Faculty of Business and Law

UNIVERSITY OF PORTSMOUTH

Working Papers in Economics & Finance
No. 2023-04

Twitter and Citations

Ho Fai Chan, Queensland University of Technology

Ali Sina Önder, University of Portsmouth

Sascha Schweitzer, Reutlingen University

Benno Torgler, Queensland University of Technology

Portsmouth Business School
https://www.port.ac.uk/about-us/structure-and-governance/organisational-structure/our-academic-structure/faculty-of-business-and-law/portsmouth-business-school

Twitter and Citations

Ho Fai Chan*

Ali Sina Önder†

Sascha Schweitzer‡§

Benno Torgler *

May 29, 2023

Abstract

Social media, especially Twitter, plays an increasingly important role among researchers in showcasing and promoting their research. Does Twitter affect academic citations? Making use of Twitter activity about columns published on VoxEU, a renowned online platform for economists, we develop an instrumental variable strategy to show that Twitter activity about a research paper has a causal effect on the number of citations that this paper will receive. We find that the existence of at least one tweet, as opposed to none, leads to $16-25\%$ more citations. Doubling the overall Twitter engagement generates up to 16% more citations.

JEL Classification: A14, J24

Key Words: Productivity; Social Media; Twitter; Citations; Economists

*Queensland University of Technology

†University of Portsmouth. *(Contact author: ali.onder@port.ac.uk)*

‡Reutlingen University

§University of Bayreuth

1 Introduction

Social media became an essential tool for scholars to share, enhance, validate, and expand their engagement and interaction within the academic community (Veletsianos, 2016). Twitter is one of the most actively used social media platforms by researchers for microblogging and interactions, especially among economists and social scientists (Fang et al., 2022, Sugimoto et al., 2017). Recent studies reveal a positive correlation between Twitter activity and academic citations: For instance, Ortega (2016) demonstrate that the number of Twitter followers a scholar has directly influences their citation count. Klar et al. (2020) find strong correlation between pushing research on Twitter and citation counts received by that research. Moreover, Peoples et al. (2016) discover that Twitter activity predicts citations more reliably than journal impact factor. These insights are noteworthy as numerous academic institutions and individual scientists now prioritize social media to showcase and promote their research. The important question is whether the observed correlations are causal: Does Twitter activity have a causal effect on academic citations?

We analyze Twitter engagement about a substantial collection of VoxEU (http://voxeu.org/) columns which we cross-reference with corresponding publications and citations. VoxEU is a notable online platform which enjoys significant recognition within the economics community and allows researchers to report about their socially and policy-relevant work. Chan et al. (2020) discovered that more renowned economists receive greater attention when disseminating their work through the VoxEU portal. To establish causality, we implement an instrumental variables (IV) strategy. Our approach capitalizes on the timing of the VoxEU column publication (both day of the week and season of the year), as such timing can influence user engagement with Twitter while not directly affecting citation frequency. Our findings suggest that Twitter activity does indeed have an effect on the number of citations an article receives.

2 Data

An important share of the VoxEU columns are based on recently circulated economics research, most of them being either recent publications or freshly minted working papers. We

complement the extensive dataset of (Chan et al., 2020) consisting of 6,086 VoxEU columns from 2008 until 2017, data about Twitter activity about these columns and an exhaustive set of controls for various personal characteristics of individual researchers with publication and citation data by matching these 6,086 columns with the Web of Science database (WoS). We stop our coverage of VoxEU columns at the end of 2017 so that associated WoS publication have sufficient time to accumulate citations.

Since there is no explicit information regarding the connection between journal articles and VoxEU columns, we implement a search and matching procedure that takes into account the metadata and titles of publications. This procedure involves several steps. First, we search for all peer-reviewed journal articles that match the authors' first and last names for each VoxEU column. We utilize an automated Python script and the WoS Advanced Search Query Builder to perform a search for the author criterion. Following the WoS search, a matching procedure is applied to identify close matches in terms of timing, content, and authorship. We calculate two scores for this purpose. The first score is for the similarity between the titles of WoS results and their corresponding VoxEU columns. The second score represents the number of shared authors between the articles.

If there is only one matching WoS article, we define it as the best match. If there are no matches, the VoxEU column is marked as not related to a journal publication. For multiple matches, we employ a two-step approach to select the best match. In the first step, we retain articles with the highest number of shared authors. In the second step, among the remaining articles, we choose the one with the highest title similarity. In case of a tie, we make a random selection. This process gets 2,731 of 6,086 VoxEU columns matched to peer-reviewed publications in the WoS.

