Issue 22(2), 2022, pp. 132-160
https://doi.org/10.18757/ejtir.2022.22.2.5702

TU Delft OPEN

EJTIR

ISSN: 1567-7141
http://ejtir.tudelft.nl/

Using location-based social media data to explain COVID-19 spread in Italy

Evgeniia Shtele[1]
Department of Architecture and Urban Studies, Politecnico di Milano, Italy

Paolo Beria[1]
Department of Architecture and Urban Studies, Politecnico di Milano, Italy

Vardhman Lunkar[1]
Department of Architecture and Urban Studies, Politecnico di Milano, Italy

During crisis, social network data is valuable information for authorities and researchers. Since the COVID-19 outbreak, a significant amount of data, including mobility-related data, has been released from various agencies to support studies. This paper aims to check the suitability of mobility-related datasets in describing and predicting new cases, about the spatial dimension during the initial phase of the outbreak. We focus on rich anonymized datasets through Facebook – Data for Good program: colocation matrices, movement matrices and stay-at-home data. However, we also compare their usability with a traditional Origin-Destination matrix. Our test case is Italy, the second country hit after China, where the infection spread irregularly from a few northern provinces to rest of Italy and abroad. Our regression models explain the number of actual new cases at the provincial level (corresponding to NUTS-3) by three groups of variables: active cases proxying local infections, interprovincial mobility proxying the arrival of cases from outside, and the degree of people staying at home proxying infections from cohabitants. The variants among the models consist of different measures of interprovincial mobility, thus allowing us to confront them. The result is the inclusion of time-dependent mobility data improving the significance of model and is effective in explaining irregular rise of cases in different parts of the country. Moreover, colocation results as the best measure. From a policy perspective, results show that mobility restrictions help reduce the geographical spread of infection at the very beginning, but once the outbreak, the interprovincial mobility becomes less relevant.

Publishing history

Submitted: 12 April 2021
Accepted: 28 April 2022
Published: 26 May 2022

Cite as

Shtele, E., Beria, P., & Lunkar V. (2022). Using location-based social media data to explain COVID-19 spread in Italy. European Journal of Transport and Infrastructure Research, 22(2), 132-160.

Keywords: *big data, covid-19, Facebook, Italy, location based social media, mobility*

[1] A: Via Bonardi 3, Milano 20133, Italy E: paolo.beria@polimi.it ; vardhman.lunkar@polimi.it ; e.shtele@gmail.com

1. Introduction

The COVID-19 outbreak has resulted in a health crisis on a global scale. Inducing stress on the national health systems, it has provoked extraordinary measures from respective governments to contain the virus. The "lockdown" measures adopted to mitigate the spread of the virus in Italy – for the first time ever at this scale, sadly followed by many other countries – had frozen almost every social and economic aspect of the nation, leaving most people locked in their homes while residents stranded in another region or country were unable to re-enter.

During crisis situations, data from social networks proves to be a valuable source of information for authorities and researchers, because of its capability to produce a nearly real-time picture of such changes, providing a unique opportunity to analyse and interpret the response to these challenging events through mobility changes.

This paper is dedicated to checking the suitability of various datasets available in predicting new cases and analysing the spatial spread of infection. To do that, we will create four different models explaining the number of actual new cases at the provincial level (corresponding to the NUTS-3 level), through three control groups: active cases, interprovincial mobility and social distancing proxies. The variants among the models consist of different source of interprovincial mobility, including traditional static OD data and time-dependent movement data coming from social media. What will be shown is that time-dependent data is effective in explaining the irregular rise of cases in different parts of the country, especially at the nascency of the outbreak, when local cases were few and alone would not be a good determinant of epidemic evolution.

The paper, of course, is based on certain hypotheses and has some limitations. Our main hypotheses are that total cases are, strictly limiting to the short term, a proxy of active cases (as people have not yet recovered or died and may infect others). We also assume that infections come from three "types" of causes: direct contacts with cohabitants (relatives, neighbours, etc.), everyday local contacts (e.g. at work) or from alien cases through non-local mobility. These sources are proxied by controls as described in Model conceptualization. The main limitation of the work is that we do not implement an epidemic model capable of predicting the evolution of disease and active cases for long periods. In addition, the model takes the active cases as an input from officially published statistics and, while not capable of predicting them, still points out the influence of mobility on observed cases.

The paper starts in Section 2, with the available literature exploring the role of mobility and social factors in such an emergency showing multi-disciplinary approaches to address the research question. Section 3 describes the timeline and phases of the pandemic in the Italian context. Model conceptualization introduces the general model conceptualisation and types of included variables. Section 5 is a concise description of the datasets, defining the metrics used, and Section 6 verifies its degree of representativeness by comparing it with available official statistics and analysing their correlation.

Section 7 represents the core of the paper, where the first model is explained in detail. Section 8 describes its results: β-coefficients are represented, interpreted, and analysed. In section 9, we build three models using alternative measures of mobility, compare them, then comment on their usability. Section 10 concludes and discusses the research work and its application into the realm of ongoing and future pandemics, pointing out also its limitations.

2. Literature review

Epidemiologists consider that the spread of a disease is best modelled by a logistic function or by other functions calculated from models like SIR or SEIR in Wu et al., 2020 and Ma et al., 2020. These models consider the time evolution of key variables, such as number of people considered as susceptible, infectious, and recovered, but often lacks a detailed spatial structure.

However, there is evidence that, to analyse the pandemic issue, a multidisciplinary approach (spatial, ecological, social analysis) is preferable as highlighted by Turner, 2002. During the pandemic, the availability of data from public institutions and large tech platforms (Google, Facebook or Twitter) is proving to be of fundamental use to the proliferation of academic work and, possibly, to coordinate responses to the outbreak and inform citizens.

The interconnection of current world and the fast and leap-frog pattern of the COVID-19 pandemic suggests that mobility (systematic and occasional long-distance) played a fundamental role in spreading coronavirus almost everywhere in the world. Therefore, we consider such mobility data to be important in studying the coronavirus spread, as it can bring significant changes in the social and economic aspects of peoples' lives.

The domain of the spread of infectious diseases considering different social and mobility factors has been explored by literature. A stochastic computational framework for the forecast of global epidemics that considers worldwide air travel infrastructure complemented with census population data was presented in Colizza et al., 2006. In another study, Troko et al., 2011, the role of public transport in spreading acute respiratory infection was investigated. Sallah, K. et al., 2017 used impedance model for predicting human mobility and calculated the probability for a person to travel in the context of infectious disease. The built model was compared with the gravity model as showed in Simini et al., 2012, and radiation models in Stefanouli, M. and Polyzos, S., 2017. Values are usually presented as average values for a certain territory divided into smaller sections and do not consider historical relations among areas, holidays, lockdowns, and other events affecting the movements. Such models are widely used, but they are less informative than real mobility data.

Initial studies trying to explain the spread of coronavirus infection were conducted in China, where Zhang, Y et al., 2020 described the factors influencing the number of imported cases from the centre of the epidemic in China and the spread of the pandemic. They explained how different modes of transport such as high-speed train, air and coach contributed to the infection spread using the gravity model. The presence of transport hubs in a city correlated strongly to the speed of the pandemic spread, but its link with the number of confirmed cases turned out to be weak. Lau, H. et al., 2020 analysed data on domestic and international passenger volume and flight routes in China and compared these to the distribution of COVID-19 cases, thereby deducing a significant correlation between them. Chen, Z. et al, 2020 used a Bayesian spatial-temporal model, determining the distribution of COVID-19 cases and their correlation with the migration of the Wuhan population in the early stages of the epidemic, which is important for early warning and prevention of future outbreaks. Kraemer, M et al., 2020 studied the real-time mobility data from Wuhan and detailed travel history to elucidate the role of imported cases across China and to ascertain the impact of control measures.

Kuchler T. et al., 2020 used Facebook mobility data to connect areas with stronger social ties to two early COVID-19 hotspots (Westchester County, NY, U.S. and Lodi province in Italy) having more confirmed cases. Tizzoni M. et al., 2014 explored official census surveys, mobility data extracted from mobile phone call records as a proxy, and the radiation model calibrated with census data to model the situation in France, Portugal and Spain.

Giuliani, D. et al., 2020 analysed the infected numbers in Italian provinces. The data is used to model the spatial-temporal distribution of COVID-19 infections at the local level. An endemic-epidemic multivariate time-series mixed effects generalized linear model has been implemented to understand and predict spatial-temporal diffusion of the phenomenon. They observed how the initially affected provinces were influenced by local endogenous transmission while in most of the northern and central provinces, a relevant number of cases is explained by the transmission from neighbouring provinces and for many of the provinces in the south; the contagions follow an endemic trend. Giordano, G. et al., 2020 propose a model with $c = 8$ compartments or stages of infection: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), collectively termed SIDARTHE. However, only one compartment is measured in the SARS-CoV-2 crisis: the number of active cases. Della Rossa, F. et al., 2020 parametrized, from real data, a network model with the 20 Italian regions as nodes linking the regions with proximity flows and long-distance movements among them to explain how the inter-regional fluxes must be carefully and selectively controlled as they can have dramatic effects on recurrent epidemic waves.

A deep analysis of human mobility and epidemic in Italy was performed by using raw data (Call Detail Records – CDR's and Extended Detail Records – XDR'S) provided by the mobile operator WINDTRE as showed in Cintia, P. et al., 2020. They highlighted a striking relation between the negative variation of movement fluxes and the negative variation of the net reproduction number of the virus. It was discovered that the reproduction number continues to decrease during lockdown and during the phase 2 (post lockdown), when the mobility begins to rise again, the reproduction number does not lead to uncontrolled growth. They explain the impact of the non-medical interventions.

In the nearly two years since Disaster Maps were launched by Facebook, the datasets have been used during major disasters to study the flow of the population and provide help during such challenging situations. Limiting to COVID-19, mobility and movement range data were, for example, used to understand the impact of the social distancing measures on the spread of the SARS-CoV-2 virus in the greater Seattle Area as explained in Burstein, R. et al., 2020. In a follow-up report, Thakkar N. et al., 2020 modelled diagnosis data, permitting them to understand how mobility decrease can be connected to decrease in COVID-19 transmission using reproduction number as a metric. Chang S. et al., 2021 adopted Facebook data (movements) to SIER models to predict new cases in the US. Beria, P. and Lunkar, V., 2021 described in detail the real mobility patterns that occurred during lockdown in Italy, including the territorial differences and analysing the relation between interprovincial flows and population presence. Facebook's research group published interesting results of building Neural Relational Autoregression to predict the growth in new cases in the US using movements data as demonstrated in Le M. et al., 2020. Chang, M. et al., 2020 explored colocation matrices for modelling the mobility impact on the spread of COVID-19 in Taiwan using metapopulation models that incorporate human movement data. Cornelia et al., 2021 find that mobility data alone is sufficient to meaningfully forecast COVID-19 infections 7–10 days ahead at all geographic scales (cities, provinces, states, countries). Their methods were evaluated for Italy, China, South Korea, France, and national data from over 80 countries. Furthermore, identical models that exclude mobility data perform substantially worse, suggesting an important role for mobility data in forecasting. Their approach uses simple and transparent statistical models to estimate the effect of NPIs (non-pharmaceutical interventions) on mobility, and basic machine learning methods to generate 10-day forecasts of COVID-19 cases. They cite that a positive edge of such an approach is that it involves minimal assumptions about disease dynamics and requires just publicly available data.

This research's added value lays in the possibility of using dynamic location-based social network data, in addition to traditional static OD matrices, to reproduce the spatial spread of the infection. While the (actually obvious) relation between mobility and infection spread has been demonstrated in literature, including in the Italian case, a quantification of the size of the effect of mobility is still unchartered waters. Ideally, our approach could be extended to become a tool to drive closures and travel bans in a more punctual way than has occurred until now.

3. Covid-19 spread in Italy

Italy was the first European country where the 2020 COVID-19 outbreak developed locally. The first local confirmed cases were found in the Northern province of Lodi and Padua on the 21st of February. Other cases probably spread in Europe even before the Italian ones, but they were unrecognised or were possibly isolated.

Following the first cases, decrees introduced more and more restrictive measures because of the exponential increase. Without going into detail about the evolution of the pandemic (see Beria, P. and Lunkar, V., 2021), the timeline of the pandemic and the government policies during its course has been illustrated in Figure 1.

