CESifo WORKING PAPERS

10087
2022

November 2022

Building Daily Economic Sentiment Indicators

Pilar Rey del Castillo

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10087*

Building Daily Economic Sentiment Indicators

Abstract

The availability of copious amounts of data produced by the increasing datification of our society is nowadays deemed an opportunity to produce timely and convenient statistical information. This paper shows the building of economic sentiment indexes from the texts of the most read economic newspapers in Spain. The data are collected through the scraping of the Digital Periodical and Newspaper Library website. To compute the sentiment, an existing emotional lexicon for Spanish words has been customized, allowing for inferring sentiment for words in texts. The resulting indexes are later compared to other well-known indicators that try to monitor similar or related phenomena.

JEL-Codes: C180, C430, C550, C890.

Keywords: index numbers, large datasets, leading indicators, proxy variables, sentiment analysis, web scraping.

Pilar Rey del Castillo
Institute for Fiscal Studies
Avda. Cardenal Herrera Oria, 378
Spain - 28035 Madrid
mpilar.rey@ief.hacienda.gob.es

November 6, 2022

1. Introduction

Monitoring the economic activity of a country or of a region is a crucial issue for policy implementations and actions of economic agents. Different statistical indicators are built with this aim, using both quantitative and qualitative approaches and by different government and private institutions.

The increasing datification of the current society, with huge amounts of data about many human, social and economic activities, is nowadays deemed an opportunity to produce timely and useful statistical information. However, the use of these sources of data is not a direct task, involving a detailed analysis about the possible outcomes and the levels of quality required. The exploration of the data is often made difficult by the fact that most of them are not publicly available, but are in the hands of private companies that avoid disclosing information about their operation. Because of this, the construction of indicators from Open Data sources, which are freely available to everyone without restrictions from copyright, patents or other mechanisms of control, becomes a motivating opportunity.

Being also aware of the influence of economic narratives on the progress of the economy itself (Shiller, 2017), newspapers are proposed as the primary source of information. As what is tried to be measured refers to the economic activity, the texts of the most read economic newspapers are considered, in particular those that are freely available at the Digital Periodical and Newspaper Library website[1]. The scraping of the pages on this website obtains a selection of texts that are later processed.

The opinions and sentiments expressed in texts written in natural language are nowadays analyzed using Natural Language Processing (NLP) techniques. There has been widespread research in the field to estimate the sentiment polarity and strength in texts. Most of it has been focused on public opinions and reviews of specific products for deriving useful knowledge to improve sales

[1] https://www.bne.es/en/catalogs/digital-periodical-and-newspaper-library

performance (Patil and Atique, 2015; Rambocas and Pacheco, 2018). But our objective is just to compute summary indexes to obtain a rough idea on the evolution across time of the average opinion over the economic perspectives in the country's newspapers. Therefore, the proposed method does not address the full possible complexity in NLP. A naïve and straightforward method, *keyword spotting* (Cambria *et al.*, 2013) is used to compute sentiment scores for the texts. It classifies texts by sentiment categories or scores based on the presence of unambiguous sentiment related words such as happy, sad or lucky. The method may be prone to errors due to its difficulties in detecting complex nuances of emotions such as irony, sarcasm and others in the texts. But, in the case of estimating an aggregate sentiment from multiple scores, the errors may cancel out and keyword spotting can be more precise than other techniques designed to optimize per-document scoring (Hopkins and King, 2010).

To apply the method, a sentiment lexicon including the sentiment polarity for words is needed. The polarity scores proposed are obtained from an existing emotional lexicon for Spanish words which has been adapted and customized to calculate a sort of sentiment polarity. These scores are later averaged and summarized for building daily sentiment indicators at the country and newspaper level.

The remainder of this paper is organized as follows: the next section describes the data collection through the process of scraping the Digital Periodical and Newspaper Library website; Section 3 then shows the proposed model for computing the sentiment polarity on the selected texts from the emotional lexicon, and how the polarities are summarized; Section 4 presents the calculated indexes; section 5 compares them to other similar or related indicators, and finally, a number of remarks and conclusions are presented in Section 6.

2. Data collection

The primary intention is to build an indicator monitoring the general sentiments and opinions about the economic perspectives as reflected in business newspaper articles, trying to obtain a rough idea on the movements over time. Thus, the main data collected consist of the texts of articles published in these newspapers.

The Digital Periodical and Newspaper Library is part of the Hispanic Digital Library project, whose aim is public reference and internet dissemination of the bibliographic heritage held in the Spanish National Library. It gives access to the digital collection of newspapers and magazines, both historical and current, and the press documents deposited daily by the editors. Open access publications can be consulted without restrictions but those with intellectual property rights in force can only be consulted from the Spanish National Library facilities. In particular, the daily publications selected, the three most read national business newspapers in Spain (*Cinco Días, El Economista* and *Expansión*) are not free of rights and have restricted access, what makes it not possible to download directly the complete content of the articles.

