

The Review: Production and Consumption of APSR*

Gaurav Sood
gsood07@gmail.com

October 3rd, 2015

NB: Preliminary draft. Please do not cite without permission.

Scientific production is affected by a variety of variables that have little to do with science. For instance, investment in research is correlated with commercialization lags — greater the lags, lower the investment (Budish, Roin and Williams, 2013).¹ On the flip side, minor changes to how research is delivered can have large effects on consumption of research. For instance, papers listed first in National Bureau of Economic Research email digests get 27% more citations than papers listed at other locations.² More generally, a broad range of factors, from availability of funding to fads to various parts of the scientific paper production pipeline, including, the number of articles editors must review, the incentives to be nice to the author, likely affect what topic is researched, the quality of the research, who produces the research, and how many read it.

Using an original dataset of rich meta data on all articles published in the The American Political Science Review (APSR), the preeminent political science journal, I shed light on some of these issues. In particular, I study whether, like top economic journals, over time the number of articles published by The Review has declined (Card and DellaVigna, 2013, 2014). I find that the

*The data, and the scripts to acquire, process and analyze the data can be found at https://github.com/soodoku/meta_apsr.

¹The article posits that later stage cancer drugs have smaller commercialization lags. It isn't clear why. But the general point likely holds.

²See <http://www.npr.org/2015/07/15/423101360/no-1-with-a-bullet-point-to-get-research-cited-make-sure-its-listed-first>.

number of articles published by the APSR has increased over time, though the number of articles per issue has plateaued over the last thirty or so years.

Next, I investigate two particular aspects of production: co-authorship and gender of authors. For a variety of reasons — ease of co-authorship, greater differentiation in skills, greater complexity of projects — the proportion of co-authored articles in a variety of fields is on the rise (Barnett, Ault and Kaserman, 1988; Card and DellaVigna, 2013; Cunningham and Dillon, 1997). I study whether that is true for the APSR. Expectedly, I find that co-authorship rates have increased significantly. Second, I estimate the proportion of women per published article in the APSR over time. While the proportion has increased over time, it still remains very low.

Next, I move to study consumption using consumption. I find that both the abstract views and full-text views follow the well-known power law distribution, with a few articles receiving a lot of views and most articles receiving no views. I end by investigating something fun. Some previous analyses have suggested that paper titles have gotten longer over time.³ I check whether this is true for the APSR. I find that it is.

Data and Results

I scrape the APSR website for the data. Scraping yields a dataset of 30,235 unique document objects. These documents include editor's notes, rejoinders, articles, research notes, book reviews and other miscellanea that have been published in the APSR. I exclude editor's notes, book reviews, etc. from the data. This leaves us with 4,385 articles, which serves as the final dataset. For each article, we know the title, the abstract and the pages spanned (though not the font or the line spacing or the page margins), the names and institutional affiliations of the authors, and the total number of times the abstract, and the full-text of the article have been viewed. In addition, we have information on the number of pages in each issue (volume). To these data, I add imputed gender of the authors — proportion of people with the name who were female according to the

³See <http://datacolada.org/2013/12/04/titleogy/>

census — using the R Package `gender` (Mullen, 2015).⁴ I use these data to answer the questions.

The number of articles published in the top economics journal⁵, The American Economic Review (AER), has declined over the years (Card and DellaVigna, 2014). This surprising pattern is a result of longer articles and constant size of the issue Card and DellaVigna (2014). (Over the past few years, the AER has capped article size.) To see whether the same holds for The APSR, I first investigate article length. Over the past 100 or so years, average article length has shown marked variability (see Figure 1). But unlike top economics journals there is no trend towards longer articles. (It is very likely, however, that the length of online appendices has grown substantially.) This likely reflects a longer tradition of capping article lengths in political science.

Moving to number of articles per issue, unlike top economic journals, over the last hundred years, the number of articles per issue has more than doubled (see Figure 2). Though, over the past thirty or so years, the number of articles per issue has plateaued. Given no trend in article length, and increase in number of articles per issue, it has to be that pages per issue would have increased. And so we find (see Figure 3).

