Mobility trends in Italy during the first wave of Covid-19 pandemic: analysis on Google data

Ilaria Bombelli[a,b] and Daniele De Rocchi[b]

[a]Italian National Institute of Statistics (ISTAT)
[b]Department of Statistical Sciences, Sapienza University of Rome;
ilaria.bombelli@uniroma1.it, daniele.derocchi@uniroma1.it

Abstract

During Covid-19 pandemic, Governments implemented policies to reduce the spread of the virus. In Italy, policies have been implemented starting from 9th March 2020, when in the whole country lock-down policies were adopted. In this study, we analyze mobility data to understand which were the main drivers of mobility during the first pandemic wave. In particular, we analyze Google mobility reports, to study the relative changes in mobility w.r.t. a specific baseline and to analyze several different mobility drivers. In addition, we implement Multilinear Principal Component Analysis to extract relevant features from a multidimensional object. Results show good performances in terms of explained Frobenious norm and two PCs are able to synthesize the trends; finally, the reconstructed trends are also similar to the true original ones.

Keywords: Covid-19, human mobility data, mpca, three-way data

1. Introduction

At the beginning of 2020, the SARS-CoV-2 virus and its related disease, known as Covid-19, started to spread all around the world. The first cases were detected in China [1] at the end of 2019; then, on December 31st, Chinese government informed the World Health Organization (WHO) about the detection of some similar cases of respiratory disease due to an undefined agent. On 7th January the virus was identified. Then, on January 31st two Chinese citizens tested positive in Rome and on 17th February the first person affected by the disease in Italy - who has not visited China the months before - was detected. The virus started spreading and the first death occurred on February 20th in Vo' Euganeo (Veneto). In such situation, in the absence of therapies and due to the cases' fast spread, Non-pharmaceutical interventions (NPI) have been implemented by many governments. In particular, in Italy, lock-down policies and mobility restrictions were mainly adopted to reduce the spread of Covid-19. Clearly, the policies had a deep and strong impact on human behavior, especially on human mobility.

The analysis of human mobility data was deeply investigated. For example, [2] compared Covid-19 data and demographic variables with the GPS data in order to study how the restriction orders affect human mobility. In addition, the relationship between Covid-19 transmission and mobility was also analyzed by [3], who focused on 52 countries around the world. Moreover, [4] provided a comprehensive overview of human mobility data and it also compared different data sources to make the researchers and policymakers aware of the nature of the available data sets. A different study has been carried out by [5], who investigated the impact of COVID-19 on the number of people involved in crashes. Finally, [6] studied the effect of the restriction policies on mobility using the geolocalized data from 13 M Facebook users in France, Italy, and the UK.

The research focused on studying mobility data is considered as a data-driven approach which is an extremely helpful tool to support the decision-making process, as the new challenges of digitalization, innovation, and sustainability require.

In this paper, we aim to analyze mobility data provided by Google (more details in Section 2) and to find indicators able to synthesize the mobility data trends related to different categories, as also [7] did for England and Wales.

2. Data

Google Covid-19 Community Mobility Reports (GCMR) provide data on human mobility. In particular, mobility reports (GCMR link) share daily mobility data on relative changes compared to a baseline, the day that represents a 'normal' value for that day of the week. In detail, the baseline day is the median value from the 5-week period Jan 3 – Feb 6, 2020. Due to privacy reasons, the absolute values of mobility are not provided. Therefore, for any reported date, the daily relative change is estimated as the percentage change w.r.t. the corresponding baseline weekday. In addition, mobility data refer to 6 different categories which indicate the reasons why the mobility occurred: *grocery, parks, residential, retail, transit, and workplaces*.

In our analysis, the statistical units are the 20 Italian regions; for each of the 6 Google categories, mobility relative changes are provided; finally, the time frame cover from $2020-02-15$ until $2020-08-29$, for a total of 27 weeks.

Therefore, the data structure can be considered as a Three-way data array, having three ways, i.e. rows, columns, and layers, and referring to three modes, i.e. Regions, time, and Google categories, respectively. For each layer (Google category), a rows-by-columns data set provides information on the mobility relative changes during the period taken into account (week 8-week 35) for 20 Italian regions.