3 Descriptive Findings

We start by estimating the following linear probability model for the likelihood of a VoxEU column to be matched with a peer-reviewed journal publication:

$$Publication_i = \alpha X + \beta(Twitter\ Activity_i) + \epsilon_i \tag{1}$$

where *Publication* is one if the VoxEU column could be matched to a peer-reviewed journal publication in WoS and zero otherwise. X is a vector of control variables containing characteristics of the author team (whether there is an Econometric Society fellow or female author on the team, highest rank PhD institutions and highest rank affiliation among authors, number of total citation to date by authors), characteristics of the VoxEU column (topic, availability of tables and figures, title length, overall word count, publication time) and number of tweets about this column by authors and the VoxEU platform or VoxEU editors (*Richard Baldwin* and *Romesh Vaitilingam*).

We capture the *Twitter activity* using three different measures: *any tweet/retweet*, *total tweets/retweets*, and *other engagement*. We refer to any Twitter account except VoxEU's official account, VoxEU editors' personal accounts and accounts of a respective column's authors as *users' accounts* or simply *users*. *Any tweet/retweet* is a binary variable that is one if any user account tweeted or retweeted at least once about the column, and zero otherwise. *Total tweets/retweets* is the logarithm of the total number of users' tweets and retweets, *other engagement* is the logarithm of the number of likes and replies by users about the respective VoxEU column. We observe $30,841$ Twitter interactions for $4,653$ VoxEU columns. $2,623$ columns received an interaction from users other than VoxEU editors or authors. $12,243$ Twitter interactions about $2,905$ columns were registered on the very same day that columns went online on *voxeu.org*.[1] The average number of users' tweets and retweets of a VoxEU column is 5.1, the average number of likes and replies is 2.5, the median for both is zero.

Columns (1), (3), and (5) in Table 1 show estimated coefficients for equation 1. Twitter activity is an insignificant predictor for whether there is an associated peer-reviewed publication about a VoxEU column. Not all VoxEU columns are based on authors' recent research papers, some may be a policy comment or an opinion column so that not all columns can be linked to peer-reviewed publications. Furthermore, even though a column is based on a working paper or project draft, it is possible that this paper never got officially published. Our regressions do not pick up a systemic correlation between Twitter activity and whether a VoxEU column could be matched to a publication in the WoS.

[1] Of $2,279$ VoxEU columns that could be merged into the WoS, $1,595$ have at least one Twitter interaction, and $1,110$ of them received these interactions on the very same day that they went online on *voxeu.org*.

Table 1: Twitter activity, peer-reviewed journal publications, and citations

	(1) Published?	(2) Cites	(3) Published?	(4) Cites	(5) Published?	(6) Cites
Any Tweet/Retweet?	0.022	0.102^b				
	[0.019]	[0.050]				
Total Tweets/Retweets			-0.001	0.036^c		
			[0.008]	[0.020]		
Other Engagement					-0.001	0.042^c
					[0.009]	[0.025]
Observations	6086	2279	6086	2279	6086	2279
R^2	0.059	0.475	0.059	0.474	0.059	0.474
F	7.886	47.584	7.809	47.518	7.810	47.451

Standard errors in brackets. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Provided there is a peer-reviewed journal publication with a VoxEU column, how are its citations related to Twitter activity of this VoxEU column? When VoxEU columns are matched to WoS, we find 2,731 match and there are 452 publications that were published before their VoxEU columns. We drop them from our analysis in order to make sure that a VoxEU column was online before its associated journal publication because the mechanism how VoxEU column may be associated with citation accumulation will differ if there is already a journal publication before the VoxEU column and this aspect is especially crucial for our subsequent analysis in the next subsection. We regress the number of total citations of matched peer-reviewed publications as of 2021 on their corresponding VoxEU column's Twitter activity captured by the above mentioned three measures using the following specification:

$$E(Citations_i|Publication_i = 1) = \gamma X + \delta(Twitter\ Activity_i) + \epsilon_i \tag{2}$$

where $Citations_i$ are expressed in logarithm. The average number of citations is 23.6, whereas the median is 8. Columns (2), (4), and (6) in Table 1 reveal a positive and statistically significant relation between Twitter activity associated with a VoxEU column and citations accumulated by the corresponding WoS publication. Doubling the number of tweets and retweets or likes and replies is associated with a 4% increase in incoming citations of the associated publication.

