A fractional model for the COVID-19 pandemic: Application to Italian data

Elisa Alòs\textsuperscript{a,b},\textsuperscript{†}, Maria Elvira Mancino\textsuperscript{c}, Raúl Merino\textsuperscript{d,e}, and Simona Sanfelici\textsuperscript{f}

\textsuperscript{a}Dpt. Economia i Empresa, Universitat Pompeu Fabra, Barcelona, Spain; \textsuperscript{b}Barcelona Graduate School of Economics, Barcelona, Spain; \textsuperscript{c}Department of Economics and Management, University of Florence, Florence, Italy; \textsuperscript{d}Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain; \textsuperscript{e}VidaCaixa S.A, Market Risk Management Unit, C/Juan Gris, Barcelona, Spain; \textsuperscript{f}Department of Economics and Management, University of Parma, Parma, Italy

\textbf{ABSTRACT}

We provide a probabilistic SIRD model for the COVID-19 pandemic in Italy, where we allow the infection, recovery and death rates to be random. In particular, the underlying random factor is driven by a fractional Brownian motion. Our model is simple and needs only some few parameters to be calibrated.

\section{1. Introduction}

The modeling of infectious diseases and their propagation is essential in the field of mathematical epidemiology. These models help us understand the disease behavior as well as being a useful tool to control its spread. There is extensive literature. For example, in the case of autonomous models, which does not take into account age structure nor environmental fluctuations, we can see [1–3] and [4]. But, many diseases show seasonal behavior. Therefore, non-autonomous models are usually more realistic. For example, we can refer to [5–8] and [9]. In recent years, models based on Ordinary Differential Equations have been studied, some examples are [1–4, 10, 11] and [12].

Since the start of the COVID-19 pandemic, many researchers have been reviewing the literature to explain its dynamics. Most of these approaches are done by using compartmental models, where the population is divided into compartments that describe the different situations regarding the infectious disease (such as susceptible, infected, etc.), and people’s progress between them.

The basic reference compartmental model is the SIR model where the population is divided into susceptible (S), infected (I), and recovered (R). As this model is too simple...
to describe the complexity of several epidemics, some extensions have been proposed in
the literature. For example, the SIRD model also considers the deaths compartment (D),
and the SEIR introduces exposed (E). Some other classical models, such as SEIRS, also
allow to model the loss of immunity after recovery. Some recent approaches pay special
attention to non directly observable compartments as asymptomatic (A) (see, for
example, [13] and the references therein) that, even not being directly observable, play a
crucial role in the pandemic. Other extensions consider time-dependent parameters, as
in [14], where the kinetic of the rates of infection ($\beta$) and death ($\mu$) are modeled by
exponential functions, while the recovery rate ($\gamma$) is of logistic type. Recent literature
also exploits stochastic models, where the main variables account for a noise
(Brownian) component, or branching processes [15–18] and [19].

In this paper, we introduce a probabilistic SIRD model, where we allow the coeffi-
cients to be stochastic processes determined by a few number of parameters. This
randomness is a way to accommodate the effect of several unobservable factors, as
the loose of immunity or the existence of asymptomatic. Our construction of the
model is motivated by the descriptive analysis of the parameter time series, where
we observe that the $\beta$ returns have negatively correlated increments, an observation
that suggests to model them by a fractional Brownian motion (fBm) with a Hurst
parameter $H < \frac{1}{2}$. We recall that this approach does not need to consider a drift for
$\beta$, but this drift arises simply by the properties of the fBm. Once modeled $\beta$, we see
that simple relationships between the evolution of diagnosed and infected, allow us
to model $\gamma$ and $\mu$.

The model is calibrated in such a way that the mean paths of infected, death and
recovered fit the corresponding observed values, as well as the variability of $\beta, \gamma, \mu$.
Our approach is not only able to adjust observed data, but also to study the different
possible scenarios according to its stochastic behavior.