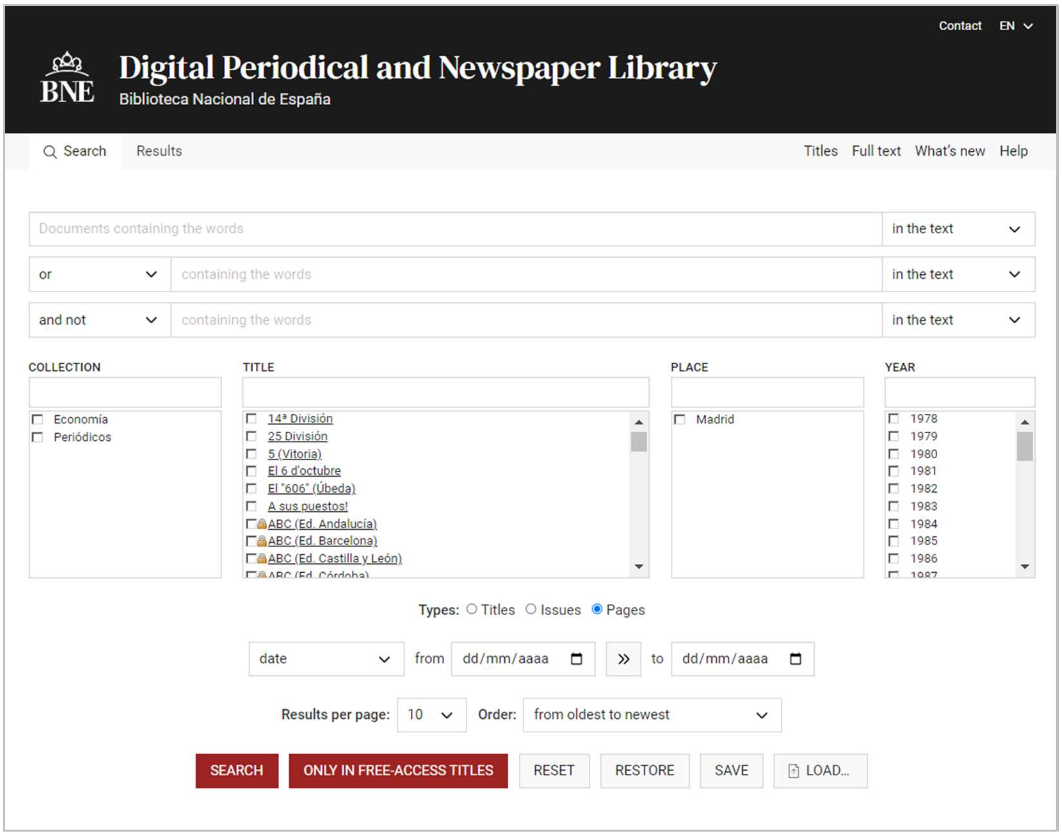
But the application that manages the collection allows searches for any term in the texts and provides, for these restricted publications, some information for the users to assess if they are interested in moving to consult them, request a copy or locate the copy in other sources.

Therefore, once some terms for restricting the topic of interest have been introduced in the search page (Figure 1), for each result, among others, the following information is presented:

- Title of the newspaper
- Date
- Page
- Fragments of text that precede and follow the search terms and clarify its meaning.

The keywords chosen to restrict the searches are terms related to the economic situation such as (in Spanish language): "fiscalidad" or "de hacienda" or words starting by "economi" or "tribute", and similar items. The software Selenium (Sharma, 2020) for web scraping (automated processes of data extraction) has been used for collecting the data from the pages of search results.

Figure 1. Digital Periodical and Newspaper Library website



Once these data are conveniently extracted, the small text fragments obtained are subject to sentiment analysis and daily indicators are built in a way that is described in the next section.

3. Model for computing the indexes

The idea is that the fragments of text preceding and following the topic terms should concentrate the sentiments and opinions on the subject. The method has been previously described in a paper that applies it to texts from Twitter (Rey-del-Castillo, 2019). Thus, the indexes are built following three steps:

3.1. Assignment of sentiment scores to words

The Emotional Norms for Spanish Words (ENSW) (Stadthagen-Gonzalez *et al.*, 2017) is a lexicon developed for the research on the emotional aspects of the language. It presents details on the affective properties of 14,031 Spanish words collected by the subjective ratings of a number of undergraduate psychology students. The ratings reflect the experiences and associations with particular objects, events and abstract words. In particular, apart from other features, two fundamental dimensions are rated for each word:

- Valence: reflects the level of pleasantness associated with a word, and can be linked to the direction or quality of the sentiment (ranging from pleasant 9 to unpleasant 1).
- Arousal: shows the level of activation or intensity a word elicits (ranging from quiet 1 to active 9).

The ENSW shows the mean values and the standard deviation of the ratings. From this information, the lexicon is customized to quantify the polarity of words using their affective characteristics. For this purpose, the first step is to lineally transform the arousals to make them take values between 0 and 1, from minimum to maximum intensity, to use them as weights later. Secondly, 5 (median value for the range of valence) is subtracted from the original valences to move them to the interval [-4, 4], trying to reflect the opposition between negative and positive values. Multiplying then the new valences by the corresponding previously computed weights, an evaluation of the strength of the sentiment for each word is obtained. Thereby, the weight modifies the value of the valence in the appropriate direction, from minimum to whole intensity. The last step is to move these measures to the interval [0, 100] to obtain the final sentiment values. Thus, the expression of the computed score for each word is:

$$s(w) = \frac{\left[\left(\frac{arousal(w)-1}{8}\right).(valence(w)-5)\right]+4}{8}.100 \tag{1}$$

were the valence and arousal for the word *w* are represented by the mean values in ENSW. The scores proposed combine and encompass the two dimensions –quality (valence) and intensity (arousal) of the sentiment– in a single number between 0 and 100. With this procedure, polarity scores close to 50 represent neutral sentiments; words expressing positive feelings or opinions are assigned higher scores, and words with a negative connotation are assigned lower.

The text fragments collected by scraping are studied as emotionally loaded opinions and are accordingly subject to sentiment analysis.

3.2. Assignment of polarities to text fragments

What this step accomplishes is the assignment of a score between zero and a hundred to each text reflecting its degree of sentiment. The fragments of text are previously processed by tokenizing into words, converting to lowercase, removing special characters, and removing stop words (Liu, 2020). The assignment of the polarity is then simply executed through the detection of the presence of ENSW words. Text fragments not including any of these words are discarded. For each combination of newspaper, day and page in the search results, the score values of the detected words in the corresponding text fragments are later combined by computing their average, as it is usually done (Gupta, Abhinav and Annapa, 2013; Hogenboom *et al.*, 2013) when combining sentiment values of words in texts. Table 1 shows some examples of the operation of the polarity assignment to text fragments.

