

An estimate of the Italian Consumer Confidence Index at regional level using Google Trends data

Josep Domenech¹, Andrea Marletta²

¹Department of Economics and Social Sciences, Universitat Politècnica de València, Spain

²Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy.

Abstract

Data about consumer confidence indices are often used as a gauge of the entire economy of a country. In Italy, this information is collected by Istat and it is available at national level and at the first sub-level, the geographic area, but not at the regional level. Previous research has demonstrated that the volume of some Google searches are correlated with the consumer confidence. Since Google Trends data are available both at national and regional level, the aim of this paper is to explore they can be combined with the data offered by Istat to obtain an estimate for consumer confidence indices at the second sub-level, i.e., the regional area. To this end, a set of search topics and words have been selected as potential predictors to acquire more information about consumer confidence indices for 20 Italian regions from 2007 to 2022. The obtained regional estimates are in line with the geographic area being successful to identify the periods of economic crisis due to the 2008 financial crisis and the 2020 Covid-19 pandemic.

Keywords: *Consumer confidence, Google Trends, Non-traditional data sources*

1. Introduction

The Consumer Confidence Index (CCI, henceforth) is a very important indicator to measure the health condition of the economy of a country. Many contributions showed the strict link between these indices and the economic activity (Golinelli and Parigi, 2004; Kilic and Cankaya, 2016). Other contributions focused the attention on the predictive power of these measures to forecast the consumption spending (Malgarini and Margani, 2007; Dees and Brinca, 2013).

For these reasons for all the countries in the European Union, the survey on consumer confidence is part of the joint harmonized EU program of business and consumer surveys. In Italy, this survey is realized monthly by the Italian National Institute of Statistics (ISTAT) with the CATI (Computer Assisted Telephone Interviewing) technique. The observed phenomena are consumer assessments and expectations and the final data dissemination is realized producing a final indicator called CCI and a group of 4 sub-indicators measuring the personal, economic, current and future confidence of the interviewers. Finally, these indicators are published as index numbers (reference year 2010 =100).

From a geographic point of view, the CCI is available at national level and at the first sub-level, a modified version of NUTS-1 (Nomenclature of Territorial Units for Statistics). Usually, in Italy, the first level of NUTS involves 5 areas, North-West, North-East, Centre, South and Islands. For CCI, data are available in 4 areas: North-West, North-East, Centre and South (also including Islands). No information is at NUTS-2 level, that is to say, the 20 single regions.

An index about regional consumer confidence is missing and the aim of this paper is to verify whether the existence of Google Trends (GT) time-series about some economic terms related to this issue could help in the estimation of such index. The birth of GT data in 2004 gave the start to a literature in combining these data with macro-economic variables mixing official and non-traditional data sources. GT data have often been associated to predictions for unemployment rates at national level (Naccarato et al., 2018; Simionescu and Cifuentes-Faura, 2021), other authors underlined the attention on smallest area at regional level (Falorsi et al., 2017; Bartha and Bontempi, 2022). Finally, some papers tried to explain the relation between economic indicators and GT data during Covid-19 pandemic (Lee, 2020; Simionescu and Raisiene, 2021).

The availability of GT time-series on real time at regional level for Italy could be exploited to control what is the relationship between official estimates for CCI at NUTS-1 level and the searches in Google for terms and topics related to CCI at NUTS-2 level.

The paper is organized as follows: after the introduction, Section 2 is devoted to the methodology and the data collection, in Section 3 some preliminary results have been presented and finally in Section 4 some conclusions will follow.

2. Methods and data collection

The theoretical background of the methodology is related to the conjoint use of data from different sources, from a statistical point of view it involves a linear regression based on some techniques of multivariate analysis. In the linear regression model, the dependent variable is the estimate for regional CCI and the independent variables are the results of the multivariate analysis. To extract the similarities between official and GT data, the following process is described. The procedure for computing a regional estimate of CCI could be resumed in 4 steps, similar to the process described by Eichenauer et al. (2022) and Woo and Owen (2019).

Firstly, the search volumes for some terms potentially correlated with the CCI are retrieved and seasonally adjusted; secondly, among all terms, only those highly correlated with the CCI are selected; thirdly, a Principal Component Analysis (PCA) of the selected series is conducted to extract the common signal in them (Jolliffe, 2002); finally, the first Principal Component is used to estimate the parameters of a linear regression model on the CCI.

After the model for the NUTS-1 level CCI is estimated, the search volumes at the regional level for the same terms are retrieved from GT seasonally adjusted. Using the PCA factor loadings of the national data, factors at the regional level can be computed. Then, using the model parameters estimated, it is possible to estimate the CCI at a regional level.

The first principal component was used to estimate:

$$CCI_t = \beta_0 + \beta_1 GT_t + u_t$$

where CCI is the Consumer Confidence Index at Nuts-1 level and GT is the first principal component extracting the common signal of the selected search term volumes.

