

# Modeling the first wave of the COVID-19 pandemic in the Lombardy region, Italy, by using the daily number of swabs

## *Modellazione della fase iniziale della pandemia da COVID-19 in Lombardia, con l'utilizzo del numero giornaliero di tamponi*

Claudia Furlan and Cinzia Mortarino

**Abstract** The daily fluctuations in the released number of Covid-19 cases played a big role both at the beginning and in the most critical weeks of the outbreak, when local authorities in Italy had to decide whether to impose a lockdown and at which level. Public opinion was focused on this information as well, to understand how quickly the epidemic was spreading. In this work, we propose a nonlinear asymmetric diffusion model, which includes information on the daily number of swabs, to describe daily fluctuations in the number of confirmed cases in addition to the main trend of the outbreak evolution. The proposed model is compared with alternative model structures in the application to data of the Lombardy region.

**Abstract** *Le oscillazioni giornaliere nel numero di contagi diagnosticati di Covid-19 sono state al centro dell'attenzione nelle fasi iniziali della pandemia, quando le autorità avevano scarsi elementi per decidere quali restrizioni adottare. Anche l'opinione pubblica e i media erano costantemente focalizzati su questo dato quotidiano, per cercare di ricavarne elementi sull'evoluzione del contagio. In questo lavoro, proponiamo un modello di diffusione nonlineare asimmetrico per descrivere, oltre al trend dei contagi diagnosticati, le oscillazioni giornaliere. Il modello utilizza come input il numero di tamponi processati quotidianamente. Il modello proposto è sottoposto a comparazione con altri cinque modelli sui dati della regione Lombardia.*

**Key words:** nonlinear models, Generalized Bass model, logistic, diffusion

## 1 Introduction

Italy was the first nation to be affected by Covid-19 after China, and the epidemic has mainly been located in Northern Italy. On February 21st, 2020, an infected patient was detected in the small town of Codogno, which is located in the Lombardy

---

Claudia Furlan and Cinzia Mortarino  
Department of Statistical Sciences, Padova, Italy,  
e-mail: furlan@stat.unipd.it, mortarino@stat.unipd.it

region. During the *first wave*, among the Italian regions, Lombardy is the most affected by the epidemic, with a death toll three times greater than that in China [1]. It is apparent that, in Italy, the regional autonomy regarding health policy has resulted in services with different levels of quality [1], such as the number of beds and the capacity of processing swabs. With regard to the number of beds in Italy, the forecasts of hospitalisations was faced by [2] for the bordering Veneto region, while [3] modeled the intensive care unit occupancy.

The capacity of processing swabs is of particular importance for detecting the state of the epidemic, measuring the lockdown effects and, most importantly, reducing the outbreak. Our opinion is that it is necessary to include the number of swabs to describe the local fluctuations in the epidemic evolution in addition to detecting the main trend. At the beginning of the outbreak, the curve of confirmed cases was usually modeled through an exponential [4] or a logistic growth model [5]. When the data collection window became long enough, the models were usually of two types: the compartmental and ARIMA models.

We made an effort to describe the cumulative number of confirmed cases in the Lombardy region, based on the combination of a nonlinear model and the number of completed swabs. In the class of growth models, we propose a new version of the dynamic potential model [6], where the novelty consists of the formulation of a new intervention function with the number of daily swabs as an explanatory variable. The model is particularly parsimonious since the intervention function has only one additional parameter. The base of the dynamic potential model was chosen since a) it has an asymmetric shape and makes it possible to model a ‘saddle’, which is a rather common nonlinear pattern; b) it gives an estimate of the total number of confirmed cases at the end of the epidemic; and c) the total number of confirmed cases is not fixed throughout the outbreak, but it is allowed to change over time. Since the capability of processing swabs increased over time and, consequently, the meeting criteria for people for being tested were enlarged with the aim of detecting a larger number of asymptomatic positive subjects, it is sensible to suppose that the number of diagnosed cases increases with time.

The proposed model was compared with five alternative growth models described in Section 2. Three-week forecasts of the spreading dynamics were provided for each model as well. The models were compared in terms of  $R^2$  and BIC values, for the cumulative values. The squared linear correlation coefficient between observed and fitted daily values was evaluated as well.

## 2 Models

A general diffusion of innovations model can be defined through a nonlinear regression model as follows:

$$y(t) = z(t, \vartheta) + \varepsilon(t), \quad (1)$$

where  $y(t)$  are the cumulative sales of a product at time  $t$  and  $z(t, \vartheta) = z(t)$  is a specific structure to be used to describe an evolution process. Here,  $\varepsilon_t$  are assumed

Title Suppressed Due to Excessive Length

to be i.i.d. Gaussian with variance  $\sigma^2$ . The components of the parameter vector  $\vartheta$  are jointly estimated using nonlinear least squares (or, equivalently, likelihood estimation).

