

Scholarly Citations and Twitter Followers: An Insight Into the Case of Economists*

Jorge Vladimir Garcia Perez[†]

April 14, 2018

Abstract

The present study aims to explore the patterns of Twitter usage among economists. I gather data from the website of RePec IDEAS, which contains information on registered economists around the world, their twitter account information with the total number of followers, the number of citations they are mentioned in, and their h-index (a metric of publication impact). I run a regression analysis to assess the effects of a percentage increase of citations in the percentage increase of Twitter followers. I find that citations, have a positive effect of 1.5% increase in Twitter followers, and this result is statistically significant. On the other hand, the h-index has negative effect of 0.05% in Twitter followers, which is statistically significant at the 5% level. As proposed by previous authors, I hypothesize that increase in Twitter followers is due to greater utility levels the audience gets from following their preferred authors.

*Many thanks to Professor Tyler Ransom from the University of Oklahoma Economics Department for providing the research idea and without whom this work would not be possible.

[†]Department of Economics, University of Oklahoma. E-mail address: jorgevladimir.garcia@ou.edu

1 Introduction

The emergence of social media, and its subsequent permeation into the fabric of society, has substantially transformed the way individuals interact with one another. Scholarly practice is not exempt from this phenomenon, and thus, the adoption of social media by professors and students is quite evident. It is not far from truth that Twitter is an online medium popularized among economists. Several researchers, many of them who are considered pioneers in the field, have a Twitter account. For instance, 2017 Nobel Prize awardee in Economics, Richard H. Thaler ¹, has a Twitter account with more than 120,000 followers, a huge number considering that he is not an avid user, with no more than 6,000 tweets since he created his account in August 2011. In this sense, physical limitations no longer impose constraints on the transmission of ideas, access to research, and information on their findings; but on the contrary, social media enables researchers to reach larger audiences who might not necessarily be affiliated with an academic institution. At least in an idealistic way, social media serve as platforms for the diffusion of research and direct interaction with researchers, or the democratization of access to knowledge, being Twitter (a micro-blogging social media) one of them.

Nonetheless, the manner in which academics use Twitter could greatly differ from the manner that is utilized by larger audiences, from the manner that is utilized by other scientists in different fields, and there could even be differences from one economist to another, or differences within sub-fields of Economics. In this sense, different questions emerge. To what extent economists differ in the way they interact with online audiences when compared to other scientists in different fields? What is the effect of scholarly activity, such as citations, in the number of Twitter followers

¹ @R.Thaler

a scholar acquires? And what determines an increase of Twitter followers among economists? To answer some of these questions, I collect previous findings of other studies and conduct my own regression analysis and summary statistics based on data available online. I study the effects of citations on the number of Twitter followers, taking the h-index into consideration as an additional variable for my proposed model.

2 Literature Review

Technology has made academia more open to data sharing, democratization of expertise, and to alternative models of peer review and reputation management (Greenhow and Gleason, 2014). Social media allows scientists to interact with larger audiences in real time, through different matrices on research and practice impact evaluation, which can consequently enhance or degrade their own career advancement (Preston, 2018). Previous literature on the relation between scholarly practice and Twitter has found that researchers, as a whole, tend to share more links and retweet more than the average Twitter users, and within them, economists share the most links whilst also engaging in less scholarly communication (as opposed to biochemists or astrophysicists), which could be due to more general Economics-related tweets rather than research practices (Holmberg and Thelwall, 2014).

Giusta, Jaworska, and Greetham (2018), from a sample study of 64,121 tweets from the top 25 economists and top 25 scientists, find that economists on Twitter tweet less, mention fewer people, use less accessible language, with words that are more complex and have more abbreviations, and with a tone that is more distant and less personal. Moreover, when compared to scientists in other fields, economists engage in less conversations with strangers. Also, highly inclusive pronouns

such as 'we' and 'our' are used twice less often by economists, as opposed to other scientists, suggesting that this factor could lead to a poor understanding and traction of economics among the larger public (Giusta, Jaworska, and Greetham, 2018). The way economists are using Twitter, in terms of interaction with larger non-academic audiences, promote the formulation of further inquiries: Does the public not trust economists? Do they not understand what economists do and how they work? Is their work misrepresented in the media? And how do economists themselves interact with public opinion (Royal Economic Society, 2018).