Table 1. Examples of the polarity sentiment assignment

Text Fragment	Tokens	Tokens in ENSW	Scores	Final Polarity
"… recuperar estas horas será muy difícil, y han alertado de la situación de quiebra económica en que se…"	[recuperar, económica, situación, alertado, horas, complicado, quiebra]	[recuperar, situación, difícil, económica, quiebra]	[60.4, 50.0, 40.6, 54.2, 25.5]	46.1
"… especial, es, pues, un merecido tributo a lo mejor de nuestra clase empresarial y un mensaje de optimismo…"	[tributo, mensaje, optimismo, merecido, empresarial, especial, clase, mejor]	[tributo, mensaje, optimismo, empresarial, especial, clase, mejor]	[55.0, 57.9, 75.2, 47.4, 74.4, 54.1, 66.7]	61.5
"… Ecofin será la fiscalidad digital y el debate para consensuar en la OCDE y el G-20 esta tasa…"	[digital, fiscalidad, ecofin, consensuar, ocde, tasa, debate]	[digital, fiscalidad, consensuar, tasa, debate]	[53.2, 44.1, 56.7, 47.2, 51.7]	50.6
"… habrá oído o leído que España tiene, dentro de la Unión Europea, una baja fiscalidad, es decir que los españoles pagamos menos…"	[unión, dentro, oído, España, fiscalidad, europea, menos, decir, leído, pagamos, españoles, baja]	[unión, dentro, oído, fiscalidad, europea, menos, decir, baja]	[56.9, 55.8, 53.8, 44.1, 58.3, 41.0, 56.7, 45.2]	51.5

3.3. Calculation of summary indexes

Summary indexes are calculated to obtain information about the trend of the economic sentiment during the observation period, for each one of the newspapers and also for the whole newspapers scope. They are calculated for each day in the simplest way as the mean values of the text fragments polarities computed at the preceding step.

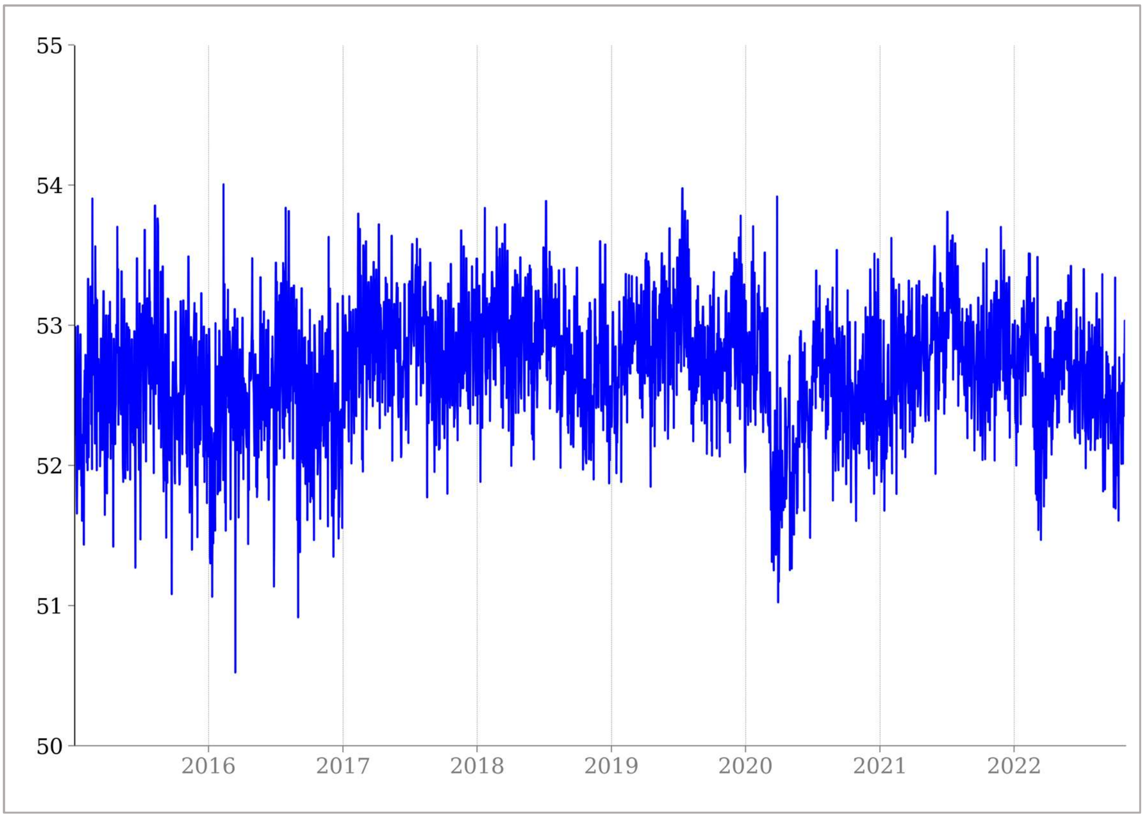
Other papers have previously shown the computation of aggregate sentiments over time based on sentiment scores for texts. For example, O'Connor et al. (2010) build an aggregate opinion index on a topic counting positive and negative tweets containing a topic keyword, based on a list of positive and negative words; Das and Chen (2004) compute an aggregate sentiment index to analyze stock behavior based on text from investor message boards; and Darena et al. (2018) compute sentiment indexes from documents online to study the association of online text related to a company and the movements of the stock prices of that company. The novelty of our approach lies in using a bigger lexicon with gradual polarities allowing for more subtleties and nuances. It is also relevant that the sentiment refers to the feelings towards the economy based on business newspapers articles.

4. Resulting indicators

Following the procedures described in the preceding section, sentiment indexes are calculated daily for each one of the newspapers from 2015, having on average more than 500 text fragments for each day. The indexes for the whole scope are computed similarly in a direct way, using the texts of the articles of each day for all newspapers, and not as an average of the corresponding newspapers indexes. All data processing and analysis have been performed using several packages of Python software (Pedregosa *et al.*, 2011; Python Core Team, 2018).