In this paper, we will compare the performance of alternative evolution structures. The basic model is a logistic one (LOG):

$$z(t) = m \frac{e^{\frac{t-\lambda}{\eta}}}{1 + e^{\frac{t-\lambda}{\eta}}}, \quad (2)$$

where  $\lambda$  is the mode, median and average of the distribution, while  $\eta$  is a shape parameter. Parameter  $m$  is the market potential, which is the limiting value for  $z(t)$ , as  $t$  goes to infinity.

The Generalized Bass Model corresponds to:

$$z(t) = m \frac{1 - e^{-(p+q) \int_0^t w(\tau) d\tau}}{1 + \frac{q}{p} e^{-(p+q) \int_0^t w(\tau) d\tau}}, \quad (3)$$

where  $m$  is the market potential,  $p$  is the innovation coefficient,  $q$  is the imitation coefficient and  $w(t)$  can be any integrable function. The effect of the intervention function  $w(t)$  is to accelerate or decrease diffusion with respect to a symmetric unimodal path, which would arise in (3) for  $w(t) = 1$  for all  $t$  values. For  $t$  values such as  $w(t) > 1$  diffusion is accelerated, while  $w(t) < 1$  corresponds to time periods with decreased diffusion speed. Below, we examine the model (GBM<sub>RECT</sub>) arising when  $w(t)$  is specified by the so-called *rectangular* shock:

$$w_R(t) = 1 + cI_{a \leq t \leq b}. \quad (4)$$

This allows us to describe the diffusion of a product for which we observe a constant shock with intensity  $c$ , either positive or negative, in the time interval  $[a, b]$  [7].

Due to the asymmetric path observed for almost every region, we also examine the more flexible Bemmaor model, in an extended version with a rectangular shock (BeGBM<sub>RECT</sub>):

$$z(t) = m \frac{1 - e^{-(p+q) \int_0^t w_R(\tau) d\tau}}{[1 + \frac{q}{p} e^{-(p+q) \int_0^t w_R(\tau) d\tau}]^A}, \quad (5)$$

where  $A$  is a further parameter allowing for asymmetry (positive asymmetry for  $A > 1$ , negative asymmetry for  $A < 1$ ), with function  $w_R$  specified as in (4).

A different way to provide flexibility to the evolutive structure can be obtained through a dynamic market potential model [6], eventually perturbed by shocks

$$z(t) = m \sqrt{\frac{1 - e^{-(p_c+q_c)t}}{1 + \frac{q_c}{p_c} e^{-(p_c+q_c)t}}} \frac{1 - e^{-(p+q) \int_0^t w(\tau) d\tau}}{1 + \frac{q}{p} e^{-(p+q) \int_0^t w(\tau) d\tau}}, \quad (6)$$

where  $p_c$  and  $q_c$  are two parameters to describe how fast the dynamic market potential approaches its maximum value,  $m$ , while  $w(t)$  is a general intervention function. If in model (6) we use, as proposed in [8], the following intervention function

$$w_s(t) = 1 + \alpha_1 \cos\left(\frac{2\pi t}{s}\right) + \alpha_2 \sin\left(\frac{2\pi t}{s}\right), \tag{7}$$

we allow the model to incorporate cyclic seasonal fluctuations of width  $\alpha_1$  and  $\alpha_2$  with period  $s$  (DMPseas). If  $w(t) = 1$ , we obtain dynamic market potential (DMP).

Here, we propose to assess the usefulness of a dynamic market potential model as in (6), but with an intervention function depending upon the number of swabs analyzed at day  $t$ ,  $B(t)$  (DMPsw). In particular, we suggest using

$$w_B(t) = 1 + \xi \left( \frac{B(t) - \mu_B}{\sigma_B} \right), \tag{8}$$

where  $\mu_B$  and  $\sigma_B$  are the average and the standard deviation, respectively, of the  $B(t)$  values recorded during the observation period. It is easy to appreciate that such a structure accelerates, with respect to an underlying trend described by a DMP, the number of cases whenever  $B(t)$  exceeds its average, while cases are reduced with a below-average number of swabs.