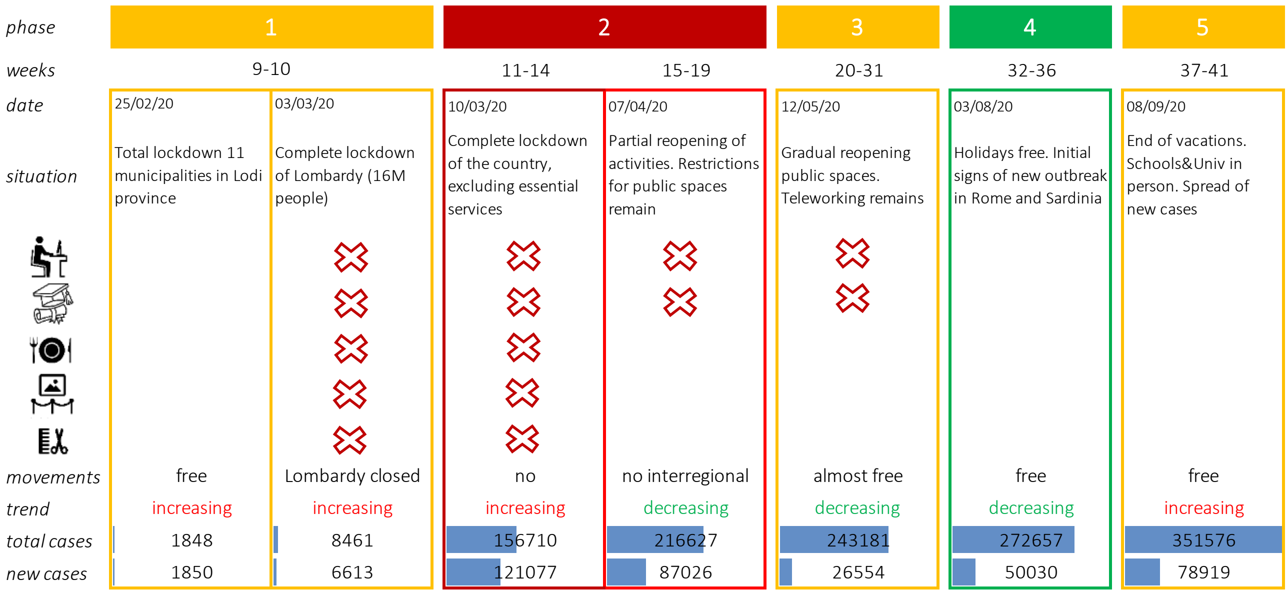


Figure 1. Timeline of the COVID-19 outbreak in Italy and the government policies. Source: our elaborations. Icons from https://thenounproject.com/

4. Model conceptualization

We will now introduce the model used to assess the impact of mobility on the spread of the SARS-CoV-2 virus in Italy and the effect of the restrictive administrative measures adopted.

The idea behind the model is that new cases in a province do not depend simply on locally active cases but (especially at the beginning and before restrictions are imposed) also on contacts with alien cases from other provinces. The more an area has mobility ties with others, the more likely cases are imported and, consequently, the earlier the outbreak develops. When lockdown and mobility restrictions are in place, interprovincial mobility is reduced significantly, but local mobility is too. In this phase, where the virus is already widespread, the evolution of the outbreak will remain fast and severe because infections still occur "at home" (meaning among relatives and strict contacts). In areas without many cases, the lockdown instead prevents the spread of the virus and keeps numbers relatively low.

Figure 2 summarizes the model structure, including the controls. What we expect is that at the beginning of the outbreak, new cases depend largely on interprovincial mobility, meaning that they are imported. Interprovincial mobility is described by means of different measures, described in detail in Section 5. On the initiation of the outbreak, exogenous cases still occur, but are overwhelmed by infections due to local contacts (for example, at school or in workplaces) depending on the usual variable of active cases. Then, if severe mobility restrictions are in place, a higher stay-at-home rate (described in Section 5) could explain new cases more than active cases since infections become more and more a family matter.

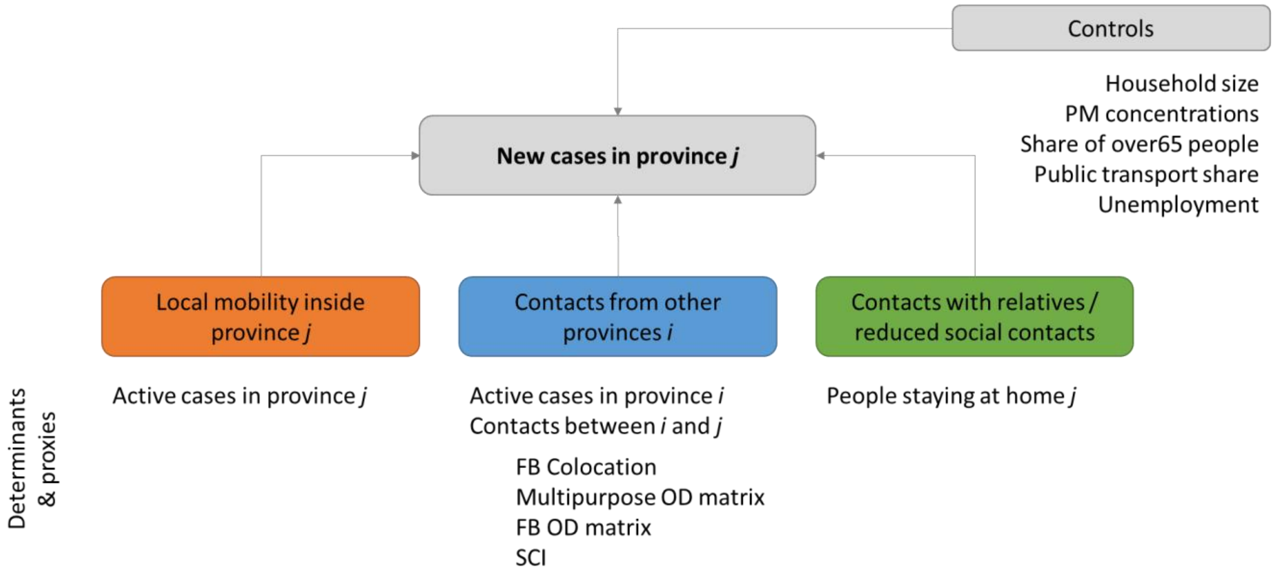


Figure 2. Model conceptualization

We include a two-week lag in the time variation of variables, because the incubation period usually does not exceed 14 days as mentioned in Backer, J. A. et al., 2020, and to observe the effect of movements: even with zero mobility, some transmission will continue within homes and tight personal circles as explained in Burstein, R., 2020.

5. Dataset description and general trends

Our research period ranges between the 9th and 41st weeks of 2020 and starts in correspondence with the publication of official statistics of COVID-19 cases. For the sake of simplicity and representation, the weeks are aggregated into phases as shown in Figure 1, in accordance with the restrictions imposed by the government that would ultimately determine and group the mobility characteristics of the population.

The panel data sample includes 110 observations across the 33 weeks, which corresponds to the number of Italian provinces (corresponding to NUTS-3 level) and available time range, respectively. The dependent variable is new COVID-19 cases for each week. In the following sections, the various datasets, their trends, and relevance will be described.

5.1 Epidemic Data

We obtain daily cumulative total cases by province, as provided by the Italian Department of Civil Protection, and calculate total cumulative and new cases by week to make the data comparable and to reduce the noise of the very first days of outbreak. Assuming people are contagious within two weeks after testing positive, we calculate new cases as follows:

$$\boldsymbol{New_{t-1}} + \boldsymbol{New_{t-2}} = \boldsymbol{Tot_{t-1}} - \boldsymbol{Tot_{t-3}} = \boldsymbol{Tot_{s-7}} - \boldsymbol{Tot_{s-21}} \quad \text{(Equation 1)}$$

where t is the week number and s the day number.

An assumption here that is reasonably valid, especially in early outbreak phases, is that active cases (which cause the infections) are calculated from new and total cases but without considering the dead and recovered, because of their unavailability at the provincial level[2].

Despite the scarce quality and lack of punctuality of initial infection data because of the unforeseen stress on the entire system, we can observe the initial exponential increase and a gradual flattening in the number of new cases after the introduction of the lockdown, as illustrated in Figure 3.

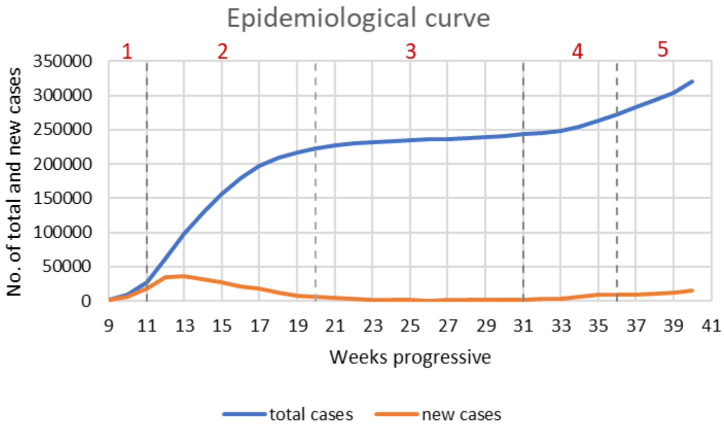


Figure 3. Total and New cases of infected people during the study period (Week 9 – Week 41)

5.2 Movement Data

In addition to epidemic data, the model is fed by movement data to proxy the effect of mobility on infections. Other than traditional OD sources (see paragraph 5.7), in this paper we make use of the *Facebook disaster prevention maps* (part of the *Facebook – Data for Good program*). In both cases, data is available coherently with epidemic data at the provincial level (NUTS-3).

Facebook Data for Good partners with over 450 organizations across nearly 70 countries in the world to help combat humanitarian crisis. The data[3] is available for researchers who approach with a proposal of how the data will be applied and investigated to benefit the available literature and help local governments give insight into the developing situation. The datasets utilise anonymized and aggregated data focussing on current and historic location sensing and information on cell connectivity. Comparing the public response to the interventions in terms of mobility on a spatial-temporal scale (measured relative to the pre-crisis conditions) can provide an insight into the effectiveness of emergency interventions and consequently help affected communities in case of a resurgence of emergency situations.

5.3 Movement Range (Facebook)

Movement Range trend datasets include two types of metrics.

a. *Travel Range*: is a metric representative of the average number of *level 16 Bing tiles* (0.6km x 0.6km) that a Facebook user (with location services on) was present on during a 24-hour period relative to the pre-crisis levels. Thus, it can be used to measure the degree of change of mobility ranges of people on the move.
b. *Staying at home*: this metric explains the percentage of Facebook users (with location services on) that stay on one such tile at three different hours of the day. Thus, a stationary user is analogous to staying put or staying home. The trend of *Stay at home* in the analysed period is represented in Figure 4.

---

[2] They are available at the regional level at https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-covid19-italia.md
[3] https://dataforgood.fb.com/research/

5.4 Colocation Maps (Facebook)

Colocation indexes (matrices) represent the probability that a random person from province A and a random person from B are in the same level 16 Bing tile (anywhere in the world) in 5-minute intervals over the time span of a week. It is important to mention here that if a particular user has incomplete mobility data or is stationary for various reasons, the user is excluded from the dataset. Once the user's permanence has been assigned, an estimate of how often two users from two regions cross each other or are simultaneously present in a particular level 16 Bing tile is calculated.

