)	(0.041)
Political Analysis	66.663	49.252	0.715
	(5.413)	(5.972)	(0.044)
Politics and Gender	39.734	30.767	0.422
	(6.181)	(6.264)	(0.046)
SMR	55.682	38.717	0.609
	(5.652)	(6.156)	(0.046)
Observations	1906	1906	1906
R^2	0.303	0.318	0.774
Adjusted R^2	0.297	0.312	0.772

Results are coefficients from OLS regression (standard errors in parentheses). All models include year fixed effects (coefficients omitted from table). Constant is not included in any model. *APSR*, *American Political Science Review*; *SMR*, *Sociological Methods and Research.*

Our overall conclusion is that there is no apparent relationship between author gender and citation count in this data set once author count is held constant.

The importance of the author count variable suggests that confounding might explain at least some of the gender discrepancies found in Dion, Sumner, and Mitchell (2018). Women are underrepresented as solo authors in general (West *et al.* 2013), tend to collaborate more on average than men (Kumar and Ratnavelu 2016), and tend to have a greater number of co-authors (Bozeman and Gaughan 2011).[10] However, in the Dion, Sumner, and Mitchell (2018) data set, articles with *only* women as authors have on average 0.8 fewer authors (with a standard error 0.05) compared to articles with male authors or mixed-gender author teams. Again, this is sensible: for nearly any fixed probability that an author is female, a greater number of co-authors implies a lower probability of an all-female author team.[11] Moreover, it is plausible that there is a positive association between the number of authors of an article and the number of papers cited by that article (Corbyn 2010; Davis 2010); citing more papers presumably means a greater chance of citing at least one paper authored by women. Consequently, there is a possible back-door causal pathway between the probability that all authors on a published article are female and the probability that the article cites another paper with only women authors.

It is also possible that papers with female authors cite women more often because all authors (including women) often cite *themselves* (Fowler and Aksnes 2007). Even if women cite themselves less often than men (King *et al.* 2017), a positive propensity for authors to self-cite could explain

---

10 However, McDowell, Singell, and Stater (2006) find that women and men are equally likely to co-author.

11 If the probability of any author being a woman is p, then the probability that at least one author on a paper is a woman is $1 - p^n$, where n is the number of authors. This quantity is strictly increasing in n for $p \in (0, 1)$.

why a paper with only female authors would be more likely to cite another paper with only female authors.

Although we find that female-authored papers are not any more or less likely to be cited on average once we control for the journal and number of authors, our results do not imply that there is no gender representation problem in the political methodology community. Both *Political Analysis* and *Econometrica* have a substantially lower proportion of articles with at least one female author than *Sociological Research and Methods*, with the latter journal having over 9 percentage points more articles with a female author compared to the former two in this data set. Given the very high citation rates of *Political Analysis* and *Econometrica*, both of which are broadly comparable[12] to *American Political Science Review* in impact factor, our results suggest that boosting the representation of women in citations could be achieved by increasing the diversity of the methods community through active recruitment and mentoring of, and co-authorship with, people from underrepresented groups (including women) who have an interest in methodology.

References

Abramo, Giovanni, and Ciriaco Andrea D'Angelo. 2015. The relationship between the number of authors of a publication, its citations and the impact factor of the publishing journal: Evidence from Italy. *Journal of Informetrics* 9(4):746–761.

Bozeman, Barry, and Monica Gaughan. 2011. How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy* 40(10):1393–1402.

Cameron, Elissa Z., Angela M. White, and Meeghan R. Gray. 2016. Solving the productivity and impact puzzle: Do men outperform women, or are metrics biased? *BioScience* 66(3):245–252.

Chamberlain, Scott, Carl Boettiger, Ted Hart, and Karthik Ram. 2017. *rcrossref: Client for various 'CrossRef' 'APIs'*. R package version 0.8.0. https://CRAN.R-project.org/package=rcrossref.

Corbyn, Zoë. 2010. An easy way to boost a paper's citations. *Nature*, http://dx.doi.org/10.1038/news.2010.406 (last accessed 4/16/2018).

Davis, Phil. 2010. Reference list length and citations: A spurious relationship. *The Scholarly Kitchen* (blog). https://goo.gl/qJb7S2 (last accessed 4/16/2018).

Dion, Michelle, Jane Lawrence Sumner, and Sara McLaughlin Mitchell. 2018. Gendered citation patterns across political science and social science methodology field. *Political Analysis* 26(3):312–327.

Esarey, Justin, and Kristin Bryant. 2018. Replication data for: Are papers written by women authors cited less frequently? https://doi.org/10.7910/DVN/PCYBPY, Harvard Dataverse, V1.

Fischer, Joern, Euan G. Ritchie, and Jan Hanspach. 2012. Academia's obsession with quantity. *Trends in Ecology & Evolution* 27(9):473–474.

Fowler, James H., and Dag W. Aksnes. 2007. Does self-citation pay? *Scientometrics* 72(3):427–437.

Franceschet, Massimo, and Antonio Costantini. 2010. The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics* 4(4):540–553.

King, Molly M., Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, and Jevin D. West. 2017. Men set their own cites high: Gender and self-citation across fields and over time. *Socius* 3:1–22.

Kumar, Sameer, and Kuru Ratnavelu. 2016. Perceptions of scholars in the field of economics on co-authorship associations: Evidence from an international survey. *PLoS ONE* 11(6): e0157633.

Leimu, Roosa, and Julia Koricheva. 2005. What determines the citation frequency of ecological papers? *Trends in Ecology and Evolution* 20(1):28–32.

Marini, Giulio. 2017. New promotion patterns in Italian universities: Less seniority and more productivity? Data from ASN. *Higher Education* 73(2):189–205.

McDowell, John M., Larry D. Singell, and Mark Stater. 2006. Two to tango? Gender differences in the decisions to publish and coauthor. *Economic Inquiry* 44(1):153–168.

Thomson Reuters. 2016. Journal citation reports. https://jcr.incites.thomsonreuters.com/ (last accessed 3/28/2018).

West, Jevin D., Jennifer Jacquet, Molly M. King, Shelley J. Correll, and Carl T. Bergstrom. The role of gender in scholarly authorship. *PLoS ONE* 8(7):e66212.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.

---

12 The 2016 impact factors for these journals are as follows. *Econometrica*: 3.379; *American Political Science Review*: 3.316; *Political Analysis*: 3.361 (Thomson Reuters 2016).