

Research Article

Forecasting the COVID-19 Diffusion in Italy and the Related Occupancy of Intensive Care Units

Livio Fenga 

Italian National Institute of Statistics, ISTAT, Rome 00184, Italy

Correspondence should be addressed to Livio Fenga; lfenga@ucsd.edu

Received 13 April 2020; Accepted 16 September 2020; Published 16 January 2021

Academic Editor: Alessandro De Gregorio

Copyright © 2021 Livio Fenga. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper provides a model-based method for the forecast of the total number of currently COVID-19 positive individuals and of the occupancy of the available intensive care units in Italy. The predictions obtained—for a time horizon of 10 days starting from March 29th—will be provided at a national as well as at a more disaggregated level, following a criterion based on the magnitude of the phenomenon. While those regions hit the most by the pandemic have been kept separated, those less affected regions have been aggregated into homogeneous macroareas. Results show that—within the forecast period considered (March 29th–April 7th)—all of the Italian regions will show a decreasing number of COVID-19 positive people. The same will be observed for the number of people who will need to be hospitalized in an intensive care unit. These estimates are valid under constancy of the government's current containment policies. In this scenario, northern regions will remain the most affected ones, whereas no significant outbreaks are foreseen in the southern regions.

1. Introduction

On March 19th, the death toll paid by Italy for the spread of the virus COVID-19 amounted to 3405 deaths, the highest paid by a single country in the world. Despite hard and relatively timely lockdown policies implemented by the government, on March 26th this figure has risen to 8165 deaths.

In such an emergency situation, a reliable forecast method for the infection development is essential for policy and decision makers to design evidence-based policies and to implement fast actions to curb the spread of the infection. In particular, predicting the number of people currently tested positive for COVID-19 (thereafter “positive cases”) could be useful to draw the epidemiological curve of the infection and therefore to predict its peak. Other than this variable, the forecasting procedure presented in this paper is used to predict the future values of another crucial variable, i.e., the number of people needing hospitalization in an intensive care unit (ICU). The Italian ICU system is at the moment severely stressed due to the spread of the disease; therefore, predictions of future ICU demand could be fruitfully considered in the design and the implementation

of operational schemes. The forecast horizon for both the variables is of 10 days starting from March 29th.

Since the Italian regions are affected in different extents by the COVID-19, it has been decided to perform the forecasting exercise for the following geographical areas: Lombardia, Piedmont, Valle d'Aosta, Veneto, Friuli Venezia Giulia, Trentino Alto Adige, Lazio, and Campania. The remaining regions have been grouped in the following macroareas: “Center” (Marche, Umbria, and Toscana) and “South” (Abruzzo, Molise, Puglia Basilicata, Calabria, Sicilia, and Sardegna). At least two other reasons justify such a break down:

- (1) The different starting times recorded for the lockdowns
- (2) The southern regions have been hit less severely and therefore, especially at the beginning of the observation period, show several zeroes or low numbers across the considered time span

In essence, in this study, the available official data, detailed in Section 2, have been employed in a three-step procedure, i.e.:

- (1) Data preprocessing, in which data anomalies are identified and corrected according to an approach of the type a Kalman filter
- (2) Univariate forecasting, based on an autoregressive moving average (ARMA) model for number of positive cases and ICU
- (3) Bootstrap-based generation of predicted values and confidence intervals

2. The Data

This paper employs the data related to COVID-19, collected and regularly updated by the Italian National Institute of Health (an agency of the Italian Ministry of Health) and by the Italian Civil Protection Department. The whole data set is freely and publicly available in a comprehensive database, accessible on the Internet at the web address <https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni> (the file name is `dpc-covid19-ita-regioni-20200323.csv`). It collects crucial data related to all the persons tested for COVID-19—from the outbreak of the pandemic (February 24th)—and, in particular,

- (1) is a collection of 21 data points—representing 19 Italian regions plus the two autonomous provinces of Trento and Bolzano—one for each day starting from the disease’s outbreak,
- (2) considers crucial variables, such as positive cases, recovered cases, deaths, number of people hospitalized, and number of people admitted to intensive care units (ICUs).

As already pointed out, in the present study, the variables of interest are the number of people who have been

- (1) tested positive for COVID-19 (in what follows denoted by the bold Latin letter \mathbf{V}),
- (2) hospitalized in an ICU (which will be denoted by the bold Latin letter \mathbf{U}).

It is worth outlining how, according to the regulations issued by the Italian government, only the people showing moderate to severe symptoms, generally associated with the infection, or who have been in close proximity with at least one positive person, are tested. Therefore, the predictions obtained are to be referred to the sample, as no attempt have been made to carry out inference procedures for the estimation of the variables at the population level.

In order to correctly process the data, all the regions showing no positive cases at the beginning of the recording period and/or low values along the whole time span have been aggregated into macroareas. This has been done to (i) give more meaningful results and (ii) save degrees of freedom (which are always precious in short time series).