### 3 Applications

Models of Section 2 were applied to the data of Lombardy, and forecasts up to May 24th are provided (three weeks ahead for each region). The six models were fitted to the cumulative confirmed cases using NLS estimation. Table 1 summarizes the values of the determination index  $R^2$  for all models: the huge values of  $R^2$  are unsurprising, given that we are working with cumulative data and any S-shaped fitting produces high determination indexes.

Lombardy is the Italian region where COVID-19 spread in the most dramatic way. The total number of infected people on May 3rd was 77528 with more than 14000 deaths (about half of the death toll up to that date in Italy as a whole). The results for Lombardy are displayed in Table 1 ( $R^2$ , BIC and  $\rho^2$ ), and in Figure 1, where observed and fitted daily values are plotted.

For this region, the logistic (Figure 1(a)) is the less effective model in describing the asymmetrical evolution of the epidemic.

A positive ( $\hat{c} > 0$ ) rectangular shock is significantly diagnosed at the beginning of the time series, both in the  $GBM_{RECT}$  and the  $BeGBM_{RECT}$ . The  $GBM_{RECT}$  estimates the end of the shock on March 25th ( $t \simeq 34$ ), but according to Figure 1(b),

**Table 1**  $R^2$  of the nonlinear models and corresponding BIC (cumulative data as response variable) and squared linear correlation coefficient,  $\rho^2$ , between observed *instantaneous* sales and fitted *instantaneous* sales.

	LOG	$GBM_{RECT}$	$BeGBM_{RECT}$	DMP	DMPseas	DMPsw
Lombardy $R^2$	0.993010	0.999629	0.999900	0.999834	0.999860	0.999919
BIC	1143.348	941.8574	850.6930	878.7936	879.2002	830.5268
$\rho^2$	0.593900	0.690817	0.826706	0.803006	0.820098	0.902698

this is not perfectly matching with the data. This is the reason why, for this model,  $\rho^2$  is particularly small (0.690817).

Conversely, the BeGBM<sub>RECT</sub> better identifies the end of the shock three days later, on March 28th, when we observe a relevant stable decrease. For this region, the lockdown policy had a delayed effect compared to what happened in Veneto, as the decrease in the number of cases was registered 20 days after March 8th, while the incubation period is up to 14 days. One reason for such a wider interval could be possible delays in taking and processing the swabs; in fact, the health system of Lombardy experienced an unexpected overload.

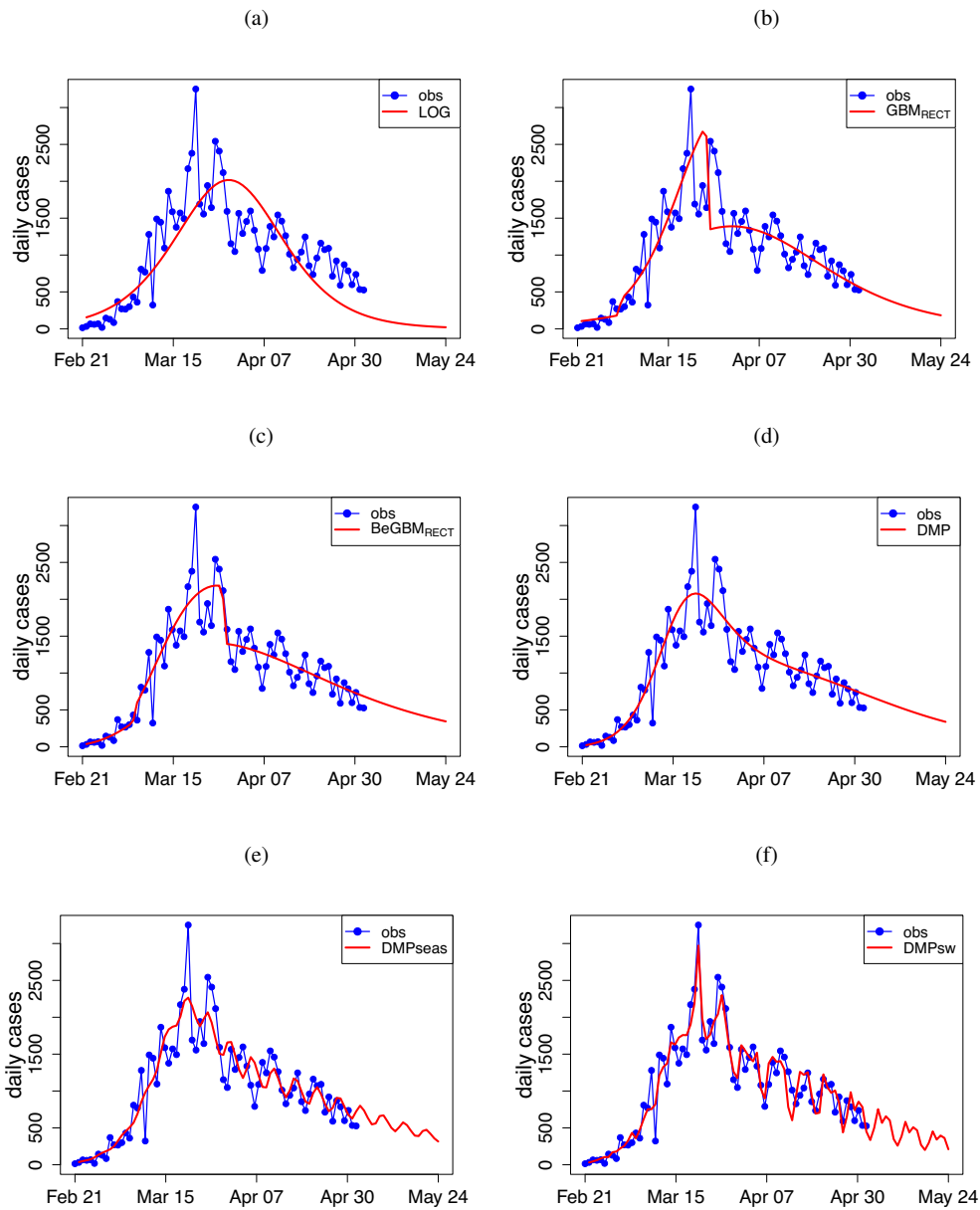
With the DMP model (Figure 1(d)), it is possible to fully appreciate the asymmetrical shape of the outbreak, especially the slow decrease in the number of cases in this region. However, its performance in terms of  $R^2$ ,  $\rho^2$  and BIC is worse than that of the BeGBM<sub>RECT</sub>.

