CoViD–19: An automatic, semiparametric estimation method for the population infected in Italy

Livio Fenga Corresp. 1

1 ISTAT, Rome, Italy

Corresponding Author: Livio Fenga
Email address: fenga@istat.it

To date, official data on the number of people infected with the SARS-CoV-2 - responsible for the CoViD–19 - have been released by the Italian Government just on the basis of a non representative sample of population which tested positive for the swab. However a reliable estimation of the number of infected, including asymptomatic people, turns out to be crucial in the preparation of operational schemes and to estimate the future number of people, who will require, to different extents, medical attentions. In order to overcome the current data shortcoming, this paper proposes a bootstrap– driven, estimation procedure for the number of people infected with the SARS-CoV-2. This method is designed to be robust, automatic and suitable to generate estimations at regional level. Obtained results show that, while official data at March the 12th report 12.839 cases in Italy, people infected wiyh the SARS-CoV-2 could be as high as 105.789.

CoViD–19: An Automatic, Semiparametric Estimation Method for the Population Infected in Italy

Livio Fenga

Italian National Institute of Statistics
ISTAT, Rome, Italy 00184
livio.fenga@istat.it

Abstract: To date, official data on the number of people infected with the SARS-CoV-2 - responsible for the CoViD–19 - have been released by the Italian Government just on the basis of a non-representative, heavily skewed, sample of population. Such a bias is due to the fact that *ad hoc* lab tests are administrated only to those showing flu-related symptoms. However, a reliable estimation of the number of infected, including the asymptomatic people, is a vital information for the implementation of policies and actions aimed at counteracting the spread of the virus. Therefore, this paper proposes a bootstrap–driven estimation procedure for the number of people infected with the CoViD–19. This method is designed to be robust, automatic and suitable to generate estimations at a national and regional level. The result obtained show that, while official data at March the 12th report 12.839 cases in Italy, the number of people infected with the CoViD–19 could be as high as 105.789.

KEYWORDS: Autoregressive metric; CoViD–19; maximum entropy bootstrap; model uncertainty; number of Italian people infected.

1. Introduction

COVID-19 epidemic has severely hit Italy and its diffusion throughout Europe is expected soon. In such a scenario, the availability of reliable information related to its spread plays a significant role in many regards. In fact, many targeted measures, such as the coordination among emergency services or the implementation of operative actions (e.g. hard or light lock-downs or even curfew) can only be efficiently taken when reliable estimates of the epidemic spread are available at the population level.

At the moment, official data on the infection in Italy are based on non-random, non-representative samples of the population: people are tested for CoViD-19 on the condition that some symptoms related to the virus are present. These data can ensure a proper estimation of the number of both deaths and hospitalizations due to the virus and are crucial for the optimization of the available resources. Nonetheless, from a statistical point of view, the number of people tested positive for CoViD-19 represents a simple count which is not suitable to provide a reliable assessment of the "true", unknown, number of infected people (thereafter "positive cases"). In addition to the strong bias components induced by this testing strategy, there is at least another mayor obstacle to the construction of a valid estimator: the small sample size available. Unavoidable at the early stages of an epidemic, the shortness of the time series of interest might lead to a strong bias in the asymptotic results and therefore to the construction of incorrect confidence intervals.

29

30   Based on the number of the deaths and of the observed positive cases and im-
31 proving on the methodology originally proposed by Pueyo (2020), this paper aims at
32 estimating the "true" number of people infected by the CoViD-19 in each of the 20 Ital-
33 ian regions. In more details, the presented procedure is designed to reduce the impact
34 of the biasing components on the parameter estimations, by employing a resampling
35 scheme, called Maximum Entropy Bootstrap and proposed by Vinod et al. (2009). In
36 addition to that, a distance measure – based on the theory of stochastic processes and
37 proposed by Piccolo (1990) – has been used to find pairs of similar regions. As it will
38 be explained later, this has been done to maintain the same methodology in those cases
39 where a variable employed in the model – i.e. the number of deaths – was missing.

40 **2. The proposed method**

41 In small data sets it is essential to save degrees of freedom (DOF) which are inevitably
42 lost in a amount correlated with the complexity of the statistical model entertained.
43 With this in mind, the proposed method is of the type semiparametric and consists of
44 two parts: a purely non-parametric and a parametric one. While the former does not
45 pose problems in terms of DOF, the latter clearly does. However, the sacrifice in terms
46 of DOF is very limited as an autoregressive model of order 1 (employed in a suitable
47 distance function, as below illustrated) has proved sufficient for the purpose. DOF–
48 saving strategy is also the driving force behind the choice to not consider exogenous
49 variables such as the regions geolocation or their population – e.g. in a regression-
50 like scheme – but to implicitly assume these (and other) variables embedded in the
51 dynamic of the time series considered.