In details, the prediction exercise will be performed on the following regions/macroareas:

- (A) Nothtern regions
 - (1) Lombardia
 - (2) Piedmont

- (3) Valle d’ Aosta
- (4) Veneto
- (5) Friuli Venezia Giulia
- (6) Emilia-Romagna
- (7) Liguria
- (8) Macroarea “Trentino Alto Adige” (Trento and Bolzano)

- (B) Center regions

- (1) Lazio
- (2) Macroarea “Center”: Marche, Umbria, and Toscana

- (C) Southern regions

- (1) Campania
- (2) Macroarea “South”: Abruzzo, Molise, Puglia, Basilicata, Calabria, Sicilia, and Sardegna

The north Italy regions—at the moment the more severely affected by the pandemic—have been treated separately along with two other regions, i.e., Lazio and Campania, since their major cities—Rome and Naples—deserve special attention for the institutional role played and their population density. On the other hand, the regions showing less worrying figures have been aggregated into macroareas according to their geolocation. The only exception is Valle d’Aosta, which has been left separated as no aggregation options could be found.

To simplify notation, for both the variables of interest \mathbf{V} and \mathbf{U} , the following convention is introduced:

$${}^K\mathbf{V}_j. \quad (1)$$

Here, the upper left superscript (denoted by the upper case Latin letter K) refers to the geographical areas (i.e., North, Center, and South), whereas the subscript j is associated with the number the different regions or macroareas are codified with, as above detailed. For example, by the symbols ${}^A\mathbf{V}_6$ and ${}^B\mathbf{V}_2$ the number of positive cases for the Emilia-Romagna region and the Center macroarea “Marche-Umbria-Toscana” are respectively identified.

3. Data Preprocessing

Missing data and other anomalies become the first challenge when designing predictive models, as statistical methods, in general, are designed and tested under the assumption of no missing observations [1]. Before delving into the details of the proposed procedure, a word of caution is needed since, unfortunately, a visual inspection of the data suggests the presence of a number of anomalous data, both at a regional and a country level. The detected anomalies might be associated with the biological sample collecting process and the related testing procedures. In fact, the typical lab workflow is governed by a set of rigid protocols which might be critically affected by factors such as the availability of manpower, swabs, reagents, and other laboratory materials. In emergency situations, such a workflow can be disrupted, and temporal inconsistency might appear as a result. For

example, a set of samples might be delivered to a laboratory with longer than usual delays with respect to the time of collection, or a given lab can only complete the screening process for a certain number of samples. In both the cases, a shift of one day (or more) in the release of the lab results can be reasonably expected. A further source of anomalies is represented by the data entry and data editing processes, carried out in working environment likely affected by the risk of contagious and under rigid deadlines.

An example of such anomalous data is given in Figure 1, where the series ${}^1\mathbf{V}_{t,1}$ (Lombardia) is depicted. Here, some data points showing values inconsistent with the overall pattern are clearly noticeable. Given the (very small) available sample size, the relative weight of such data is almost surely not negligible and can introduce severe distortions in the model parameter inference procedures and thus in the predicted values.

In order to correct those data, a Kalman smoother state-space model [2] has been applied. In particular, the Kalman smoother adopted is of the type fixed-point smoothing. This algorithm is designed to obtain the estimate of a realization \hat{w}_N (the time t_N is fixed $N < K$) of a given random variable W_t , given a set of observations $Z_k = \{z_k \mid 0 \leq N \leq k\}$. A thorough explanation of this method goes beyond the scope of the paper; therefore, the interested reader is referred to the excellent paper by Sage and Melsa [3].

In Figure 2, the corrected version of the series ${}^1\mathbf{V}_{t,1}$ —resulting by applying the Kalman smoother—is depicted. Not only this series lends itself to a better visual inspection but, more importantly, is more suitable to be processed by the adopted prediction model.

4. Theoretical Framework

The approach used in this paper relies on (i) the theory of stochastic process and (ii) a resampling method. While the former is necessary to generate the input (predicted values) of the bootstrap algorithm, as well as to justify the employment of the outlier correction method, the latter serves the purpose of

- (1) generating the final predictions, which are affected by a reduced amount of uncertainty (with respect to those generated by the stochastic model),
- (2) yielding the related confidence intervals.

4.1. The Stochastic Processes Paradigm. The approach proposed in the present paper relies on the assumption that the (transformed) time series ${}^K\mathbf{V}_{j,t}$ and ${}^K\mathbf{V}_{j,t}$ are approximately a realization of a process of the type ARMA (autoregressive moving average) [4].