The first point to be made is that *Cinco Días* and *Expansión* are not published on Sundays, and *El Economista* is not published on Sundays or Mondays. Consequently, indexes cannot be computed on these days for the corresponding newspapers. Once the indexes are calculated for the days of the week available, what has been done is to impute values for missing Sundays and Mondays. By doing so, the series might potentially have up to 4 periodic components: a weekly cycle, a monthly cycle, a quarterly cycle and an annual cycle, and they can be compared to other daily series. The method for imputing the missing indexes is "Last Observation Carried Forward" (LOCF) (Ahn, Sun and Pio Kim, 2021). Another point to note is that monthly indexes are calculated by simply averaging the indexes for the days of the week available, not including imputed values.

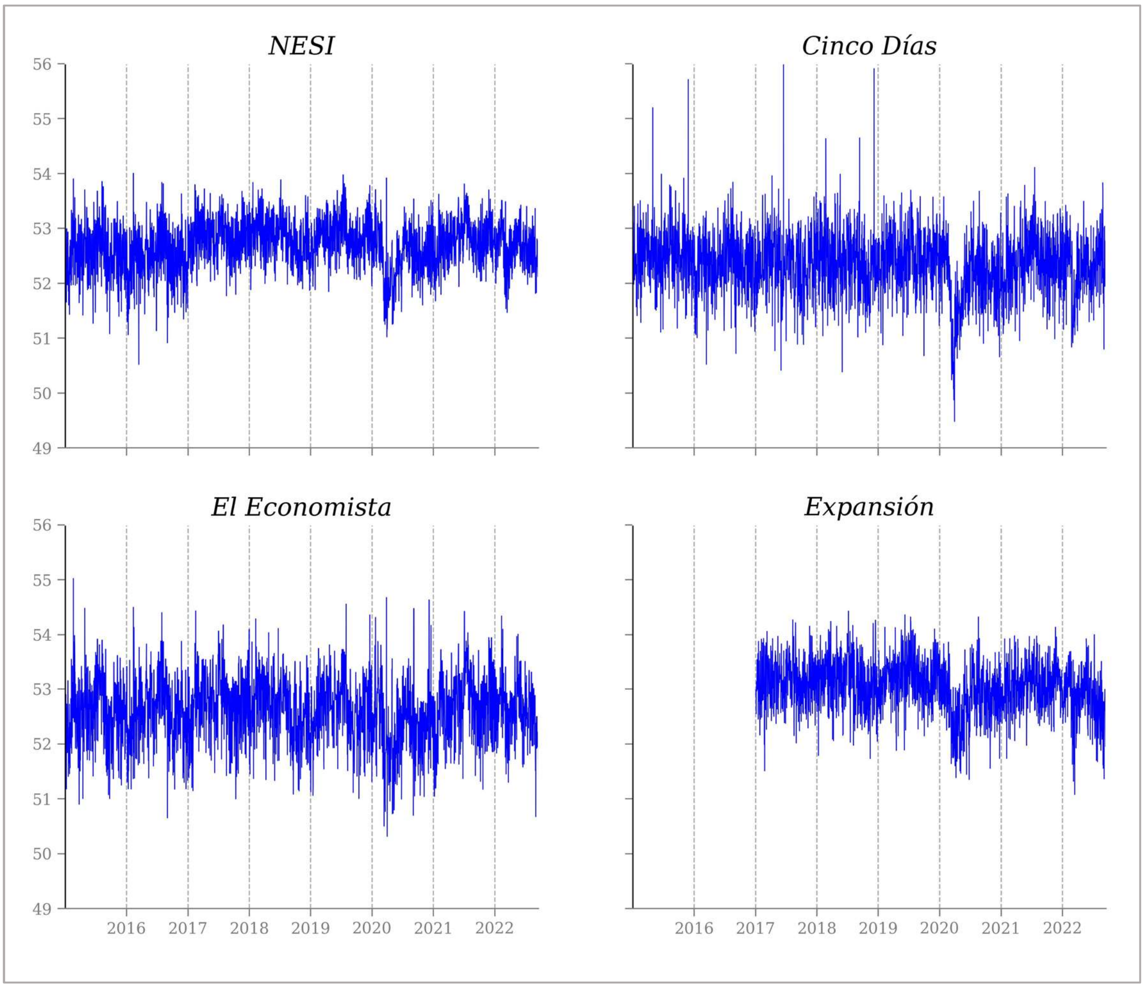
Figure 2. Newspapers Economic Sentiment indicator (*NESI*)



The resulting Newspapers Economic Sentiment Indicator (*NESI*) series for the whole newspapers scope and the period between January 2015 and October 2022 is shown in Figure 2. It can be seen that the day-to-day movement has a lot of noise −with a great number of peaks and troughs− especially at the beginning, where the number of text fragments were lower.

Figure 3 presents the series for all newspapers, using the same scale to compare their evolutions (*Expansión* is only available from 2017). One common characteristic for all of them is the declines and rises due to the coronavirus pandemic and the Russian invasion of Ukraine, with different degrees of incidence. Another fact to observe is that the aggregate *NESI* is smoother when *Expansión* newspaper enter as a component from 2017.

Figure 3. Newspapers economic sentiment indicators



A general idea about their behavior may also be extracted from Table 2.