The paper is organized as follows. In Section 2 we present a descriptive analysis of $\beta,
\gamma$ and $\mu$ corresponding to the evolution of the pandemic in Italy. Section 3 is devoted to
present our stochastic SIRD model for the Italian COVID-19 outbreak. The model is
calibrated in Section 4, while we simulate some different scenarios in Section 5. Finally,
a discussion of the results and proposals of future research are presented in Section 6.

2. A descriptive analysis of the SIRD model in the Italian COVID-
19 outbreak

Let us consider a stochastic SIRD model of the form

\[
\begin{align*}
S_{n+1} &= S_n - \beta_n I_n S_n / N \\
I_{n+1} &= I_n (1 + \beta_n S_n / N - \gamma_n - \mu_n) \\
R_{n+1} &= R_n + \gamma_n I_n \\
D_{n+1} &= D_n + \mu_n I_n,
\end{align*}
\]

(1)

where $S, I, R, D = \{S_n, I_n, R_n, D_n, n = 1, \ldots, 150\}$ denote the number of daily observed
susceptible, infected, recovered and death, $N$ is the population size and $\beta, \gamma, \mu =
{\beta_n, \gamma_n, \mu_n, n = 1, \ldots, 150}$ represent the rates of infection, recovery and death. As
$S/N \approx 1$ in all the data set, we consider the following simpler version of the above model
\begin{align*}
S_{n+1} &= S_n - \beta_n I_n \\
I_{n+1} &= I_n (1 + \beta_n - \gamma_n - \mu_n) \\
R_{n+1} &= R_n + \gamma I_n \\
D_{n+1} &= D_n + \mu I_n,
\end{align*}
\tag{2}

The data used to study the model has been downloaded from the Github repository: https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv. Every day this file is updated with all the variables necessary to calibrate a SIRD model.

Now we observe the behavior of this model in Italy in the period from the 24/2/2020 to the 28/7/2020. In Figure 1 we can observe the paths of \( I \) (totale positivi in the data set), \( R \) (dimessi-guariti), \( D \) (deceduti) and the total number of diagnosed \((I + R + D, \text{totale casi})\).

### 2.1. The process \( \beta \)

Now let us observe \( \beta, \gamma, \mu \). In Figure 2 we can see the behavior of \( \beta \), that is a decreasing function of time. This fits what expected due to the lock-down, that reduced the number of contacts between individuals. In Figure 3, we can see the corresponding increments (i.e., \( \Delta \beta = \beta_{n+1} - \beta_n \)), that are more variable at the beginning of the sample, when \( \beta \) is higher.

Moreover, in Figure 4 we can see the returns \( \frac{\Delta \beta}{\beta} \) that have a stationary mean. After deleting the outliers at days 115 and 119 (making them equal to zero), the corresponding autocorrelation function, see Figure 5, reveals a negative correlation between increments. The empirical analysis suggests to model \( \beta \) as a process with negative correlated increments with a decreasing mean. This process is designed in Section 3.

### 2.2. The process \( \mu \)

The time behavior of \( \mu \) can be seen in Figure 6. We observe that it is similar to the behavior of \( \beta \), but with a delay. Moreover, Figures 7 and 8 show a clear relationship
between the daily new diagnosed and deaths, with a delay of 4 days\(^1\). In Figure 9, we can see a strong link between total diagnosed and deaths. In Figures 10 and 11 we can see that this relationship is linear. Then, as the number of new infected is given by \(\beta I\), the daily increments of deaths (\(\Delta D\)) would be modeled simply as \(c_\mu \beta_{n-4} I_{n-4}\), for some positive constant \(c_\mu\). This leads to the following model for \(\mu\):

\[
\mu_n = c_\mu \beta_{n-4} I_{n-4} / I_n,
\]

\(^1\)This finding appears consistent with data by Istituto Superiore di Sanità https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_17_marzo-v2.pdf.
2.3. The process