52 **3. Data and contagion indicator**

53 The paper makes use of official data, published by the Italian Authorities, related to
54 the following two variables employed in the proposed method, i.e. the number of

55   1. deaths from CoViD–19 (denoted by the Latin letter M)

56   2. currently positive cases which have been recorded as a result of the administra-
57   tion of the test (denoted by the Latin letter C).

58   The data set includes 18 daily data points collected at regional level during the
59 period of February 24th to March 12th. The total number of Italian regions considered
60 is 20. However, one special administrative area (Trentino Alto Adige) is divided in two
61 subregions, i.e. Trento and Bolzano. Therefore, the set containing all the Italian regions
62 – called Ω – has cardinality $|\Omega| = 21$ (the cardinality function is denoted by the symbol
63 $|\cdot|$). Two different subsets are built from Ω i.e. $\Omega^{\bullet}$ – containing the regions for which
64 at least one death, out of the group of tested people, has been recorded and Ω° (no
65 recorded deaths). Those two sets are now specified:

66   1. $\Omega^{\bullet} \equiv Piemonte, Lombardia, Veneto, Friuli, Liguria, Emilia, Toscana, Marche,$
67       $Lazio, Abruzzo, ValleAosta, Bolzano, Campania, Puglia, Sicilia$

68   2. $\Omega^{\circ} \equiv Trento, Umbria, Molise, Basilicata, Calabria, Sardegna,$

69   being $\Omega \equiv \Omega^{\bullet} \cup \Omega^{\circ}$. In what follows, the two superscripts $\bullet$ and $\circ$ will be always
70 used respectively with reference to the regions $\{r_1, r_2, \ldots r_{15}\} \in \Omega^{\bullet}$ and $\{s_1, s_2, \ldots s_6\} \in$
71 Ω°. The time span is denoted as $\{1, 2, \ldots, T\}$.

72

73   In the case of the regions included in the set $\Omega^{\bullet}$, following Pueyo (2020), the
74 total number of positive is estimated as follows:

$$y_{j,T}^{\bullet} = w_T * 2^{\frac{\tau}{\delta}}, \tag{1}$$

$$w_T = \frac{C_T}{M_T}. \tag{2}$$

Here, w_T (Eqn. 2) is the ratio between the current positive cases (C) and the number of deaths (M) whereas, in Eqn. 1, τ is the average doubling time for the CoViD–19 (i.e. the average span of time needed for the virus to double the cases) and δ the average time needed for an infected person to die. These two constant terms have been kept fixed as estimated according to the data so far available and reported in Pueyo (2020). They are as follows: $\tau = 17.3$ and $\delta = 6.2$.

The case of the regions belonging to Ω° is more complicated. The related estimation procedure has been carried out as below detailed (the subscript t will be omitted for the sake of simplicity):

1. given the series $s_j \in \Omega^{\circ}$, a series $c^{\pi} \in \Omega^{\bullet}$ minimizer of a suitable distance function – denoted by the Greek letter $\pi(\cdot)$ – is found. In symbols:

$$c^{\pi} = \operatorname*{argmin}_{(c \in \Omega^{\bullet})} \pi(s,c); \tag{3}$$

2. the estimated number of positives at the population level – already found for c^{π}, say $I_{c^{\pi}}$ – becomes the weight for which the total cases recorded for s_j, are multiplied. Therefore, the estimate of the variable of interest for this case becomes

$$y_{j,T}^{\circ} = \frac{I_{c^{\pi}} * C_{s_j}}{C_{r_j}} \tag{4}$$

The distance function adopted $\pi(\cdot)$ (Eqn. 3), called AR-distance, has been introduced by Piccolo (2007)). Briefly, this metric can be applied if and only if the pair of series of interest are assumed to be realizations of two (possibly of different orders) ARMA (Autoregressive Moving Average) models (see, e.g. Makridakis and Hibon (1997)). Under this condition, each series can be expressed as an autoregressive model of infinite order, i.e. $AR(\infty)$, whose (infinite) sequence of AR parameters is denoted by $\{\alpha\}_j^{\infty} \equiv \alpha_1, \alpha_2, \dots$.

Without loss of generality, the distance between the series s and c, i.e. $\pi(s_t, c_t)$ (Eqn 4), under $(s_t, c_t) \sim \text{ARMA}(\alpha, \beta)$, being α and β respectively the autoregressive and moving average parameters, is expressed as

$$\pi(s,c) = \sqrt{(\sum_{j=1}^{\infty} \alpha_j(s) - \alpha_j(c))}. \tag{5}$$

For the asymptotic properties of Eqn. 5 the reader is referred to Corduas and Piccolo (2008) and Piccolo (2010).