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a real 2^{nd} order stationary process, and it is said to admit a ARMA(p, q) representation ($p, q \in \mathbb{Z}$) if, for some constant $a_1, \dots, a_p, b_1, \dots, b_q$, will be

$$\sum_{j=0}^p (a_j X(t-j)) = \sum_{j=0}^q b_j \varepsilon(t-j), \quad (t \in \mathbb{Z}), a_0 = b_0 = 1, \quad (2)$$

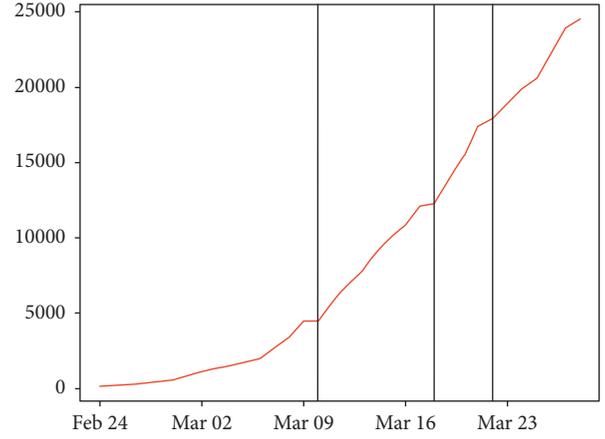


FIGURE 1: Number of people tested positive (Lombardia) (original data).

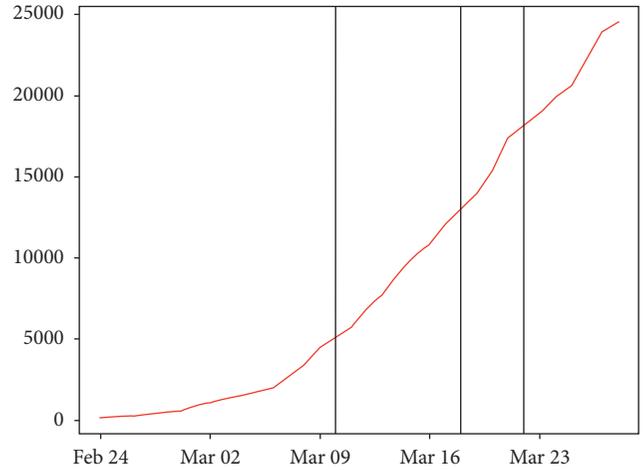


FIGURE 2: Number of people tested positive (Lombardia) (data adjusted via Kalman filter).

under the following conditions:

$$\begin{aligned} E\{\varepsilon(t) \mid F_{t-1}\} &= 0, \\ E\{\varepsilon^2(t) \mid F_{t-1}\} &= \sigma^2, \\ E\varepsilon^4(t) &< \infty, \\ \sum_{j=0}^p a_j Z^j &\neq 0, \\ \sum_{j=0}^q b_j Z^j &\neq 0, \quad |Z| \leq 1. \end{aligned} \quad (3)$$

Here, F_t denotes the sigma algebra induced by $\varepsilon(j), j \leq t$, and $\sum_{j=0}^p a_j Z^j$ and $\sum_{j=0}^q b_j Z^j$ are assumed not to have common zero.

The above conditions assure that X_t can be represented as

$$\begin{aligned} X(t) &= \sum_{j=0}^{\infty} (\beta_j \varepsilon(t-j)), \\ \sum_{j=0}^{\infty} (\beta_j Z^j) &= \sum_{j=0}^{\infty} (a_j Z^j)^{-1} \sum_{j=0}^q b_j Z^j, \end{aligned} \quad (4)$$

with β_j decreasing to 0 at geometric rate.

The dynamics of the series under investigations are not suitable for this theoretical framework as it requires 2nd-order stationarity and homoscedasticity; those conditions are simultaneously achieved by preprocessing the series according to the following filter: $\log(\nabla^d)$, being the symbol ∇ the difference operator and the exponent d indicating the order of the difference. To fully understand the role played by ∇ , the backward operator B is now introduced. In essence, B moves the time index of an observation back by p time intervals, i.e., $B^p x_t = X_{t-p}$, and thus we have

$$\nabla^d \log(X_t) = (1 - B)^d \log(X_t). \quad (5)$$

4.2. The Resampling Method. In order to extract valuable information from our data and, at the same time, decrease the total amount of uncertainty associated with the outcomes of the ARMA model, a resampling procedure has been employed. Among the several resampling methods for dependent data available—many of which are freely and publicly available in the form of powerful routines working under software packages such as Python® or R®—the adopted resampling method is of the type maximum entropy bootstrap (MEB). Proposed by Vinod [5] and subsequently improved (see, e.g., Vinod [6]), it is based on basic assumptions which are different from those usually followed by standard schemes. In more details, while in the classic bootstrap an ensemble Ω represents the population of reference the observed time series is drawn from, in MEB a large number of ensembles (subsets), say $\{\omega_1, \dots, \omega_N\}$ becomes the elements belonging to Ω , each of them containing a large number of replicates $\{x_1, \dots, x_j\}$.