To explore the mobility trends, we plot them separately per Google category. In addition, we aggregate the dates referring to 'days' into 'weeks'. In Figure 1, the trends are colored differently, according to the region they refer to. As Figure 1 shows, mobility due to grocery, retail, parks, transit, and workplaces was characterized by a negative change w.r.t. the baseline period, given that lockdown restrictions mainly forbid people to move for other reason than basic needs. In particular, we notice that for transit, retail, and grocery, the relative change w.r.t. the baseline is null starting from week 28; for parks, instead, the mobility trends were set back equal to the baseline starting from week 20 and at week 25 all the regions have the same mobility as the baseline; in addition, from week 25 onward, the mobility towards parks experiences a huge increase in the relative change w.r.t. the baseline, reaching even a positive 500 % of relative change at week 33 in some regions. For what concerns workplaces, the trend, after having experienced a consistent rapid reduction from the starting week until week 15, starts increasing; however, unlike other trends, it remains below 0, meaning that during the analyzed period mobility due to reasons linked to work never came back to the level of the baseline. Totally different is the trend related to residential mobility: in the first weeks, it started a rapid increase in positive relative change w.r.t. the baseline. After that, between weeks 13 and 20, the relative change remain more or less constant at a level of 30 %, then it started slowly decreasing until reaching again level 0 between weeks 25-35, meaning that the mobility during those weeks came back to the baseline level.

3. Statistical Analysis

It can be of interest to synthesize the information on mobility data provided by GCMR, by retaining as much information as possible. However, classical techniques, such as the Principal Component Analysis (PCA), cannot be applied, since the three modes (regions, categories, and time) should be taken into account simultaneously.

As [7] noticed, for such data a multilinear PCA (MPCA) is suggested. MPCA has its origin in the well-known Tucker decomposition of a K-Tensor [8], which is used to reduce the dimension of a tensor object, or three-way data structure; the main contributor to the method is [9]. While PCA is used

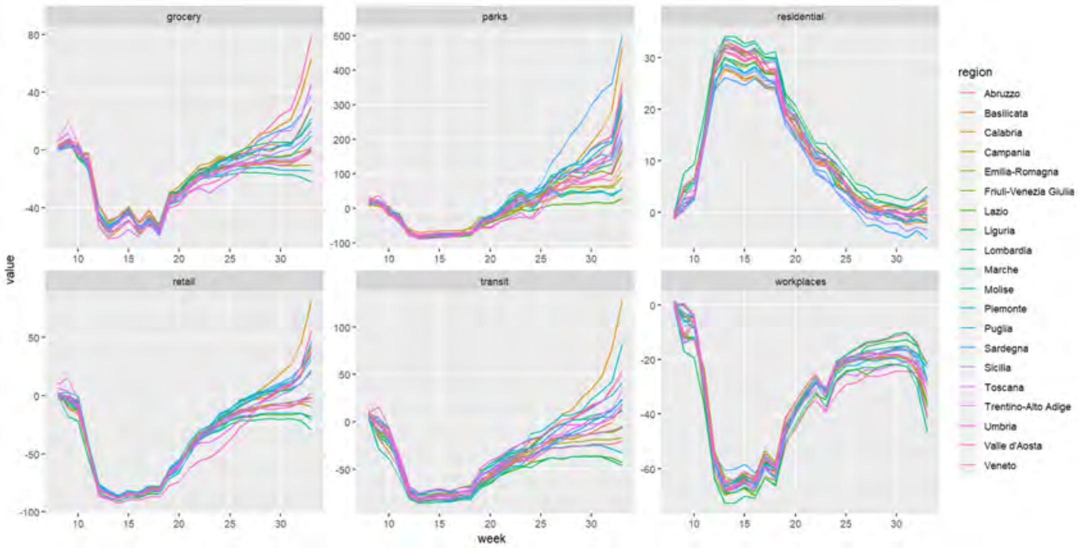

Figure 1: Mobility relative changes w.r.t. baseline separated by Italian region and reason (Google Covid-19 Community Mobility Reports (GCMR))

to reduce the dimension of the variables in a units-by-variables data set, MPCA is used to extract the relevant features of a multiway object.