Finally, let us observe the path of $\gamma$ in Figure 12. This recovery rate seems to move between an upper and a lower bound, that we can interpret as the rate under stressed conditions of the health system, and the rate when this system adapts to the new scenario and the epidemic curve decays. This is connected to the approach in [14], where $\gamma$ is assumed to be described by a logistic function, and the rate of recovery increases from a initial value $\gamma_0$ and it stabilizes at some higher level.
A measure of the stress of the model can be defined as the quantity

\[ \frac{(I + R + D) - 30n}{C_0^n} = \frac{(I + R + D)n}{C_0^n}, \]  

(3)

where \((I + R + D) - 30n\) is the average of the diagnosed people the last 30 days. If there are no new infections, the number of diagnosed remains stable, and then the above quantity is near 1. Otherwise, a high increment of new cases translates into a decrease of the value of this ratio. In Figure 13, we can see the behavior of this process.

We can also observe in Figure 12 a big variance, as expected due to the fact that not everybody recovers at the same speed. Then we would like to add some noise, multiplying this time series by an adequate random factor, as we detail in the following section.
3. A stochastic SIRD model

We propose a stochastic SIRD model where randomness is embedded by modeling the parameters of infection ($\beta$), mortality ($\mu$) and recovery ($\gamma$). In the following subsections, we will describe the model for each one of the parameters.

3.1. A stochastic model for $\beta$

As the returns of $\beta$ are negatively correlated, it is natural to construct a model based on the fractional Brownian motion (fBm). The fBm is a Gaussian process with...
stationary increments, which depends on a parameter $H \in (0, 1)$ called the Hurst parameter. More precisely, a process $B^H = B^H_t, t \in [0, T]$ is called a fractional Brownian motion (fBm) if

$$E(B^H_0) = 0.$$ 

$$E(B^H_t B^H_s) = \frac{1}{2} (s^{2H} + t^{2H} - |t - s|^{2H}).$$

Except in the Brownian motion case $H = 1/2$, the fBm increments are correlated. Denote $X_n = B^H_n - B^H_{n-1}$ and define $\rho_H(n) := Cov(X_1, X_{n+1})$. Then we have

Figure 10. Linear dependence between $\Delta$ diagnosed (with 4 days lag) and $\Delta$ Death.

Figure 11. Linear dependence between total number of diagnosed (with 4 days lag) and Deaths.
\[ q_n = E(B_1^H(B_{n+1}^H - B_n^H)) = \frac{1}{2} \left( 1^{2H} + (n+1)^{2H} - n^{2H} \right) - \frac{1}{2} \left( 1^{2H} + n^{2H} - (n-1)^{2H} \right) = \frac{1}{2} \left( (n+1)^{2H} + (n-1)^{2H} - 2n^{2H} \right) \] (4)

Notice that this quantity is positive for \( H > \frac{1}{2} \) and negative if \( H < \frac{1}{2} \).
Then, a natural candidate to model \( \beta \) is given by

\[
\hat{\beta}_{n+1} = \hat{\beta}_n (1 + k(B^H_{n+1} - B^H_n)),
\]

(5)

where \( B^H \) is a fractional Brownian motion with \( H < \frac{1}{2} \) and \( k \) is a constant. Apart from the negative correlated returns, this process has also a decreasing mean, as we prove in the following result.

