

The Italian Social Mood on Economy Index and some Covid-19 Crisis Implications

Righi A., Iannaccone R., Zardetto D., Zurlo D. (ISTAT)

2020 Banca d'Italia and Federal Reserve Board Join Conference

«Nontraditional Data and Statistical Learning with Application to Macroeconomics»

11-12 november 2020

- In October 2018 Istat released of the Social Mood on Economy Index (SMEI), an experimental high-frequency sentiment index based on Twitter data. The daily time series of the index starts on February 10th 2016 and is continuously updated
- This new index, derived from samples of public tweets in Italian captured in real time, has already gained a good spread among economic analysts for short term analysis or forecasting models
- This paper is aimed at presenting this new statistical tool for economic analysis and to verify its relationships with some daily and monthly macroeconomic indicators also coming from traditional and nontraditional sources during the Covid-19

The construction of the SME index (1)

- Twitter's Streaming API is used to collect samples of public tweets matching a filter made up of 60 relevant keywords
- Filtered tweets reported in a single day are processed as a single block to compute daily index values
- Messages are cleaned and normalized and then undergo a sentiment analysis procedure, which calculates positive and negative sentiment scores for each tweet
- Message words are matched against entries of an Italian *sentiment lexicon* and the atomic scores of matched words are averaged to yield tweet-level scores
- Tweets are then clustered into three mutually exclusive classes (Positive, Negative and Neutral) and the daily index value is derived as an appropriate central tendency measure of the score distribution of the tweets belonging to the Positive and Negative classes. Lastly, the daily index is linearly transformed

The construction of the SME index (2)

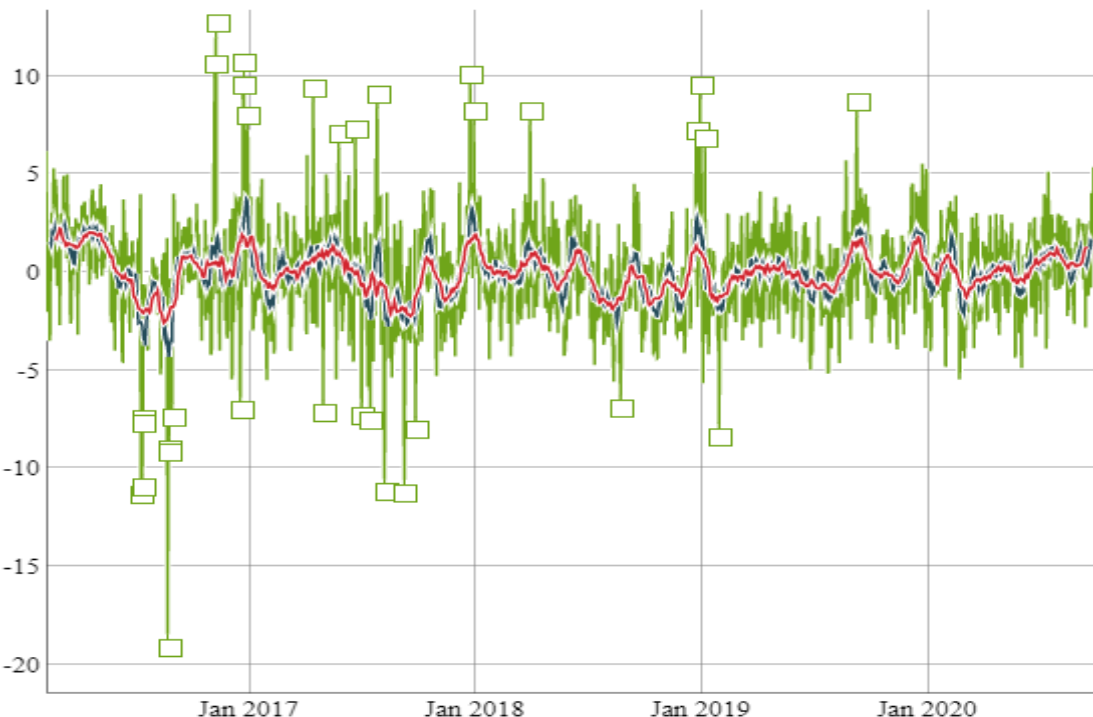
- To make the index robust against possible contaminations by off-topic tweets that might pass the filter an outlier detection procedure is set. Daily index values classified as truly anomalous are eventually imputed via nearest-neighbor interpolation
- A transformation makes the SME long-run mean, referred to the period 10 February 2016 – 30 September 2018, equal to zero

- ❖ Although, in late 2018, about 9 million Italian users were active on Twitter on a monthly basis (source: [Agcom](#)), Twitter users cannot be considered a representative sample of the Italian population, thus, the Social Mood on Economy Index cannot be referred to the overall Italian population

The SMEI series during the Covid Crisis

Social Mood on Economy - Daily Index and Moving Averages ✓

— Daily Index — 15 Days Mov. Avg. — 30 Days Mov. Avg.