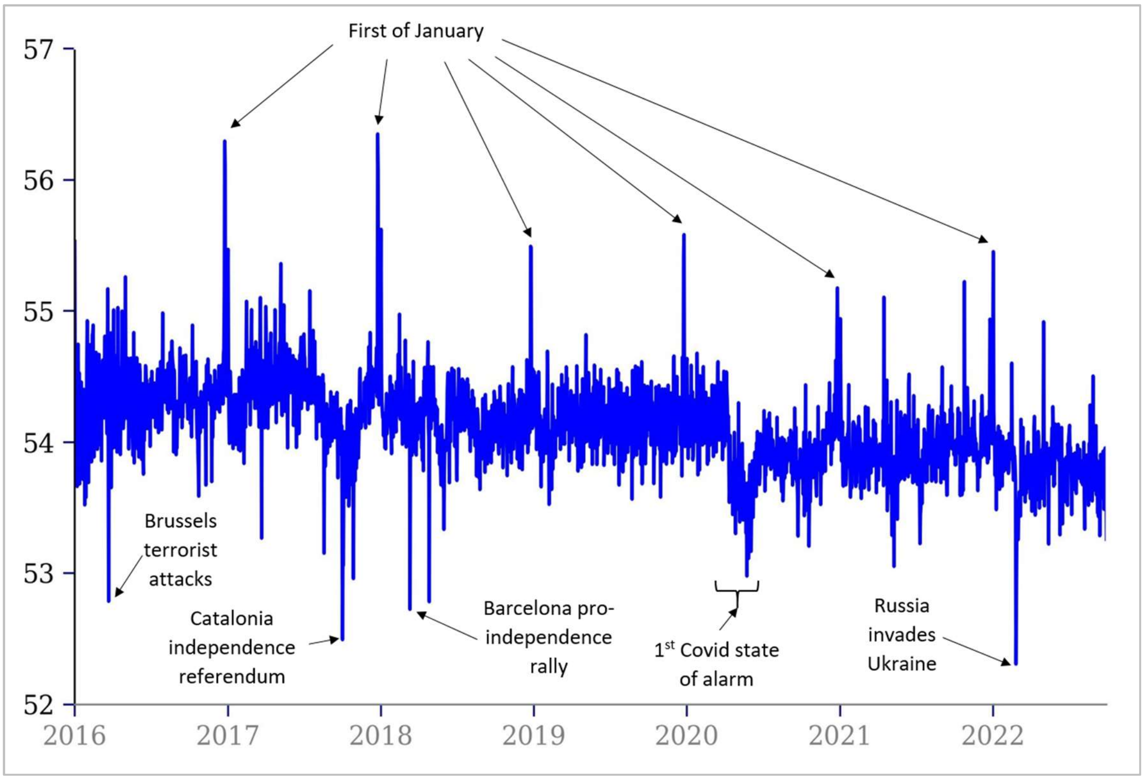
Table 2. Summary characteristics of the indicators

Index	Mean	Std	Range	Min	Median	Max
Cinco Días	52.4	0.60	6.7	49.5	52.4	56.2
El Economista	52.7	0.64	4.7	50.3	52.7	55.0
Expansión	53.1	0.51	3.4	51.1	53.1	54.4
NESI	52.7	0.47	3.5	50.5	52.7	54.0

It can be seen the diverse newspapers series range, spanning from *3.4* for *Expansión* to *6.7* for *Cinco Días*, suggesting different levels of moderation or extremism in the way to express their opinions. Likewise, the most negative average opinion is obtained for *Cinco Días* and the most positive for *Expansión*, being all of them very similar.

It is difficult to evaluate the quality of the indicators produced because the concept whose evolution we are trying to observe –the average sentiment towards the economy shown in the economic newspapers– is an abstract one, lacking a precise definition. This raises several issues about its accuracy and potential bias. Next section compares the *NESI* with the *GDP* evolution and other indicators to assess its usefulness. Nevertheless, as another way to evaluate the general ability of the method to detect the sentiments expressed in natural language texts, it may be interesting to analyze and compare the results of using the same model with a different data source.

Figure 4. Twitter sentiment indicator



A daily general sentiment index exploiting messages posted in Spain has been built from the popular social networking Twitter (Rey-del-Castillo, 2019). Figure 4 presents the constructed sentiment index from 2016. Its aim is to monitor the people's general mood, sentiments and opinions in Spain, with no specific topic in mind (the series is clearly noisier at the beginning because there had less tweets).

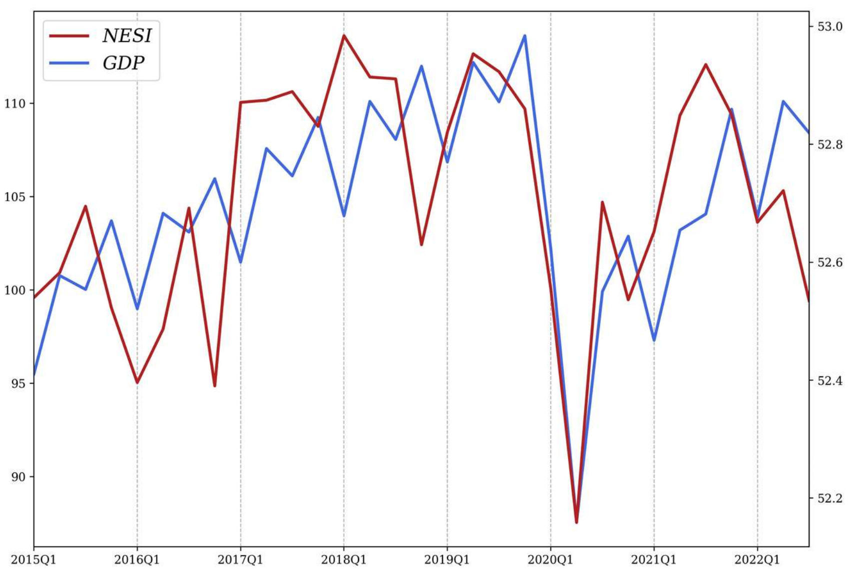
The three steps described in Section 3 have been followed using the texts of the messages, known as *tweets*, which have a 280 characters limit. This restriction forces users to extract what they want to express in a small number of sentences and words. The *keyword spotting* method to detect sentiments can be especially pertinent for these texts of reduced size, the same way it happens with the fragments of text close to the search keywords in our case.

Although the Twitter sentiment index has about 10 times more texts than *NESI* has, the first presents more outliers, with values significantly larger or smaller than the rest of the values. Most of them correspond to special events as indicated in the figure. *NESI* shows no large spikes in connection with these major events but has little immediate reactions and sometimes begins to respond later. Without going into more details, a quick outcome is that individuals react harder and more swiftly than economic newspapers to specific social, political or economic events. And, something to be expected, that newspapers refrain from expressing extreme views.