It is worth mentioning that we exclude from the analysis data from parks. Indeed, on the one hand, the analysis performance, measured by the explained Frobenious norm, is significantly reduced if we include this category (less than 50%); on the other hand, the trend of mobility towards parks is very different in terms of behavior and in terms of magnitude, as aforementioned.

By applying the `mpca` function (`Rtensor` package) on the tensor (or three-way data) object (built by using `as.tensor` function in the `Rtensor` package), we obtain a good synthesis by using 2 PCs, as the percent of Frobenius norm explained by the approximation is nearly equal to 79%. In addition, the estimated data, namely the estimates of all the tensor values after compression, are quite close to the true ones. Indeed, Figure 2 shows that points (true observations) are almost gathered by the central solid lines (estimated trend), even if during the last weeks of the considered period, points spread a bit far from the estimations, meaning that the model is only nearly able to capture the variability of the trends in those weeks.

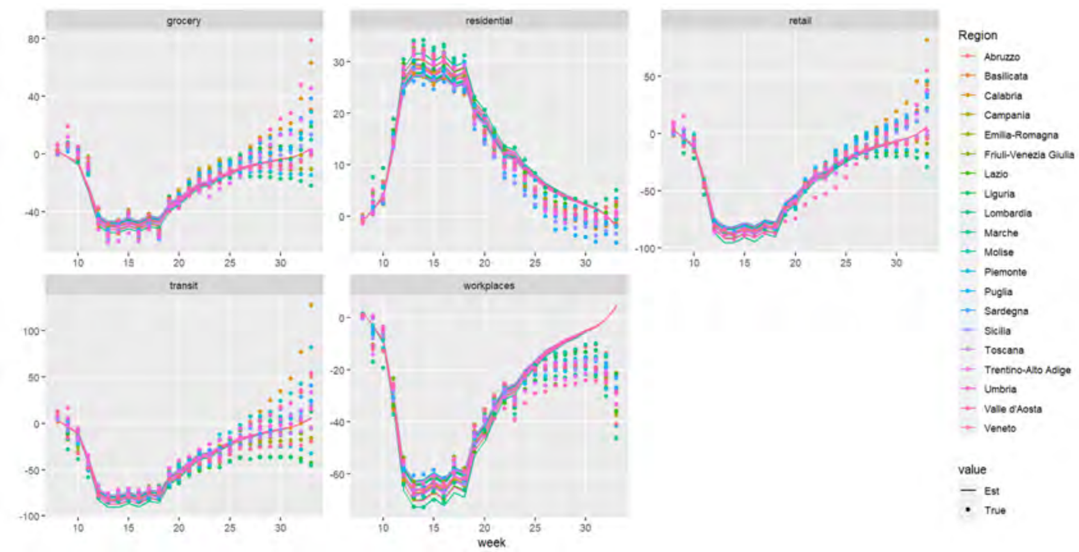

Figure 2: Estimated by mpca (2 components) and true observed mobility trends separated by Italian region and reason

By looking at the first two PCs in Figure 3, we realize that the former completely describes mobility for a residential reason, while the latter is a good synthesis of the mobility patterns for the other reasons.