4. The Resampling Method

The bootstrap scheme adopted proved to be adequate for the problem at hand. Given the pivotal role played in the proposed method, it will be briefly presented. In essence, the choice of the most appropriate resampling method is far from being an easy task, especially when the identical and independent distribution (*iid*) assumption (used in Efron's initial bootstrap method) is violated. Under dependence structures embedded

in the data, simple sampling with replacement has been proved – see, for example Carl-stein et al. (1986) – to yield suboptimal results. As a matter of fact, *iid*–based bootstrap schemes are not designed to capture, and therefore replicate, dependence structures. This is especially true under the actual conditions (small sample sizes) where the selection of the "right" resampling scheme becomes a particularly challenging task. Several *ad hoc* methods have been therefore proposed, many of which now freely and publicly available in the form of powerful routines working under software package such as Python® or R®. In more details, while in the classic bootstrap an ensemble Γ represents the population of reference the observed time series is drawn from, in *MEB* a large number of ensembles (subsets), say $\{\gamma_1, \ldots, \gamma_N\}$ becomes the elements belonging to Γ, each of them containing a large number of replicates $\{x_1, \ldots, x_J\}$. Perhaps, the most important characteristic of the *MEB* algorithm is that its design guarantees the inference process to satisfy the ergodic theorem. Formally, recalling the symbol $|\cdot|$ to denote the cardinality function (counting function) of a given ensemble of time series $\{x_t \in \gamma_i; \ i = 1, \ldots, N\}$, the *MEB* procedure generates a set of disjoint subsets $\Gamma_N \equiv \gamma_1 \cap \gamma_1 \cdots \cap \gamma_N$ s.t. $\mathbb{E}\Gamma_N \approx \mu(x_t)$, being $\mu(\cdot)$ the sample mean. Furthermore, basic shape and probabilistic structure (dependency) is guaranteed to be retained $\forall x_{t,j}^* \subset \gamma_i \subset \Gamma$.

MEB resampling scheme has not negligible advantages over many of the available bootstrap methods: it does not require complicated tune up procedures (unavoidable, for example, in the case of resampling methods of the type Block Bootstrap) and it is effective under non-stationarity. *MEB* method relies on the entropy theory and the related concept of (un)informativeness of a system. In particular, the Maximum Entropy of a given density $\rho(x)$, is chosen so that the expectation of the Shannon Information $\mathcal{H} = \mathbb{E}(-\log \rho(x))$, is maximized, i.e.

$$\max_{(\rho)} \mathcal{H} = \mathbb{E}(-\log \rho(x)).$$

Under mass and mean preserving constraints, this resampling scheme generates an ensemble of time series from a density function satisfying (4). Technically, *MEB* algorithm can be broken down, following Koutris et al. (2008), in 8 steps. They are:

1. a sorting matrix of dimension $T \times 2$, say $\mathcal{S}_1$, accommodates in its first column the time series of interest x_t and an Index Set – i.e. $I_{ind} = \{2, 3, \ldots, T\}$ – in the other one;

2. $\mathcal{S}_1$ is sorted according to the numbers placed in the first column. As a result, the order statistics $x_{(t)}$ and the vector I_{ord} of sorted I_{ind} are generated and respectively placed in the first and second column;

3. compute "intermediate points", averaging over successive order statistics, i.e. $c_t = \frac{x_{(t)} + x_{(t+1)}}{2}, \quad t = 1, \ldots T-1$ and define intervals I_t constructed on c_t and r_t, using *ad hoc* weights obtained by solving the following set of equations:

   i)
   $$g(x) = \frac{1}{r_1} \exp(\frac{[x - c_1]}{r_1}); \quad x \in I_1; r_1 = \frac{3x_{(1)}}{4} + \frac{x_{(2)}}{4}$$

   ii)
   $$g(x) = \frac{1}{c_k - c_{k-1}}; \quad x \in (c_k; c_{k+1})],$$
   $$r_k = \frac{x_{(k-1)}}{4} + \frac{x_{(k)}}{2} + \frac{x_{(k+1)}}{4}; \ k = 1, \ldots, T-1;$$

iii)

$$g(x) = \frac{1}{r_T} \exp \frac{\left([c_{T-1} - x]\right)}{r_T}; x \in I_T; \quad r_T = \frac{x_{T-1}}{4} + \frac{3x_T}{4};$$

4. from a uniform distribution in $[0,1]$, generate T pseudorandom numbers and define the interval $R_t = (t/T; t + 1/T]$ for $t = 0, 1, \ldots, T-1$, in which each p_j falls;

5. create a matching between R_t and I_t according to the following equations:

$$\begin{aligned} x_{j,t,me} &= c_{T-1} - |\theta| \ln(1 - p_j) & \text{if } p_j \in R_0, \\ x_{j,t,me} &= c_1 & - |\theta||ln(1 - p_j)| & \text{if } p_j \in R_{T-1}, \end{aligned}$$

so that a set of T values $\{x_{j,t}\}$, as the jth resample is obtained. Here θ is the mean of the standard exponential distribution;

6. a new $T \times 2$ sorting matrix $\mathcal{S}_2$ is defined and the T members of the set $\{x_{j,t}\}$ for the jth resample obtained in Step 5 is reordered in an increasing order of magnitude and placed in column 1. The sorted I_{ord} values (Step 2) are placed in column 2 of $\mathcal{S}_2$;

7. matrix S_2 is sorted according to the second column so that the order $\{1, 2, \ldots, T\}$ is there restored. The jointly sorted elements of column 1 is denoted by $\{x_{\mathcal{S},j,t}\}$, where $\mathcal{S}$ recalls the sorting step;

8. Repeat Steps 1 to 7 a large number of times.

5. The application of the maximum entropy bootstrap

In what follows, the proposed procedure is presented in a step-by-step fashion.