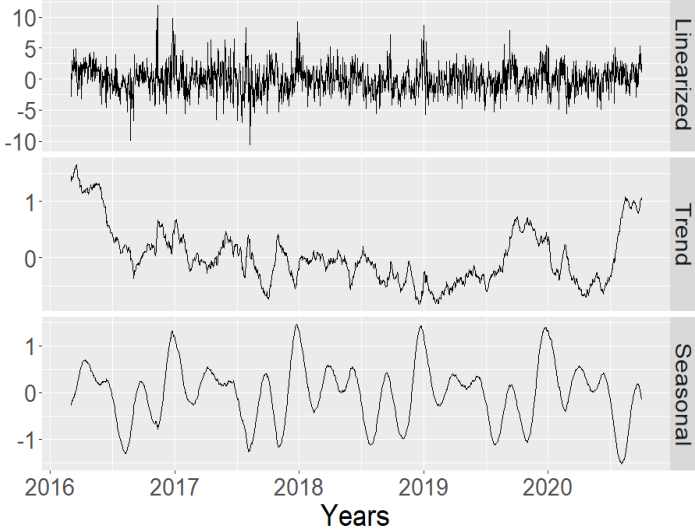


5

The SMEI, from the beginning of March 2020, showed a strong increase in the volume of tweets, which in a few days more than doubled (from around 67 thousand tweets on 3rd March 2020 to around 115 thousand ones on 5th March 2020) to reach 144 thousand one on 28 March; then, from the end of May, they returned to a volume comparable to the one before the Covid-19 crisis

Seasonal Adjustment of SME index

Social Mood on Economy Index

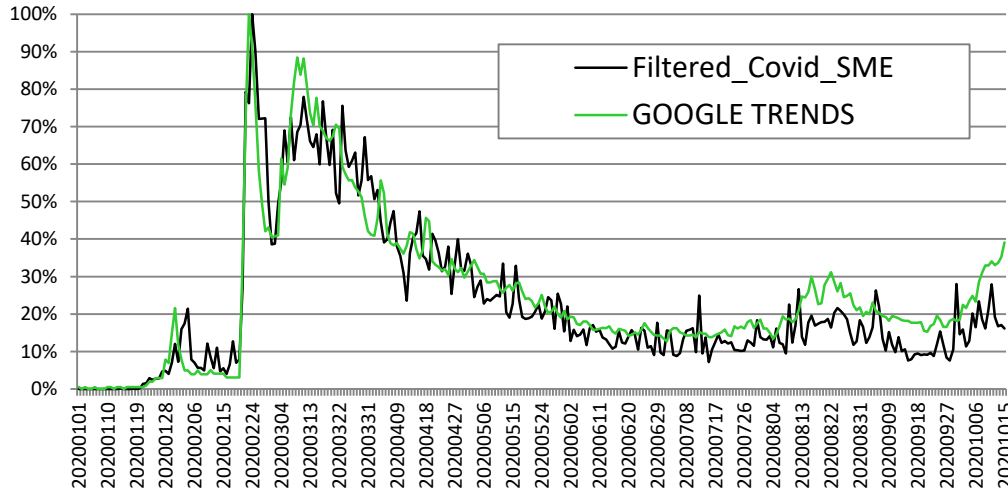


- The identification of the seasonal component for daily time series presents difficulties due to the presence of multiple seasonalities
- We follow a recent approaches, which applies modified exponential smoothing models compared to the standard models (De Livera, Hyndman, Snyder, 2011)
- The seasonal adjustment procedure is in two phases: estimation of deterministic effects through the introduction of appropriate dummy variables and identification of the components of the linearized series

- ✓ **Trend component**, recorded a strong decline from 18 February to the second half of April and in 2020 Q2 it showed a fluctuating trend, marking an upward trend until the end of May, and then undergoing a new decline in the first 10 days of June. From mid-June, it showed an upward trend peaking in September 2020
- ✓ **Seasonal component** showed positive peaks in December 2019 and from the end of March to early April 2020, whereas a negative peak started in the summer months and peaks in September

