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 matric of publication impact). I run a regression analysis to asses 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.

---

†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, students, research institutions, or journals and academic associations is quite evident. Although several social media platforms are available on the internet, it is not far from truth that Twitter is a popular medium among economists. Researchers, many of them who are considered pioneers in the field, have twitter accounts where they publicize their findings, comment on current economic events, and network with colleagues who share similar interests as them. For instance, 2017 Nobel Prize in Economics awardee, Richard H. Thaler[1], has a Twitter account. Based on his total number of followers, more than 120,000; we could possibly speculate that he is not avid user, given that he has no more than 6,000 tweets since he created his twitter account in August 2011.

The usage of social media facilitates the transmission of ideas, access to new research, and scientific findings, mainly because one-to-one interaction is no longer the only way to spread knowledge. Furthermore, social media, and platforms such as Twitter, enable researchers to reach larger audiences who might not be necessarily affiliated with a particular academic institution. At least, in an idealistic way, Twitter serves as a medium for the diffusion of research and the democratization of information.

Nonetheless, the manner in which academics use Twitter could greatly differ from the manner that is utilized by others. One would assume that leisure-related usage prevails among the average Twitter user, as opposed to academics in Twitter, though this can be hard to discern. Similarly,

---

[1] @R_Thaler

patterns of Twitter usage could also differ among academics in other fields, and even within sub-fields of the same branch. Lastly, the impact of social media in society can greatly shape public perceptions. Specifically for Economics, it would be important to determine how much Twitter influences the public perception of economists and Economics in light of the rise of populism in Western countries, and based on research that calls out on the subjective sense of authority and entitlement (Fourcade, Ollion, and Algan, 2015), characteristic of economists.

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 a scholar acquires? And what determines an increase of Twitter followers among economists? To answer some of these questions, I collect findings of previous studies and conduct a 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. The aim is to determine whether academic success will increase the number of Twitter followers an economist has.

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, practice, and 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 explained by how economists touch upon more Economics-related topics online, via commentaries on the current social panorama, rather than sharing tweets with research information (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, and 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). For these researchers, 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 the public opinion (Royal Economic Society, 2018).

On the other hand, 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 predict citations with certain limitations, 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 research 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 for each of the twenty Spanish professional soccer teams playing in *La Liga BBVA*, with the exception of *Real Madrid Club de Ftbol* and *Futbol Club Barcelona*, find 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 these findings, I utilize citations, and higher rates of citations, as a metric of scholarly success that will 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 natural log of Twitter followers an author has, which represents the percent point increase in Twitter followers. 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 the information I obtained, I took the natural logarithm of this variable in order to normalize the distribution and achieve a more linear derivation. 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; therefore my data contains information of this variable only for one time period.

The first independent variable in this study is the natural log of citations, which represents the percent point change of the total number of citations for an author, as per provided by the IDEAS database in March 2018. This information was parsed from the 'Top 5% Authors, Number of Citations' listing at the IDEAS website, and tracks citations on the publication of an author. 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, from the 'Top 5% Authors, h-index' listing with information updated as of March 2018. Although my final data set includes raking information based on the number of Twitter followers and citations, I decided not to include this variables to avoid colinearity with other variables of interest.

4 Methods

I conduct my regression analysis based on the following model:

$$\ln t.fllwrs_{it} = \alpha_0 + \alpha_1 \ln cit_{it} + \alpha_2 h.ind_{it} + \varepsilon, \tag{1}$$

where $\ln t.followers$ is an outcome variable representing the natural log of Twitter followers (*t.followers*), and $\ln citations$ is an independent variable representing the natural log of citations (*cit*) for an author, and H-index is a metric for publication impact (*h.ind*). I aim to estimate the parameter α_1, and for this purpose I use data from the IDEAS website.

IDEAS is the largest online bibliographic database dedicated to Economics based on RePEc [2], 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 provides a 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

---

[2]RePec stands for "Research Papers in Economics."

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 on 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 Citec[3], 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 and required more sophisticated coding practices. 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 a 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 of these observations remained incomplete and therefore were excluded from the final data set for the linear regression. 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 on authors' citations and h-index had more than 2,000 observations each, but only few matched

---

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

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 academic 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 Economists on Twitter, the median value of Twitter followers equals to 10,782; 2,280 is the median value of citations; and 23 is 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 would indicate that the respective distributions of these variables were highly skewed to the right, especially for our variable on 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 mesukortic for h-index, but leptokurtic for Twitter followers and citations given that tails are fatter than the normal distribution. Such finding was an indication for me to take estimate the linear regression with the natural logarithms of my variables of interest. 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; however, this value is not statistically significant at any level. The 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 change in citations will have

9

an effect of 1.48 percent change increase in the number of Twitter followers. The parameter α_2, corresponding to the H-index equals to -0.059 and it is statistically significant at the 5% level. This finding indicates that an increase of one percent point change in the H-index will have a negative effect of 0.05 percent point change in the number of Twitter followers. The adjusted R-squared and the R-squared values, 0.18 and 0.2 respectively, suggest that at least 18 to 20 percent of the variability of Twitter followers for Economists can be explained by our model. 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. As we can see in the figure, the higher the percent change in citation for a given economist, the higher the percent change of Twitter followers the author will have. The H-index is also associated with the percent change of Twitter followers, as we can see in the graph those authors with higher H-indexes (coded light blue) will seemingly tend to be at the upper right hand side of our graph, though its association in reality is much more disperse as it was found by our α_2 parameter.