1. For each time series $y_t^\bullet$ and y_t° the bootstrap procedure is applied so that B= 100 "bona fide" replications are available as a result, i.e. $\tilde{y}_{t,b}^\bullet; b = 1, 2, \ldots B$ and $\tilde{y}_{t,b}^\circ; b = 1, 2, \ldots B$;

2. for both the series, the row vector related to the last observation T is extracted, i.e. $\{v^\circ = \tilde{y}_{T,1}^\circ, \tilde{y}_{T,2}^\circ \ldots \tilde{y}_{T,B}^\circ\}$ and $\{v^\bullet = \tilde{y}_{T,1}^\bullet, \tilde{y}_{T,2}^\bullet \ldots \tilde{y}_{T,B}^\bullet\}$;

3. the expected values, i.e. $\mathbb{E}(v^\bullet)$ and $\mathbb{E}(v^\circ)$, are then extracted along with the $\approx$ 95% confidence intervals ($CI^\bullet$ and CI°), which are computed according to the t–percentile method. In essence, through this method, suitable quantiles of an ordered bootstrap sample of t–statistics are selected and, as a result, the critical values for the construction of an appropriate confidence interval become available. A thorough explanation of the t–percentile method goes beyond the scope of this paper, therefore the interested reader is referred to the excellent paper by Berkowitz and Kilian (2000).

In particular, the lower (upper) CIs will be the lower (upper) bounds of our estimator while the quantities $\mathbb{E}(v^\bullet)$ $\mathbb{E}(v^\circ)$ are estimated through the mean operator, i.e.

$$\mu^\circ = \sum_{j=1}^{6} v_j^\circ \tag{6}$$

and

$$\mu^{\bullet} = \sum_{j=1}^{6} v_j^{\bullet} \qquad (7)$$

At this point, it is worth emphasizing that the procedure not only, as just seen, requires very little in terms of input data (only the time series of the positives and the deaths are required) but also can be performed in an automatic fashion. In fact, once the data become available, one has just to properly assign the time series to the subsets Ω° and $\Omega^{\bullet}$ and the code will process the new data in an automatic way. The procedure is also very fast, as the computing time needed for the generation of the bootstrap samples requires – for the sample size in question – less than two minutes. Both code and data-set employed in this paper are freely available upon request. However, the data can also be downloaded free of charge at the following web address: https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni (the file name is dpc-covid19-ita-regioni-20200323.csv).

6. Empirical evidences

In order to give the reader the opportunity to gain a better insight on the different epidemic dynamical behaviors, in Figure 2 – 5 the time series of the variable C (as defined in Eqn. 2) is reported for each region. Note that the sudden variations noticeable in Figure 5 (Bolzano), Figure 4 (Valle D'Aosta) and Figure 3 (Molise and Campania) are due to the little number of tests administrated (i.e. the denominator of the variable C_T (2)) for these cases.

That said, the main result of the paper is summarized in Table 2, where three estimates of the number of positives are reported by region. The regions belonging to the set Ω° (no deaths) are in Italics whereas all the others, belonging to the set $\Omega^{\bullet}$, are in a standard format. In the columns "Mean" and "Lower (Upper) Bounds", the bootstrap estimates computed according to Eqn 6 and 7 and the Lower (Upper) Bounds the lower (upper) bootstrap CIs are respectively reported. The column denominated "Official Cases" accounts for the number of positives cases released by the Italian Authorities, whereas the column "Morbidity" expresses the percentage ratio between $\mu^{\bullet}$ (6) or μ° (7) and the actual population of each region, as recorded by the Italian National Institute of Statistics. The latter source of data can be freely accessed at the web address $http://dati.istat.it/Index.aspx?DataSetCode=DCIS\_POPRES1$.

By examining the data for the whole Country, it is clear how the data collected by the Italian Authorities on the positive cases severely underestimate the current situation by a factor of about 8. As expected, the top three regions in terms of number of infected persons are Lombardia, Emilia Romagna and Veneto, where the estimated infected population is respectively (bootstrap mean) around 45,020, 12,299 and 9,343.

On the other hand, the risk of contagion is relatively low in some regions – mostly located in the Southern part of Italy – and in the island of Sardinia.