It remains unclear whether Twitter usage has an effect on publications and citations, or the opposite scenario, in which publications and citations influence Twitter usage, is true. While some evidence suggest that a low correlation exists between the number of citations and the number of tweets a particular article receives, meaning that tweets and retweets of a published academic article do not predict the citations said article will receive (Haustein et al., 2013); Eysenbach (2011) find that Twitter users can, with certain limitations, predict citations though a perfect correlation between these two variables should not be expected. Nevertheless, both studies point out that Twitter-based metrics encompass different dimensions of impact. On one hand, Twitter-based indicators reflect types of impact that are not comparable to traditional citation indicators (Haustein et al., 2013); given that citations are a metric that primarily mirrors the impact of scholarly work, and Twitter could be a metric for social impact and knowledge translation (how quickly new knowledge is taken up by the public), as well as a metric to measure public interest in a specific topic (what the public is paying attention to) (Eysenbach, 2011). Pertaining to the number of Twitter followers, some econometric evidence indicates a positive relationship between (sport) success, and the number of new Twitter followers a soccer team acquires. The estimates of Perez (2013), based on the number of Twitter followers between January 13, and June 26, 2012 for each of the

twenty Spanish professional soccer teams playing in *La Liga BBVA*, excluding *Real Madrid Club the Ftbol* and *Futbol Club Barcelona*, finds that the most successful teams have the highest rate of recruitment of new Twitter followers, a phenomenon the author attributes to the increased utility received by followers from the victories their teams of preference achieve. Based on the latter, I hypothesize that larger citations are a metric of scholarly success, and we expect this variable to have a positive effect on the number of followers.

3 Data

For the present research project, I ran a descriptive statistics analysis and a regression analysis using data from IDEAS. The dependent variable is the percentage increase of Twitter followers an author has. This information is parsed from the 'Top 25% Economists by Twitter Followers' list, and contains updated data as of April 24 of 2018. I initially scraped information on the total number of Twitter followers, but given the high skewness of this data, I took the natural logarithm of this variable. Although the possibility of performing a time-series analysis is doable, and could potentially become a new research project, I do not conduct such analysis due to time constraints.

The first independent variable in this study is the percentage change of citations, based on the natural logarithm of the total number of citations an author has, as per provided by the IDEAS database in March 2018. The second independent variable is the h-index, which is defined as the highest number of publications of a scientist that received h or more citations each while the other publications have no more than h citations each (Schreiber, 2008). For example, a scholar with an h-index of 5 had published 5 papers, each of which has been cited by others at least 5 times (Wortman, 2018). This information was also extracted from the IDEAS website. Although I had

information on the ranking based on the number of Twitter followers and citations, I do not include these variables to avoid multicollinearity with other variables of interest.

4 Methods

I conduct my regression analysis based on the following model:

$$\ln t.followers_{it} = \alpha_0 + \alpha_1 \ln citations_{it} + \alpha_2 h.index_{it} + \varepsilon, \quad (1)$$

where $\ln t.followers$ is an outcome variable representing the percentage change in Twitter followers, and $\ln citations$ is an independent variable representing the percentage change in citations for an author, and $h.index$ is a metric for publication impact. I aim to estimate the parameter α_1 , and as mentioned previously, I use data from the IDEAS website.

IDEAS is the largest online bibliographic database dedicated to Economics based on RePEc², which is a group working on the provision of electronic papers that indexes over 2,500,000 items; it includes bibliographic metadata over 1,900 participating archives, and compiles information on authors (publishing and non-publishing) who are invited to register an online profile in the system (IDEAS, 2017). IDEAS offers material by serial, by classification and by author; and its server is hosted by the Federal Reserve Bank of St. Louis.

The website also creates rankings and listings that indexes registered economists based on a variety of parameters. I utilize the listings on the 'Top 25% Economists by Twitter Followers,' the 'Top 5% Authors, Number of Citations, as of March 2018,' and 'Top 5% Authors, h-index, as of March 2018.' The Twitter listing indexes economists who are also registered on Twitter, and a

²RePec stands for "Research Papers in Economics."

ranking based on the number of followers; this information gets updated daily. The second listing, as explained by its name, includes information on the top 5% authors based on their number of citations, their ranking, total citations, institution of affiliation, city and/or state, as well as country. The last listing, on the other hand, contains information on the Top 5% authors based on their h-index, which is a measure of productivity and citation impact of the publication of a scholar.