6 Conclusion

The parameter $_1$ on the natural logarithm of citations is positively related to our outcome variable, the natural log of Twitter followers. This finding implies that an increase of one percent in citations will have a positive effect in Twitter followers, increasing it by one percent. Citations, or the percent change increase in citations, as hypothesized earlier, becomes a metric of academic success that translates into a greater percentage point change in the number of Twitter followers. 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 fashion, we could extend this argument

and suggest that the success of an author, measured in the percentage point increase of citations, is reflected in the percentage point increase of Twitter followers. New followers acquire greater utility levels from keeping themselves informed about the academic and online activity of the author of their preference.

Different methodologies would help us further the applications of this research project to answer more inquiries, especially the ones brought up by previous literature. For instance, by gathering data in a time series format, we would be able to see the evolution of Twitter followers, citations, and the H-index over time; and not at a single time period as in this project. Certainly, given that I already have the coding and data structure that would facilitate the scraping of this information, I could continue with this project. Moreover, I attempted to evaluate the reverse causality dilemma, the natural log of citations on the natural log of Twitter followers and H-index (the opposite of my proposed econometric model), but I found no statistically significant evidence that established the effect of Twitter followers in citations. This finding can be pretty intuitive since we are dealing with a sample of only Top economists (based on citations and Twitter followers), who recruit more followers due to their academic work (or success), and not necessarily economists who have more citations as a function of their Twitter followers. Nonetheless, it is important to keep in mind the bias of my population sample, and more accurate findings could be estimated 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. It could be the case that economists are popular exclusively among economists, because of their distant tone in online platforms, their lesser degree of interaction with audiences outside the field, and the complex language they use. If this pattern happens to be true, there is a tendency in the field to remain in an eco chamber that feeds into the 'superiority of economists' argument suggested by (Fourcade, Ollion, and Algan, 2015); which is not the case of other scientists in different feels. Finally, by making a more extensive comparison in the online practices of scientists from different fields with academic as all as non-academic audiences, we could also hypothesize on the existence of a larger structural phenomena that reinforces mistrust toward the scientific community (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`. Accesseed 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 Larivire. 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.
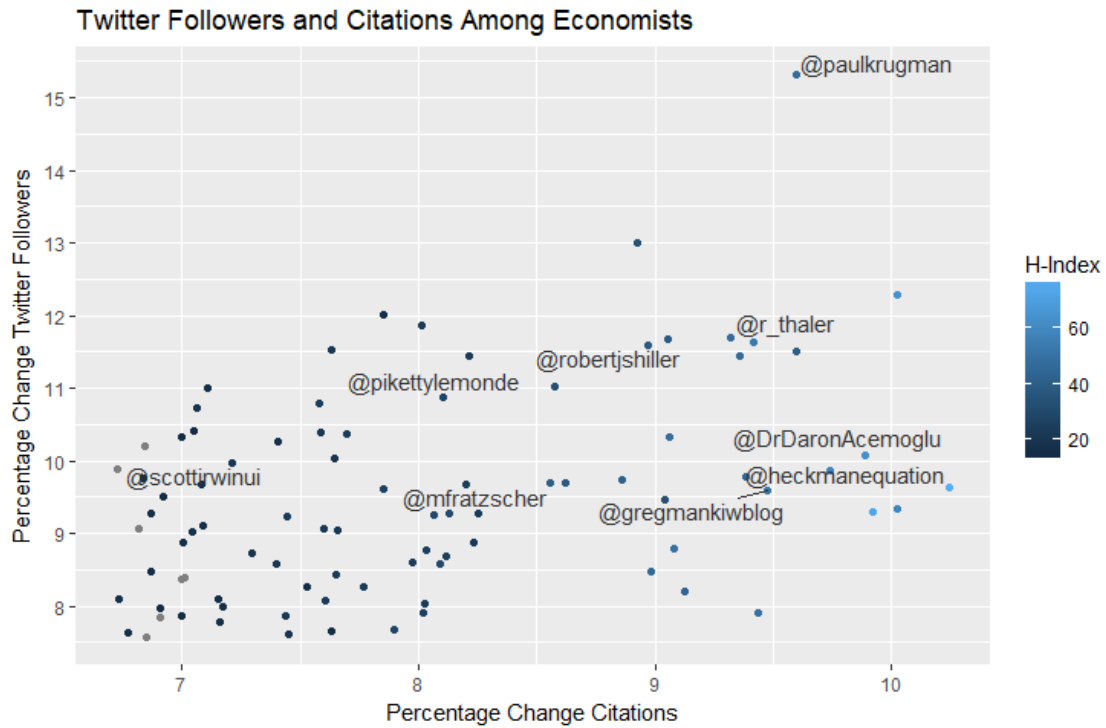
Figures and Tables

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	N
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