Thus, colocation between people of different i and j provinces (hereinafter *external colocation*) are calculated as:

$$\boldsymbol{Col_{ij}} = \frac{\boldsymbol{NumberOfColocations_{ij}}}{\boldsymbol{User_i * User_j * NumberOfIntervals}} \quad \text{(Equation 2)}$$

Colocation between people inside j province (hereinafter *internal colocation*) are equal to:

$$\boldsymbol{Col_{jj}} = \frac{\boldsymbol{NumberOfColocations_{jj}}}{\boldsymbol{User_j * (User_j - 1)/2 * NumberOfintervals}} \quad \text{(Equation 3)}$$

Where $NumberOfColocations$ is the number of 5-minute colocations in a week,
User is the number of FB users in each province,
Number of Intervals is the number of 5-minute intervals in a week.
Finally, what we obtain is a colocation matrix that contains the probability that two random users assigned to those NUTS-3 are co-located on a random minute during the week. The colocation matrix is given by:

$$\boldsymbol{C_t} = \begin{pmatrix} \boldsymbol{Col_{11t}} & \cdots & \boldsymbol{Col_{1Nt}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{Col_{1Nt}} & \cdots & \boldsymbol{Col_{NNt}} \end{pmatrix} = (\boldsymbol{Col_{ijt}}), \quad \text{(Equation 4)}$$

where: i - origin province; j - destination province; t – week number; Col_{ijt}- colocation between origin province i to destination province j. Matrix C_t is symmetrical i.e., $Col_{ijt} = Col_{jit}$.

It is interesting to observe the colocation trends within provinces during the lockdown period (Figure 4). One would expect to see a downward trend in weeks 9 to 15, when the lockdown was stricter, but it happens only in some provinces. In many others we see the opposite, despite the share of people staying at home being indisputably increased (see Stay-at-home trend) and despite the communication between the provinces decreasing noticeably as demonstrated later in Figure 6. This apparent paradox can be explained by the fact that colocation refers to mobile users only. During lockdown, the increasing number of staying put users are excluded both from numerator and denominator (Equation 2), which consequently becomes a representation of the few moving around.[4] And those moving during lockdown are not necessarily having less contacts than the many before the lockdown, explaining the increase in colocation probability in some provinces. Therefore, colocation alone cannot be a proxy of people meeting, until it is weighted with the moving population.

[4] The movement maps get the most common location of the user (tile) within the first-time window, and then the most common location (tile or administrative region) within the second time window. This defines the starting and ending of the vector. The vector time is then assigned to the starting time (in UTC) in the second window. For example, a user's start coordinates for time window 1 are the tile centre for their modal Bing tile. The end coordinates are obtained the same way, but for time window 2. The user's start and end coordinates are then used to assign them to a vector. Then the user's start and end coordinates are averaged along with everyone else assigned to that vector to yield the value surfaced in the data set. This means that users staying in one place and never moving will appear in the rows with same origin and destination the same as those that would be moving but not enough to change the modal tile or province.

The above inference will be supported later by the fact that internal colocations turned out to be a weak indicator for new cases of Covid-19 (corresponding models were built), as can be seen in the results in Table 5.

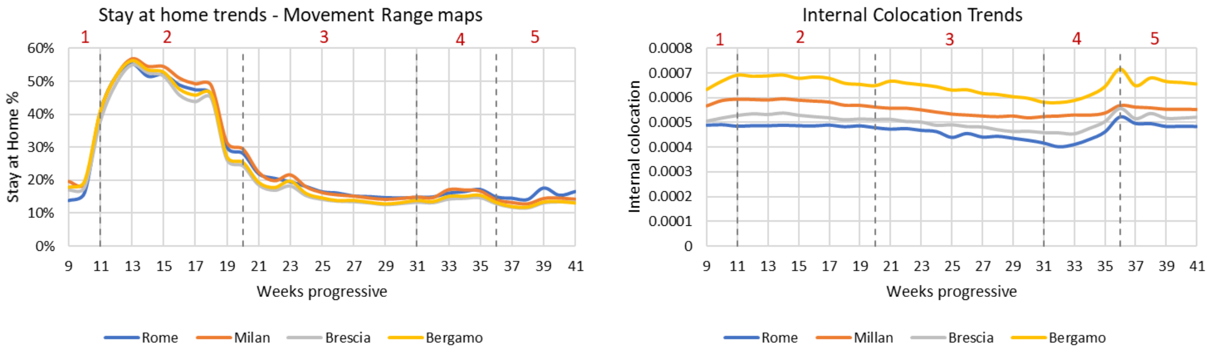


Figure 4. An example of stay-at-home trends and internal colocation values for selected Italian provinces

The graphs shown in Figure 5 illustrate the product of colocation and new cases each week[5]. Differently from colocation alone, this number is proportional to the average number of potential meetings with infected people per inhabitant of the given province. It is calculated using the formula:

$$\boldsymbol{NewPerOne_{jt}} = \sum_{i=1}^{N} \boldsymbol{New_{it} * Col_{ijt}} = \boldsymbol{New_{jj} * Col_{jj}} + \sum_{\substack{i=1 \\ i\neq j}}^{N} \boldsymbol{New_{it} * Col_{ijt}}. \qquad \text{(Equation 5)}$$

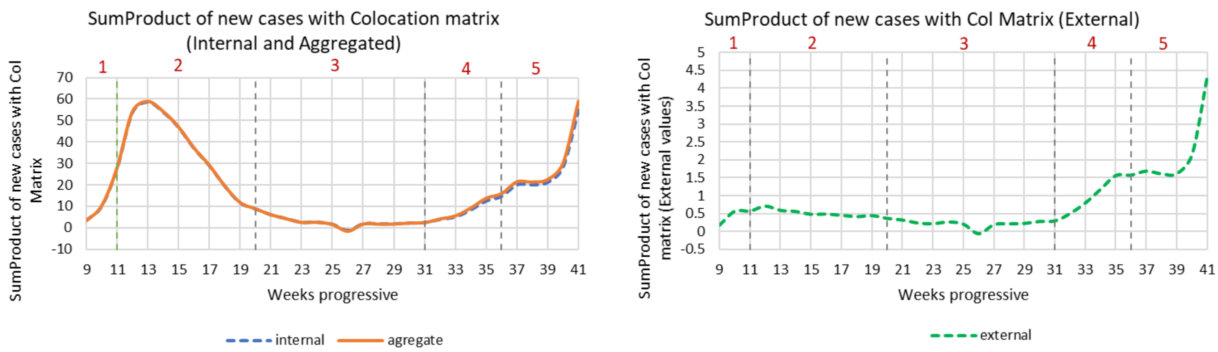


Figure 5. Representation of sum of the product of new cases by the colocation matrix (Equation 5) for the entire country.

The charts show the total number of potential meetings with contagious people. The blue dashed line corresponds to the first part in Equation 5 and is equal to the number of potential meetings with infected residents of their province. The green dashed line, represented at a smaller scale, shows the number of potential encounters with infected people from other provinces.

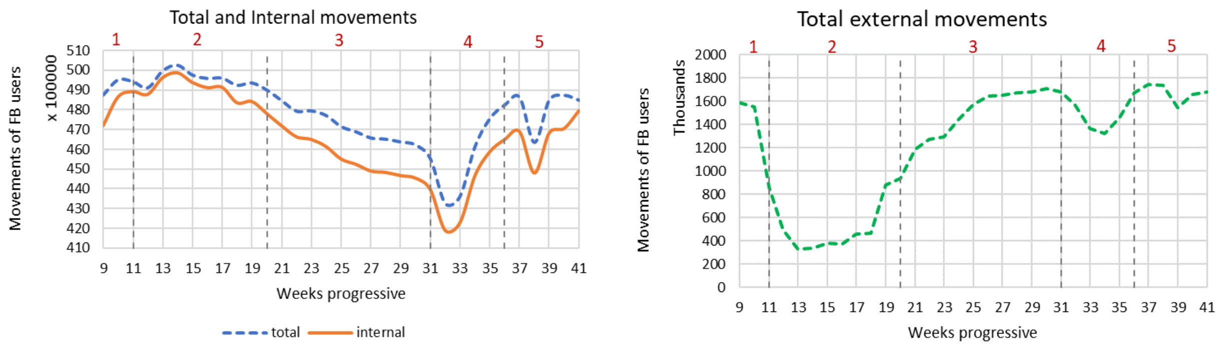
From Figure 5, the contribution of internal encounters is generally dominant and thus, globally, the likelihood of contracting the infection from visitors is small. However, this is true only at the aggregate level, while locally the contribution of *external colocation* becomes relevant, as shown in the following. The *external colocation*, in fact, embedding the long-distance mobility of Italians, can describe how infection spread thorough Italy in different periods of time. The drop and plateau from weeks 11-20 reflect the introduction of a lockdown, blocking movements across provinces, reopened since week 31.

[5] New cases as a proxy of active cases, as already explained above.

5.5 Movement between Administrative Regions (Facebook)

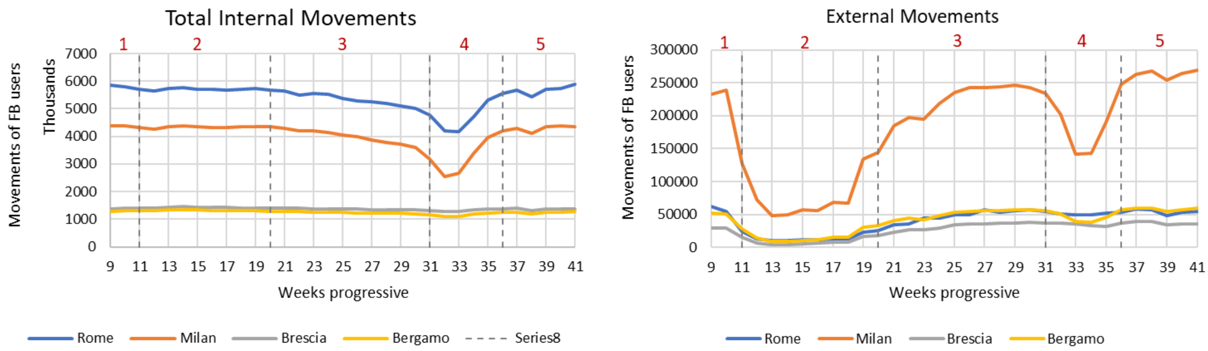
Movement Maps illustrate aggregate patterns of movement of Facebook users with location history turned on over an 8-hour interval. The maps are prepared at two different levels of aggregation. The *movement between tiles* dataset shows patterns of movement between single Bing tiles. The *movement between administrative regions* maps the movements between NUTS-3. We are here referring to the latter, which provides us with a real-time provincial level O-D matrix[6]. It is worthwhile to notice that an entry is present in the dataset only if this value is greater than a privacy threshold (10 movements of Facebook users in an 8-hour period). This introduces an underestimation of flows, especially between sparse or very far destinations. For the same reason, the *movement between tiles*, which are smaller, is much more incomplete than the one among administrative regions and is thus ignored for the current analysis.

Movement data complements the same idea as colocation indexes: movements inside the provinces did not reduce as much as the interprovincial ones, despite the restrictions. The variability of the internal data is low, while we can observe substantial decline in external values. This is observed in Figure 6 for the entirety of Italy and for selected provinces in Figure 7.[7]



a) b)

Figure 6. Facebook movements: a) total and internal to the provinces; b) interprovincial [Note: different scale].



a) **b)**

Figure 7. Facebook movements for selected Italian provinces: a) internal flows; b) external flows. [Note: different scale]

[7] Rome province is much larger in size than Milan one, which instead is extremely connected with neighbouring provinces. This explains the higher number of internal movements in Rome vs. the much higher number of interprovincial movements of Milano.

5.6 Social Connectedness Index (Facebook)

The propensity of displacements from one province to another could also be proxied by the *social connectedness indexes* (SCI), which is a relative measure of social connections between two NUTS-3 regions, weighted by the two populations. The denominator of the formula represents the total possible connections between the two populations of regions *i* and *j*. The numerator, instead, is the actual number of connections. It is given by:

$$\boldsymbol{Social\ Connectedness_{ij}} = \frac{\boldsymbol{FB\_Connections_{ij}}}{\boldsymbol{FB\_Users_i * FB\_Users_j}} \quad \text{(Equation 6)}$$

SCI is neither a measure of trips nor of real connections, but of "virtual" social media connections. $Social\ Connectedness_{ij}$ measures the relative probability of a Facebook friendship link between a given Facebook user in location *i* and a given user in location *j*.

Theoretically, it measures the strength of social connectedness between two geographic areas through Facebook friendship ties. These connections can reveal insights about economic opportunities, social mobility, and trade, but the link with physical mobility is weak. Moreover, the dataset is static and was last updated by Facebook in 2017. For these reasons, in Section 9 we do not expect it to be very accurate to model the spread of the pandemic.

