5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16432

Assessing the spread of Keynesian ideas in the economic policy debate: a Text Mining approach on Twitter

Chiara Perfetto[1], Antonella Rancan[1], Giuliano Resce[1]
[1] Department of Economics, University of Molise, Italy

Abstract

This paper proposes a methodology for examining the presence of Keynesian ideas in the economic debate. To this aim we use Twitter as a source of data to monitor the debate in real time. We quantify the presence of Keynesian and anti-Keynesian thought in tweets about the economy and we qualify the emotional tone of these tweets. Our preliminary results show that the 20 per-cent of total English tweets about #economy contain words related to Keynes while about 8 per- cent contain words referring to anti-Keynesian policies. The monthly analysis of the tweets shows a certain heterogeneity. The distribution of Keynes-related tweets is much more uneven than the distribution of anti-Keynesian tweets. Our evidence suggests that the methodology we applied to understand how much of the Keynesian thought is still around in the economic debate can be promising. The next step will be to focus on georeferenced tweets to detect heterogenity across countries and to understand how country-level trends reflect the economy cycle. This study still has some limitations that will be faced in future research such as the classification of topics and the focus on English texts for the moment.

Keywords: *Text Mining; Twitter; Keynesian thought.*

1. Introduction

Social media interactions, from commenting on a post to liking a photo, leave digital traces that can be used to extract patterns of individual, group and social behaviours. In this paper we address how social data created by users are useful indicators for providing insights about economic patterns. The political science literature shows how the feedback mechanism reinforces the entrenchment of existing policies in the case of 'positive' feedbacks or subvert current patterns due to 'negative' feedbacks (Pierson & Skocpol, 2002).

Social scientists are often unable to systematically assess the significance and impact of ideas and research outputs due to extensive research portfolios, time and resource constrains. Therefore, the dynamics of knowledge dissemination are not fully understood. The integration of digital research methodologies, with text mining techniques offers an innovative and comprehensive approach to deal with such challenges. In this study we suggest Twitter as a source to study how public opinion relies on Keynesian or 'anti-Keynesian' view when debating economic issues. Data from Twitter make it possible to monitor the onset and spread of phenomena in real time (Resce & Maynard 2018). In fact tweets have a reliable timestamp and for that they can be analyzed from a time perspective and are accessible to researchers, unlike most social networking sites (Fujiwara et al., 2021). For these reasons, Twitter has found many applications among social scientists for many different purposes as detecting tourism preferences (Chang, Chu, 2013), analysing political trends (Rill et al., 2014, Seabold et al., 2015), or studying socio-economic problems (Resce, Maynard, 2018).

With this research, we aim to identify a methodology to quantify the presence of Keynesian and anti-Keynesian ideas in tweets about the economy and to qualify the emotional tone of these tweets (Misuraca et al., 2020). In this paper the terms "Keynesian" and "anti-Keynesian" are broadly defined. For Keynesian thought or Keynesian view we mean an approach to economic problems such as unemployment and economic downturn which relies on public interventions for their overcoming. On the contrary, anti-Keynesian view is defined in terms of a free-market approach, as supported by neoclassical economics. Twitters which mention government spending policies and expansionary monetary actions are Keynesian, while restrictive, anti-inflationary policies and deregulation policies are considered anti-Keynesian.

2. Data and method

The inclusion of Twitter data to understand the economic trend involves web scraping techniques on existing official Twitter accounts. The scraping was based on preliminary criteria which come from searching of a specific hashtag. We downloaded 7.255.518 tweets from 2008 to March 2022 regarding the hashtag of interest, that is #economy. For every tweet, the following fields have been considered: URL of the tweet; text of the tweet (along with the extraction of hashtags and mentions); timestamp, i.e., time of tweet creation; username of the publisher; number of likes; number of retweets; number of replies; geolocation (if available); language.