Regarding the regions included in the subset Ω°, the application of the Piccolo distance (π) has generated the associations reported in Table 1.

7. Conclusions

It is widespread opinion in the scientific community that current official data on the diffusion of SARS-CoV-2, responsible of the correlated disease, COIVD-19,among population, are likely to suffer from a strong downward bias.

Table 1. Association found between the regions belonging to Ω° and those in $\Omega^{\bullet}$ according to the minimum distance π

Ω°	$\Omega^{\bullet}$	π
Basilicata	Veneto	0.0389
Calabria	Campania	0.6211
Molise	Lazio	0.4212
Sardegna	Abruzzo	0.0157
Trento	Abruzzo	0.00186
Umbria	Sicilia	0.01398

227   In this scenario, the aim of this paper is twofold: on one hand, it generates
228 realistic figures on the effective number of people infected with SARS-CoV-2 at a na-
229 tional and regional level; on the other hand, it provides a methodology representing a
230 viable alternative to those interested to apply inference procedures on the diffusion of
231 epidemics.

232

233   Following Pueyo (2020), this paper proposes a methodology which uses simple
234 counts, i.e. the number of deaths and the number of people tested positive to the virus
235 for Italy, to

236   1. provide an estimation at the national and regional level of the number of infected
237   people and the related confidence intervals;

238   2. extend Puejo's methodology to those regions exhibiting no deaths as a conse-
239   quence of the contraction of the CoViD-19.

240   The entire procedure has been written in the programming language R® and
241 uses official data as published by the Italian National Institute of Health. The whole
242 code is available upon request.

243

244   The results obtained show that, while official data at March 12th report, for
245 Italy, a total of 12.839 cases, the people infected with the SARS-CoV-2 could be as
246 high as 105.789. This result, along with the estimated average doubling time for the
247 CoViD-19 (≈ 6.2 days), confirms that this pandemic is to be regarded as much more
248 dangerous than currently foreseen.

249

Table 2. Estimation of the number of people infected from CoViD–19 by Italian regions. Lower and Upper Bounds are computed through the Bootstrap t–percentile method whereas the mean values is computed as in (6) and (7). The regions belonging to the set $\Omega°$ are in Italics