5.7 Movement between Administrative Regions (i-Tram model)

As an alternative to Facebook data, we use a traditional O-D matrix. Given the lack of a national Italian O-D[8], we use the one made available by the project QUAINT[9] referred to in Beria P. et al, 2019. The matrix is multipurpose (in addition to systematic trips) and includes an estimation of overnight trips. It is a simulated matrix through a full multimodal model, with census data as input and other existing local matrices. It is calibrated through observations on road and air flows, plus train station uses for regional transport. The modelling is based on a detailed zoning and is aggregated for our purposes at the same scale (provinces) of the other measures. Differently from SCI, this matrix represents actual movements, but is a static one calibrated for year 2016 unlike FB real time movement data.

5.8 Other control variables

The other control variables are described in Table 1 and are introduced to improve the models results taking into account other elements that could be relevant. For example, the control of population over-65 should consider that known cases are proportionally more among the elderly because they develop more severe symptoms differently from young people, often asymptomatic and not tested.

[8] In Italy, the only national matrix available includes only commuting trips and dates back to 2011 census. It is therefore useless for our purposes. Multipurpose and more recent matrices are available in some regions, but not covering the entire country.

[9] The dataset is available at: http://www.quaint.polimi.it/dataset/ (downloaded 1/6/2020). The matrix has been simulated through the Italian National transport model (i-TraM https://metaplanning.it/atlante/).

Table 1. Description of control variables

Variable	Identifier	Time-dependency	Source	Notes
Family members	Family	Constant	census data (ISTAT) [10]	The number of families in a province divided by the number of residential units in that province[11]
Population	Pop	Constant	ISTAT	population of province
Over 65 years	Over65	Constant	ISTAT	number of people aged 65 and above[12]
Public transport use	PubTrans	Constant	ISTAT	share of people who use public transport[13]
PM-10	PM	Constant	Legambiente (2019).	Average of the annual average values of PM10[14] (µg/mc) recorded by the urban stations in 2017, provincial scale.
Unemployment	Unemployment	Constant	ISTAT	the percent of the labour force that is unemployed[15]
Population Density	Density	Constant	ISTAT	population/km² per province[16]
%positive swab tests	Positive	Weekly	Protezione Civile	% of positively tested people[17]

6. Representativeness of the dataset

These datasets have the potential to replicate a picture of crisis situations with a certain level of spatial detail and global coverage. However, such type of location-based Social Network data has its caveats, including its actual representativeness of the population.

We compared the population structure in detail according to official statistics and Facebook data in our previous article, Beria, P. and Lunkar, V., 2021. Both the sexes are well represented in the age range 18-54 while the population >54 is underrepresented. Spatially, the data showed a relatively homogenous spatial population representativeness[18]. In most Italian provinces (80%, more if considering population), the share of population with an active Facebook profile and location services on ranges between 6% - 9%.

We also checked the representativeness of the Facebook matrices among themselves and the correlation with a recent matrix that also includes non-systematic trips – that are dominant at the national scale. The chosen matrix is described in Section 5.7. The metric correlated here is the average daily flow (excluding

---

[10] http://dati-censimentopopolazione.istat.it/Index.aspx?lang=en

[11] The size of households tests the hypothesis that larger families, during lockdown, may increase the circulation of the virus. Similarly

[12] Over-65 is the group of people most affected by infection and the most useful for our purposes, because of the biased testing in the beginning of the infection period. The lack of test led to the fact that the virus was detected only in people admitted to the hospitals who had severe health problems mostly elderly people.

[13] The variable tests the hypothesis that infections are more where public transport is more used.

[14] An air pollutant consisting of small particles with an aerodynamic diameter less than or equal to a nominal 10 micrometer. The variable has been included because the severity of the outbreak seemed initially linked to the pollution level.

[15] Unemployment is introduced to control differences across provinces among infections at work.

[16] Population density verifies the hypothesis that denser areas involve more interpersonal relations. The variable is not included in the final model, because of its insignificance.

[17] The percent of positive swab tests is not included in the final model, because of its insignificance. Moreover, swab tests data are available at the NUTS-2 level only.

[18] The Facebook users are not fully representative of all demographic and socio-economic groups within a population, as well as their territorial distribution. Thus, an effort is made to calculate the territorial population penetration of Facebook users (with location services ON) relative to the official socio- demographic data provided by ISTAT (National Institute of Statistics) as of 1st January 2019. The data representativeness has been explained in much detail in Beria, P., Lunkar, V., 2021. Presence and mobility of the population during the first wave of Covid-19 outbreak and lockdown in Italy. *Sustainable Cities and Society, 65,* p.102616..

weekends) and the check is separate for the movements between administrative regions and the internal movements. We multiplied the normalized matrixes (colocation in week 9 and SCI) by the number of Facebook users (in week 9) to show their correlation. We chose week 9 because it represents the mobility pattern before the imminent restrictions. Table 2 tabulates their correlation values, showing a very high coincidence among them, especially between the two representing movements (i-TraM and Movements).

Table 2. The correlation of mobility matrixes and SCI.

Indicators	*Colocation*, week 9*	*Movements, week 9*	*SCI**	*i-TraM*
Colocation, week 9*	1.000			
Movements, week 9	0.931	1.000		
SCI*	0.906	0.952	1.000	
i-TraM	0.881	0.985	0.940	1.000

7. Facebook Colocation Model description

The general models' structure has been described in section 4. In this section, we describe in detail how the first model, the *Facebook Colocation* one, is implemented. The other models using alternative internal and external mobility variables other than *Facebook Colocation* data are discussed and compared in Section 9.

7.1 Local mobility ("internal")

As shown in Figure 4, the *colocation* metric explains that most of the encounters are inside the province of residence[19]. Since the average duration of infection is 2 weeks[20], only two previous periods of positive cases were left as predictors of new cases in this period. The main predictor of internal infection will then be the total infected people in each province:

$$\boldsymbol{NewFromInside_{jt}} = \boldsymbol{New_{jt-1}} + \boldsymbol{New_{jt-2}} \qquad \text{(Equation 7)}$$

We tested different options that quantify the rate of local contacts. An alternative measure including internal mobility is based on *internal colocation* values. The impact of local cases per inhabitant of province *j* and per tile:

$$\boldsymbol{NewColPerCapita_{jt}} = \sum_{\substack{i=1 \\ i=j}}^{N} \boldsymbol{New_{it}} * \boldsymbol{Col_{ijt}} \qquad \text{(Equation 8)}$$

where i – origin province; j – destination province; t – week number; N – number of provinces, in our case is equal to 110.[21]

[19] Actually, the large part of meetings is probably even more local, but the dataset is not available at a smaller scale than NUTS 3.

[20] https://www.who.int/news-room/q-a-detail/q-a-schools-and-covid-19

[21] At the initial stage, the epidemic did not affect all the provinces of Italy. In the 9th week, the epidemic was observed in 100 provinces, in the 10th and 11th - in 109. Only from the fourth week on, the epidemic was in all 110 Italian provinces. In order not to explain zero new cases of infection and to obtain unbiased estimates of the parameters of equation (8), only New_jt> 0 were included in the equations. Only at t≥12 the sample was complete and comprised all 110 observations.

7.2 Contact among provinces ("external")

The epidemic in each province started with incoming flows. We expect that cases will rise mostly in provinces strictly linked with already infected ones. We can quantify their impact using Facebook *external colocation* data. The impact of imported cases per inhabitant of province *j* and per tile will be accounted for using the variable

$$\boldsymbol{NewFromRestPerCapita_{jt}} = \sum_{\substack{i=1 \\ i \neq j}}^{N} \boldsymbol{New_{it} * Col_{ijt}} \qquad \text{(Equation 9)}$$

When, by multiplying the product by population of the province and the number of tiles in it, we obtain a value that is proportional to the new Covid-19 cases contacts in province *j* from the rest of Italy:

$$\boldsymbol{NewFromRest_{jt} = Pop_j * Tiles_j} \sum_{\substack{i=1 \\ i \neq j}}^{N} \boldsymbol{New_{it} * Col_{ijt}} \qquad \text{(Equation 10)}$$

7.3 Contact with relatives

As a result of the restrictions, the intensity of contacts between people had decreased. However, people may still get infected due to contacts with people living in the same environment. Since infection is not instantaneous and may take a week or more, infections continue to increase even when lockdown is in effect. Moreover, there were delays in the reporting of new cases. One of the predictors of the effect of restrictive measures will be the variable:

$$\boldsymbol{StayHome_{it} = Pop_{it} * StayH_{it}}, \qquad \text{(Equation 10)}$$

where $StayH_{jt}$ is proportional to per capita number of people in province *j* in week *t* that adhered to lockdown restrictions. Identically like active cases, Equation 9 and Equation 10 are included with two lags. The expected sign of this variable is not univocal: more people confined at home means fewer social contacts, but also more probability of transmission from cohabitants (e.g., grandparents from nephews).

7.4 The model

To assess the influence of the factors listed in equations (7), (9) and (10), as well as to further select those significantly affecting the number of new cases of Covid-19 infection, a linear regression model is proposed as follows:

$$New_{jt} = \beta_0 + \beta_1 \sum_{\tau=1}^{2} New_{jt-\tau} + \beta_2 \sum_{\tau=1}^{2} NewFromRest_{jt-\tau}$$

$$\boldsymbol{+\beta_3 \sum_{\tau=1}^{2} StayHome_{jt-\tau} + \beta_4 Family_j + \beta_5 Over65_j + \beta_6 PM_j + \beta_7 Unemployment_j + \epsilon_{jt}},$$

(Equation 11)

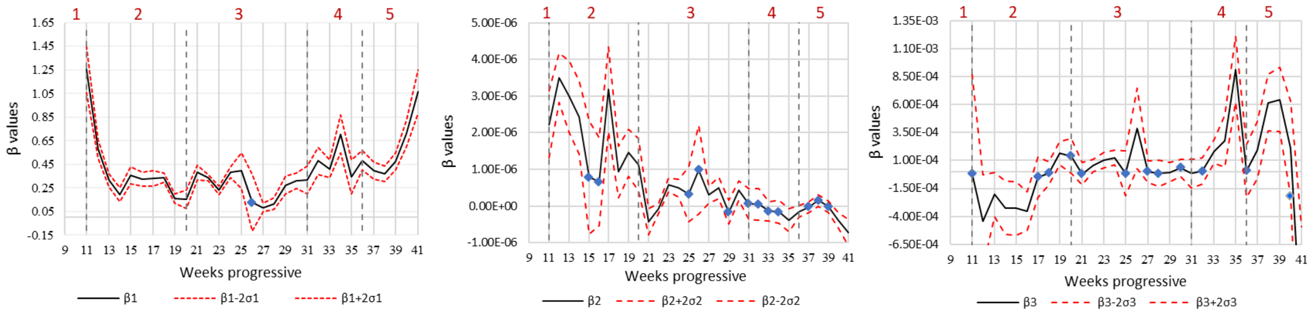
where ϵ_{jt} – random errors.

Control variables were selected from the longer list of Table 1, using stepwise selection procedures.

A second group of models is built substituting $\sum_{\tau=1}^{2} New_{jt-\tau}$ internal infections with previously introduced Equation 9, $\mathrm{NewColPerCapita_{jt}}$.

8. Facebook Colocation Model results

The equations were estimated by the ordinary least squares method separately for each week[22]. The dynamics of the spread of infection and the number of influencing factors do not allow for obtaining beta-coefficients suitable for the entire period of observation. Figure 8 and Figure 9 illustrate the obtained beta-coefficients as functions of time, showing when each of the three factors is more influential.



a) b) c)

Figure 8. Beta-coefficients graphs of (Equation 11), lag 1 week: a) internal infections, b) external infections, c) stay-at-home effect. The dots mean that the coefficient for that week is not significant, as also understandable from the confidence intervals.