5. Comparison to other indicators

An evaluation of the quality of the computed indexes and its possible usefulness can be done by comparing them to others measuring the same or similar concept.

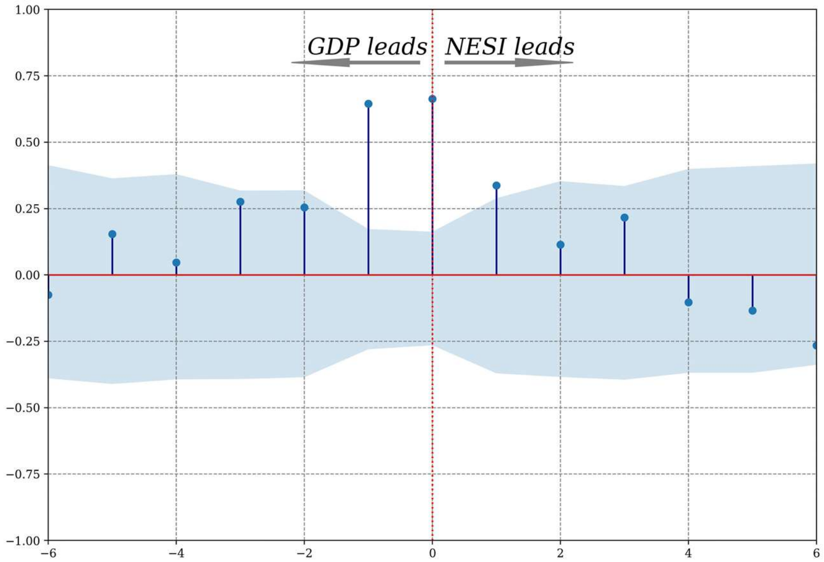
Figure 5. Quarterly *NESI* and *GDP*



The main indicator of economic activity is the Gross Domestic Product (*GDP)*, available with quarterly frequency. Figure 5 presents the quarterly *NESI* (computed as the monthly *NESI* average) jointly with the *GDP* volume growth (in %).

A certain association between them can be seen. The results of the Augmented Engle-Granger test of co-integration (Engle and Granger, 1987) also support that there is a long run relationship between both indicators. This is confirmed in Figure 6, where its cross-correlations (Box, Jenkins and Reinsel, 1994) are shown. The graph includes a shaded area where cross-correlation values are not significantly different from 0 at the 95% confidence level. Significant positive values at lags near to 0 can be observed, being especially notable the contemporary correlation (0.7). Therefore, *NESI* provides a good proxy for economic activity.

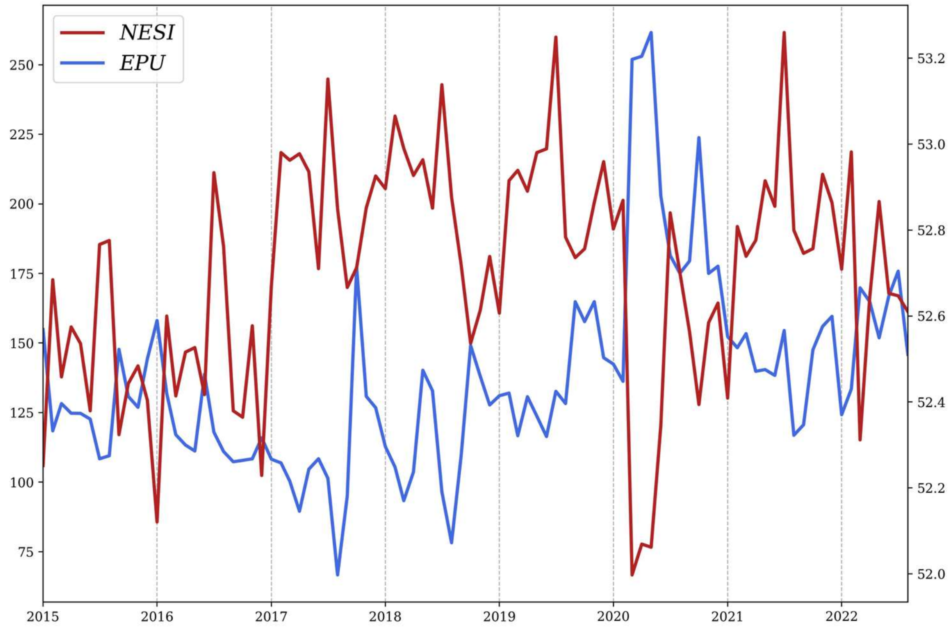
Figure 6. Cross-correlations between quarterly *NESI* and *GDP*



There are also some possible candidates to compare on a monthly frequency basis (there are no known indicators quantifying anything equivalent at a daily granularity level). For example, the European Policy Uncertainty (*EPU*) monthly index (Ghirelli, Pérez and Urtasun, 2019), developed at the Bank of Spain, is a well-known indicator of economic policy uncertainty based on newspaper coverage frequency of 7 relevant Spanish national newspapers (the four most read generalist newspapers: *ABC, El Mundo, El País*, and *La Vanguardia*, and the three headline