Unlike standard bootstrap schemes, in the MEB case the resample set Ω mimics the observed realization of the underlying stochastic process, in MEB a large number of subsets, say $\{\omega_1, \dots, \omega_N\}$ becomes the elements belonging to Ω , each of them containing a large number of replicates $\{x_1, \dots, x_j\}$. Among the important features of the MEB scheme, it is worth mentioning the consistency of its bootstrap samples with the ergodic theorem (see, e.g., Birkhoff [7]) and with the probabilistic structure of the observed time series. In Figure 3, an example of the application of MEB for the variable ${}^1V_{t,1}$ is given.

5. The Forecasting Method

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

- (1) Equation (2) is estimated for both \mathbf{V}_t and \mathbf{U}_t so that the model orders (equation (2)) M_1 and M_2 become available.
- (2) For each time series \mathbf{V}_t and \mathbf{U}_t , the MEB procedure is applied so that the sets V and U —each containing $B = 500$ “bonafide” replications—are available, i.e., $V \equiv [V_1^*, V_2^*, \dots, V_B^*]$ and $U \equiv [U_1^*, U_2^*, \dots, U_B^*]$ (in Figure 4, the set V for the variable ${}^1V_{t,1}$ is given).

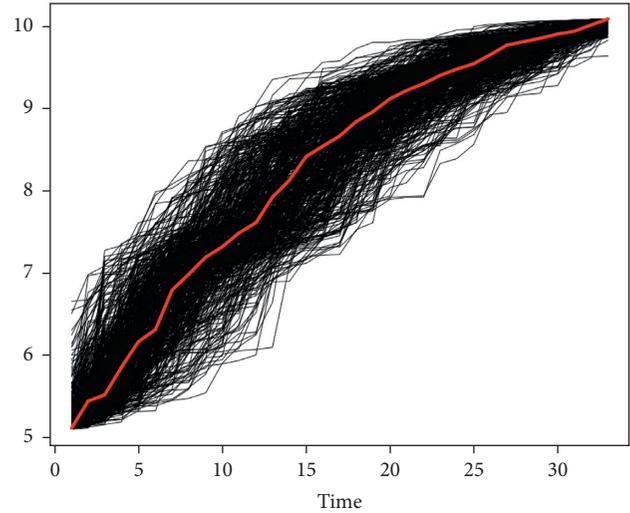


FIGURE 3: Lombardia: $B = 500$ bootstrap replications performed via the MEB algorithm on the adjusted, log-transformed, data (in red, the original time series is depicted).

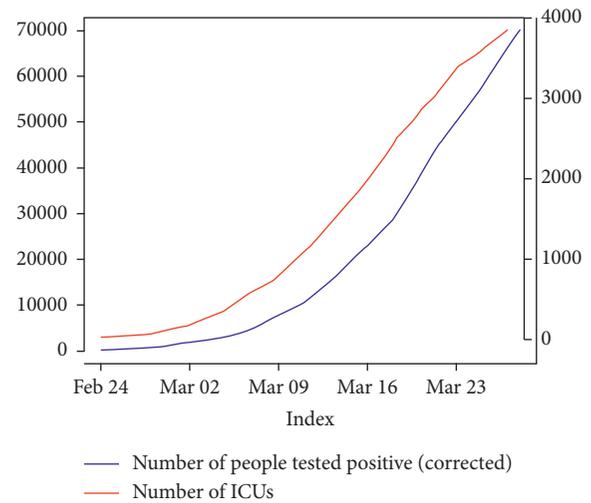


FIGURE 4: Italy, time series data of positive (data corrected via Kalman filter, left side axis) and of the number of people hospitalized in ICUs (right side axis).

- (3) For each of the replications stored in V , equation (2) is estimated according to the model order selected, i.e., M_1 , and the 1- to 10-step-ahead predictions—as well as the 5% and 95% bootstrap confidence interval—are generated.
- (4) The B predictions and the confidence intervals obtained in the previous step are stored in the $B \times 3$ matrix $[\mathbf{F}_V(h), h = 1, 2, \dots, 10]$, whose columns are lower bootstrap confidence interval, bootstrap prediction, and upper bootstrap confidence interval, respectively, denoted by the symbols $CI_{L,b}^*(h)$, $\mathbf{V}_{t,b}^*$, and $CI_{U,b}^*(h)$ $b = 1, \dots, B$.
- (5) The median value $\hat{V}^* = X(\mathbf{V}_{t,b}^*)$ is then extracted along with the $\approx 95\%$ confidence intervals $CI_{L,b}^*(h = 1)$ and $CI_{U,b}^*(h = 1)$, computed according to the

t-percentile method. The explanation of this procedure goes beyond the scope of this paper; therefore, the interested reader is referred to the excellent paper by Berkowitz and Kilian [8].