To quantify how many Tweets reflect Keynesian or anti-Keynesian thought we built a taxonomy taking into consideration a subsample of tweets containing the word "Keynes". From this subsample of 3,253 tweets we have extracted single words, bigrams, and trigrams which have a frequency higher than 3% of the tweets containing at least one Keyn-based word. For the extraction of the single words a Term Document Matrix was produced, with tweets id by column and stemmed words by row. The Term Document Matrix indicates the number of times each word appears in each tweet. For the bi-grams and tri-grams we used the function to tokenize in consecutive sequences of words, called n-grams. As one might expect, a lot of the most common bi-grams are “stop-words”, as "of the", "to be", etc. For that we removed cases where one or two of the two topics is a stop-word. The same was done for the tri-grams. Then the n-grams extracted were manually labelled (a strategy used by recent studies, such as Angelico et al., 2022) as Keynesian policy related or anti-Keynes policy related as in the Table1. Unigram, bigram and trigram where used, in the case of anti-Keynesian n-grams trigram where not connected to the topic and this is the reason why they were not considered.

Table 1. Taxonomy

PRO KEYNES			ANTI KEYNES		
TOPIC	BIGRAM	TRIGRAM	TOPIC	BIGRAM	TRIGRAM
Keynesian	maynard keynes	john maynard keynes	hayek	neoclassical economics	
Keyn	john maynard	wethepeople economy obama	auster	miltonfriedman rocks	
Invest	keynesian economics	economics rukvtbw2 dems	inflat	money economy	
Obama	keynesian economic	proof deficit spending	right	hayek round	
Stimulus	keynesian economists	feed finance economy	neoclass	austrian economy	
Recess	economy keynes	dems econ economy	teaparti	hayek rap	
Debt	keynes economics	maynard keynes economy	friedman	tcpt teaparty	
Deficit	keynes economy	maynard keynes considered	trump	modern economics	
Johnmaynardkeyn	keynesian economy	investing economy guest	miltonfriedman	milton friedman	
Nyt	economy keynesian	keynesian economics economy	neoliber	economy inflation	
economic	investing economy		republican	economy austerity	
Keynesianeconom	paul krugman		libertarian	keynesian fail	
Marx	economy obama		reagan	political economy	
Marxism	deficit spending		freemarket	ron paul	
Skidelski	rocks keynesian		monetarist	anti keynesian	
Democrat	keynesian proof			zerohedge kenesian	
Nytim	keynesian stimulus			economy teaparty	
	growing economy				
	economic growth				
	proof deficit				
	keynesian theory				
	economix blog				
	keynesian policies				
	capitalism money				
	keynes bit				
	century keynes				
	economy jobs				
	economy recession				
	obama economy				
	economic recovery				
	economy debt				
	government economy				
	government spending				
	keynes considered				
	post keynesian				
	neo keynesian				
	keynesian model				
	economy investing				
	keynesianism economy				
	nyt economix				
	economy keynesianeconomics				
	keynesian depression				
	cambridge corridor				
	debt economy				
	economics keynes				
	economy stimulus				
	jm keynes				
	keynesian economist				
	keynesian money				
	keynesian multipler				
	krugman keynes				
	obama kenesian				
	robert skidelsky				
	socialism marxism				
	stimulus economy				
	general theory				

To quantify what each tweet is about and if the tweet is connected to Keynesian or anti-Keynesian thought and policy, we used text mining techniques. For text mining we reduced the dataset to only 6.457.704 English tweets to facilitate some functions such as stemming and sentiment. Using R language, we adopted extensive customization of existing tools and algorithms to conduct such analyses. The text corpus of analysis was prepared using functions from the R package "tm" (Feinerer and Hornik, 2018; Feinerer, Hornik, and Meyer, 2008): punctuation, stop words (i.e., in English, words like "the", "is", "of", etc), special characters and numbers were removed from the corpus. The words were then converted to lowercase and stemmed. Also, the topics in the taxonomy were converted to lowercase and stemmed.

The words of the taxonomy are identified and counted in the tweets through the functions of the "stringr" package on R (Wickham, 2019). Keynes related tweets are defined as the sum of tweets featuring one or more words contained in the list of Keynesian topics while tweets related to anti-Keyensian policies result from the sum of tweets that contain one or more words contained in the list of anti-Keynesian topics.

To identify the semantic orientation of each line of text downloaded from Twitter we have adapted the functions of the "sentimentr" package developed by Rinker (2017) which uses a dictionary-based approach, i.e. based on a predefined polarized word list. Once the sentiment has been estimated, we produce graphs capable of highlighting the trend of sentiment over time. The set of information obtained by the text mining plus the number of tweets combined with the sentiment could be combined to build a sentiment index.