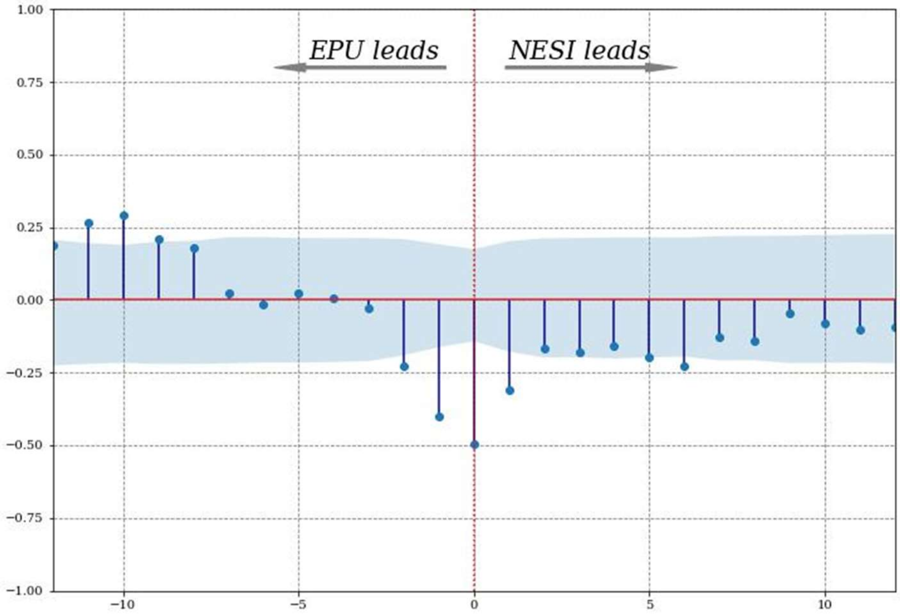
Spanish business newspapers: *Cinco Días, El Economista,* and *Expansión*). Its construction follows closely a previous paper (Baker, Bloom and Davis, 2016). It counts the number of articles containing simultaneously at least one keyword related to the categories of "uncertainty", "economy", and "policy". Figure 7 presents the *EPU* jointly with the monthly *NESI* index.

Figure 7. Monthly *NESI* and *EPU*



The *EPU* and the monthly *NESI* index have the information of three common newspapers but, on the other side, they use completely different instruments to measure, one counting the number of articles containing simultaneously keywords related to the uncertainty, and the other assigning economic sentiments to articles using a polarized lexicon. Nevertheless, the Figure 7 shows that their lines are almost symmetrical (especially from 2017, when *Expansión* starts to be available), suggesting they measure two opposing phenomena. Besides, the results from the Augmented Engle-Granger test of co-integration imply a long run relationship between both indexes, confirmed also in Figure 8 by the significant negative cross-correlations at lags near to 0. Based on this, *NESI* can be interpreted as an economic policy confidence indicator, with a behavior in opposition to the uncertainty measured by *EPU.*

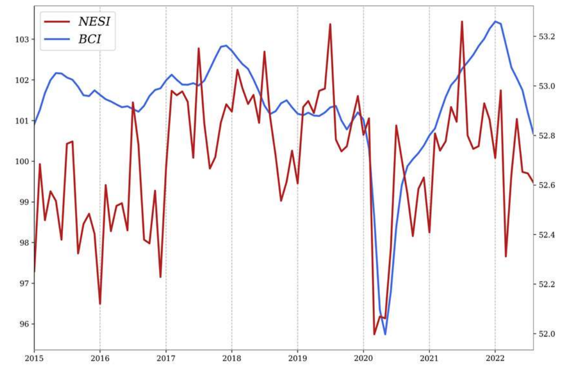
Figure 8. Cross-correlations between monthly *NESI* and *EPU*



The Organization for Economic Cooperation and Development (OECD) elaborates two monthly indicators that can have also a relationship with monthly *NESI*: the Business Confidence Index (*BCI*) and the Consumer Confidence Index (*CCI*). The *BCI* is based on opinion surveys about production, orders and stocks of finished goods in the industry sector, and monitors economic activity (OECD, 2022a). For its part, the *CCI*, based on household opinion surveys, provides an indication of household consumption and saving behavior (OECD, 2022b).

Figure 9. Graphs of monthly *NESI* series jointly with *OECD* series

(a). *BCI* **(b). *CCI***



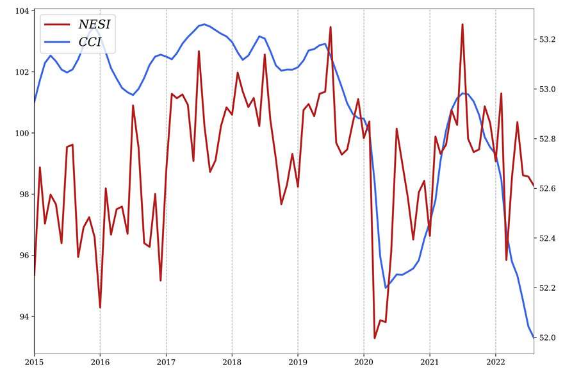
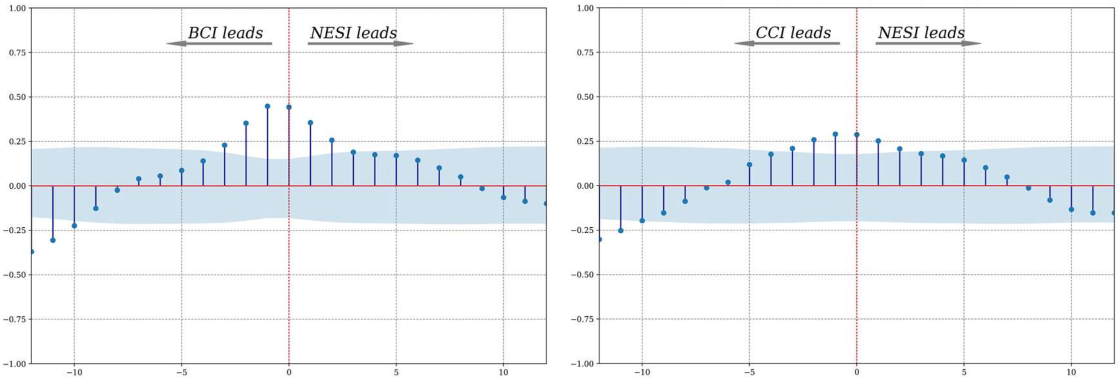


Figure 9 displays the graphs of monthly *NESI* jointly with *BCI* and *CCI.* A certain association between monthly *NESI* and both of them can be seen, having some common ups and downs. The Augmented Engle-Granger tests give also evidence of co-integration relationships and, lastly, in Figure 10, it is shown that the association of monthly *NESI* with both indexes is positive, being stronger with the *BCI,* as would be expected.