The method described above has been applied to the CCI in Italy. A list of terms, comprising keywords and topics related to the economic sentiment, was requested to GT. Those with a Pearson's correlation coefficient higher than 0.7 were used as input for the PCA.

Data for CCI and GT time-series are referred to the period from 2007 to 2022. Since the volatility of GT data, multiple extractions have been achieved in January and February and the final result is the average time series of multiple extractions.

GT data have been collected using the library *gtrendsR* and *trendecon* in the statistical software R. The dataset referred to 33 searches (15 topics and 18 terms) related to the domain of the macroeconomic phenomena as the CCI. The list of the terms is the following: “production”, “unemployment”, “employment”, “labour”, “recession”, “inflation”, “public debt”, “economic crisis”, “recovery”, “minimum income guaranteed”, “unemployment allowance”, “failure”, “purchase”, “saving”, “bankrupt”, “mortgage”, “sales”, “construction”. The time span covered from January 2007 to December 2022 using monthly data for 21 geographical areas, the national searches and the 20 regional searches. Data about CCI have been achieved using the Istat datawarehouse available at “dati.istat.it”. This dataset is available as monthly data from January 2005 to December 2022 for 5 geographical areas, the national level and the 4 sub-national levels (North-West, North-East, Centre and South (also including Islands)).

3. Results

In this section, preliminary results are shown for Central Italy and its regions coloured in Figure 1. The geographic area named Central Italy is composed of 4 regions: Tuscany, Umbria, Marche and Lazio. From a geographical point of view, it represents about 20% of the total population of Italy and 19% of the total area of Italy. The most representative cities of this area are Rome in Lazio and Florence in Tuscany. From an economic point of view, this area is characterized by a strong industrial activity realized by small and medium enterprises and by an important touristic sector.

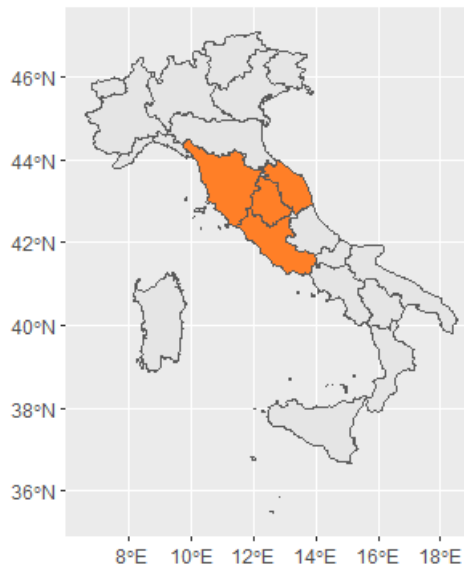


Figure 1. Map of Italy and Central Italy (coloured).

This area has been selected as a referring point for this work because it is probably the most homogenous area able to better represent the entire Italian territory. This statement is confirmed by the Figure 2, in which CCI time series from 2007 to 2022 is displayed for Italy and Central Italy. It is possible to note that the perception is very similar with a correlation coefficient equal to 0.98, this means that in Central Italy the consumer confidence is essentially very close to the national one. For the data collection process of Istat, no information is available about how this confidence is perceived in the 4 regions and the motivation of this work is trying to give a representation of CCI at regional level using an alternative data source.

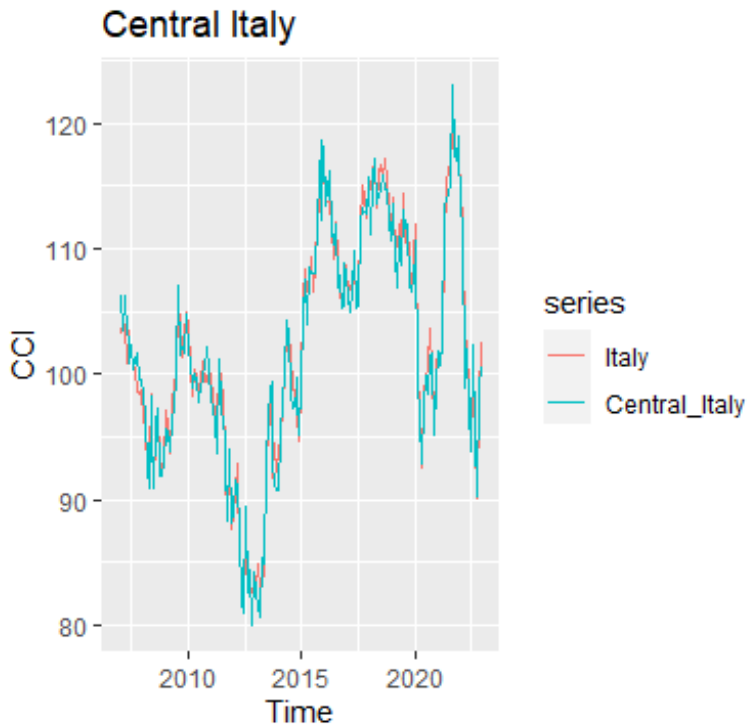


Figure 2. CCI for Italy and Central Italy (2007-2022). Source: Istat

Using the methods described in the previous section, an estimate for CCI based on GT data is displayed in Figure 3. Among all the terms used, the terms with the higher correlation with the CCI in Central Italy are: economic crisis (topic), sales (topic), unemployment allowance (topic), flights (topic), economic crisis (term).

