Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Can news help measure economic sentiment? An application in COVID-19 times

Pablo Aguilar, Corinna Ghirelli \(^\ast\), Matías Pacce, Alberto Urtasun

Bank of Spain (Eurosystem), Spain

**Abstract**

We construct a new newspaper-based sentiment indicator for Spain that allows to monitor economic activity in real-time. As opposed to survey-based confidence indicators that are released at the end of the month, our indicator can be constructed on a daily basis. We compare our index with the popular Economic Sentiment Indicator of the European Commission and show that ours performs significantly better in nowcasting the Spanish GDP. Moreover, it proves to be helpful to predict the current COVID-19 recession from an earlier date.

1. Introduction

Benchmark data to assess economic activity are available with some publication lag. In this context, textual analysis has enabled the development of new sources of data, including news media, to monitor real-time economic activity (e.g. Shapiro et al., 2020; Thorsrud, 2020; Combes et al., 2018). Following this literature, we propose a new daily economic news sentiment indicator (DENSI) based on newspaper data for Spain.

To construct our measure, we rely on the procedure used for the influential economic policy uncertainty index by Baker et al. (2016)\(^1\) exploiting a large database of Spanish press. In a nutshell, our indicator reflects the balance between articles that contain keywords related to upturns and downturns of the Spanish business cycle.

Our index is shown to closely follow the dynamics of one of the most popular survey-based confidence indicator – i.e. the Economic Sentiment Indicator (ESI) of the European Commission – while it displays three major advantages when compared to these kind of indicators: (1) it does not suffer from bias along business cycles, which happens if agents adjust their growth expectations during recessions (Gayer and Marc, 2018); (2) it can be constructed on a daily basis and in real-time – each day, the index can be computed using data which refer up to the previous day – while the surveys refer, in general, to the first weeks of the month of reference and are published in the last days of month\(^3\); (3) it does not depend on the survey’s response rate which could lead to sampling problems. The usefulness of the

---

\(^1\) We thank Silvia Albrizio, Angel Gaviñan, José Gonzalez, Danilo Leiva-León, Eva Ortega, Javier Pérez, and Diego Torres for their comments and suggestions. We also thank all participants at the 40\(^\ast\) International Symposium on Forecasting and at the internal seminar of the Bank of Spain for their helpful comments. The views expressed in this paper are our own and do not necessarily reflect the views of the Bank of Spain or the European System of Central Banks (ESCB).

\(^\ast\) Correspondence to: Economic Developments Department, Bank of Spain, Calle Alcalá 48, 28014 Madrid, Spain.

E-mail addresses: pablo.aguilar@bde.es (P. Aguilar), corinna.ghirelli@bde.es (C. Ghirelli), matias.pacce@bde.es (M. Pacce), aurutasun@bde.es (A. Urtasun).

\(^2\) This could happen, for example, if business managers and consumers adapt their economic expectations to a more modest growth. Such a bias became evident after the end of the Great Recession, when average economic growth was significantly lower than that observed during the pre-crisis period but the ESI stood at similar average levels in both periods.

\(^3\) During March 2020, three ESI surveys out of five were carried out between the 2nd and the 12th, while the other two surveys were accomplished between the 2nd and the 23rd.
last two mentioned advantages are particularly relevant since the emergence of the COVID-19 crisis.

To analyze whether the DENSII could do a better job than survey-based confidence indicators regarding the assessment of the economic sentiment, we perform two exercises. First, we show that a model including both the DENSII and GDP significantly improves the GDP forecast accuracy when compared to a model that includes the ESI and GDP as the main variables. This result suggests that the DENSII could provide a better signal for an earlier evaluation of the economic situation than the ESI. Second, we use both indicators to compute the implied probability of recessions at short-term horizons. Results show that the DENSII does better than the ESI in predicting the economic crisis due to the COVID-19 outbreak.

Therefore, on top of providing evidence that newspaper-based indicators can better deal with some of the limitations of survey-based confidence indicators, we contribute to the nowcasting literature by proposing a new newspaper-based sentiment indicator for Spain, which could be obtained on a daily basis at any point during the month. This becomes critical when important shocks occur in the middle of the month as it was the case with the COVID-19 outbreak. Moreover, we show that, in line with the literature, newspaper-based confidence indicators can improve classical confidence indicator measures when assessing the more recent economic developments.

In the rest of the paper we describe our DENSII (Section 2), two empirical applications (Section 3), and offer some concluding remarks (Section 4).