- (6) $CI_L^*(h = 2, \dots, 10)$ and $CI_U^*(h = 2, \dots, 10)$ (the subscript b is omitted for brevity) are computed conditional to a subset of V , say \tilde{V} , made up of the bootstrap replications whose range falls between the minimum and maximum values of the values of the confidence intervals computed for $h = 1$. In symbols,

$$\min(CI_U^*(h = 1)) \leq [\tilde{V} \subset tV] \leq \max(CI_U^*(h = 1)). \quad (6)$$

- (7) Steps 1–6 are repeated for U_t , so that a new matrix of prediction of dimension $B \times 3$ is built, i.e., $[F_V(h), h = 1, 2, \dots, 10]$, whose columns are as in $F_U(h)$ and denoted by the symbols $CI_U^*(h)$, \hat{U}_t^* , and $CI_U^*(h)$.

Unfortunately, the whole procedure cannot be considered fully automatic since the estimation of equation (2) (step 1) is required.

5.1. The Adopted Models. The stochastic model structures identified for both V_t and U_t are almost always of the type ARMA(1, 0), with the exception of Campania (ARMA(0, 1), for both the variables V_t and U_t) and Emilia-Romagna, for which the best model for the variable U_t is of the type ARMA(1, 1). The most suitable prefilter (equation (5)) has been always of the type $d = 3$ difference of the natural log of the variables of interest.

6. Empirical Evidences

At the national level (data have been plotted in Figure 4), the peak in the number of COVID-19 positives will be reached on April 2nd, with a number of predicted positive close to 77,000. The maximum forecasted value for the occupied ICU—expected for April 4th—will be 4280. These values have been calculated using an indirect methodology, i.e., by summing up the estimates obtained at a disaggregated level. The results related to COVID-19 positives and the ICU occupancy are reported, respectively, in Tables 1 and 2, where the bootstrap standard deviations of the quantities \hat{V}_t^* and \hat{U}_t^* —respectively denoted with the symbols $[\hat{\sigma}(V_t)]^*$ and $[\hat{\sigma}(U_t)]^*$ —are reported along with their confidence intervals, i.e., $CI_L^*(h)$ (lower) and $CI_U^*(h)$ (upper). In what follows, the main results reported in these tables are commented:

- (i) Lombardia—the most affected region—will reach the peak of positive cases (25963) and of the demand of ICUs (1425), respectively, on April 2nd and 4th.
- (ii) Emilia-Romagna is the second most affected region by COVID-19 but still shows a very high number of victims. The trend of infected people will reach its peak on April 5th, whereas the

TABLE 1: 10-step-ahead predictions for the variable V_t (number of persons tested positive).

		Italy			
		$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
<hr/>					
29 March		66504	71586	76647	6911
30 March		69753	73409	79373	12785.3
31 March		70760	75696	82773	16989.7
1 April		70817	76723	86430	20353.4
2 April		69925	77811	90800	28812.7
3 April		69309	76952	96806	35454.1
4 April		68480	75927	102946	42234.7
5 April		66531	74860	109265	58238.4
6 April		64940	75006	117550	72145.3
7 April		60197	74875	126205	91586.4
<hr/>					
		Lombardia			
<hr/>					
29 March		22538	25214	26384	5001.4
30 March		23296	25456	26790	5010.8
31 March		23969	25792	27565	5167.0
1 April		23864	25842	27634	5301.3
2 April		23662	25963	28230	6248.7
3 April		23209	25675	28342	7123.5
4 April		22709	25717	28781	8307.2
5 April		21964	25067	29524	9972.6
6 April		20802	24431	30131	12368.5
7 April		19619	23752	30628	14375.8
<hr/>					
		Piedmont			
<hr/>					
29 March		6115	6635	7017	1136.4
30 March		6264	6610	7193	1253.0
31 March		6209	6568	7481	1759.2
1 April		6079	6401	7942	2531.6
2 April		5837	6152	8271	3357.0
3 April		5502	5898	8590	4025.1
4 April		5134	5566	8753	4985.6
5 April		4666	5135	8995	5893.4
6 April		4178	4732	9279	6809.7
7 April		3707	4289	9417	7986.3
<hr/>					
		Liguria			
<hr/>					
29 March		2026	2062	2217	220.9
30 March		2007	2048	2242	299.7
31 March		1955	2019	2361	476.0
1 April		1891	1969	2416	634.1
2 April		1804	1900	2515	803.5
3 April		1701	1805	2580	988.9
4 April		1589	1709	2645	1178.1
5 April		1470	1590	2748	1445.7
6 April		1344	1440	2917	1878.3
7 April		1214	1311	3090	2158.2
<hr/>					
		Valle d'Aosta			
<hr/>					
29 March		404	480	512	111.7
30 March		410	491	524	121.6
31 March		417	501	546	142.2
1 April		420	500	551	146.0
2 April		411	494	551	152.3
3 April		382	483	542	179.3
4 April		359	462	526	188.9
5 April		339	443	541	222.6
6 April		301	412	530	254.5
7 April		267	383	516	280.3
<hr/>					
		Veneto			
<hr/>					
29 March		6346	7145	7559	1730.7
30 March		6569	7213	7749	1745.4

TABLE 1: Continued.