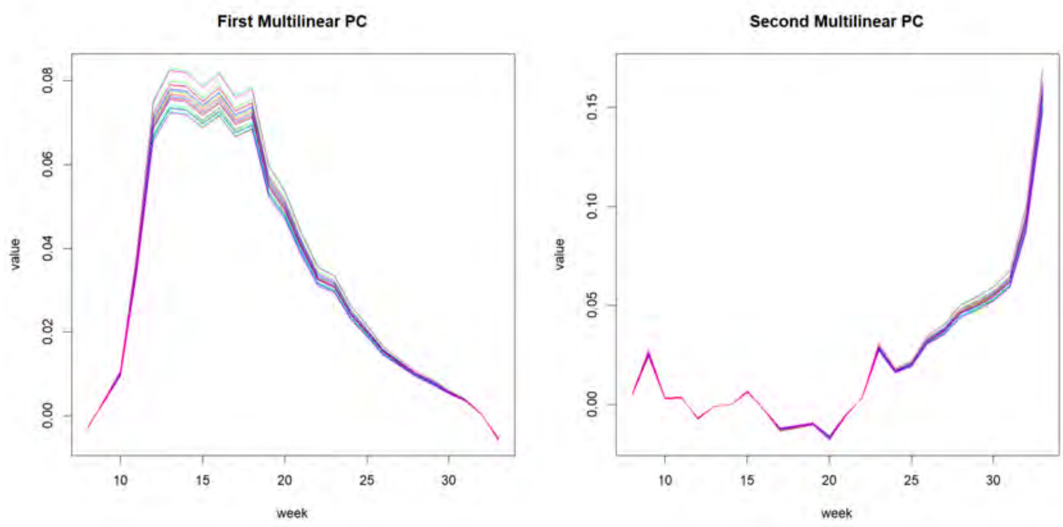

Figure 3: Mobility patterns of two PC resulting from mpca

To further describe the mobility trends, we apply the multilinear PCA with three components. By analyzing the results, we obviously obtain an increase in explained variance percentage, moving from 79% to approximately 80%. However, we realize that this gain in terms of explained variance is too small to balance the increase in model complexity. Anyway, for the sake of completeness, some details on the obtained 3 PCs are provided below. From Figure 4, we notice that the third PC allowed us to better capture the variability of the trends in the last weeks of the considered time frame, especially the trends of mobility toward workplaces.

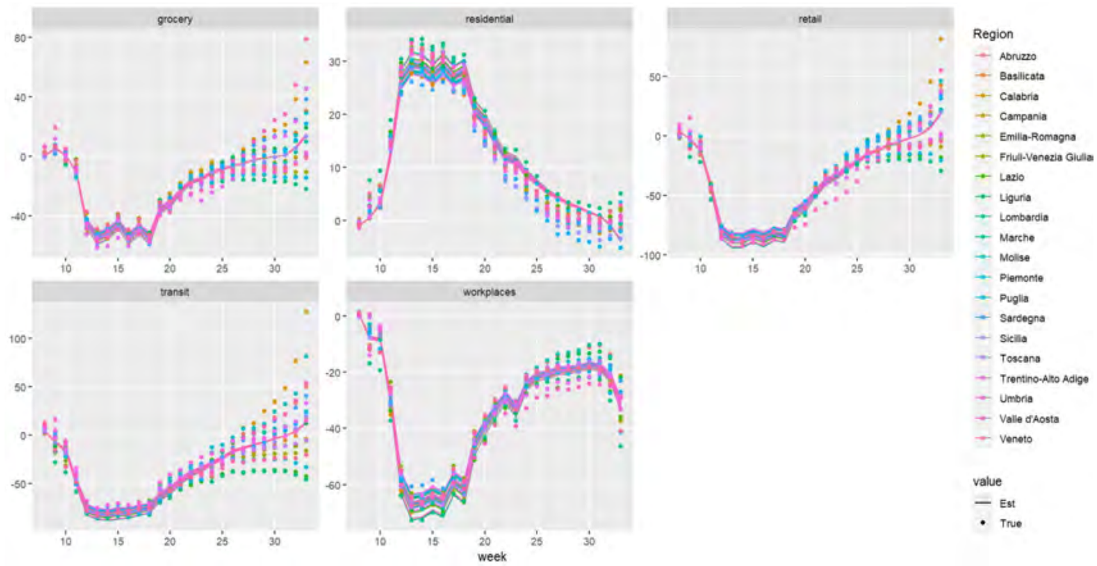

Figure 4: Estimated by mpca (3 components) and true observed mobility trends separated by Italian region and reason

In addition, by looking at the first three PC in Figure 5, we realize that the former completely describes the mobility for a residential reason, the third almost describes the mobility towards workspaces, while the second one is a good synthesis of the mobility patterns for the other reasons.