2. Description of the index

We build a sentiment indicator capturing the economic tone of news published in the Spanish press. In a nutshell, it reflects the balance between the number of articles containing keywords related to upturns and downturns in the Spanish business cycle. We consider 7 relevant Spanish national newspapers: El País, El Mundo, La Vanguardia, ABC, Expansión, Cinco Días, and El Economista. The first 4 are the most-read generalist newspapers in Spain, while Expansión, Cinco Días, and El Economista are the three main Spanish business newspapers.

All searches are carried out using the Dow Jones’ Factiva service. For each newspaper, we conduct queries from the first month the newspaper is collected in the Dow Jones’ Factiva database, starting from January 1997. We restrict all queries to the following articles: (i) in the Spanish language; (ii) with content related to Spain, based on Factiva’s indexing; (iii) about corporate news, economic news, or news about financial markets, according to Factiva’s indexing. We then perform two types of queries.

First, we count the number of articles that, in addition to satisfying the aforementioned conditions, contain upswing-related keywords: i.e., they contain the word recuperación or one of the following words, if preceded or followed by either económico/economía within a distance of 5 words: aeler*, cres*, expans*, increment*, aumento*, recuper*, mejora*. To ensure that the news items are about the Spanish business cycle, we also require the article to contain the word España.

Second, we count the number of articles that, in addition to satisfying the aforementioned conditions, are about downswings. In particular, the articles must contain the word recesión or crisis or one of the following words, if preceded or followed by either económico/economía within a distance of 5 words: descen*, disminu*, redu*, ralentiz*, decrec*, desaceler*, contracción*. The articles should also contain the word España.

Therefore, on top of providing evidence that newspaper-based indicators can better deal with some of the limitations that typical suffer survey-based confidence indicators, we contribute to the nowcasting literature by proposing a new newspaper-based sentiment indicator for Spain, which could be obtained on a daily basis at any point during the month. Moreover, we show that, in line with the literature, newspaper-based confidence indicators can improve classical confidence indicator measures when assessing the more recent economic developments.

To construct the index, we follow the procedure used by Baker et al. (2016). We scale the difference between upturn and downturn counts by the total number of economic articles in the same newspaper/period — either a daily or a monthly period, standardize the series of scaled counts, average them across newspapers, and rescale the resulting index to mean 0.

Fig. 1 shows the 7-day moving average of the daily DENSII. A major advantage of newspaper-based indexes compared to the traditional monthly confidence indicators is that each day they incorporate new information in the construction of the index. This becomes crucial when important shocks occur in the middle of the month or far from the date at which the monthly confidence indicators are released, i.e. when Spain was locked down on the 14th of March due to the COVID-19 outbreak.

Fig. 2 shows our monthly DENSII against the ESI. Both indexes are highly correlated. However, their behavior differs dramatically in March 2020 after the COVID-19 outbreak in Spain. The DENSII signals the outbreak correctly and drops to values reached during the Great Recession, while the ESI remains at the positive
levels characterizing the recent months. This is because the ESI relies on answers mostly collected in the first half of the month, which do not capture the start of the economic crisis that occurred from the lockdown implemented on March 14th to fight against the spread of COVID-19.

3. Empirical analysis

In a first application we carry out the pseudo-real-time nowcasting exercise for GDP. We rely on mixed-frequency bi-variate vector autoregressive (MF-BiVAR) model, as in Mariano and Murasawa (2010), and compare the GDP forecast accuracy of a model that alternatively includes the ESI or the DENSi as monthly variables.\(^4\) We estimate this model by means of maximum likelihood estimation.

For our exercise, we assume it is the last day of each month \(t\) and use the information that would have been available at that moment (we use real-time vintages for GDP). Since the flash estimates for GDP are published 30 days after the end of the reference quarter, within a quarter, we predict the one-quarter-ahead GDP growth no matter whether it is the first, second, or third month of the quarter. We start the exercise in January 2013, and for each month until December 2019, we conduct the nowcast of quarter-on-quarter GDP growth.

The nowcasts are shown in Fig. 3. The blue and red lines represent the predictions based on the model with the ESI and DENSi, respectively, while the horizontal green line is the target value (second release of GDP).

To compare the forecast accuracy of both models, we rely on the forecast root mean squared error (RMSE). The RMSE of the Bi-VAR model that contains the DENSi is about 20% lower than that obtained under the model that includes the ESI, the difference being significant at the 1% confidence level.\(^5\) This indicates that the DENSi significantly improves the nowcast of GDP when compared with the ESI.

In our second application, we test whether our text-based sentiment indicator has predictive content in terms of business-cycle turning points. We start from a very standard model, where recession probabilities at future horizons are estimated as a (probit) function of the present value of the slope of the yield curve—i.e., the difference between short-term and long-term yields (e.g. Wright, 2006).