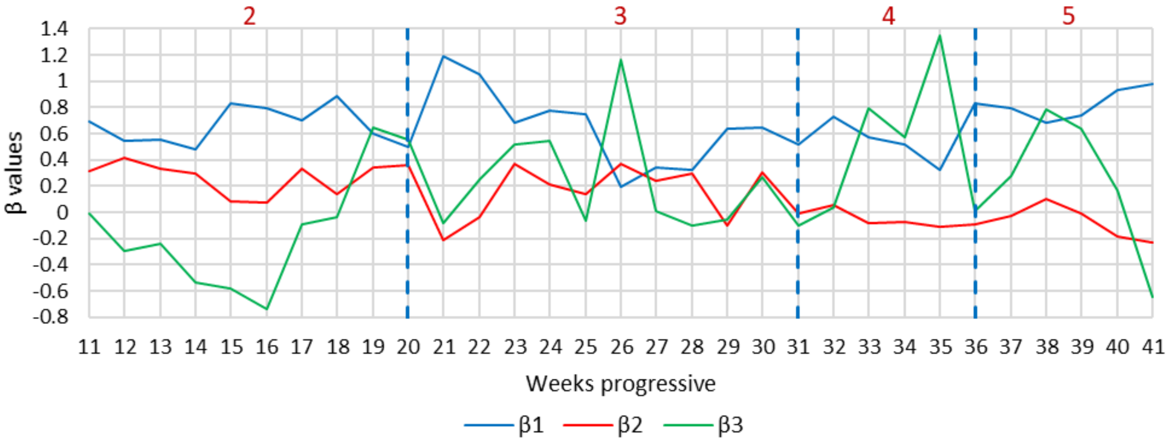


Figure 9. Scaled (standardized) beta-coefficients graphs (Equation 11)

8.1 The effect of internal and interprovincial mobility

The graph of the first beta-coefficient (β_1) in Figure 8a shows the influence of active cases within the province. These beta-coefficients are the approximate values of infection speed in Italy. During the very first weeks, the value is the highest because of the uncontrolled diffusion of the outbreak in Northern Italy. Moreover, during the first weeks, the new cases were underestimated, and this might also explain the out-of-scale result. The effect of lockdown (phase 2) is evident as betas fall and remain quite steady until phase 3 (post lockdown). Interestingly, betas resume growing following summer, anticipating the second wave of outbreak.

[22] As the unit of the analysed period is a week and epidemic curve is split into small segments so that each one can be better described by a linear function.

The second coefficient in Figure 8b is the effect of *colocation* among different provinces (β_2), describing the contribution of infections imported from outside. The coefficient is significant in most equations and phases. The beta-coefficients of interprovincial mobility are higher at the beginning of outbreak and their contribution to new cases is rather high, even at the national scale (Figure 9's standardized coefficients are comparable).

This coefficient must be interpreted spatially: if interactions among two areas are many, its contribution to new infections of the second province is high (for example, between Bergamo and Milan: many cases in Bergamo and many relations with Milan). To the contrary, if interprovincial interactions are few, even a high number of cases in the origin province does not cause many infections (for example, between Bergamo and any province of Southern Italy: very few contacts and, consequently, a low probability of spreading the virus). In other words, the fact that this coefficient is significant means that the mobility between provinces measured in terms of colocation can explain the dynamics of infections in Italian provinces. Ignoring this mobility component would result in a worse prediction of the evolution of the outbreak.

The β_2 sees low values during phase 4, the vacation period, when the beta coefficients become lower than zero and mostly insignificant. This apparently counterintuitive fact is because during the summer, Italians left their provinces and moved to holiday areas in other provinces (colocation is linked to the province of usual residence, not that of current position).[23]

8.2 The effect of staying at home

The product of population and stay-at-home of the province proxies the number of people who reduced their mobility and limited their contacts to cohabitants. Beta-coefficients are mostly negative (Figure 8c) and significant during the quarantine (phase 2), which confirms that the local lockdown was effective in reducing infections: the more people stay at home, the less new cases are registered. *Stay-at-home* is seldom significant after lockdown, which is reasonable.

8.3 Control variables

Control variables provide further insights into the determinants of the outbreak. Their coefficients can be checked in Table 6, Table 7 and Table 8 in the Appendix.

The variable *family density* has mainly significant and negative coefficients. In the urban provinces, people live in smaller families, which probably makes them more mobile and likely to live a freer lifestyle. In rural context, families are slightly larger, and the system of relations is more closed with respect to cities, preventing a rapid spread of the infection.

Detailed study of the effect of age on infection rates was not performed. However, the *Over 65* coefficient representing the share of elderly in provincial populations is significant in most periods and positive during the quarantine (phase 2).

Pollution, which is a known problem in Milan and Lombardy in general, has been initially associated with more virus circulation, as it was found to be significant in Coccia, M., 2020 [24]. In our model, we find that *PM* is significant in few weeks and does not have a great influence on the dependent variable. PM concentration assumes significance after multiplying by population, thereby proxying the number of people. The scarce significance found could suggest higher mortality rate due to pollution, but not the transmissibility of the virus.

[23] An example could help. A resident in Milan and one in Turin both move to Sicily for holidays. They meet. Colocation measures a contact between Milano and Turin, but the infection happens and is probably registered in Sicily. So, we apparently have a relatively high number of contacts between Milano and Torino, but no infections occur in the two, as they are elsewhere, at the seaside.

[24] The further spread of COVID in most of the world, not necessarily following pollution patterns, is a proof that the initial correlation was not causation.

We tested the influence of the *unemployment rate* to characterize the role of workplaces during the pandemic. Afterwards, it turned out to be insignificant in most periods, suggesting that workplaces are places of infection like others.

A further control variable has been tested because it is claimed as relevant in the public debate: the *share of public transport* use in the province. Being a static variable (not time-dependent in the time series), its value could be useful at the very beginning before the lockdown altered the mobility choices. Later, its value is insignificant and including it in the model does not improve results; moreover, the variable has negative coefficients, which is mathematically incoherent. For this reason, it has been excluded in the final version.

8.4 Overall interpretation

The three sources of infections assumed in the model appear to be significant and coherent with expectations. Their overall interpretation helps in interpreting the phenomenon into two parts: during the first phase of outbreak and during the lockdown.

Before the lockdown, the stay-at-home metric is not significant, and the new cases are mainly determined by active cases in the province (according to standardized regression coefficients).

Colocation indexes, representing interprovincial mobility, vary substantially within provinces and time. It means different provinces have different probabilities to undergo outbreaks of infection due to their reciprocal links. This intuitive result explains itself in terms of outbreak modelling capacity. It permits us to improve the correctness of outbreak models in unaffected (or slightly affected) provinces, owing to an explicit consideration of real mobility patterns instead of modelled ones.

During the lockdown, stay-at-home has a greater effect together with *over-65*, as most people who were ill (tested in the hospitals) at that period belonged to this age group. Links with other cities, instead, remain mostly significant but have a lower effect on the dependent variable. This represents the fact that long-distance mobility, heavily limited, happened by more secure private car or extensively using highly protective devices.

The policies limiting people's movements showed their impact on new cases, which is confirmed by the results: *stay-at-home* beta coefficients are negative mostly during the quarantine.

8.5 Residuals

The model's capability to predict the spatial pattern of the outbreak in the short term can be realised through the residuals between modelled and actual dependent variable. The new infections in every province and in the different weeks have a large variability (from zero or few units, to thousands). Nevertheless, the model forecasts their size quite well. The concentration of the virus mostly in Northern provinces in the beginning of the lockdown is clear, while already in week 19, it became quite homogenous all over the country.

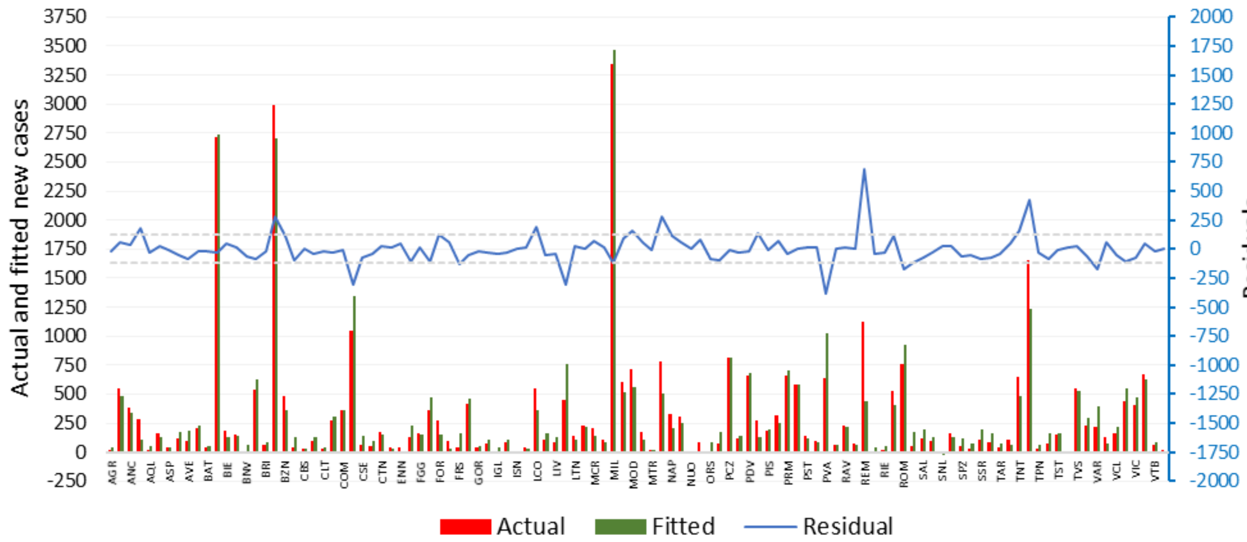


Figure 2. Figure 10. Residuals of colocation (Equation 11) model, week 12

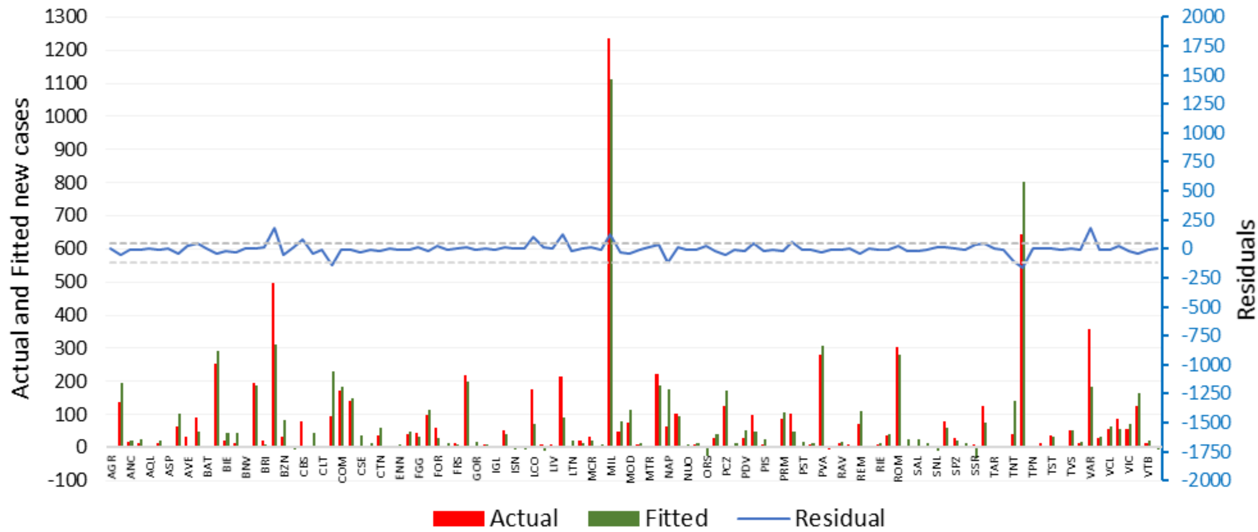


Figure 11. Residuals of colocation (Equation 11) model, week 19

Overall, residuals' plots demonstrate high forecast accuracy one week ahead, with few outliers as confirmed from Figure 10 and Figure 11**.** Figure 12 compares the 1-week forecast with those 2 and 3 weeks ahead. The quality of forecasts reduces significantly if the model, which is linear and fed with just 2 periods in advance, is used to forecast cases farther in time during the exponential phase. Better results are obtained when the rise slows down or is linear, but of course this is not particularly useful to set containment measures. To perform longer forecasts, a different approach should be used, typically using artificial recurrent neural network models such as LSTM (Luca et al, 2020). These models perform very well in terms of total cases, but their non-parametric nature makes them unsuitable for our purposes.