	Lower Bound	Mean	Upper Bound	Official Cases	Population	morbidity
Abruzzo	526	600	807	78	1.311.580	0,06
Basilicata	48	54	70	8	562.869	0,01
Bolzano	697	730	795	103	531.178	0,15
Calabria	182	238	493	32	1.947.131	0,03
Campania	988	1292	2676	174	5.801.692	0,05
Emilia Romagna	10980	12299	14897	1758	4.459.477	0,33
Friuli Venezia Giulia	983	1201	2514	148	1.215.220	0,21
Lazio	1485	1680	2089	172	5.879.082	0,04
Liguria	1346	1608	1995	243	1.550.640	0,13
Lombardia	37744	45020	49723	6896	10.060.574	0,49
Marche	3151	3891	4593	570	1.525.271	0,30
Molise	119	134	167	16	305.617	0,05
Piemonte	3216	3703	4217	554	4.356.406	0,10
Puglia	490	670	1292	98	4.029.053	0,03
Sardegna	244	278	375	39	1.639.591	0,02
Sicilia	776	865	1098	111	4.999.891	0,02
Toscana	2352	2755	3965	352	3.729.641	0,11
Trento	670	764	1028	102	541.098	0,19
Umbria	432	481	611	62	882.015	0,07
Valle Aosta	139	183	356	26	125.666	0,28