All listings were extracted and parsed into data sets through the assistance of the programming language 'R' and the 'rvest' package, wherein I was also able to extract the identification profile associated with each author within the IDEAS database, the link to each authors' profile within said database, and the link to some authors' profiles within Citec³, and their correspondent Twitter handle. Although my population sample is heavily biased, parsing data from these already-made listings simplified the extraction of information. Initial attempts to gather data on all registered authors with Twitter accounts were unsuccessful, as the 'rvest' package works more efficiently by selecting html_nodes from websites, which were strangely not available in the complete listings I desired to scrape data from. Extracting this information required further understanding of the xml_paths and the table structure of websites, which could potentially be continuation of this research project. The different data sets were merged based on the authors identification profiles within the IDEAS database, a combination of no more than six characters between letters and numbers, though several observations remained incomplete and therefore were excluded from the final data set for the linear model. I prioritized the maintenance of observations with a Twitter account and a Twitter handle, since it contained our outcome variable of interest, but unfortunately not all authors from other listings had a Twitter account. For instance, the listing with information

³Citec indexes documents distributed on the RePEc database, and provides indicators of impact factor for journals and working paper series <http://citec.repec.org/>

on citations and h-index had more than 2,000 observations each, but only few matched the IDs of those authors in the Twitter listing. Furthermore, when trying to extract the link to the Citec profile of each author, which also has data on citations and the h-index, my code returned links for some of the authors but not all of them. Better understanding of the package 'rvest' and websites is essential to estimate the effects of citations on Twitter followers with a more representative sample size, or complete population, of *tweeting* economists.

5 Research Findings

As summarized in 1, for our population sample of Economics authors on Twitter, the median value of Twitter Followers equals to 10,782; 2,280 as the median value of citations; and 23 as the median value for the h-index. The kurtosis and skewness of the independent variables were also calculated. Twitter followers, citations and h-index had a skewness estimation of 9.32, 1.2, and 1.3; respectively, which indicates that the distribution of these variables was highly skewed to the right, especially for our variable 'Twitter followers.' On the other hand, our measurements on kurtosis are 89.11 for Twitter followers, 6.3 for citations, and 3.4 for the h-index. This also indicates that the distribution is mesokurtic for h-index, but leptokurtic for Twitter followers and citations. The standard deviations are 464,782.6 for Twitter followers, 25,849.3 for citations, and 15.4 for the h-index.

The estimation for the constant intercept coefficient α_0 is -0.671, but is not statistically significant. The parameter α_2 , corresponding to h-index, equals to -0.059 and it is statistically significant at the 5% level. This finding would imply that an increase of one point in the h-index will have a negative effect of 0.05% in the number of Twitter followers. Finally, our parameter of interest

α_1 equals to 1.48 and it is statistically significant at the 1% level implying that an increase of one percent in citations will have an effect of 1.48% increase in the number of Twitter followers. This model explains at least 18% to 20% of the variability of Twitter followers among Economist, based on the results of the r-squared and adjusted r-squared. Table 2 displays the findings of my regression analysis, and Figure 1 provides a graphing plot of our variables of interest, and includes the Twitter handles of some authors who were chosen at random.

6 Conclusion

The parameter β_1 of the natural logarithm of citations is positively related to our outcome variable, the natural log of Twitter followers. This finding implicates that an increase of one percent in citations will have positive effect in Twitter followers, increasing it by one percent. Citations, or the percentage increase of citations more specifically, as hypothesized earlier, becomes a metric of academic success that translates into a greater percentage increase in the number of followers on Twitter. Perez (2013) proposes that the increase in Twitter followers is due to the higher utility received by 'fans' from the success of a particular soccer team. In a similar way, we could extend this argument and suggest that the success of an author, measured in the percentage increase of citations, is reflected in the percentage increase of Twitter followers, who acquire greater utility from keeping themselves informed about the academic and online activity of their preferred author.