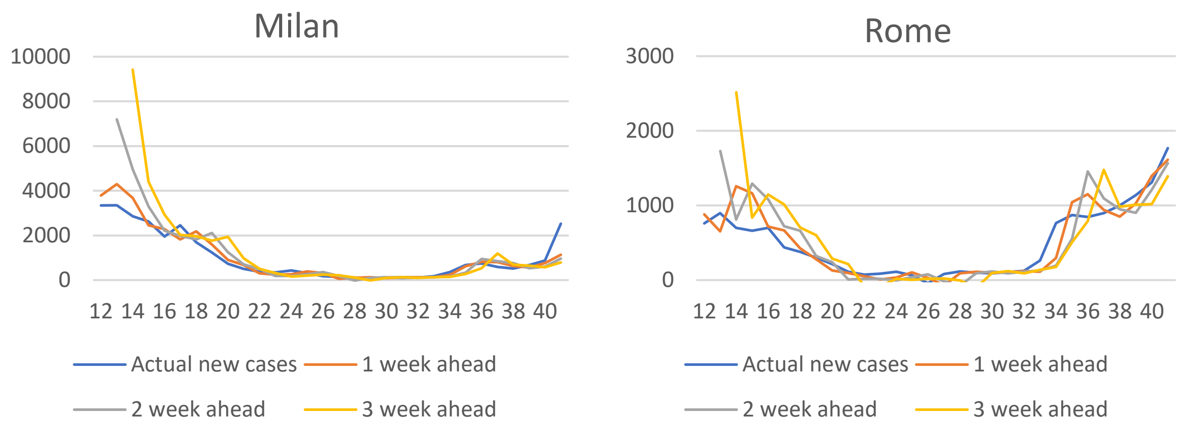


Figure 12. 1, 2 and 3 steps ahead forecasts compared with actual cases for Milan and Rome provinces.

9. Other models and comparison

The model described above uses *Facebook colocation* to measure the interactions between provinces, which caused the outbreak. The model proved to be capable of describing both the size and timing of infections. However, *colocation* is not the only proxy of movements at the national scale and could be unavailable on other occasions. The aim here is to build similar models using alternative measures of movements, namely: the *Facebook movements*, the *Facebook Social Connectedness Index (SCI)*, and a traditional O-D matrix, then test their performances to verify which better describe the spatial spread of the virus.

The measures differ, especially in two dimensions: the availability or not of a time series, and the origin of the data which can be based on movements or other kind of interactions.

Additional models differ from (Equation 11) based on *colocation*, by third term (β_2) that describes external sources of infection. The third terms of these models are shown in Table 3.

Table 3. Models with different external movements terms

Model	The third term (contacts among provinces)	Equation
Colocation	$\beta_2 Pop_j Tiles_j \sum_{\tau=1}^{2} \sum_{\substack{i=1 \\ i\neq j}}^{N} New_{it-\tau} Col_{ijt-\tau}$	(11) original model
Movements	$\beta_2 \sum_{\tau=1}^{2} \sum_{\substack{i=1 \\ i\neq j}}^{N} \frac{New_{it-\tau} Mov_{ijt-\tau}}{User_{it-\tau}}$	**Equation 12**
SCI	$\beta_2 \ Pop_j \sum_{\tau=1}^{2} \sum_{\substack{i=1 \\ i\neq j}}^{N} New_{it-\tau} SCI_{ijt-\tau}$	**Equation 13**
i-TraM	$\beta_2 \sum_{\tau=1}^{k} \sum_{\substack{i=1 \\ i\neq j}}^{N} New_{it-\tau} ITraM_{ij}/Pop_i$	**Equation 14**

In addition, we built a second set of the four models, replacing the second term (β1) initially defined in (8), to verify the hypothesis that *internal colocation* is not a reliable proxy of local contacts. If these models perform worse than the corresponding ones not considering intra-provincial mobility in β1 (see section 7.1), the original model formulation (section 7.4) will be confirmed.

9.1 Facebook movements model

The *Facebook Data for Good* programme provides an O-D matrix representing the real daily movements by province. Like *colocation*, it is based on real movements and has been available since February 24th, 2020. The shortcoming of this matrix is that it includes only O-D pairs beyond the threshold of 10 users. Consequently, during the lockdown, we only found about 800 O-Ds instead of 12100 (all combinations among 110 Italian provinces) found in the *colocation*. Nevertheless, the data is not normalized and represents live data and could, in principle, be significant.

Unlike colocation (Equation 2 and Equation 3), movements show the absolute number of FB users Mov_{ijt}, who move from province *i* to province *j* at week *t*. Dividing the movements from *i* to *j* by the number of users in *i*, we get the share of people who went from province *i* to province *j*. Multiplying this share by the number of new cases *I*, we get the expected number of infected people that will travel at week *t* from province *i* to province *j*:

$$\sum_{\substack{i=1 \\ i \neq j}}^{N} New_{it} Mov_{ijt} / User_{it-\tau}$$

9.2 Facebook SCI model

The third proxy of potentially infected contacts (interprovincial) is based on social connectedness. The hypothesis is that people usually visit places where they have friends or relatives. Based on this assumption, *Social connectedness index* can also be a proxy to explain the people's movements.

SCI is built in the same way as *colocation* (compare Equation 2 and Equation 6). That is why our SCI-based proxy is built in the same way (Equation 11). However, using SCI for modelling in the same manner as time-varying data, we expect a lower significance due to its static nature, not representative of the mobility changes in the lockdown. Adopting the same logic of the previous models, we obtain equation with third term shown in Table 3 (Equation 13).

9.3 Static OD matrix models

We redefine a model based on the traditional O-D matrix and use the one described in 4.6. This represents the same idea as movements but considers estimated average number of daily trips among provinces based on real Italian populations. Dividing the number of trips per population (as in Equation 12) we obtained trips per capita and define the model with a third term (Equation 14) in Table 2, where $ITraM_{ij}$ is the number of people that move from province *i* to *j* according to i-TraM data (weekly) and $\sum_{\substack{i=1 \\ j \neq i}}^{N} New_{it} * I - TraM_{ij} / Pop_i$ is the number of potentially infected people that move from the rest of Italy to province *j* at week *t*.

9.4 Variants including internal movements

From (11) to (14) we used only interprovincial values (*i≠j*) because including the internal movements in the equations showed worse results. In this section, we want to verify explicitly the appropriateness of this assumption. To do so, we substituted the second terms (□1) of equations according to Table 4, leaving other terms in (Equation 11 – Equation 14) unchanged.

Table 4. Models with different internal movement terms.

Model	The second term (internal mobility by active cases) in Equation (11)	Equation
Colocation	$\beta_1 \sum_{\tau=1}^{2} New_{jt-\tau}$	(11) original model
Colocation	$\beta_1 Pop_j Tiles_j \sum_{\tau=1}^{2} New_{jt-\tau} Col_{jjt-\tau}$	**Equation 15**

Movements	$\beta_1 \sum_{\tau=1}^{2} \frac{New_{jt-\tau} Mov_{jjt-\tau}}{User_{jt-\tau}}$	**Equation 16**
SCI	$\beta_1 Pop_j \sum_{\tau=1}^{2} New_{jt-\tau} SCI_{jjt-\tau}$	**Equation 17**
i-TraM	$\beta_1 \sum_{\tau=1}^{k} New_{jt-\tau} Meta_{jj}/Pop_j$	**Equation 18**

9.5 Models comparison

We initially compare the *betas* of the four types of models (for a total of 124 equations based on Equation 11 – Equation 14) in turn based on the four matrixes.

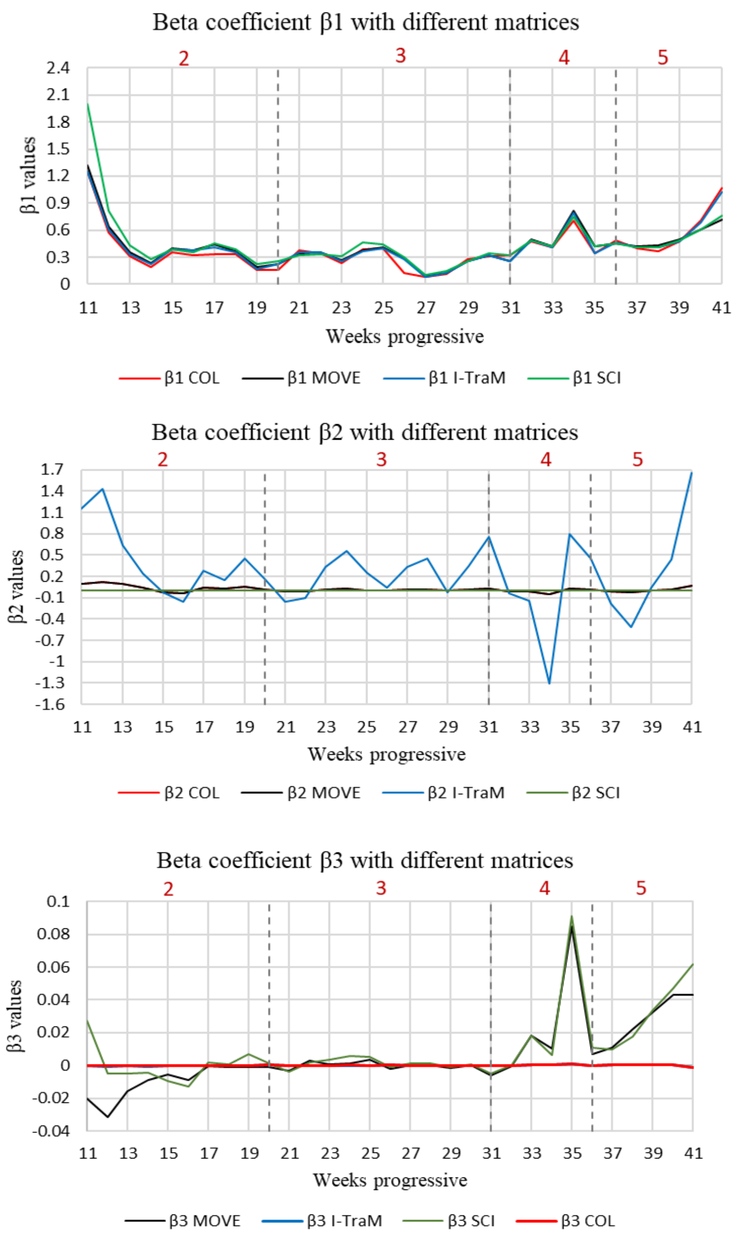


Figure 13 The comparison of beta-coefficients of the equations with different mobility matrixes Colocation (11), Movements (12), SCI (13), i-TraM (14).

Figure 13 plots the coefficients. $\beta 1$ presents a similar trend in all four models and the differences in estimation are because of the $\beta 2$, which changes in the formulation. $\beta 2$ is obviously the most variable. The beta for i-TraM matrix is the most different, but even the one for SCI deviates from the *colocation* and *movements* ones. Also, $\beta 3$, which is quite constant in the *colocation* model, varies significantly from the others.

Similar elaborations are done for the other four variants of the models (Equation 15 to Equation 18).

To compare all eight types of models, the predicted values of new infections were calculated for one period ahead of the sample on which the model was built. Then, the square deviations of the predicted new cases from the actual ones were computed. The performed calculations allow us to find the pseudo-determination coefficient:

$$\widetilde{\boldsymbol{R}}^2 = \mathbf{1} - \sum_{t=11}^{40} \sum_{j=1}^{110} \left(\boldsymbol{New}_{jt+1} - \widehat{\boldsymbol{New}}_{jt+1}\right)^2 / \sum_{t=11}^{40} \sum_{j=1}^{110} \left(\boldsymbol{New}_{jt+1} - \overline{\boldsymbol{New}_j}\right)^2, \qquad \text{(Equation 19)}$$

where $\widehat{New}_{jt+1}$ step ahead forecast according to ((Equation 11) – (Equation 14)) and Table 4.

$\overline{New_j} = \frac{1}{30} \sum_{t=12}^{41} New_{jt}$ is the historical mean new cases forecast for province j.

The pseudo-determination coefficient is also calculated for each province separately using the formula,

$$\widetilde{\boldsymbol{R}}_j^2 = \mathbf{1} - \sum_{t=11}^{40} \left(\boldsymbol{New}_{jt+1} - \widehat{\boldsymbol{New}}_{jt+1}\right)^2 / \sum_{t=11}^{40} \left(\boldsymbol{New}_{jt+1} - \overline{\boldsymbol{New}_j}\right)^2 \qquad \text{(Equation 20)}$$

The comparison of the equations' results based on different types of mobility data in two modes (with internal values and without them) are in the first four columns of Table 5.