3. Preliminary results

The 19.3% of the economy total English tweets contain words related to Keynesian ideas while 7.7% contain words referring to an anti-Keynesian policies. The monthly analysis of the tweets (see Figure 1) shows a greater homogeneity of the distribution of anti-Keynesian tweets up to July 2016 with a subsequent increase in the volume of Tweets until reaching the peak in March 2019.

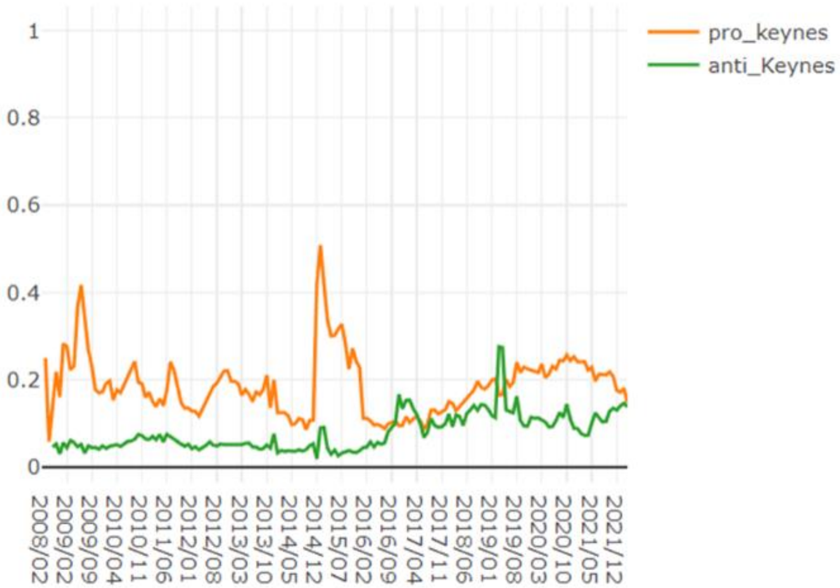

Figure 1. Distribution of tweets in timeline

As we can see from Figure 2 (top left) the distribution of Keynes-related tweets is much more uneven than the distribution of anti-keynesian tweets. The first peak is reached in June 2009, when 40% of the economic tweets written in that period contain words referring to Keynes. The largest volume of tweets referring to Keynes is reached in January 2015. This may be connected to the desire to abandon neoclassical theories following the 2007/2008 financial crisis and the subsequent Great Recession. If we look at the remaining part of Figure 2, we can also see an irregular distribution of likes, retweets and replies.

As we can see from Figure 3, the increase in the volume of anti-Keynesian tweets after 2016 corresponds to an increase in likes and retweets while the most replies tweets are those of 2009 and 2010.