- We further exploited the information coming from the SME index to produce a new daily series, that is, the normalized share of tweets containing the terms “Coronavirus” or “Covid” in the text (among those used in the measure of the SME) out of the total tweets used to measure the SME index from the 1st January 2020 to 15th October 2020
- Using Google Trends, we calculated a daily series computing a normalized weighted average of the query shares referring to the keywords “Coronavirus” and “Covid” for the Italian territory, using as weights the relative “popularity” of the queries for the two keywords reported by Google Trends
- We used also the daily series on the number of newly infected persons and deaths coming from the “health risk associated with Coronavirus

Filtered Covid SME compared with Google Trends



Signals extracted are quite similar, in both cases series peak in 23 -24 February, they present a negative peak at the beginning of March and then another increase before mid-March

After March, the interest of the keywords reduced by 50% and the decrease continued until June, when both the shares remain between 10% and 20%

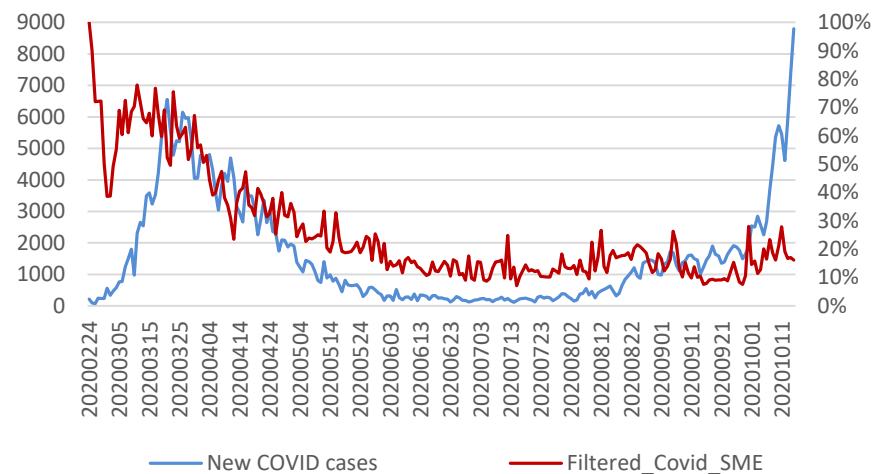
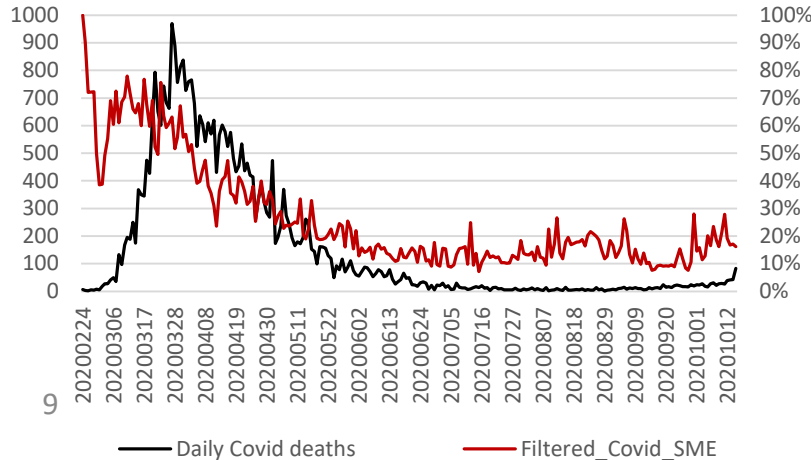
The increase of GT series restarted in August (peak 22/8), after there is a decline and a new rapid increase in October

The effect is less evident for Covid SME, even if we observe peaks at the beginning and at the end of September, when the series start a new increase

Filtered Covid SME compared with Civil Prot. Data

Filtered Covid SME SME and Covid deaths series behave more similarly after 24 March 2020,

The interest in the Covid-19 issue among the Italians declined with the reduction of the severity of the sanitary emergency and the increase in the new cases in September and October 2020 caused a slightly increase of the number of the filtered Covid SME conversations