Figure 10. Cross-correlations between monthly *NESI* and *OECD* series

(a). *BCI* **(b). *CCI***



6. Final remarks and conclusions

The preceding sections have presented the building of economic sentiment indexes based on the scraping of the Digital Periodical and Newspaper Library website to collect fragments of text of business economic articles in Spain. It has been shown that the proposed *NESI* indicator contains useful information about the economic activity because it is highly correlated with the quarterly *GDP* volume growth.

Drawing causal inferences for predicting future economic activity from our index can be extremely challenging. But *NESI* has the advantage that it can be immediately calculated for the preceding day or month and can provide valuable data on economic activity in advance of the *GDP* publication date. Another interesting feature of our indicator is that it is built from public information, without the need to proprietary data.

The comparison of *NESI* to *EPU* index (an indicator of economic policy uncertainty) provides an interpretation of our index as an economic policy confidence indicator. The reason is that it operates over time in opposition to the uncertainty measured by *EPU*, as shown by the strong negative relationship between them.

With respect to other proxies for economic activity such as the ones elaborated by the OECD from opinion surveys, *NESI* presents a relevant association with *BCI,* while it is less noticeable with *CCI*.

References

Ahn, H., Sun, K. and Pio Kim, K. (2021) 'Comparison of missing data imputation methods in time series forecasting', *Computers, Materials & Continua, vol. 70, no.1*, 70(1), pp. 767–779. Available at: https://doi.org/10.32604/cmc.2022.019369.

Baker, S.R., Bloom, N. and Davis, S.J. (2016) 'Measuring Economic Policy Uncertainty*', *The Quarterly Journal of Economics*, 131(4), pp. 1593–1636. Available at: https://doi.org/10.1093/qje/qjw024.

Box, G.E., Jenkins, G.M. and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control 3rd edition*. Edited by Prentice Hall.

Cambria, E. *et al.* (2013) 'New Avenues in Opinion Mining and Sentiment Analysis', *Intelligent Systems IEE*, 28(2), pp. 15–20. Available at: https://doi.org/10.1109/MIS.2013.30.

Darena, F. *et al.* (2018) 'Machine Learning-Based Analysis of the Association between Online Texts and Stock Price Movements', *Inteligencia Artificial*, 21(61), pp. 95–110. Available at: https://doi.org/10.4114/intartif.vol21iss61.

Das, S.R. and Chen, M.Y. (2004) 'Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web', *Management Science*, 53(9), pp. 1375–1388.

Engle, R.F. and Granger, C.W.J. (1987) 'Co-Integration and Error Correction: Representation, Estimation, and Testing', *Econometrica*, 55(2), pp. 251–276. Available at: https://doi.org/10.2307/1913236.

Ghirelli, C., Pérez, J.J. and Urtasun, A. (2019) 'A new economic policy uncertainty index for Spain', *Economics Letters*, 182, pp. 64–67. Available at: https://doi.org/https://doi.org/10.1016/j.econlet.2019.05.021.

Gupta, N., Abhinav, K.R. and Annapa (2013) 'Fuzzy sentiment analysis on microblogs for movie revenue prediction', in *International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)*, pp. 1–4. Available at: https://doi.org/10.1109/CS2PA.2013.6749350.

Hogenboom, A. *et al.* (2013) 'Exploiting Emoticons in Sentiment Analysis', in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pp. 703–710.

Hopkins, D. and King, G. (2010) 'A Method of Automated Nonparametric Content Analysis for Social Science', *American Journal of Political Science*, 54(1), pp. 229–247.

Liu, B. (2020) ‘Studies in Natural Language Processing’, in Bing Liu (ed.) *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 2nd edn. Cambridge: Cambridge University Press (Studies in Natural Language Processing), pp. ii–ii. Available at: https://doi.org/DOI: undefined.

O’Connor, B. *et al.* (2010) ‘From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series’, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Washington, pp. 122–129.

OECD (2022a) *Business confidence index (BCI) (indicator)*, *Business confidence index (BCI) (indicator)*. Available at: https://doi.org/10.1787/3092dc4f-en.

OECD (2022b) *Consumer confidence index (CCI) (indicator).* , *Consumer confidence index (CCI) (indicator).* Available at: https://doi.org/10.1787/46434d78-en.

Patil, H.P. and Atique, M. (2015) ‘Sentiment Analysis for Social Media: A Survey’, in *2015 2nd International Conference on Information Science and Security (ICISS)*, pp. 1–4. Available at: https://doi.org/10.1109/ICISSEC.2015.7371033.

Pedregosa, F. *et al.* (2011) ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Python Core Team (2018) ‘Python: A dynamic, open-source programming language’, *Python Software Foundation* [Preprint].

Rambocas, M. and Pacheco, B.G. (2018) ‘Online sentiment analysis in marketing research: a review’, *Journal of Research in Interactive Marketing*, 12(2), pp. 146–163. Available at: https://doi.org/10.1108/JRIM-05-2017-0030.

Rey-del-Castillo, P. (2019) ‘A sentiment index based on Spanish tweets’, *BEIO Boletín de Estadística e Investigación Operativa*, 35(2), pp. 130–147.

Sharma, P.R. (2020) *Selenium with Python - A Beginner’s Guide*. BPB Publications.

Shiller, R.J. (2017) ‘Narrative Economics’, *American Economic Review*, 107(4), pp. 967–1004. Available at: https://doi.org/10.1257/aer.107.4.967.

Stadthagen-Gonzalez, H. *et al.* (2017) ‘Norms of valence and arousal for 14,031 Spanish words’, *Behavior Research Methods*, 49(1), pp. 111–123. Available at: https://doi.org/10.3758/s13428-015-0700-2.