Veneto	8382	9343	12028	1297	4.905.854	0,25
Totale Italia	74.950	87.789	105.789	12.839	60359546	0,18
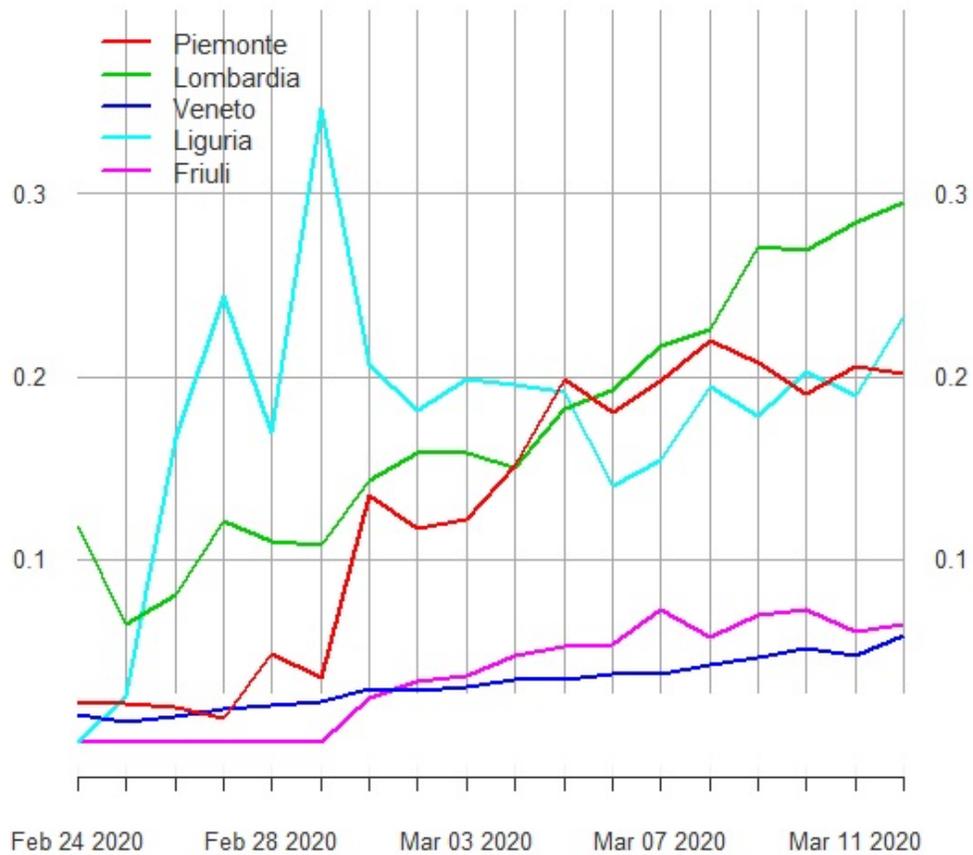
Fig. 1. Percentage ratio deaths / new cases for the following Italian regions: Piemonte, Lombardia, Veneto, Liguria and Friuli-Venezia-Giulia
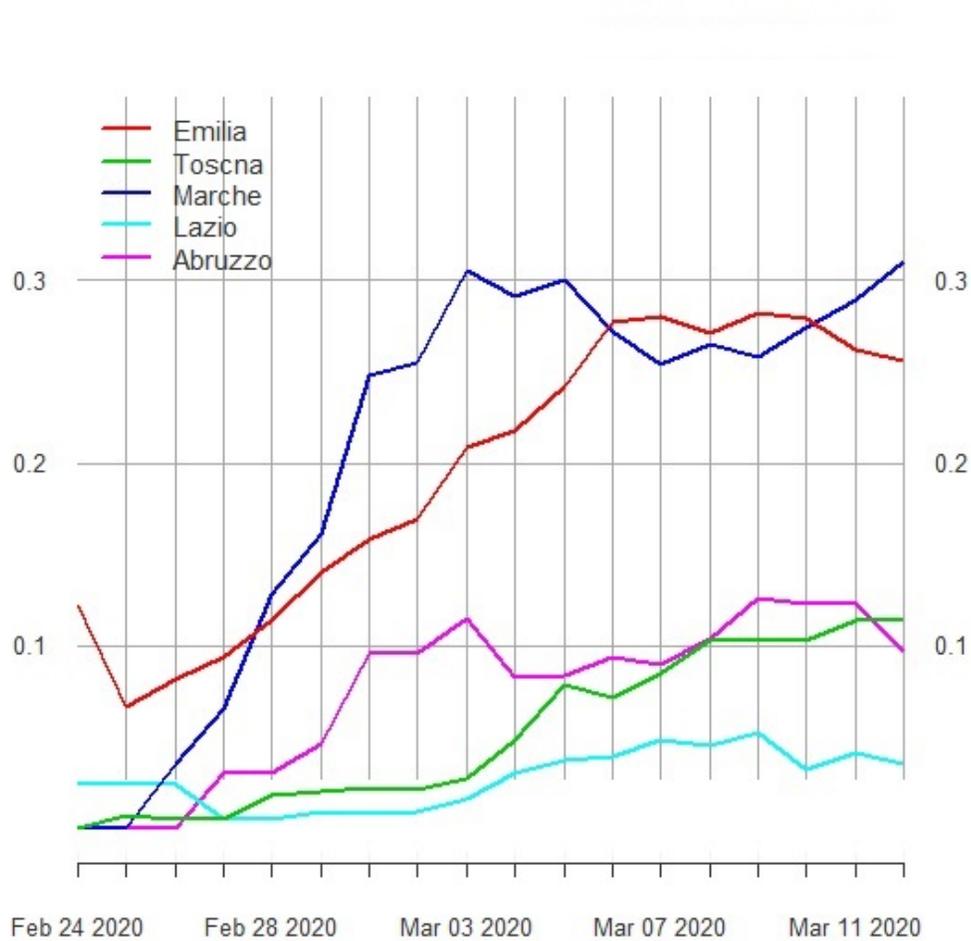
Fig. 2. Percentage ratio deaths / new cases for the following Italian regions Emilia, Toscana, Marche, Lazio and Abruzzo
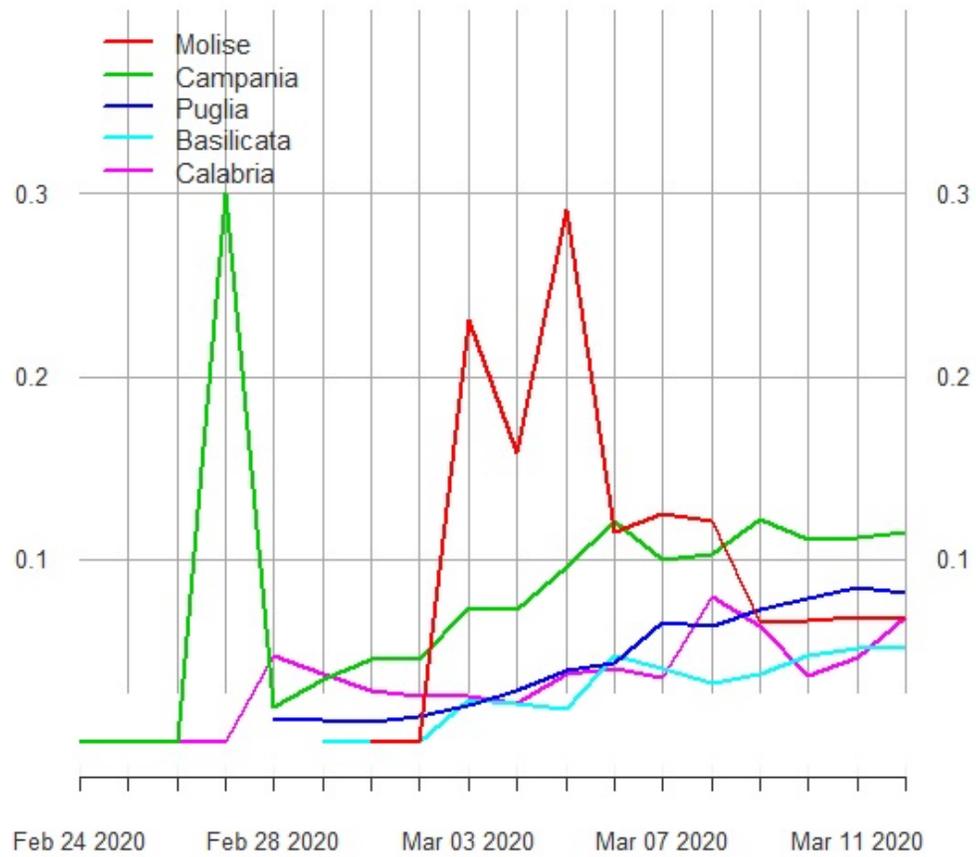
Fig. 3. Percentage ratio deaths / new cases for the following Italian regions: Molise, Campania, Puglia, Basilicata and Calabria
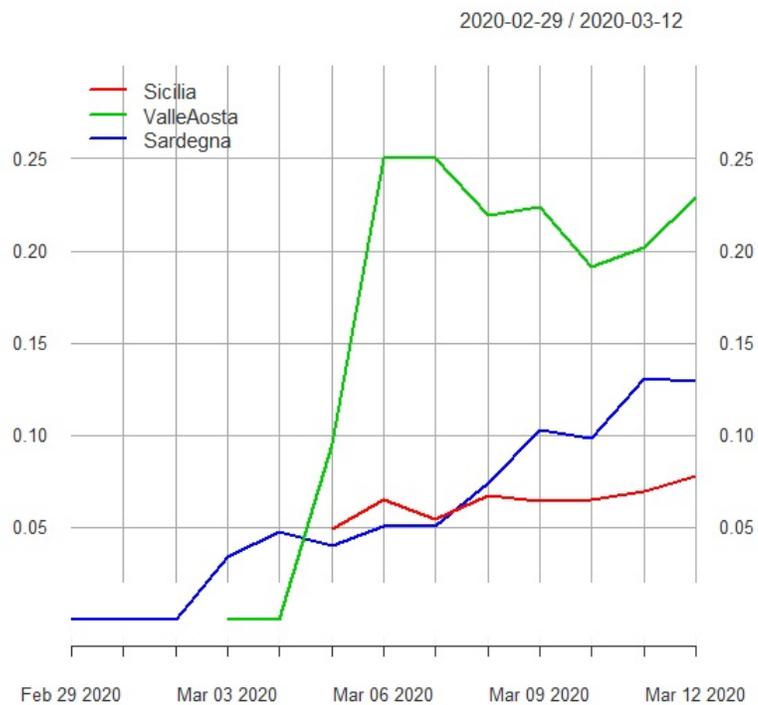
Fig. 4. Percentage ratio deaths / new cases for the following Italian regions: Sicilia, Valle d'Aosta, Sardegna)
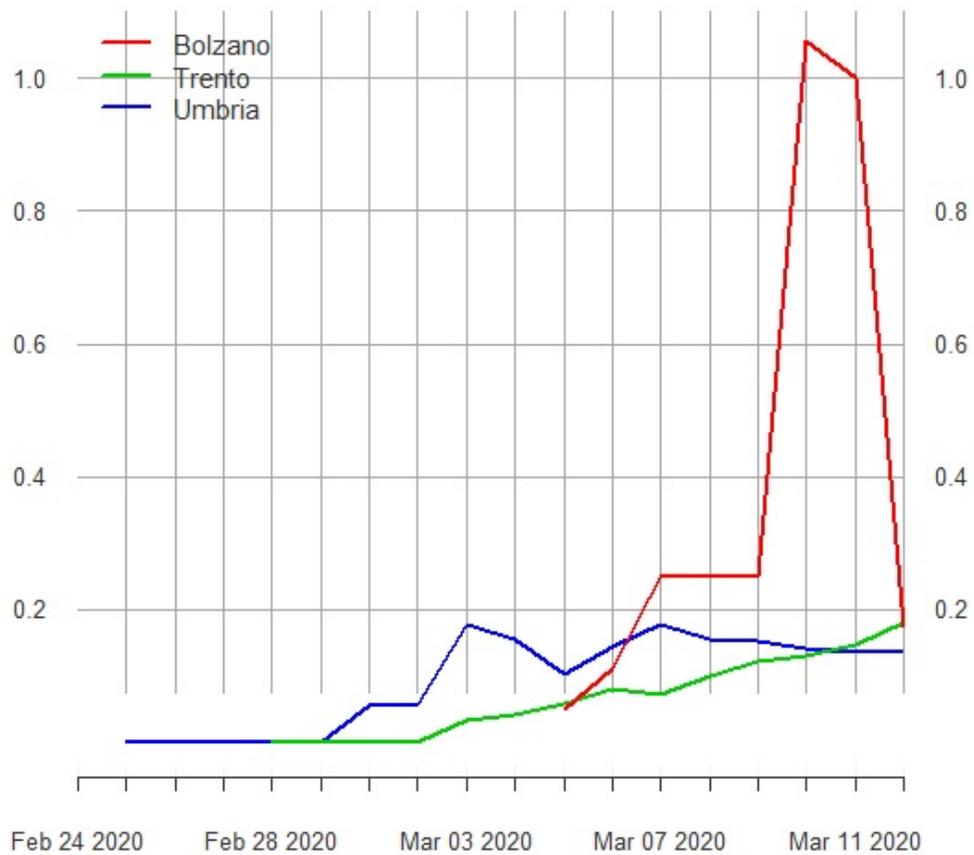
Fig. 5. Percentage ratio deaths / new cases for the following Italian regions: Bolzano, Trento, Umbria)