These preliminary results are satisfactory because for Tuscany, Marche and Lazio the correlation coefficient between the regional estimate and the official CCI for Italy and Central Italy is higher than 0.5 (see Table 1).

Table 1. Correlation between regional CCI and Central Italy (2007-2022). Source: Istat and GT data

<i>Region (Central Italy)</i>	<i>Correlation with Central Italy</i>	<i>Correlation with Italy</i>
<i>Tuscany</i>	0.551	0.566
<i>Umbria</i>	0.448	0.464
<i>Marche</i>	0.543	0.559
<i>Lazio</i>	0.581	0.597

Probably, results for Umbria are less stable, because it is the smallest region of the Central area and GT data for very small areas are not reliable.



Figure 3. Estimated regional CCI compared to Central Italy (2007-2022). Source: Istat and GT data.

From a graphical point of view, all GT time series have a fall in 2009, probably due to the financial crisis. With respect to the decrease of CCI in 2012 due to the spread crisis, according to GT data, it is well recognized in Lazio. After this period, there is a slight increase for all the regions without the peak of 2016 and 2018 amounted by the real data. Relatively to the Covid-19 pandemic, the fall of the consumer confidence is present in Tuscany and Lazio.

4. Conclusions

In this work an approach to obtain an estimate for Consumer Confidence Index at regional level has been presented for Italy from 2007 to 2022 using Google Trends data. Istat provided monthly data for CCI at national and NUTS-1 level but not at regional level. On the other hand, GT data have been collected using some terms and categories related to the consumer confidence for the same period at national and regional level. Combining these two official and unofficial sources, and using a simple approach based on principal component analysis, a first estimate for CCI have been achieved for the 4 regions of Central Italy. The regional estimates are affected by the typical volatility of GT data, but especially for Lazio and Tuscany, the regional CCI presents good values of correlation with official CCI for Central Italy and a fair capability to catch sudden drops due to some crisis.

Some future works could regard the introduction of some error measurements to evaluate the goodness of the model, or a different procedure to select the terms to insert in the estimate model. Finally, some enhancements could involve the integration of some prediction methods for the time-series analysis of the estimated values.

Acknowledgments

This work was partially funded by MCIN/AEI/10.13039/501100011033 under grant PID2019-107765RB-I00.

References

- Bartha, S., & Bontempi, M. E. (2022). Measuring Economic Uncertainty for Poland.
- Dees, S., & Brinca, P. S. (2013). Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, 134, 1-14.
- Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google Trends. *Economic Inquiry*, 60(2), 694-705.
- Falorsi, S., Fasulo, A., Naccarato, A., & Pratesi, M. (2017, July). Small area model for Italian regional monthly estimates of young unemployed using Google trends data. In *61st World Congress of the International Statistical Institute* (pp. 16-21).

- Golinelli, R., & Parigi, G. (2004). Consumer sentiment and economic activity: a cross country comparison. *Journal of Business Cycle Measurement and Analysis*, 2004(2), 147-170.
- Jolliffe, I.T., (2002). *Principal Component Analysis*, second edition, New York: Springer-Verlag New York, Inc.
- Kilic, E., & Cankaya, S. (2016). Consumer confidence and economic activity: a factor augmented VAR approach. *Applied Economics*, 48(32), 3062-3080.
- Lee, H. S. (2020). Exploring the initial impact of COVID-19 sentiment on US stock market using big data. *Sustainability*, 12(16), 6648.
- Malgarini, M., & Margani, P. (2007). Psychology, consumer sentiment and household expenditures. *Applied Economics*, 39(13), 1719-1729.
- Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, 130, 114-122.
- Simionescu, M., & Cifuentes-Faura, J. (2022). Can unemployment forecasts based on Google Trends help government design better policies? An investigation based on Spain and Portugal. *Journal of Policy Modeling*, 44(1), 1-21.
- Simionescu, M., & Raišienė, A. G. (2021). A bridge between sentiment indicators: What does Google Trends tell us about COVID-19 pandemic and employment expectations in the EU new member states?. *Technological Forecasting and Social Change*, 173, 121170.
- Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, 38(2), 81-91.