The performance of the DMP<sub>seas</sub>, with a weekly cycle ( $\hat{s} = 7.005$  days), is not satisfactory, as it does not adequately capture the fluctuations (except for the very end of the series). Here, too, the  $R^2$ ,  $\rho^2$  and BIC values are worse than those obtained with the BeGBM<sub>RECT</sub>.

Finally, the DMP<sub>sw</sub> (Figure 1(f)) performs very well. With this model, we obtained the largest values for  $R^2$ , 0.999919, and  $\rho^2$ , 0.902698. The BIC value for this model, 830.5628, supports it with respect to the BeGBM<sub>RECT</sub> (850.6930), which was the best model up to this point. The proposed model, which is highly parsimonious, is able to describe the daily fluctuations in cases very well and proved to be the best of the models analysed here for Lombardy.

## References

1. Indolfi, C., Spaccarotella, C.: The outbreak of covid-19 in italy. *JACC: Case Reports* **2**(9), 1414–1418 (2020)
2. Gregori, D., Azzolina, D., Lanera, C., Prosepe, I., Destro, N., Lorenzoni, G., Berchiolla, P.: A first estimation of the impact of public health actions against covid-19 in Veneto (Italy). *Journal of Epidemiology and Community Health* **74** (10), 858–860 (2020)
3. Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., Lovison, G.: An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical Journal* (2020) doi:202000189
4. Remuzzi, A., Remuzzi, G.: Covid-19 and italy: what next? *Health policy* **395**, 1225–1228 (2020)
5. Shen, C.Y.: A logistic growth model for covid-19 proliferation: Experiences from china and international implications in infectious diseases. *International Journal of Infectious Diseases* **96**, 582–589 (2020)
6. Guseo, R., Guidolin, M.: Modelling a dynamic market potential: a class of automata networks for diffusion of innovations. *Technological Forecasting and Social Change* **76**(6), 806–820 (2009)
7. Guseo, R., Dalla Valle, A.: Oil and gas depletion: diffusion models and forecasting under strategic intervention. *Statistical Methods and Applications* **14**(3), 375–387 (2005)
8. Guidolin, M., Guseo, R.: Modelling seasonality in innovation diffusion. *Technological Forecasting and Social Change* **86**, 33–40 (2014)



**Fig. 1** Lombardy. Observed and fitted values with the alternative models. (a) Logistic (LOG); (b) GBM with rectangular shock ( $GBM_{RECT}$ ); (c) Bemmaor GBM with rectangular shock ( $BeGBM_{RECT}$ ); (d) Dynamic market potential (DMP); (e) Dynamic market potential+seasonal effect (DMPseas); (f) Dynamic market potential+swabs (DMPsw).