4 Identification and Empirical Results

In order to claim causality, we employ an instrumental variable (IV) strategy by making use of the timing of the VoxEU column. We use the day of the week, season of the year, and their interactions as instruments. One can expect that the appearance of a VoxEU column on a weekend may affect how Twitter users engage with it. Users may be eager to read VoxEU columns on a weekend and —since Twitter is highly popular among economists— share on Twitter more as users might have more time on weekends, or alternatively, users may have better chances to read VoxEU columns during the week if this is their office routine or if they get notified by their peers in their hall about VoxEU columns. We initially identify VoxEU columns that appeared online on a Saturday or Sunday as weekend columns. A similar argument works also for the season of the year. Columns published during summer may either get more attention because users have more free time or they might not be aware of VoxEU columns during summer as a result of diminished regular peer interactions during summer. Figure 1 documents the distribution of VoxEU columns' online appearance over days of the week and seasons of the year from 2008 to 2017.

We have no prior expectation of how timing should affect Twitter activity regarding VoxEU columns, but we expect timing to have an effect on Twitter activity. The crucial exclusion restriction is that there is no reason why this timing should affect citations received by the corresponding journal publication through any other channel than the very Twitter activity itself. Our IV strategy is carried out as follows:

$$Twitter\ Activity_i^{IV} = f(Weekend_i, Summer_i, WeekendXSummer_i; \{2ndStageControls_i\})$$

in the first stage, and then the total number of citations is regressed on the instrumented Twitter activity and control variables in the second stage to estimate

$$Citations_i = g(Twitter\ Activity_i^{IV}, AuthorControls_i, TwitterControls_i, DocumentControls_i)$$

Author controls are authors' personal characteristics such as gender and age, productivity measures such as their number of publications and total citations as well as their academic

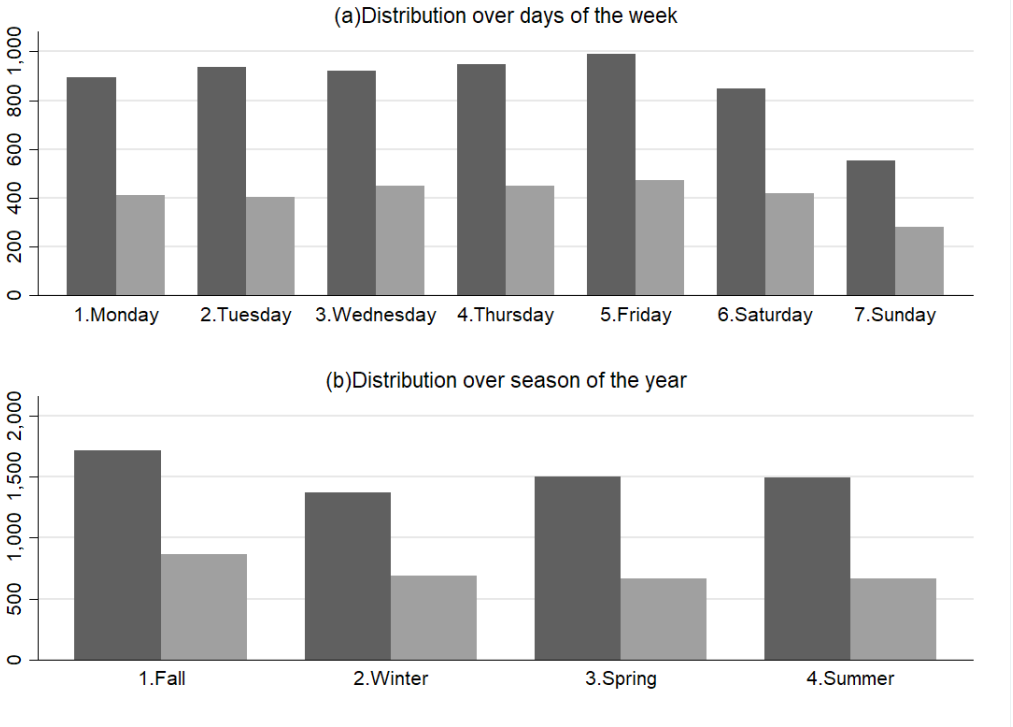

Figure 1: Distribution of VoxEU columns' time of first appearance on voxeu.org over days and seasons (dark shade). Light shade bars capture the distribution when VoxEU columns without an associated WoS publication are dropped.

background (whether they graduated from a top PhD program). Twitter controls are tweets and retweets of authors and VoxEU's official accounts. Any tweeting behavior of authors may affect journal citations as well. For instance, an author who promotes their own VoxEU column on Twitter may also be doing a good job in promoting their journal articles also elsewhere, which is very hard to capture. As a result, authors' own tweets or engagements can not be a clean instrument in this analysis. We include authors' Twitter engagement as a control variable in our analysis to proxy their tendency to self-promote their research. Twitter activity by VoxEU's official accounts (including its editors' personal accounts) may reflect positively on the quality of a VoxEU column's content, which in turn may be correlated with the quality of the journal article's content and hence its citations in the future.