To measure the stance of the business cycle, we rely on the recession dates given by the Spanish Business Cycle Dating Committee (Spanish Economic Association, 2015). Departing from this model, we check whether forecast accuracy improves by adding alternatively the DENSi or the ESI as an additional regressor. We estimate these models up to a 12-month horizon considering the period from January 1997 until February 2020. Then we predict recession probabilities using both models during the COVID-19 outbreak.

Fig. 4 shows the estimated probabilities of recession from March 2020 up to 3, 6, 9, and 12 months ahead. To show the importance of having real-time updates of the DENSi, we evaluate the model using the information available on the 5th, 10th, 15th, 20th, 25th, and 31st of March. On each date, we assume that the value of the DENSi for March is the average value of the already known days of the month. Therefore, the value computed on the 31st of March matches the monthly value of the DENSi, while the values computed on earlier dates are approximations of that monthly value. As for the ESI, there is only one date available at the end of the month.

The estimations based on the DENSi adjust correctly to the rapid economic developments due to the COVID-19 outbreak. By March 14th, when the country was immediately locked down, the probability of recession increased, at least, by a factor of two at any horizon, and reached values near 1 during the second half of the month. Since most of the ESI’s surveys took place before the lockdown, it is not surprising that the recession probability for March based on the ESI remains very low. These results indicate that in the presence of big events, the DENSi would be able to capture their impact earlier than the ESI.

4. Conclusions

We propose a new indicator of economic sentiment based on newspaper articles that allows assessing real-time economic activity in Spain. As opposed to the survey-based confidence indicators, our indicator can be constructed on a daily basis. Our index yields significantly more precise GDP nowcast than the ESI. Moreover, the DENSi does better than the ESI in predicting the business cycle turning point during the COVID-19 economic crisis.

\(4\) In a first look at the explanatory power of both indicators, we find that, with a quarterly frequency, the ESI and the DENSi are equally good explanatory variables of GDP (that is, the adjusted R-squared of both regressions reaches a value of 0.76). However, this framework does not take into account the high frequency advantage of DENSi.

\(5\) The relative RMSE is about 0.797. To test whether the difference is significant, we rely on the Diebold and Mariano test.
Fig. 3. Quarterly GDP growth nowcast. Note: The black dotted (gray solid) line represents the predictions obtained from the model that includes the ESI (DENSI) and the GDP. The black solid line depicts the target variable (second release of GDP).

Fig. 4. Recession probabilities during the COVID-19 outbreak. Note: The x-axis represents the dates in March 2020 when we compute the probability of recession at different horizons based on the model with the DENSI: 3/2020 (black solid black line with asterisks), 6/2020 (black dashed–dotted line with triangles), 9/2020 (gray solid line with circles), 12/2020 (gray dashed line with diamonds), 3/2021 (gray dotted line with squares). Recession probabilities obtained with the model with the ESI are represented by the markers only at the end of the month (we use similar markers to the ones used for the DENSI model at each of the horizons).

References

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131(4), 1593–1636.
Caggiano, G., Castelnuovo, E., Figueres, J.M., 2017. Economic policy uncertainty and unemployment in the United States: A nonlinear approach. Econom. Lett. 151, 31–34.
Colombo, V., 2013. Economic policy uncertainty in the US: Does it matter for the Euro area. Econom. Lett. 121(1), 39–42.
Combes, S., Bortoli, C., Renault, T., 2018. Nowcasting GDP growth by reading newspapers. Economics and Statistics, N.505–506, 2018. Big Data and Statistics (Part 1).
Fontaine, I., Didier, L., Razafindravaoosolonirina, J., 2017. Foreign policy uncertainty shocks and US macroeconomic activity: Evidence from China. Econom. Lett. 155, 121–125.
Gayer, C., Marc, B., 2018. A “New Modesty”? Level Shifts in Survey Data and the Decreasing Trend of “Normal” Growth. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.
Ghirelli, C., Pérez, J.J., Urtasun, A., 2019. A new economic policy uncertainty index for Spain. Econom. Lett. 182, 64–67.
Mariano, R.S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real GDP. Oxford Bull. Econom. Statist. 72 (1), 27–46.
Shapiro, A., Sudhof, M., Wilson, D., 2020. Measuring News Sentiment, Federal Reserve Bank of San Francisco, Working Paper 2017–01.
Spanish Economic Association, 2015. CF Index of economic activity. Spanish Business Cycle Dating Committee.
Thorsrud, L.A., 2020. Words are the new numbers: A newsy coincident index of the business cycle. J. Bus. Econom. Statist. 38 (2), 393–409.
Wright, J.H., 2006. The Yield Curve and Predicting Recessions 2006/07. Federal Reserve Board, https://www.federalreserve.gov/Pubs/feds/2006/200607/200607pap.pdf.