**Proposition 3.1.** Consider a fractional Brownian motion \( B^H \) with \( H < \frac{1}{2} \) and the process defined by (5). Then, provided \( \beta_0 > 0 \), \( E(\hat{\beta}_n) \) is decreasing as \( n \to \infty \).
Proof. A recursive computation gives us that
\[ \hat{\beta}_n := \hat{\beta}_0 A_n, \]
with
\[ A_n := \prod_{m=1}^{n} (1 + k \Delta B^H_m), \]
where \( \Delta B^H_m = B^H_{m+1} - B^H_m \). The above product can be written as
\[ 1 + \sum_{i=1}^{n} k^i \left( \sum_{m_1 < \cdots < m_i} \Delta B^H_{m_1} \cdots \Delta B^H_{m_i} \right) \]
(6)
Now, we can compute the expectation using Isserlis Theorem. This gives us that
\[ E(A_n) = E(\Pi_{m=1}^{n} (1 + k \Delta B^H_m)) \]
\[ = 1 + \sum_{i=1}^{n} k^i \left( \sum_{m_1 < \cdots < m_i} E(\Delta B^H_{m_1} \cdots \Delta B^H_{m_i}) \right) \]
\[ = 1 + \sum_{i=1}^{n} k^i \left( \sum_{m_1 < \cdots < m_i} \sum_{p \in \mathcal{P}_i} \Pi_{(k,j) \in p} \text{Cov}(\Delta B^H_{m_1}, \Delta B^H_{m_j}) \right) \]
\[ = 1 + \sum_{i=2}^{n} k^i \left( \sum_{m_1 < \cdots < m_i} \sum_{p \in \mathcal{P}_i} \Pi_{(k,j) \in p} \rho_H(|m_j - m_k|) \right), \]
(7)
where we have used that partitions \( \mathcal{P}_i \) exist only if \( i \) is even. Then, as \( k^i \) is positive for all even \( i \) and all the \( \rho_H(|m_j - m_k|) \) are strictly negative, \( E(A_n) \) is decreasing. This allows us to complete the proof. \( \Box \)

Even when the model (5) reproduces some empirical properties of \( \beta \), a numerical analysis shows that it has to be modified before being an adequate model. More precisely, we can see in Figure 14 that the variability of the paths is high, and that the process \( \hat{\beta} \) can even become negative (see also Figure 15). Moreover, the estimated correlation between two consecutive increments of \( \hat{\beta} \) does not coincide with the observed one for the \( \beta \) returns.

A way to reduce the variance of the paths is obviously to define a model of the type
\[ \hat{\beta}_{n+1} = \sum_{i=1}^{m} \hat{\beta}^i_n, \]
(8)
where \( \hat{\beta}^i, i = 1, \ldots, m \) are given by
\[ \hat{\beta}^i_{n+1} = \hat{\beta}^i_n (1 + c^i_p (B_{n+1}^{kH} - B_n^{kH})), \]
being \( B^{kH}, i = 1, \ldots, m, \) independent fractional Brownian motions with Hurst parameter \( H < \frac{1}{2} \) and where \( c^i_p, i = 1, \ldots, n \) are positive constants. The interpretation of (8) is intuitive: the observed process \( \hat{\beta} \) is really the sum of different stochastic \( \hat{\beta}^i \) that correspond to the particular transmission rates in different groups, locations, etc. In Figures 15 and 16 we can observe, for \( H = 0.1 \), the behavior of the cases \( m = 1 \) (where the volatility is too big), and \( m = 10 \), where the area delimited by the the maximum and minimum paths contain the observed values of \( \beta \).
3.2. A stochastic model for $\mu$ and $\gamma$

As pointed out in 2, the linear relationship between diagnosed (with 4 days lag) and death, suggests modeling $\mu$ simply as

$$\mu_n = c_\mu \beta_{n-4} I_{n-4} / I_n.$$ 

On the other hand, the observations in Section 2 lead to a model for $\gamma$ of the type

$$\gamma_n = c_\gamma^1 [(I + R + D)_{-30n} / (I + R + D)_n] \exp (c_\gamma^2 (B_{n+1}^H - B_n^H)),$$

for some positive constants $c_\gamma^1, c_\gamma^2$ and for some adequate Hurst parameter $H_\gamma$.