250 **8. Acknowledgments**

251 The author is deeply grateful to Dr. Luigi Di Landro for the generous help in the proof-
252 reading process.

253 **References and links**

254 Berkowitz, J., and Kilian, L. (2000), "Recent developments in bootstrapping time series," *Econometric*
255 *Reviews*, 19(1), 1–48.
256 Carlstein, E. et al. (1986), "The use of subseries values for estimating the variance of a general statistic
257 from a stationary sequence," *The annals of statistics*, 14(3), 1171–1179.
258 Corduas, M., and Piccolo, D. (2008), "Time series clustering and classification by the autoregressive
259 metric," *Computational statistics & data analysis*, 52(4), 1860–1872.
260 Koutris, A., Heracleous, M. S., and Spanos, A. (2008), "Testing for nonstationarity using maximum
261 entropy resampling: A misspecification testing perspective," *Econometric Reviews*, 27(4-6), 363–384.
262 Makridakis, S., and Hibon, M. (1997), "ARMA models and the Box–Jenkins methodology," *Journal of*
263 *Forecasting*, 16(3), 147–163.
264 Piccolo, D. (1990), "A distance measure for classifying ARIMA models," *Journal of Time Series Analysis*,
265 11(2), 153–164.
266 Piccolo, D. (2007), Statistical issues on the AR metric in time series analysis,, in *Proceedings of the SIS*
267 *2007 intermediate conference" Risk and Prediction*, pp. 221–232.
268 Piccolo, D. (2010), "The autoregressive metric for comparing time series models," *Statistica*, 70(4), 459–
269 480.
270 Pueyo, T. (2020), Coronavirus: Why You Must Act Now,, in *https://medium.com/@tomaspueyo/coronavirus-*
271 *act-today-or-people-will-die-f4d3d9cd99ca*.
272 Vinod, H. D., López-de Lacalle, J. et al. (2009), "Maximum entropy bootstrap for time series: the
273 meboot R package," *Journal of Statistical Software*, 29(5), 1–19.