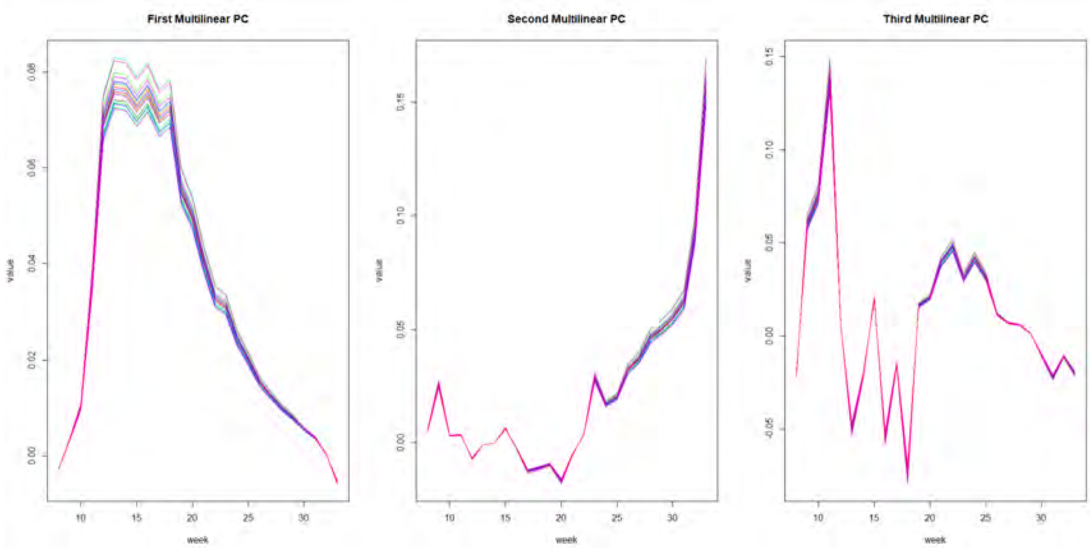

Figure 5: Mobility patterns of three PC resulting from mpca

4. Conclusion

The analysis allowed us to study how mobility during the first way of the pandemic in Italy changed with respect to the baseline period. Trends show that there was a negative relative change for mobility due to reasons other than residential, while the change was positive and rapid for residential mobility. By reducing the dimension of the three-way data object by using the Multilinear PCA, it was possible to obtain two PCs which are a good synthesis of the 5 trends (one for each mobility driver), as the percentage of Frobenious norm explained shows. In addition, the MPCA performed well also in terms of data reconstruction. Indeed, by estimating the trends and comparing them with the true ones, we experience a good model performance.

Further developments consider an analysis that involves mobility data and mortality data to study the relationship between the two.

References

[1] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 2020.

[2] Samuel Engle, John Stromme, and Anson Zhou. Staying at home: mobility effects of covid-19. *Available at SSRN 3565703*, 2020.

[3] Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Kylie EC Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F Brazeau, Lorenzo Cattarino, Laura V Cooper, et al. Reduction in mobility and covid-19 transmission. *Nature communications*, 12(1):1090, 2021.

[4] Tao Hu, Siqin Wang, Bing She, Mengxi Zhang, Xiao Huang, Yunhe Cui, Jacob Khuri, Yaxin Hu, Xiaokang Fu, Xiaoyue Wang, et al. Human mobility data in the covid-19 pandemic: characteristics, applications, and challenges. *International Journal of Digital Earth*, 14(9):1126–1147, 2021.

[5] Jie Zhang, Baoheng Feng, Yina Wu, Pengpeng Xu, Ruimin Ke, and Ni Dong. The effect of human mobility and control measures on traffic safety during covid-19 pandemic. *PLoS one*, 16(3):e0243263, 2021.

[6] Alessandro Galeazzi, Matteo Cinelli, Giovanni Bonaccorsi, Francesco Pierri, Ana Lucia Schmidt, Antonio Scala, Fabio Pammolli, and Walter Quattrociocchi. Human mobility in response to covid-19 in france, italy and uk. *Scientific reports*, 11(1):13141, 2021.

[7] Ugofilippo Basellini, Diego Alburez-Gutierrez, Emanuele Del Fava, Daniela Perrotta, Marco

Bonetti, Carlo G Camarda, and Emilio Zagheni. Linking excess mortality to mobility data during the first wave of covid-19 in england and wales. *SSM-Population Health*, 14:100799, 2021.

[8] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[9] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. *IEEE transactions on Neural Networks*, 19(1):18–39, 2008.