3.3. The global model

The above models for $\beta, \gamma$, and $\mu$ lead to the following stochastic SIRD model:

$$\begin{align*}
\beta_{n+1} &= \sum_{i=1}^{10} \beta_i^i, \text{ with } \beta_i^i = \beta_i^i (1 + c_\beta (B_{i,n+1}^H - B_{i,n}^H)) \\
\mu_n &= c_\mu \beta_{n-4} I_{n-4} / I_n \\
\gamma_n &= c_\gamma^1 [(I + R + D)_{-30n} / (I + R + D)_n] \exp (c_\gamma^2 (B_{n+1}^H - B_n^H)) \\
S_{n+1} &= S_n - \beta_n I_n \\
I_{n+1} &= I_n (1 + \beta_n - \gamma_n - \mu_n) \\
R_{n+1} &= R_n + \gamma I_n \\
D_{n+1} &= D_n + \mu I_n.
\end{align*}$$

(9)

Notice that only 7 parameters have to be calibrated: $m, H, H_\gamma$ and $c_\beta, c_\mu, c_\gamma^1, c_\gamma^2$. We see in the following section how the set of the first three parameters $m, H, H_\gamma$, that define the driving processes of the model, is chosen based on empirical observations, while the last group is calibrated by means of a classical least squares method.
4. Calibration

4.1. The steps of the calibration process

The calibration procedure is as follows.

- In the first step, we fix reasonable values of $m$ and the Hurst parameters according to observed data, and
- fixed $m$, $H$, and $H_c$, we calibrate $c_\beta$, $c_\mu$, $c_\gamma^1$, $c_\gamma^2$ by the least squares method.
Step 1 focuses on choosing adequate driving random processes for the model. As estimating with precision and robustness this group of parameters is not straightforward (see for example [20]), so we simply proceed empirically. More precisely, we have seen in Section 2 that the maximum and the minimum paths in the case $H = 0.1$ and $m = 10$ envelope the observed $\beta$ time series. Moreover, the observed correlation of beta returns (for lag $= 1$) is equal to $-0.347$, while (from a 1000 simulations sample), we estimate this quantity to be $-0.314$ for $H = 0.1$ and $m = 10$. This leads to choose $m = 10, H = 0.1$ in our model.

In order to choose $H_c$, we compute the autocorrelation function of

$$\log \left( \frac{\gamma_n}{(I + R + D)_{-30n}/(I + R + D)_{n}} \right).$$

At lag $= 1$, this autocorrelation function is equal to 0.244. Then, taking into account Equation (4), we get an estimation of $H_c = 0.657$. Then we choose $H_c = 0.6$ for the sake of simplicity.

4.2. Initial guess for $c_\beta$, $c_\mu$, $c_1$, $c_2$ and global calibration

In order to be able to start the calibration of the other parameters, we will need to have an initial guess.

To find the parameter $c_\beta$, we are going to fit the average $\hat{\beta}$ against the realized $\beta$. This gives us a value of 0.32250809.

For $\mu$, as we have seen previously, we can obtain a nice estimation by doing a linear regression between the infected and death people. The slope (0.13904755) will be the initial guess for $c_\mu$.

In the case of $\gamma$, in order to obtain the variables $c_1, c_2$, we minimize the distance between the average $\hat{\gamma}$ and the observed $\gamma$, as well as the distance between the standard deviation of the returns of $\hat{\gamma}$ and the observed $\gamma$. This gives us (0.03512639,0.5) as initial guess.

Once we have the initial guess, we find the best possible parameters to minimize the distance between the infected, the death and the recovered time series as well as the empirical variance of the $\gamma$ by OLS.

5. Results

Following the procedure in the previous section, we get the following estimation of the parameters of the model:

$$c_\beta = 0.32102563, c_\mu = 0.13981687, c_1 = 0.03383898, c_2 = 0.54395798.$$

5.1. The average paths of the global calibration

Simulating the model and taking the corresponding mean paths we fit the observed data, as we can see in Figure 17.
5.2. The average paths of $\beta$, $\gamma$, and $\mu$

In the previous section, we have seen that our model is able to fit the evolution of the pandemic in Italy. In the Figures 18–20, it is possible to observe that the average simulated path of $\beta$, $\gamma$, and $\mu$ behaves like a running average of each time series.