Next, we analyze data on two aspects of production: number of authors per article, and proportion of female authors per article. Looking at the number of authors per article over time, like in other places, co-authorship is on the rise. Though, solo authored papers still make a sizable proportion of publications in the APSR, the modal publication today has two authors (see Figure 4). As for proportion of women per article, the data are distressing. While the proportion of women on each article published in the APSR has risen consistently and sharply, the average article still has just 20% female authors (see Figure 5). (Pairing it with American Political Science Association membership data would provide one (still unsatisfactory) way of getting an estimate of gender bias.)

Flipping the lens and looking at two consumption metrics: number of abstract views and

⁴There are at least two major limitations in how the gender of the names is imputed. First, time can be correlated with gender distribution of a name, and we lack information on the age of the author. I assume that the age of an author's age is uniformly distributed between 25 and 65. While this is unsatisfactory, using different ranges yields very similar results. Second, the data on the names is only from the U.S. Expanding the database to include data from Canada and Europe are obvious next steps.

⁵The Quarterly Journal of Economics is the top-ranked economics journal by some measures today.

number of full-text views, provides a familiar power-law distribution. Most of the articles (abstracts) aren't viewed at all. And a small set of articles gets a lot of views (see Figures 6 and 8).

Lastly, some fun. Plotting the length of titles of APSR articles over time clearly shows that the titles have become longer over time. In fact, the average title length has increased by 50% – from an average of 50 to about 75 characters today.