Italy				
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
31 March	6706	7431	8067	1888.1
1 April	6721	7463	8144	2031.5
2 April	6535	7539	8415	2443.8
3 April	6414	7469	8758	3121.9
4 April	6073	7374	9168	4423.1
5 April	5703	7182	9214	4957.7
6 April	5303	6842	9593	6023.7
7 April	4798	6537	10203	7743.4
Friuli Venezia Giulia				
29 March	1029	1124	1167	180.2
30 March	1083	1138	1195	141.6
31 March	1100	1169	1226	153.6
1 April	1104	1189	1269	211.4
2 April	1109	1212	1267	201.2
3 April	1097	1209	1272	207.6
4 April	1060	1230	1287	287.8
5 April	1035	1203	1305	321.4
6 April	1005	1202	1333	413.2
7 April	976	1208	1343	445.7
Emilia-Romagna				
29 March	9163	10500	11265	2912.4
30 March	9464	10975	11717	3323.8
31 March	9535	11360	12176	3591.6
1 April	9587	11613	12562	4123.3
2 April	9476	11815	12954	4783.0
3 April	9276	11943	13291	5547.1
4 April	9002	12132	13539	6471.6
5 April	8528	12244	13720	7240.6
6 April	8034	12125	13995	8173.0
7 April	7311	11966	14189	9017.2
Trentino Alto Adige				
29 March	2021	2147	2324	325.1
30 March	2076	2158	2488	426.1
31 March	2114	2124	2617	581.6
1 April	2110	2083	2707	685.9
2 April	2097	1988	2787	761.4
3 April	2064	1868	2764	817.6
4 April	2023	1725	2742	827.9
5 April	1944	1548	2751	902.1
6 April	1851	1373	2718	991.0
7 April	1707	1200	2641	1041.3
Lazio				
29 March	2013	2208	2355	429.0
30 March	2160	2244	2436	348.2
31 March	2182	2298	2553	413.7
1 April	2153	2279	2666	612.4
2 April	2161	2256	2785	723.5
3 April	2140	2222	2991	1016.0
4 April	2100	2154	3132	1278.3
5 April	2041	2046	3233	1480.3
6 April	1908	1956	3485	1817.4
7 April	1768	1821	3782	2128.3
Macroarea Center				
29 March	6822	7499	7911	1459.1
30 March	7104	7671	8087	1403.7
31 March	7130	7864	8396	1740.8
1 April	7001	7896	8741	2316.0

TABLE 1: Continued.

Italy				
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
2 April	6758	7889	8987	3019.6
3 April	6424	7851	9156	3800.7
4 April	6078	7813	9695	5015.6
5 April	5512	7718	10081	5917.4
6 April	4925	7494	10580	7727.4
7 April	4397	7238	10941	9036.2
Campania				
29 March	1169	1453	1549	499.7
30 March	1267	1494	1662	513.7
31 March	1293	1527	1747	600.4
1 April	1287	1548	1835	709.3
2 April	1280	1582	1898	817.0
3 April	1252	1619	1953	901.6
4 April	1222	1640	1972	998.3
5 April	1160	1649	2020	1047.3
6 April	1090	1648	2063	1218.5
7 April	994	1645	2145	1418.8
Macroarea South				
29 March	4380	5226	5662	1501.9
30 March	4646	5406	5846	1504.8
31 March	4689	5648	6246	1981.0
1 April	4705	5791	6528	2312.2
2 April	4496	5931	6902	3100.9
3 April	4122	6129	7325	3973.6
4 April	3841	6169	7695	4984.0
5 April	3445	6346	8141	6118.4
6 April	2981	6355	8787	7459.2
7 April	2512	6345	9397	8918.4

TABLE 2: 10-step-ahead predictions for the variable U_t (people in ICUs).

Italy				
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
29 March	3681	3960	4086	570.4
30 March	3751	4009	4170	601.3
31 March	3730	4075	4264	784.2
1 April	3732	4141	4350	880.0
2 April	3707	4185	4386	945.2
3 April	3667	4262	4493	1012.7
4 April	3608	4280	4610	1402.6
5 April	3492	4275	4701	1611.8
6 April	3347	4249	4814	2097.0
7 April	3143	4243	4899	2418.4
Lombardia				
29 March	1272	1369	1426	202.6
30 March	1286	1389	1458	237.4
31 March	1287	1402	1482	242.4
1 April	1291	1412	1490	274.7
2 April	1289	1420	1509	304.6
3 April	1286	1427	1526	325.6
4 April	1270	1425	1537	386.4
5 April	1252	1421	1558	423.7
6 April	1227	1416	1578	501.5
7 April	1191	1412	1599	580.3

TABLE 2: Continued.