Document controls are numbers of figures and tables in the VoxEU column, the length of the abstract, title, and main body of a VoxEU column as these have strong bearing on the journal publication's content and coverage. We also control for the journal quality using

Table 2: Citations and Twitter activity –based on two-day-weekend instrument

$Dep.Var: Cites$	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Any Tweet/Retweet?	0.139^{b}	0.163^{c}				
	[0.0670]	[0.0956]				
Total Tweets/Retweets			0.0565^{b}	0.0602		
			[0.0269]	[0.0372]		
Other Engagement					0.0585^{c}	0.0803
					[0.0343]	[0.0542]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1252	1252	1252	1252	1252	1252
F	27.14	27.02	27.04	26.99	27.07	26.99
$Kleibergen-Paap$		3.00e-78		6.40e-81		1.79e-60
$Cragg-Donald\ F$		78.14		105.7		59.43
$Hansen's\ J$		0.957		0.946		0.927

Standard errors in brackets. ^{c} $p<0.10$, ^{b} $p<0.05$, ^{a} $p<0.01$

the journal quality indices obtained from Kalaitzidakis et al. (2003). We use two sets of year fixed effects for the publication year of the VoxEU column and the year of the associated journal publication in the WoS data.

We restrict weekdays to Monday and Friday in order to have a balanced distribution of observations between weekend and weekdays. This also captures the idea that a column that is published on any given day could have easily slipped to the next day so that a column that was published on a Friday could have been published on Saturday, or similarly, a column that was published on a Sunday could have easily slipped into Monday.

Table 2 shows OLS and IV coefficient estimates for three alternative measures that capture Twitter activity. OLS coefficients in columns (1), (3), and (5) differ from those in Table 1 because the sample in Table 2 is restricted to Monday, Friday, Saturday, and Sunday. OLS coefficients of Twitter activity proxies are positively and statistically significantly correlated with citations also in this subsample. Although IV coefficients for the total number of tweets/retweets and for the total engagement of users have p-values larger than 10%, coefficients' point estimates are larger than their standard errors. Existence of at least one tweet leads to 16% more citations compared to when there is no tweet at all. Doubling the number of tweets/retweets and likes/replies leads to 6% and 8% more citations, respectively.

Table 3: Citations and Twitter activity –based on one-day-weekend instrument

$Dep.Var: Cites$	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Any Tweet/Retweet?	0.170^{c}	0.245^{c}				
	[0.0903]	[0.134]				
Total Tweets/Retweets			0.0791^{b}	0.111^{b}		
			[0.0373]	[0.0521]		
Other Engagement					0.0922^{c}	0.162^{b}
					[0.0483]	[0.0750]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	666	666	666	666	666	666
F	14.52	14.44	14.43	14.48	14.41	14.52
$Kleibergen-Paap$		1.16e-36		5.18e-37		1.27e-29
$Cragg-Donald\ F$		35.16		47.30		29.73
$Hansen's\ J$		0.835		0.897		0.901

Standard errors in brackets. ^{c} $p<0.10$, ^{b} $p<0.05$, ^{a} $p<0.01$

There is, however, a rather low number of VoxEU columns getting published on Sundays, as can be seen in Figure 1, so that our initial (and conventional) definition of weekend makes our dataset light on weekend and heavy on weekdays. In Table 3, we document OLS and IV estimations for another subsample where weekend is defined as Saturday and we take Monday as the weekday. We choose these two days to emphasize the distinction between the weekend and weekday. Columns that appear on a Monday can be claimed to have a very different attention supply from potential readers compared to columns that appear on a Saturday whereas the distinction between a Friday and a Saturday may not be very strong.