Table 5. The comparison of pseudo R2 of the equations, according to three variants: the formulations for component 1 and 2 and the presence of control variables. Weeks 11- 41 (one week ahead).

Local mobility component	*Contacts from other provinces component (Matrix of movements)*				*Control variables*	
	Colocation	*Movements*	*SCI*	*i-TraM*	*Base model*	*Base model with control variables*
Basic formulation	**0.724**	0.661	0.530	0.634	0.517	0.603
With internal colocation values	-0.083	-1.230	0.261	0.578	-	-

The best results are obtained with the *colocation* and *movements matrixes*, as expected: they are the only time-continuous mobility data that can represent the evolution of the outbreak and the effect of restrictions. *SCI* and *i-TraM* matrices, being static, can be used but are less performing in describing the outbreak evolution. The four models including intra-provincial movements (2nd row) perform worse in terms of ability to forecast than the corresponding ones without, and must not be used, validating the choices made since Section 7.

To estimate the significance of the mobility factors and control variables, we created two further models, also in Table 5, which consider only the number of infected people (base model), with (equation 22) and without control variables (equation 21).

$$New_{jt} = \beta_0 + \beta_1 \sum_{\tau=1}^{2} New_{jt-\tau} + \epsilon_{jt} \qquad \text{Equation 21}$$

$$New_{jt} = \beta_0 + \beta_1 \sum_{\tau=1}^{2} New_{jt-\tau} + \beta_4 Family_j + \beta_5 Over65_j + \beta_6 PM_j + \beta_7 Unemployment_j + \epsilon_{jt},$$

Equation 22

The results of (21) and (22) are presented in the appendix. Pseudo R2 of the base model is 0.52, while the one with the control variables is 0.60. Both values are substantially lower than the one of (11) (see Table 5), certifying that the addition of interprovincial movements and stay-at-home metrics have improved the significance of the model.

Finally, based on calculations from Equation 20 with *colocation* matrix in Figure 14, we built a map illustrating the quality of forecasting for the Italian provinces according to the best-performing model. Provinces with $\tilde{R}^2 < 0.3$ are relatively few and belong to two cases. In the Centre-South, the outbreak was limited in the first COVID wave and thus residuals are calculated on very small infection numbers. The provinces of Bergamo, Lodi, Cremona and Pavia (but also Pesaro-Urbino) were the places of the very first cases and, excluding Pavia, the most hit. Here, the out-of-scale effects of the outbreak but especially the lack of testing capacity resulted in a significant underestimation of reported cases, which explains the low $\tilde{R}^2$. If reported cases were realistic (or had a similar rate of representativeness as that of other provinces), these would not look like outliers.

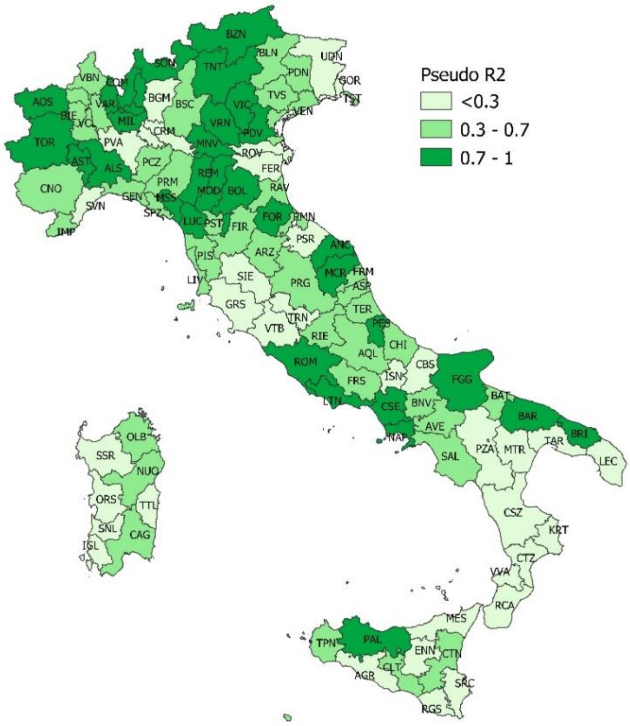


Figure 14. Pseudo R2 values of Italian provinces

10. Conclusion

The evolution of an infection depends on the number of contacts between people. Most contacts occur among people living in proximity: relatives, colleagues, schoolmates, on public transport, in shops, etc. However, the speed and the leap-frog-like diffusion of COVID-19 across the world showed the role of long-distance mobility. This kind of effect is not typically considered in traditional models, which assume population as a single pool.

The paper responds to two different goals: showing that the consideration of mobility data is significant in explaining the evolution of the outbreak and pointing out the best performing metrics to support the claim.

The availability of detailed datasets concerning mobility during the period of the first COVID-19 wave helped to model the spatial diffusion of the infection across Italian provinces. These datasets are the Facebook normalised *colocation* (probability of two people from Italian provinces being in the same tile within 5-minute intervals), Facebook normalized *movements between administrative regions*, Facebook *Social connectedness index* (illustrating the strength of virtual friendships between provinces), Facebook *Stay-at-Home* metric, and a traditional static multipurpose O-D matrix. Facebook data is available for many countries across the world starting from the COVID outbreak[25] and is shared with over 450 partner organizations across nearly 70 countries. Therefore, the potential replicability of this approach is large and goes beyond both the COVID outbreak and our case.

We built a model that explains the number of new infections in a province (NUTS-3) from three main components, each one with a potentially different weight over time. The first is the infections that occur locally from everyday contacts. Without a measure of local interactions, this component is simply proxied though the number of active cases in the province. The second component describes the imported infections, which is proxied by the above-mentioned mobility metrics. The third component refers to the immobility of people that, on one side, reduces the social contacts, but alternatively, amplifies the spread among cohabitants. To test the soundness of our choices, we created ten model variants that, when multiplied by the number of weeks modelled, make a total of more than 300 equations.

The analysis shows that the interprovincial mobility component is relevant in explaining the observed infections and mobility data and should be considered to predict the appearance of the first cases of infection in different provinces. Our results are coherent with earlier findings that mobility reductions reduce infections. Concerning the choice of the OD matrix to be used for the movement component, we found that, understandably, static matrices (traditional multipurpose ODs or *Facebook SCI*) are unable to represent the evolution of the outbreak. If available, higher forecast accuracy can be obtained in the presence of real data such as the *Facebook Movements* or *Colocation*, which is particularly important if restrictions are introduced to assess the effectiveness of the proposed measures.

The policy consequences of this model are obvious: it is possible to predict *where* new local outbreaks will happen thanks to mobility data, as the spread is not homogeneous in space, but follows mobility patterns of people. This also means that generalized restrictions, which heavily affect economy, could be made selective, but equally effective. On the contrary, once the outbreak occurs, interprovincial mobility gives a negligible effect and can thus be allowed for specific reasons, since most of the population is already self-confined at home.

The paper has some limits to be mentioned. The forecast capability is limited to one week during the initial (exponential) rise of cases. Once cases grow less, it is more precise also 2-3 weeks ahead, but clearly less useful. This temporal limit makes it a tool for understanding the drivers better than for forecasting cases. Possible improvements and future work, in fact, may include the development of properly said forecasting tools; for example, based on nonlinear models or neural networks, that also embed mobility data.

[25] The only thing to note here with respect to other countries is that the data is available from the dates when the initial COVID-19 cases were observed, which varies from country to country

References

Backer, J. A., Klinkenberg, D., & Wallinga, J., 2020. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* 25(5), p.2000062.

Beria P., Bertolin A., Tolentino S., Debernardi A., Ferrara E., Filippini G., 2019. *A model-based evaluation of national transport policies*. Paper presented at: XXI SIET conference, Bologna, September 9th-10th, 2019.

Beria, P., Lunkar, V., 2021. Presence and mobility of the population during the first wave of Covid-19 outbreak and lockdown in Italy. *Sustainable Cities and Society*, *65*, p.102616.

Burstein, R., Hu, H., Thakkar, N., Schroeder, A., Famulare, M. and Klein, D., 2020. *Understanding the impact of COVID-19 policy change in the greater Seattle area using mobility data*. Institute for Disease Modeling.

Chang, M. C., Kahn, R., Li, Y. A., Lee, C. S., Buckee, C. O., and Chang, H. H., 2020. Modeling the impact of human mobility and travel restrictions on the potential spread of SARS-CoV-2 in Taiwan. *medRxiv*.

Chang S., Pierson E., Koh P., Gerardin J., Redbird B., Grusky D., and Leskovec, J., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840), pp. 82-87.

Chen, Z., Zhang, Q., Lu, Y., Guo, Z M., Zhang, X; Zhang, W. J., Liao, C.H., Li, Q.L., Han, X.H. and Lu, J.H., 2020. Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan. *Chinese Medical Journal*.

Cintia, P., Fadda, D., Giannotti, F., Pappalardo, L., Rossetti, G., Pedreschi, D., Rinzivillo, S., Bonato, P., Fabbri, F., Penone, F and Savarese M., 2020. The relationship between human mobility and viral transmissibility during the COVID-19 epidemics in Italy. *arXiv preprint arXiv:2006.03141.*

Coccia, M., 2020. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Science of the Total Environment, 729:138474.*

Colizza, V., Barrat, B., Barthe´lemy, M., Vespignan, A., 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7), pp.2015-2020.

Della Rossa, F., Salzano, D., Di Meglio, A., De Lellis, F., Coraggio, M., Calabrese, C., Guarino, A., Cardona-Rivera, R., De Lellis, P., Liuzza, D. and Iudice, F.L., 2020. A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nature communications*, *11*(1), pp.1-9.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. and Colaneri, M., 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature medicine*, *26*(6), pp.855-860.

Giuliani, D., Dickson, M. M., Espa, G., and Santi, F., 2020. Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (COVID-19) in Italy. *Available at SSRN 3559569*.

Ilin, C., Annan-Phan, S., Tai, X.H., Mehra, S., Hsiang, S. and Blumenstock, J.E., 2021. Public mobility data enables covid-19 forecasting and management at local and global scales. Scientific reports, 11(1), pp.1-11.

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., Du Plessis, L., Faria, N.R., Li, R., Hanage, W.P. AND Brownstein, J.S., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, 368(690):4493-97.

Kuchler T., Russel D., Stroebel J., 2020. The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook (No. w26990). National Bureau of Economic Research.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., Bania, J and Khosrawipour, T., (2020). The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. *Journal of Microbiology, Immunology and Infection*. 53(3), pp.467-472.

Le M., Ibrahim M., Sagun L., Lacroix T., Nickel M. Neural Relational Autoregression for High-Resolution COVID-19 Forecasting. https://ai.facebook.com/research/publications/neural-relational-autoregression-for-high-resolution-covid-19-forecasting/

Legambiente (2019). Mal'Aria di città 2019. La mobilità urbana al centro del risanamento. Dossier di Legambiente, Roma (Italy)

Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L. (2020). A Survey on Deep Learning for Human Mobility. arXiv preprint arXiv:2012.02825.

Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number *Infectious Disease Modelling*, 5, 129-141

Sallah, K., Giorgi, R., Bengtsson, K., Lu, X., Wetter, E., Adrien, P., Rebaudet, S., Piarroux, R. and Gaudart, J., 2017. Mathematical models for predicting human mobility in the context of infectious disease spread: introducing the impedance model. *International journal of health geographics, 16(1), pp.1-11.*

Simini, F., Gonzalez, M.C, Maritan A., & Barabasi A-L., 2012, A universal model for mobility and migration patterns. *Nature*, 484(7392), pp.96–100.

Stefanouli, M., Polyzos, S., 2017. Gravity vs radiation model: two approaches on commuting in Greece. *Transportation Research Procedia*. 24, pp.65–72.

Thakkar N., Burstein R., Hu H., Selvaraj P. and Klein D., 2020. Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA. *Institute for Disease Modeling.*

Tizzoni M., Bajardi P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., Gonzales, M.C. and Colizza, V., 2014. On the human mobility proxies for modelling epidemics. *PLoS Comput Biol*, 10(7), p.e1003716.