	Italy			
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
Piedmont				
29 March	422	433	451	32.7
30 March	428	436	457	40.1
31 March	428	438	470	58.3
1 April	427	436	478	70.2
2 April	424	434	491	92.0
3 April	419	427	498	111.7
4 April	409	417	508	134.0
5 April	395	406	512	163.1
6 April	373	396	523	204.5
7 April	354	383	535	254.3
Liguria				
29 March	156	163	168	15.1
30 March	159	163	171	16.1
31 March	158	163	174	21.4
1 April	158	162	178	26.8
2 April	158	161	181	30.9
3 April	156	159	185	39.5
4 April	153	157	187	47.8
5 April	148	153	191	60.0
6 April	143	151	193	70.6
7 April	137	145	195	82.4
Valle Aosta				
29 March	24	25	27	2.4
30 March	25	25	29	3.1
31 March	24	25	30	4.0
1 April	24	24	31	4.3
2 April	23	22	32	5.5
3 April	22	21	33	6.8
4 April	21	19	34	8.2
5 April	19	16	36	12.3
6 April	18	14	36	13.3
7 April	15	12	37	18.4
Veneto				
29 March	331	357	367	48.5
30 March	335	360	375	53.7
31 March	335	365	382	65.3
1 April	334	370	391	80.6
2 April	332	375	401	96.7
3 April	324	378	409	119.4
4 April	312	375	424	156.7
5 April	298	376	440	200.4
6 April	286	370	448	231.1
7 April	267	369	461	268.0
Friuli Venezia Giulia				
29 March	56	57	60	4.3
30 March	56	58	60	4.8
31 March	56	58	63	4.5
1 April	55	58	63	5.2
2 April	54	57	65	7.0
3 April	52	56	66	8.3
4 April	51	56	67	10.2
5 April	49	55	69	11.4
6 April	48	53	71	12.0
7 April	44	52	72	15.3

TABLE 2: Continued.

	Italy			
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	$[\hat{\sigma}(V_t)]^*$
Emilia-Romagna				
29 March	306	349	379	100.5
30 March	309	365	394	116.7
31 March	309	369	409	136.5
1 April	312	372.	421	148.5
2 April	313	379	433	152.7
3 April	314	386	439	174.1
4 April	311	388	442	180.7
5 April	307	392	442	188.6
6 April	298	399	447	212.6
7 April	288	401	444	212.5
Trentino Alto Adige				
29 March	110	124	135	20.4
30 March	113	127	144	27.4
31 March	114	130	150	33.5
1 April	113	132	155	40.5
2 April	113	133	158	42.6
3 April	111	133	160	48.5
4 April	107	131	161	54.8
5 April	101	130	164	60.3
6 April	94	126	171	76.8
7 April	87	123	175	87.3
Lazio				
29 March	120	136	146	29.8
30 March	125	139	157	35.4
31 March	123	141	171	58.6
1 April	118	144	179	71.0
2 April	111	145	200	109.5
3 April	104	145	219	135.3
4 April	95	141	235	160.3
5 April	84	139	264	196.8
6 April	73	134	294	253.8
7 April	61	131	343	321.5
Macroarea Center				
29 March	476	491	509	40.6
30 March	485	492	519	45.3
31 March	484	492	534	64.9
1 April	482	489	547	85.6
2 April	479	484	565	116.2
3 April	475	478	586	142.0
4 April	468	469	608	193.6
5 April	460	460	624	224.0
6 April	448	448	649	273.3
7 April	438	438	681	331.9
Campania				
29 March	126	142	167	41.2
30 March	133	159	197	68.5
31 March	138	176	228	97.4
1 April	144	189	257	114.9
2 April	147	201	293	150.8
3 April	148	206	326	197.8
4 April	143	211	348	212.4
5 April	134	203	382	246.2
6 April	120	197	411	311.7
7 April	109	192	447	370.1

TABLE 2: Continued.

	Italy			$[\hat{\sigma}(\mathbf{V}_t)]^*$
	$CI_L^*(h)$	\hat{V}_t^*	$CI_U^*(h)$	
Macroarea South				
29 March	296	319	342	52.2
30 March	303	326	348	55.4
31 March	312	335	367	64.8
1 April	306	340	367	70.3
2 April	302	343	383	100.4
3 April	298	346	394	118.5
4 April	289	348	408	130.7
5 April	278	342	421	171.6
6 April	264	334	441	212.7
7 April	244	318	462	270.5

number of cases in intensive care will continue to grow at a progressively slower rate over the forecasting period.