Further research could touch upon more inquiries, such as the ones raised by researchers in the revised literature. For instance, by gathering data in a time-series format. Given that I have the data structure and coding to facilitate the scraping of this information, we could assess more of these research topics. Moreover, I attempted to evaluate the reverse causality dilemma in Twitter

followers and citations, and I found no statistically significant parameters that established the effect of Twitter followers in citations. However, we should keep in mind the bias of my population sample, and continue this study with better data.

Additionally, the hypotheses proposed by Giusta, Jaworska, and Greetham (2018) were not assessed. These authors argue that the distanced-tone of economists, based on the way economists interact in online platforms in comparison to scientists in other fields, feeds into the general mistrust of economists among the general public, and it explains recent social phenomena such as the emergence of economic populism in Western countries. A general pattern found in this research study is that those economists with higher percentage increase in citations and higher h-indexes will consequently present a higher percentage increase of Twitter followers; which was quite evident since the beginning of the project. It could be the case that (due to distance tone of economists in online platforms, the lesser degree of interaction with audiences outside the field, and the complex language they use), economists are popular exclusively among economists. If true, this would support the 'superiority of economists' argument suggested by (Fourcade, Ollion, and Algan, 2015), and to a certain extent the echo chamber situation of scholars within the field. Some of these factors are not the case among scientists in other fields, who engage online with scientific as well as non-scientific audiences, although a more extensive comparison in the online practices of scientists needs to be made as the mistrust toward the scientific community could be part of a larger structural phenomenon (Gawande, 2016).

References

- Eysenbach, Gunther. 2011. "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact." *J Med Internet Res* 13 (4):e123. URL <http://www.jmir.org/2011/4/e123/>.
- Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29 (1):89–114. URL <http://www.aeaweb.org/articles?id=10.1257/jep.29.1.89>.
- Gawande, Atul. 2016. "The Mistrust of Science." *The New Yorker* URL <https://www.newyorker.com/news/news-desk/the-mistrust-of-science>. Accessed April 23, 2018.
- Giusta, Marina Della, Sylvia Jaworska, and Danica Vukadinovic Greetham. 2018. "Tweeting Economists: Antisocials in the Socials?" Unpublished working paper, last version 23 March 2018.
- Greenhow, Christine and Benjamin Gleason. 2014. "Social scholarship: Reconsidering scholarly practices in the age of social media." *British Journal of Educational Technology* 45 (3):392–402. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12150>.
- Haustein, Stefanie, Isabella Peters, Cassidy R. Sugimoto, Mike Thelwall, and Vincent Lariviere. 2013. "Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature." *Journal of the Association for Information Science and Technology* 65 (4):656–669. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23101>.
- Holmberg, Kim and Mike Thelwall. 2014. "Disciplinary Differences in Twitter Scholarly Communication." *Scientonometrics* 101 (2):1027–1042.
- IDEAS. 2017. "RePEc/IDEAS rankings." URL <https://ideas.repec.org/top/>. Accessed April 24, 2018.
- Perez, Levi. 2013. "What Drives the Number of New Twitter Followers? An Economic Case Study of Professional Soccer Teams." *Economics Bulletin* 33 (3):1941–1947.
- Preston, Jennifer. 2018. "If Twitter Is a Work Necessity." *The New York Times* URL <https://www.nytimes.com/2012/03/01/education/digital-skills-can-be-quickly-acquired.html?pagewanted=all>. Accessed April 23, 2018.
- Royal Economic Society. 2018. "Tweeting Economists are Less Effective Communicators than Scientists." URL <http://www.res.org.uk/details/mediabrief/10921411/TWEETING-ECONOMISTS-ARE-LESS-EFFECTIVE-COMMUNICATORS-THAN-SCIENTISTS.html>. Accessed April 14, 2018.
- Schreiber, Michael. 2008. "An empirical investigation of the gindex for 26 physicists in comparison with the hindex, the Aindex, and the Rindex." *Journal of the American Society for Information Science and Technology* 59 (9):1513–1522. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20856>.

Wortman, Sue. 2018. "Research Impact Metrics: Citation Analysis." *University of Michigan Library* URL <http://guides.lib.umich.edu/c.php?g=282982&p=1887449>. Last Updated: March 5, 2018. Accessed April 24, 2018.