Other nontraditional and traditional data used

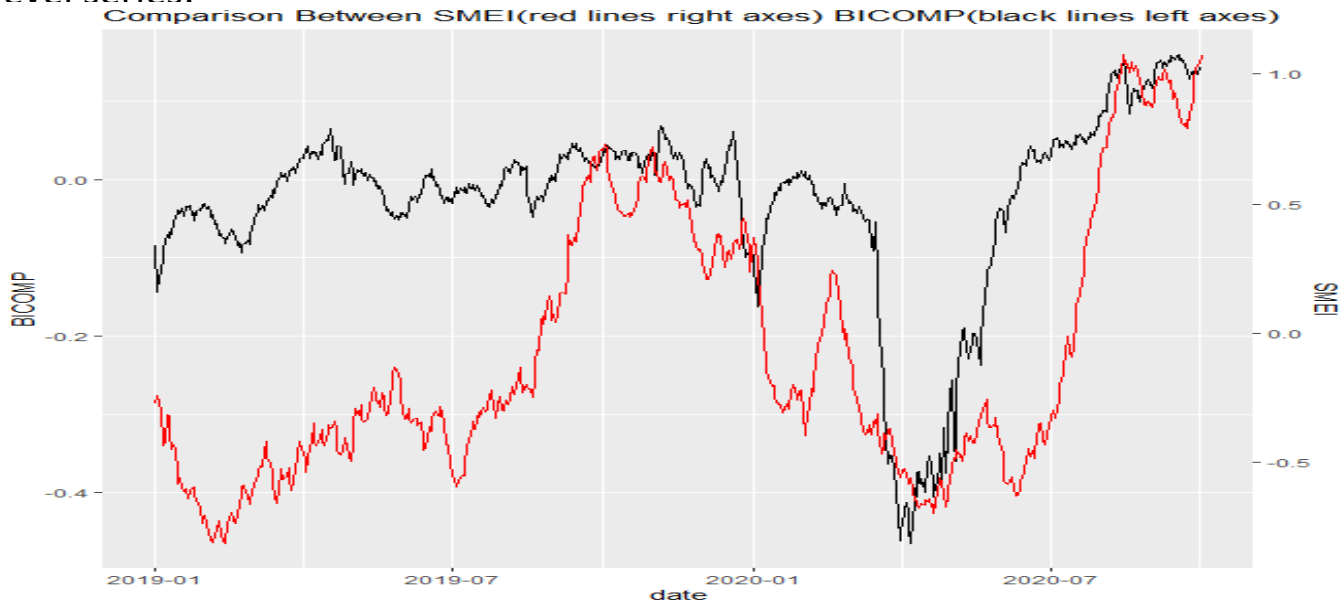
The SME index has been compared with other economic indicators both on monthly and daily frequency:

1. Monthly indicators: **consumer confidence indicators** (total and main components), **BI card transaction** (from the issuer side) series shares of the daily transaction by Merchant category code groups (e.g., clothing, hotels and restaurants, home, work, retail, services, telecom, web, travels and transport) in terms of indices (2014=100), available since 2014, **e-commerce monthly transaction** (from the acquirer side, namely, the expenditure of Italians on Italian firms) in terms of indices (2014=100), available since 2014
2. Daily indicators: BICOMP series referring to **Pos** and **Atm** transactions and **Target2** series

Low correlation with monthly indicators compared to the daily indicators

Relationships between SMEI with other economic nontraditional series

- Pre-treatment of the daily series by the implementation of outlier detection procedure and the removal of other deterministic effects
- Seasonal adjustment through the TBATS procedure in R (exponential smoothing model)
- The variable showing a stronger relationship with the Twitter SMEI is the BICOMP-POS level series:



Relationships between SMEI with other economic nontraditional series

1. Series are both integrated and through a cointegration analysis they show a cointegration relationship
2. Using a vector error correction model (VECM) we estimated the relationship between the two variables

$$\Delta POS_t = -0.011(POS_{t-1} - 0.15SME_{t-1}) + 0.14\Delta POS_{t-1} - 0.001\Delta SME_{t-1} + 0.07\Delta POS_{t-2} - 0.02\Delta SME_{t-2}$$

$$\Delta SME_t = 0.037(POS_{t-1} - 0.15SME_{t-1}) - 0.15\Delta POS_{t-1} + 0.3\Delta SME_{t-1} - 0.04\Delta POS_{t-2} - 0.03\Delta SME_{t-2}$$

- ❑ We could verify that SMEI, a relevant Istat experimental indicator for the analysis and the short-term forecasting, and other traditional macroeconomic indicators have very different trends during the Covid period
- ❑ We checked if the signals on the tracking of terms Covid and Coronavirus in the SMEI tweets were comparable to the evolution of the search for the same terms in Google trends and to the evolutions of the new Covid positive case and deaths and we found several similarities so that we argued that the SMEI could correctly catch the worried of twitterers for the pandemic
- ❑ New nontraditional sources allowed us better understanding the explicative economic power of SMEI and the correlation between SMEI and the BI Pos daily transaction series could allow us using SMEI in forecasting models with other traditional macroeconomic indicators