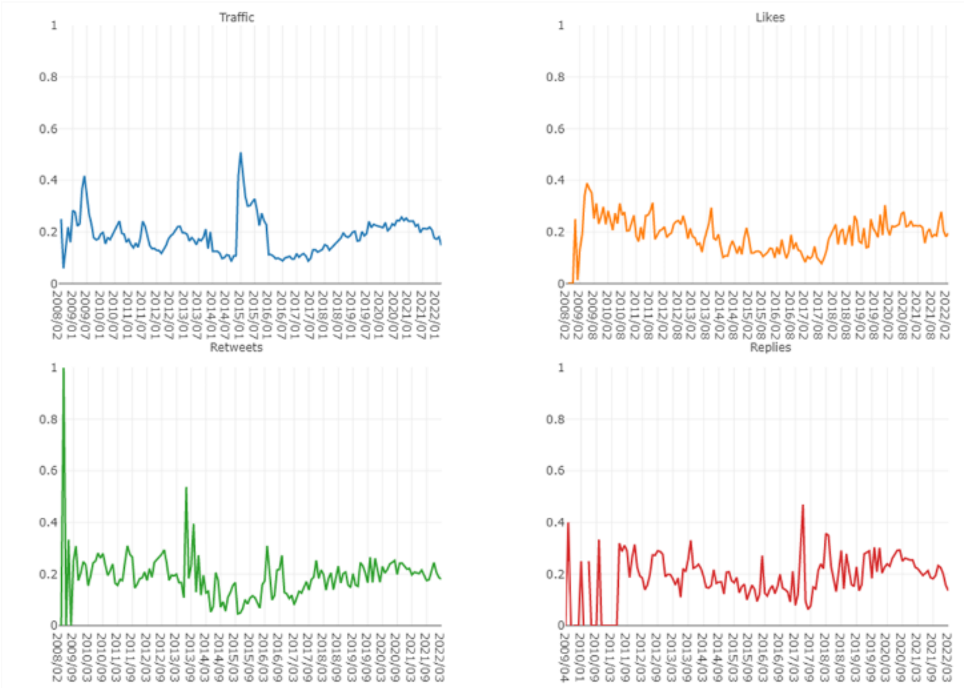

Figure 2. Statistics on Keynes-related tweets

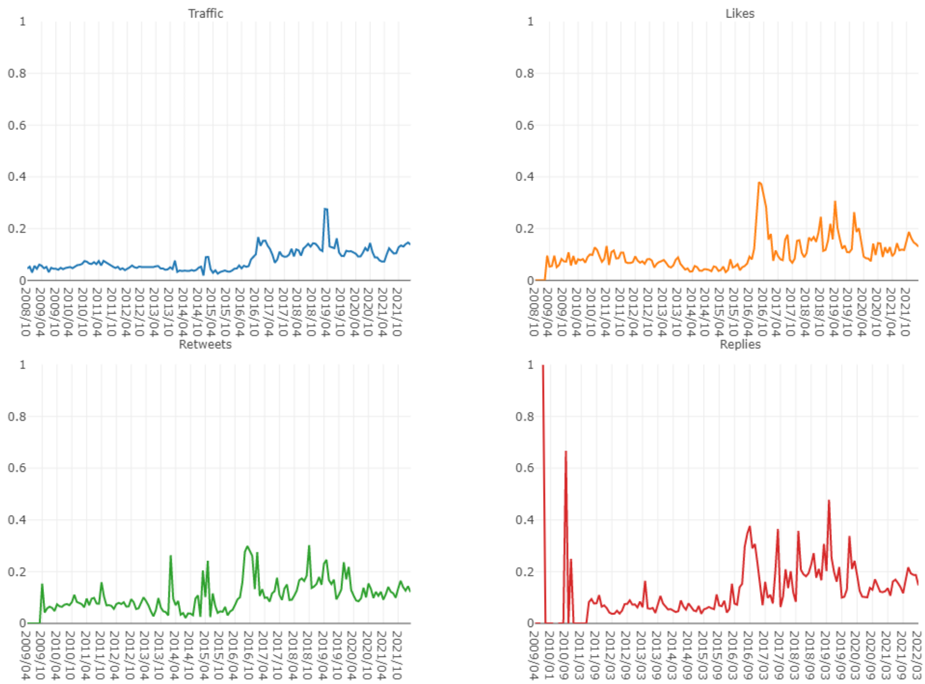

Figure 3. Statistics on antiKeynes-related tweets

As in the Figure 4, the sentiment relating to anti-Keynesian tweets is almost positive. Keynes-related tweets are associated with more volatile sentiment through 2014. Post-2014 the average sentiment of tweets is positive. Combining the results from supervised text mining with the sentiment generates an index that allows to compare the average sentiment trend of Keynesian or anti-Keynesian tweets (Figure 5).