Figures

Figure 1: Number of Pages per Article Over Time



Figure 2: Number of Articles per Issue Over Time

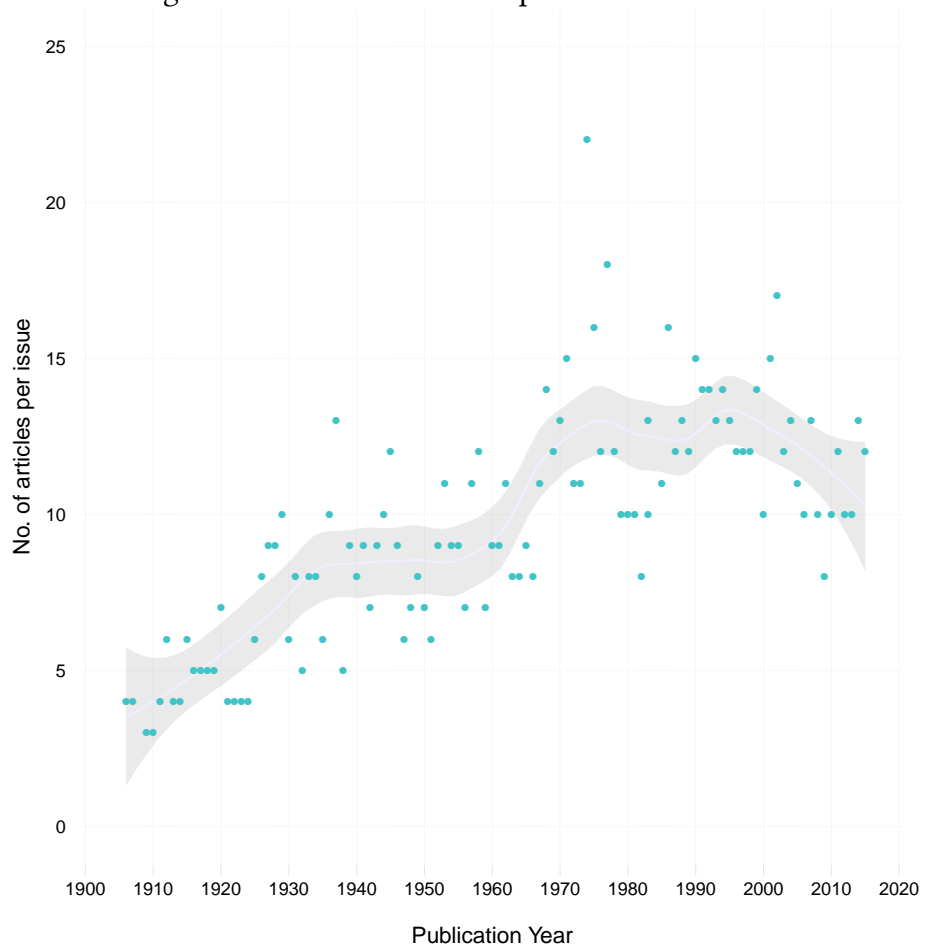


Figure 3: Pages per Issue Over time

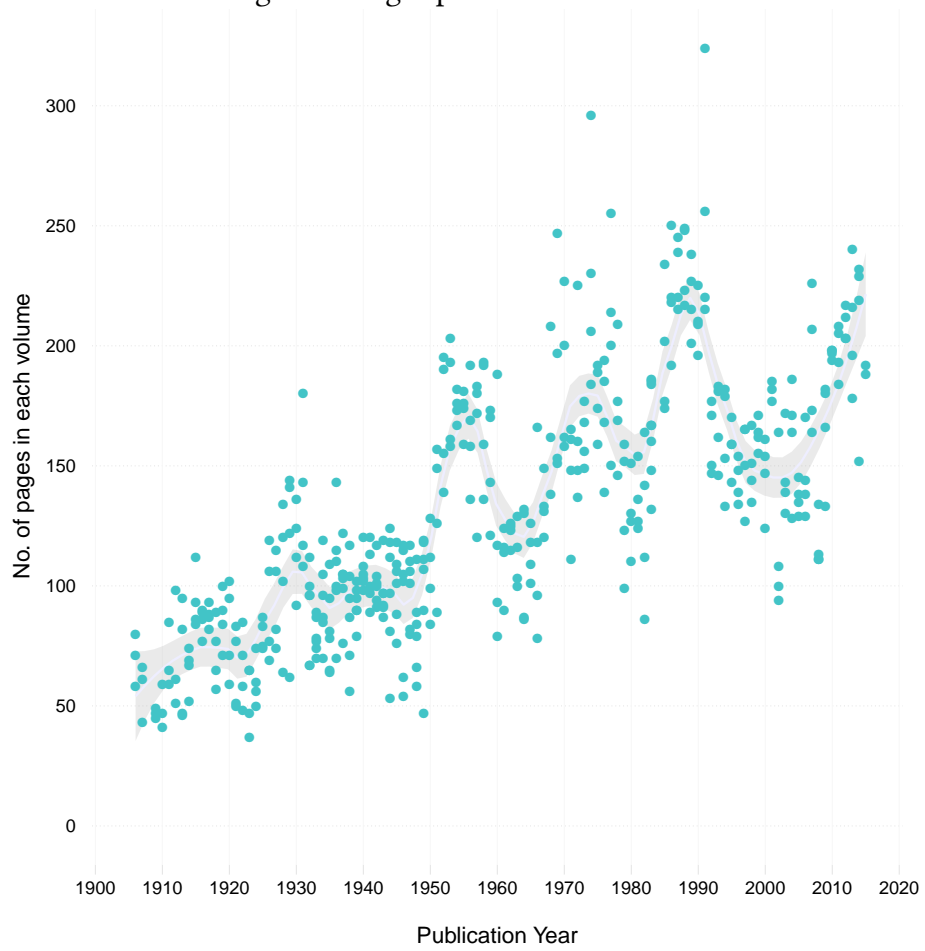


Figure 4: Number of Authors per Article Over Time

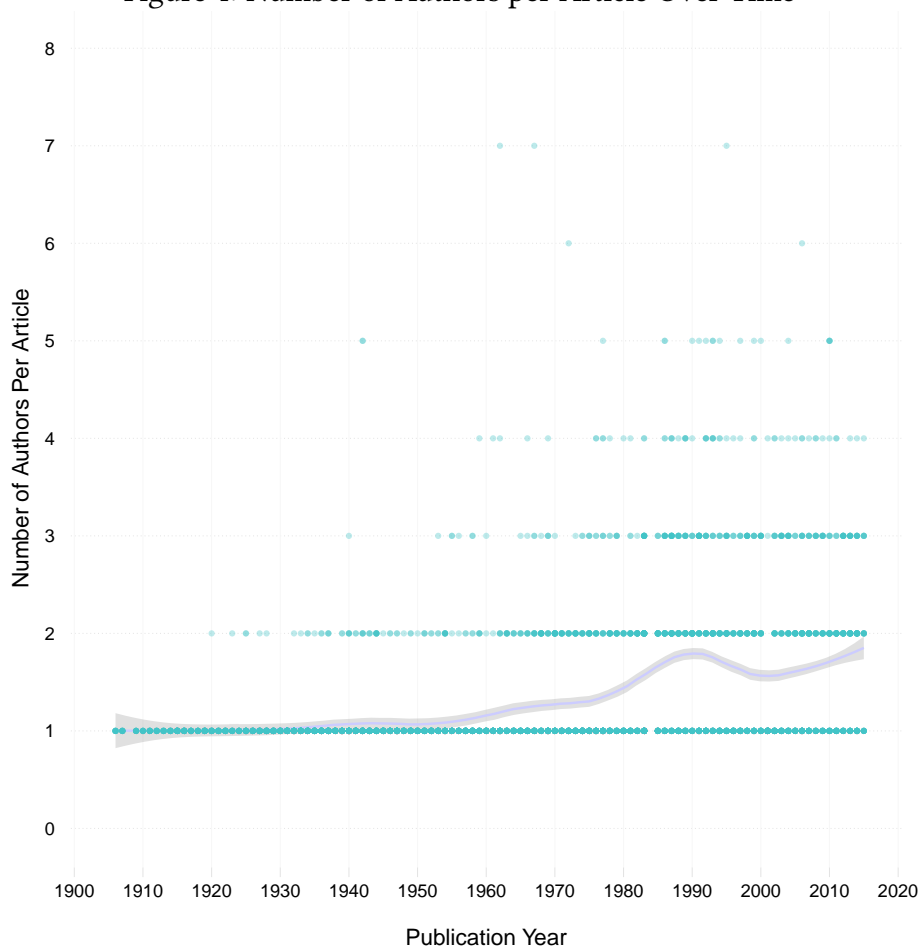


Figure 5: Proportion of Women Per Article Over Time



Figure 6: Abstract Views

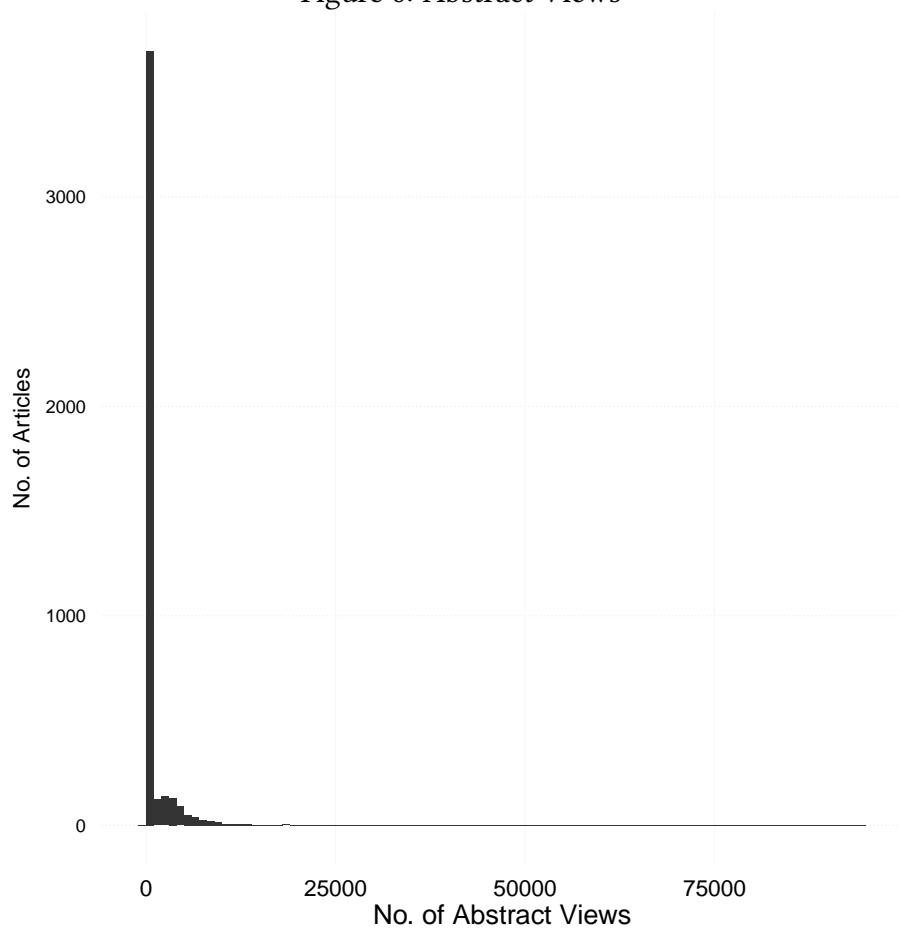


Figure 7: Full-Text Views

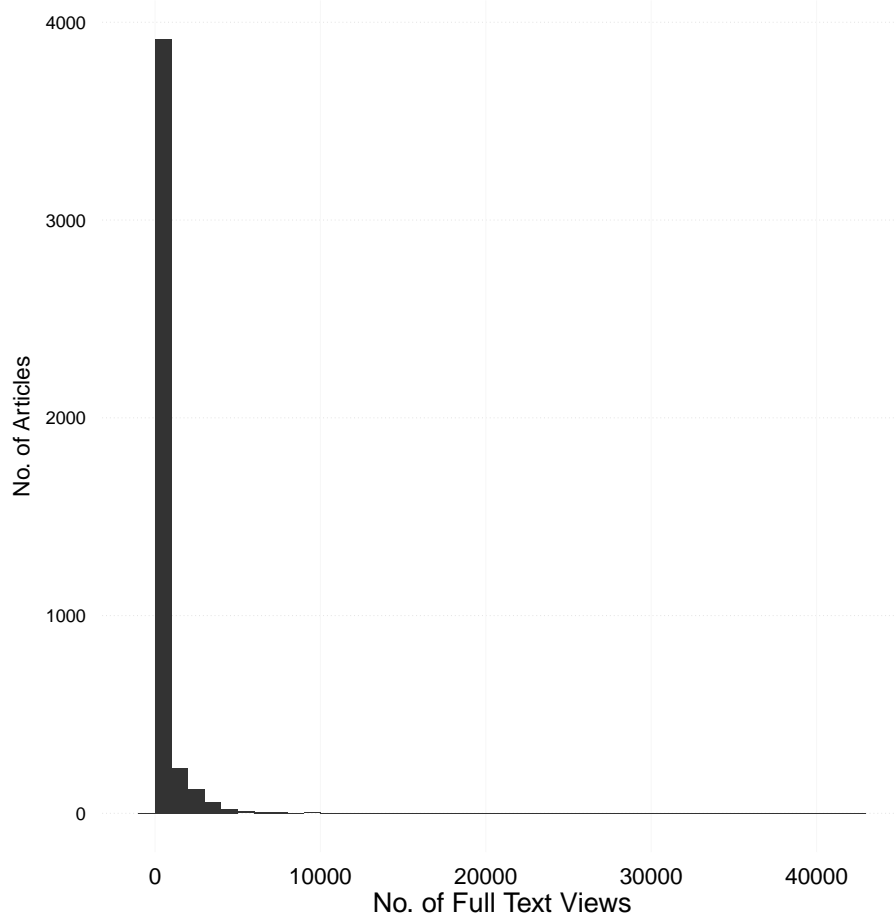


Figure 8: Title Length Over Time



References

- Barnett, Andy H, Richard W Ault and David L Kaserman. 1988. “The rising incidence of co-authorship in economics: Further evidence.” *The review of Economics and statistics* pp. 539–543.
- Budish, Eric, Benjamin N Roin and Heidi Williams. 2013. Do firms underinvest in long-term research? Evidence from cancer clinical trials. Technical report National Bureau of Economic Research.
- Card, David and Stefano DellaVigna. 2013. Nine facts about top journals in economics. Technical report National Bureau of Economic Research.
- Card, David and Stefano DellaVigna. 2014. “Page Limits on Economics Articles: Evidence from Two Journals.” *The Journal of Economic Perspectives* pp. 149–167.
- Cunningham, Sally Jo and Stuart M Dillon. 1997. “Authorship patterns in information systems.” *Scientometrics* 39(1):19–27.
- Mullen, Lincoln. 2015. *gender: Predict Gender from Names Using Historical Data*. R package version 0.5.1.9000.
URL: <https://github.com/ropensci/gender>