We obtain positive and statistically significant OLS as well as IV coefficients for all three proxies of Twitter activity. Existence of at least one tweet leads to 25% more citations. Doubling the number of tweets/retweets and likes/replies leads to 11% and 16% more citations, respectively. IV coefficients are larger than their OLS counterparts in this case. This is likely to occur due to the higher local average treatment effect resulting from the restrictive definition of weekend that we use in Table 3. Since most of the Twitter activity about a VoxEU column takes place on the same day of its publication on *voxeu.org*, one can argue that those who read and tweet about a VoxEU column on a Monday are more likely to remember that research and cite it eventually compared to those who read it on a

Table 4: First stage estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	Weekend is Sat. & Sun.			Weekend is Saturday		
	Any Tw.	Total Tw.	Other Eng.	Any Tw.	Total Tw.	Other Eng.
Weekend	-0.0335	-0.104^{b}	-0.0571	-0.0693^{b}	-0.136^{c}	-0.0750
	[0.0233]	[0.0518]	[0.0458]	[0.0344]	[0.0766]	[0.0659]
Summer	0.00245	-0.0447	-0.00999	0.0459	0.0464	0.0692
	[0.0332]	[0.0722]	[0.0662]	[0.0493]	[0.103]	[0.0886]
WeekendXSummer	0.0725	0.253^{b}	0.219^{b}	0.0670	0.171	0.0872
	[0.0455]	[0.104]	[0.0986]	[0.0649]	[0.146]	[0.128]
2nd St.Cont.&FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1252	1252	1252	666	666	666
R^2	0.578	0.649	0.533	0.556	0.635	0.551
F	95.51	43.90	18.54	47.31	26.43	12.24

Standard errors in brackets. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Saturday. It is also possible that those who read VoxEU on a Monday read it for specific research purposes compared to those who read on a Saturday rather casually. Either way, the IV picks up on this selection so that IV coefficients are larger than the OLS coefficients in Table 3. Under and over-identification as well as weak identification tests reveal plausible results for both sets of IVs employed in Table 2 and Table 3. Corresponding first stage estimations for both definitions of weekend are reported in Table 4. Coefficients of $Weekend$ and $WeekendXSummer$ have larger point estimates than their standard errors. Since early years of our data contain early diffusion of Twitter, it is important to include time controls for VoxEU columns' online appearance. The three instruments with time trend turn out jointly significant in all specifications.

5 Conclusion

We investigate whether Twitter activity has an impact on citations of published research and we deliver causal evidence for it. Causality is established based on the exclusion restriction that timing of first appearance of a VoxEU column could not affect incoming citations to its associated journal publication via any other channel than what this timing means for its

visibility and Twitter activity. Hence Twitter or possibly other wide-spread social media platforms can be instrumental in reaching out to wider parts of the scientific community.

Our analysis is based on economists' contributions to VoxEU, which is and has long been a very popular online platform to communicate research briefs or policy discussions. Since Twitter is a highly popular social media platform among economists, we could connect the Twitter activity around VoxEU to economists' citations reasonably well. For other disciplines, other social media channels may be more influential but the main implications will not diverge from our findings about how Twitter boosts economists' citations. A piece of research must be seen by the relevant audience so that it can get cited at all and apparently Twitter provides this visibility for economists. An important policy implication of our findings is that researchers as well as academic institutions should not be shy when it comes to sharing and showcasing their research on social media because such activity actually leads to higher number of citations.

References

Chan, H. F., S. M. Bodiuzzman, and B. Torgler (2020): "The power of social cues in the battle for attention: evidence from an online platform for scholarly commentary," *Journal of Informetrics*, 14, 101077.

Fang, Z., R. Costas, and P. Wouters (2022): "User engagement with scholarly tweets of scientific papers: a large-scale and cross-disciplinary analysis," *Scientometrics*, 127, 4523–4546.

Kalaitzidakis, P., T. P. Mamuneas, and T. Stengos (2003): "Rankings of academic journals and institutions in economics," *Journal of the European Economic Association*, 1, 1346–1366.

Klar, S., Y. Krupnikov, J. B. Ryan, K. Searles, and Y. Shmargad (2020): "Using social media to promote academic research: identifying the benefits of Twitter for sharing academic work," *Plos One*, 15, e0229446.

Ortega, J. L. (2016): "To be or not to be on Twitter, and its relarelation with the tweeting and citation of research papers," *Scientometrics*, 109, 1353–1364.

Peoples, B. K., S. R. Midway, D. Sackett, A. Lynch, and P. B. Cooney (2016): "Twitter predicts citation rates of ecological research," *Plos One*, 11, e0166570.

Sugimoto, C. R., S. Work, V. Larivière, and S. Haustein (2017): "Scholarly use of social media and altmetrics: a review of the literature," *Journal of the Association for Information Science and Technology*, 68, 2037–2062.

Veletsianos, G. (2016): *Social Media in Academia: Networked Scholars*, New York, Routledge.