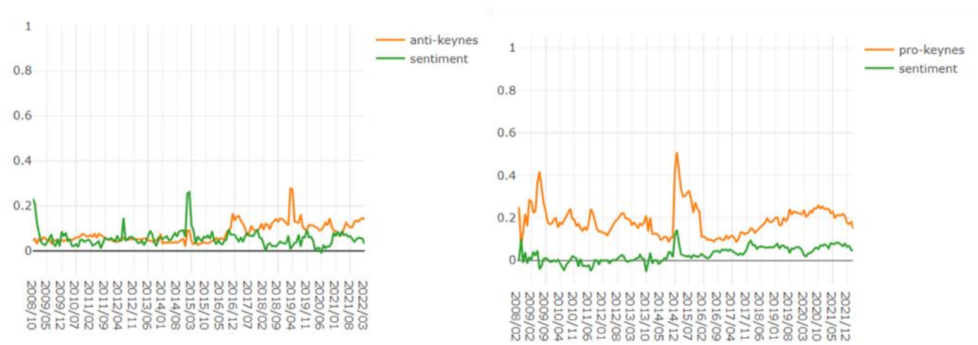

Figure 4. Sentiment analysis trends

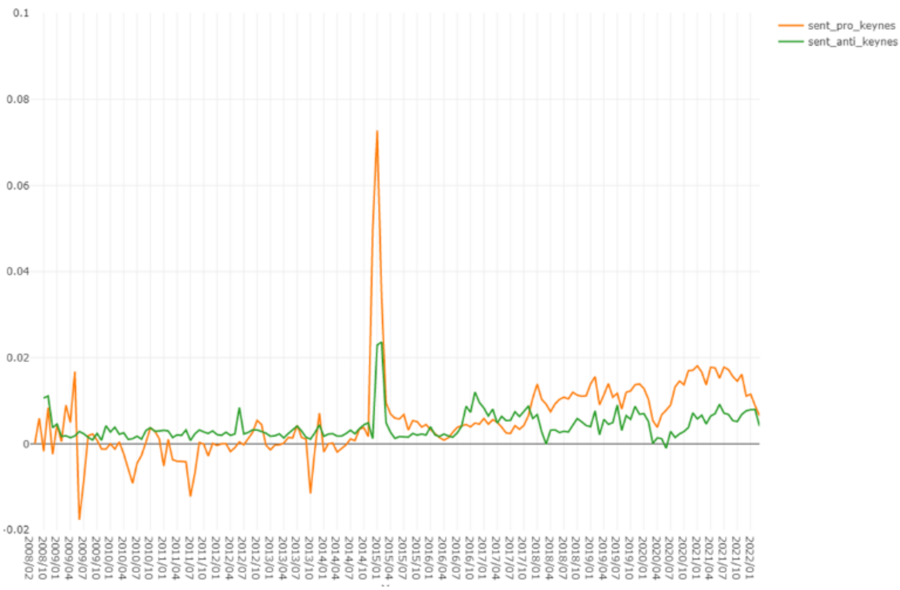

Figure 5. Sentiment index

The preliminary evidence of this research suggest a methodology for an understanding of the main references to which the economic policy debate carried out throug social media relies on. In line with previous literature, our results suggest that the integration of text mining techniques with Twitter data has a great potential to add knowledge to social science (Chang, Chu, 2013; Rill et al., 2014; Seabold et al., 2015; Resce, Maynard, 2018). For our future research directions, we plan to reduce the set of tweets at only georeferenced tweets to focus these trends across countries and understand how these trends reflect the economic cycle.

This study also includes some limitations that should be taken into account. When designing a study based on the taxonomy, there may be subjectivity in the classification of topics. Also the reduction to English texts only, necessary above all for the sentiment, could generate loss of information from the general dataset.

References

Angelico, C., Marcucci, J., Miccoli, M., & Quarta, F. (2022). Can we measure inflation expectations using Twitter?. *Journal of Econometrics*, 228(2), 259-277.

Chang, C. C., & Chu, K. H. (2013, June). A recommender system combining social networks for tourist attractions. In 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks (pp. 42-47). IEEE.

Feinerer, I. (2008). An introduction to text mining in R. R News, 8(2), 19-22.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of statistical software*, 25(5), 1-54.

Fujiwara, T., Müller, K., & Schwarz, C. (2021). The effect of social media on elections: Evidence from the United States (No. w28849). *National Bureau of Economic Research*.

Misuraca, M., Forciniti, A., Scepi, G., & Spano, M. (2020). Sentiment Analysis for Education with R: packages, methods and practical applications. arXiv preprint arXiv:2005.12840.

Pierson, P., & Skocpol, T. (2002). Historical institutionalism in contemporary political science. *Political science: The state of the discipline*, 3(1), 1-32.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, 89, 14-46.

Resce, G., & Maynard, D. (2018). What matters most to people around the world? Retrieving Better Life Index priorities on Twitter. *Technological Forecasting and Social Change*, 137, 61-75.

Rill, S., Reinel, D., Scheidt, J., & Zicari, R. V. (2014). Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. Knowledge-Based Systems, 69, 24-33.

Rinker, T. (2017). sentimentr: Calculate Text Polarity Sentiment. R package version 2.9.0.

Seabold, S., Rutherford, A., De Backer, O., & Coppola, A. (2015). The pulse of public opinion: using Twitter data to analyze public perception of reform in El Salvador. World Bank Policy Research Working Paper, (7399).

Wickham, H. (2019). stringr: Simple, consistent wrappers for common string operations. R package version 1.4. 0.