Figures and Tables

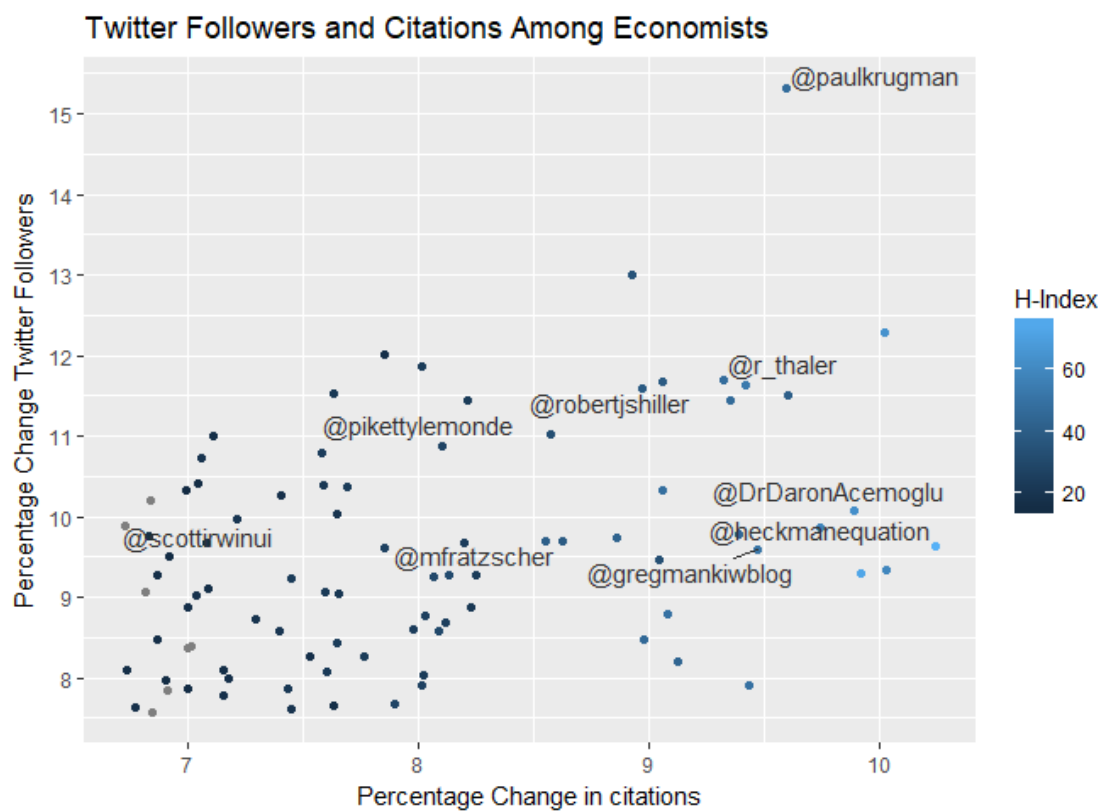


Figure 1: Percentage change in Twitter followers as a function of percentage change in citations. H-index is color-coded from dark blue, low h-index, to light blue, high h-index; and the Twitter handle of some authors chosen at random are included.

Table 1: Summary Statistics for Twitter followers, Number of citations and H-index

	Median	Std. Dev.	Min	Max	Skweness	Kurtosis	<i>N</i>
Twitter Followers	10782	464782.6	1948	4501778	9.32	89.11	94
Citations	2280	5849.3	834	28047	1.2	6.3	88
h.index	23.00	15.40655	14	79	1.3	3.4	87

Note: Table 1. Includes information on the summary statistics of our variables of interest. Their estimates on skweness and kurtosis are also included.

Table 2: Regression Model of Twitter Followers

Dependent variable: ln.followers	
ln.citations	1.484*** (0.456)
h.index	−0.059** (0.028)
Constant	−0.671 (2.912)
Observations	81
R ²	0.201
Adjusted R ²	0.181
Residual Std. Error	1.315 (df = 78)
F Statistic	9.818*** (df = 2; 78)

Note: *p<0.1; **p<0.05; ***p<0.01