- (iii) Veneto is the third region for number of deaths. Here, the number of positive cases, as well as the number of cases in ICU, will reach the peak on April 3rd.
- (iv) For Piedmont—the fourth region for number of victims—the predicted positive cases will reach the peak on March 29th (6635), whereas the persons in ICU will be 431 on March 31st, when the peak is predicted.
- (v) Liguria will begin a process of relative reduction of positive cases as early as March 29th. The number of cases in intensive care, after a period of stability (lasting until March 31st), will start a slow decreasing path.
- (vi) Positive cases in Trentino Alto Adige—which incorporates the cities of Trento and Bolzano—are projected to be 2158 on March 30th and then a decreasing trend is expected. The ICU beds occupied in this region will reach its peak on around April 3rd.
- (vii) The positive cases in Friuli Venezia Giulia show a relatively stable trend in the first half of the prediction interval with a peak around April 4th, after that the absolute number of cases will start decreasing. The number of cases in ICU will reach the peak between March 30th and April 1st.
- (viii) Valle d'Aosta is a small region which has been relatively less impacted by the virus. Here, a downward trend is expected to start on March 31st (for the positive cases) and around March 31st (cases in ICUs).
- (ix) The upward trend in the number of positive cases of Lazio is estimated to stop on March 31st and to reach the minimum at the end of the forecasting people (1821 cases). The number of ICU cases is estimated to reach its peak on the period 1–3 April.
- (x) The macroarea Center will reach its peak at the very beginning of the month of April (for the variable \mathbf{V}), whereas for the variable \mathbf{U} the estimated peak day is around 31st March.

- (xi) Campania will reach the peak of contagions on April 5th, whereas ICU cases will do on the previous day.
- (xii) The remaining southern regions (Abruzzo, Molise, Puglia, Basilicata, Calabria, Sicily, and Sardinia) will show an upward trend in the number of future positive cases lasting until April 6th, where 6355 cases are predicted. The number of persons requiring an ICU will reach the peak on April 4th (348 is the estimated number of cases).

7. Conclusion

A forecasting method for two variables typically crucial during a pandemic—i.e., the number of positives and the future occupancy of ICU beds—has been proposed. The whole procedure has been designed to fulfill such a need-to-know goal using a minimal set of data; that is, the time series related to the positives and the ICU occupancy. This is a point of strength, as, especially in the initial stages of a pandemic, the available time series are limited to only basic variables (such as the ones considered in this paper) and are necessarily short, a fact that in general rules out multivariate approaches. In addition to that, this procedure uses two powerful tools, i.e., ARIMA models and the MEB resampling scheme, to generate estimates and confidence intervals of those quantities which are less affected by uncertainty components than it would be without using the bootstrap step. Finally, the procedure includes a filter of the type Kalman, which proved to be effective in correcting irregularities and anomalies (e.g., outliers) typically found in this type of data. At least two are the points of weakness of the proposed method: firstly, the assumption that both the time series of interest are both realizations of (unknown) data generating processes of the type ARIMA is arbitrary and entails the introduction of not negligible amount of uncertainty into the analysis (order selection uncertainty). Secondly, once the “best” model order is found, the inference process inevitably leads to the loss of precious degrees of freedom. Future research directions include exploring different prediction models (e.g., of the type exponential smoothing) and combining the predictions generated by them.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics.

Conflicts of Interest

The author has no conflicts of interest to declare. In particular, he has no affiliations with or involvement in any organization or entity with any financial interest or

nonfinancial interest in the subject matter or materials discussed in this manuscript.

References

- [1] S. Moritz, A. Sarda, T. Bartz-Beielstein, M. Zaefferer, and J. Stork, "Comparison of different methods for univariate time series imputation in R," 2015, <http://arxiv.org/abs/1510.03924>.
- [2] D. Simon, "Kalman filtering," *Embedded Systems Programming*, vol. 14, no. 6, pp. 72–79, 2001.
- [3] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*, Vol. 496, McGraw-Hill, New York, NY, USA, 1971.
- [4] S. Makridakis and M. Hibon, "ARMA models and the Box-Jenkins methodology," *Journal of Forecasting*, vol. 16, no. 3, pp. 147–163, 1997.
- [5] H. D. Vinod, "Maximum entropy ensembles for time series inference in economics," *Journal of Asian Economics*, vol. 17, no. 6, pp. 955–978, 2006.
- [6] H. D. Vinod, "New bootstrap inference for spurious regression problems," *Journal of Applied Statistics*, vol. 43, no. 2, pp. 317–335, 2016.
- [7] G. D. Birkhoff, "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 17, no. 12, pp. 656–660, 1931.
- [8] J. Berkowitz and L. Kilian, "Recent developments in bootstrapping time series," *Econometric Reviews*, vol. 19, no. 1, pp. 1–48, 2000.