Troko, J., Myles, P., Gibson, J., Hashim, A., Enstone, J., Kingdon, S., Packham, C., Amin, S., Hayward, A. and Van-Tam, J.N., 2011. Is public transport a risk factor for acute respiratory infection? *BMC Infectious diseases, 11*(1), pp.1-6.

Turner, B. L., 2002. Contested identities: Human-environment geography and disciplinary implications in a restructuring academy. *Annaals of the American Association of Geographers*. 92 (1), pp.52–74.

Wu, K., Darcet, D., Wang, Q., and Sornette D., 2020 Generalized logistic growth modelling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear dynamics* 101(3), pp.1561-1581.

Zhang, Y., Zhang, A., and Wang, J., 2020. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transport Policy*, 94, pp.34–42

Appendix: Parameters of models by weeks based on colocation indexes (Equation 11)

Table 6. Parameters of models by weeks based on colocation indexes (Equation 11)

	New 11	New 12	New 13	New 14	New 15	New 16	New 17
New(t-τ)	1.253723***	0.576923***	0.313398***	0.191816***	0.356698***	0.323925***	0.330481***
NewFromRest(jt-τ)	0.000002***	0.0000035***	0.0000029***	0.00000243***	7.82E-07	7.03E-07	0.0000033***
StayHome(jt-τ)	2.50E-06	-0.000438**	-0.000214*	-0.000328**	-0.000329**	-0.000348***	-2.67E-05
Family	-45.58453	0.027619	15.07648	-140.5685	3.065413	-0.285733	-33.4466
Over65	0.000163	0.001538***	0.001468**	0.002196***	0.001786***	0.001878***	2.19E-05
PM	-1.145582	4.405566**	2.430826	1.56968	-0.004018	1.725757	-1.270805
Unemployment	-2.955082	-3.486955	-3.829632	0.019407	0.042211	1.439483	4.691632**
C	180.2977	-50.0839	-45.28942	293.8814	-39.72054	-95.92899	30.84881
R2-adj	0.820614	0.943944	0.941497	0.890497	0.893011	0.922099	0.929819
	New 18	**New 19**	**New 20**	**New 21**	**New 22**	**New 23**	**New 24**
New(t-τ)	0.335524***	0.162667***	0.156391***	0.381766***	0.340071***	0.2301***	0.383653***
NewFromRest(jt-τ)	0.000001***	0.0000014***	0.000001***	-0.0000004**	-5.72E-08	0.00000058**	0.0000005***
StayHome(jt-τ)	-1.01E-05	0.000165***	0.000141*	-2.06E-05	0.0000533*	0.000102***	0.000127***
Family	-9.292308	-21.91558	-47.44696	-9.037685	-11.11351	3.265425	2.59708
Over65	2.18E-05	-0.000658***	-0.000477*	1.29E-06	-0.000166**	-0.000217***	-0.000216**
PM	-1.391802*	-0.09912	0.530573	0.165357	0.093143	-0.130307	-0.372612*
Unemployment	1.380988	-0.633576	-0.401399	0.230823	0.411423	-0.186842	-0.371883
C	19.83437	67.53567	118.424	18.58326	24.89368	0.009637	6.870216
R2-adj	0.959387	0.902248	0.787283	0.899286	0.960277	0.918097	0.9494
	New 25	**New 26**	**New 27**	**New 28**	**New 29**	**New 30**	**New 31**
New(t-τ)	0.393801***	0.127804	0.084074***	0.117185***	0.273988***	0.310147***	0.316947***
NewFromRest(jt-τ)	3.18E-07	9.68E-07	0.0000003***	0.0000005***	-1.76E-07	4.02E-07	6.31E-08
StayHome(jt-τ)	-1.53E-05	0.000383*	9.86E-07	-1.48E-05	-0.0000068*	3.07E-05	-8.57E-06
Family	-10.27261	61.92711	-22.28097**	-30.13475**	-6.447622	6.347326	4.958551
Over65	-3.48E-05	-0.000754**	5.55E-05	8.22E-05	6.13E-05	-2.89E-05	6.70E-05
PM	-0.135372	0.868929	0.406871*	0.315601	0.30487	-0.15632	0.254809
Unemployment	-0.493209	-0.519308	-0.001312	-0.080296	-0.161557	-0.287646	-0.039932
C	34.27772	-146.7599	45.20195*	65.63814**	9.732251	-6.226753	-18.71875
R2-adj	0.641325	0.167271	0.61574	0.585542	0.651611	0.773792	0.551603
	New 32	**New 33**	**New 34**	**New 35**	**New 36**	**New 37**	**New 38**
New(t-τ)	0.476128***	0.412121***	0.704438***	0.341297***	0.480222***	0.394892***	0.372716***
NewFromRest(jt-τ)	3.12E-08	-1.33E-07	-1.57E-07	-0.0000004**	-0.00000015*	-2.68E-08	0.00000017**
StayHome(jt-τ)	-7.28E-07	0.000179***	0.000272**	0.000878***	1.00E-05	0.000197	0.000603***
Family	5.837445	-7.946454	-23.05728	3.594745	24.84749	-5.05734	-30.14617
Over65	-1.45E-05	-7.37E-05	2.01E-05	-0.000525**	0.000219	-7.79E-05	-0.00061***
PM	0.882143***	-0.484192*	-1.95792***	-1.468131**	0.101989	-0.801433	-0.616265
Unemployment	-0.245856	0.008549	-0.189709	-0.436311	-0.87352	-0.271633	0.541711
C	-24.40204	31.32725	85.41803	32.17657	-53.95717	42.55745	93.53047*
R2-adj	0.656163	0.846407	0.850537	0.880032	0.931256	0.928158	0.939253
	New 39	**New 40**	**New 41**				
New(t-τ)	0.466983***	0.697425***	1.052242***				
NewFromRest(jt-τ)	-2.34E-08	-0.0000004*	-0.0000007**				
StayHome(jt-τ)	0.000673***	0.000255	-0.001441***				
Family	19.10178	92.23089***	89.25825				
Over65	-0.0005***	1.27E-05	0.002816***				
PM	-0.184907	0.662792	0.289956				
Unemployment	-0.762803	-1.570887	-1.92494				
C	-38.30716	-227.9817**	-249.661				

	New 11	**New 12**	**New 13**	**New 14**	**New 15**	**New 16**	**New 17**
New(t-τ)	1.56586***	0.955287***	0.526907***	0.356149***	0.407613***	0.388203***	0.450229***
C	49.08372***	98.12261***	76.20002***	49.94426*	-2.626010	-8.542444	-36.99605*
R2-adj	0.732169	0.806147	0.869915	0.823950	0.886430	0.908174	0.907832
	New 18	**New 19**	**New 20**	**New 21**	**New 22**	**New 23**	**New 24**
New(t-τ)	0.371493***	0.250531***	0.274223***	0.304603***	0.315099***	0.317341***	0.477246***
C	-21.448110	1.615088	4.346142	0.172837	-2.019112	-2.318430	-1.864254
R2-adj	0.956732	0.872134	0.764702	0.892708	0.954933	0.883590	0.932614
	New 25	**New 26**	**New 27**	**New 28**	**New 29**	**New 30**	**New 31**
New(t-τ)	0.422695***	0.224053***	0.151111***	0.195979***	0.339885***	0.411253***	0.453504***
C	-3.810496	-4.739084*	9.749293**	9.852965***	3.965391***	4.51068***	3.700222*
R2-adj	0.641033	0.113956	0.363837	0.291355	0.618081	0.732870	0.520940
	New 32	**New 33**	**New 34**	**New 35**	**New 36**	**New 37**	**New 38**
New(t-τ)	0.517116***	0.598853**	1.162224***	0.931645***	0.563987***	0.479467***	0.515433***
C	7.176608**	6.597648*	-8.562269	0.774525	8.970167*	11.26716*	6.060358*
R2-adj	0.634196	0.682956	0.717196	0.757372	0.927940	0.926886	0.921561
	New 39	**New 40**	**New 41**				
New(t-τ)	0.624552***	0.735907**	1.024633***				
C	-9.07002*	-2.933993	28.23704*				
R2-adj	0.943586	0.933727	0.889054		**Pseudo R^2= 0.517365**		
R2-adj	0.954483	0.94357	0.914773		**Pseudo R2-0.74**		

Table 7. Parameters of base model, which include new cases only by weeks (Equation 12)

Table 8. Parameters of base model, which include new cases and control variables by weeks

	New 11	New 12	New 13	New 14	New 15	New 16	New 17
New(t-τ)	1.44152***	0.787159***	0.430967***	0.27311***	0.384224***	0.357458***	0.467278***
Family	-62.67911	-72.6346	-32.52159	-226.359**	-83.22089	-81.30813	2.47869
Over65	0.000623**	0.001112***	0.000931***	0.000797***	0.000195	0.000185*	-1.78E-05
PM	-0.681488	4.778519	2.723423	1.446995	-0.3466	1.264288	-0.77601
Unemployment	-4.367974	-6.061836	-5.225755	-0.344022	0.50503	2.039184	3.603582
C	191.3273	107.5558	64.54266	514.6574	191.6034	127.6434	-63.74858
R2-adj	0.778749	0.881548	0.915599	0.865782	0.885075	0.909364	0.908489
	New 18	**New 19**	**New 20**	**New 21**	**New 22**	**New 23**	**New 24**
New(t-τ)	0.383138	0.229887***	0.257021***	0.318906***	0.332457***	0.316405***	0.465005***
Family	0.817249***	27.62989	-8.576136	-16.87127	-7.376491	17.24778	13.06051
Over65	-1.02E-05	0.00011**	2.41E-05	-0.000059*	-0.000046**	1.58E-05	0.0000373**
PM	-1.235084	0.391127	0.688639	0.232234	0.081092	-0.124241	-0.328203
Unemployment	0.940677	-1.543419	-0.955439	0.295531	0.441602	-0.357912	-0.440227
C	-3.761651	-68.0955	16.775	37.68991	13.5824	-39.0995	-24.72726
R2-adj	0.956954	0.874838	0.760969	0.89568	0.959415	0.883929	0.937991
	New 25	**New 26**	**New 27**	**New 28**	**New 29**	**New 30**	**New 31**
New(t-τ)	0.445351**	0.289597***	0.109345***	0.145854***	0.255315***	0.342156***	0.320544***
Family	-11.72565	82.47472*	-21.60612*	-29.8816*	-6.228771	5.686254	3.965196
Over65	-4.73E-05	-0.000101*	0.0000785**	0.0000946*	0.0000407**	0.0000408***	0.0000612**
PM	-0.076806	1.054249	0.395586	0.259737	0.271182	-0.044077	0.2711
Unemployment	-0.520197	-0.692878	-0.169881	-0.264548	-0.119212	-0.399468	-0.057314
C	36.66114	-213.1253*	43.99825*	66.53566*	10.1073	-7.564986	-16.40221
R2-adj	0.64537	0.140876	0.594894	0.546021	0.654634	0.758799	0.559581
	New 32	**New 33**	**New 34**	**New 35**	**New 36**	**New 37**	**New 38**
New(t-τ)	0.478522***	0.412707***	0.746411**	0.519292***	0.491083***	0.431042***	0.450894***
Family	5.654756	3.517241	-5.353836*	55.97336	22.19668	0.639329	1.890016
Over65	-1.26E-05	0.00015***	0.000311**	0.000491***	0.000134**	0.000123**	0.000157**
PM	0.888303**	-0.443344	-1.892771	-0.631565	0.207419	-0.712734*	-0.603049
Unemployment	-0.254328	-0.0047	-0.23857**	-0.672477	-0.760107	-0.151283	0.827385
C	-24.03967	-2.060443	33.01755	-139.2643	-49.72297	21.79545	0.15283
R2-adj	0.662696	0.82526	0.839991	0.81434	0.930046	0.927797	0.92535
	New 39	**New 40**	**New 41**				
New(t-τ)	0.576923***	0.748572***	0.893042***				
Family	47.2773*	80.01488**	-8.236381				
Over65	0.000109	-5.99E-05	0.000452**				
PM	-0.054367	1.119319	1.811821				
Unemployment	0.119185	-0.451868	-1.096768				
C	-127.8722*	-215.0971*	-12.08717				
R2-adj	0.946096	0.935421	0.892734		**Pseudo R^2=0.603**		