5.3. Some simulated paths

Before, we have seen that the model performs well on average. Both the dynamics of the model and the adjustment of the parameters. In particular, the simulated mean path
of each parameter behaves as a moving average of the time series. Even so, we are also interested in the fact that for each parameter, the dynamics of each simulated path has a behavior similar to the observed one in real life. Therefore, we will compare the real data of $\beta$, $\mu$ and $\gamma$ with any simulated path.

In Figure 21, we observe that the random path decreases similarly to the real data. We have also observed this behavior also in Figure 14. In fact, it is a property of the selected process.

In Figure 22, we observe that the random path follows the trend of the original data. In Figure 23, it is possible to visually observe that the simulations behave similarly to the real data. In particular, we can mimic the peaks and valleys of $\gamma$ as can be seen in Figure 23.
As might be expected, being able to simulate different scenarios for $\beta$, $\mu$ and $\gamma$ give us the chance to generate different scenarios of the evolution of the pandemic. All the scenarios of the pandemic have a similar shape as it can be seen in Figures 24–26. But notice each simulation of the pandemic has a different magnitude.

6. Discussion

In the present work, we have extended the classic SIRD model including stochastic parameters and we obtain an easy-to-calibrate pure probabilistic model. We have been able to reproduce the evolution of the parameters of the Italian outbreak using only seven
parameters. The properties of the fractional Brownian motion and the relationship between the parameters of the model are able to reproduce the exponential and logistic trends observed for example in [14]. Therefore, we have been able to fit the empirical data as well as imitate the noise of the variables. One of the advantages of having a stochastic model is that we are capable of generating a wide variety of scenarios.

One challenging problem is now to model the evolution after a lockdown. This translates into a new dynamics of $\beta$, probably driven by a fBm with $H > \frac{1}{2}$. Moreover, the comparison between different countries that applied different policies during the pandemic would be of great interest.
References

[1] Kermack, W. O., McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. Lon. Series A.* 115(772):700–721.

[2] Anderson, R. M., May, R. M. (1992). *Infectious Disease of Humans, Dynamical and Control.* New York: Oxford University Press.

[3] Diekmann, O., Heesterbeek, J.A.P. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* Wiley Series in Mathematical and Computational Biology. Chichester: Wiley.

[4] Thieme, H.R. (2003). *Mathematics in Population Biology.* London: Princeton University Press.

[5] London, W. P., Yorke, J. A. (1973). Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *Am. J. Epidemiol.* 98(6):453–468. DOI: 10.1093/oxfordjournals.aje.a121575.

[6] Dowell, S. (2001). Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg. Infect. Dis.* 7(3):369–374.

[7] Earn, D., Dushoff, J., Levin, S. (2002). Ecology and evolution of the flu. *Trends Ecol. Evol.* 17(7):334–340.

[8] Stone, L., Shulgin, B., Agur, Z. (2000). Theoretical examination of pulse vaccination policy in the sir epidemic model. *Math. Comput. Modell.* 31(4-5):207–215.

[9] Stone, L., Olinky, R., Huppert, A. (2007). Seasonal dynamics of recurrent epidemics. *Nature.* 446(7135):533–536. DOI: 10.1038/nature05638.

[10] Capasso, V. (2008). *Mathematical Structure of Epidemic Systems,* vol. 97. Berlin: Springer.

[11] Mena-Lorca, J., Hethcote, H.W. (1992). Dynamic models of infectious diseases as regulators of population sizes. *J. Math. Biol.* 30(7):693–716. DOI: 10.1007/BF00173264.

[12] Brauer, F., Driessche, P. v. D., Wu, J. (2008). *Mathematical Epidemiology.* Berlin: Springer-Verlag.

[13] Ivorra, B., Ferrández, R.M., Vela-Pérez, M., Ramos, A.M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (covid-19) taking into account the undetected infections. the case of china. *Commun. Nonlinear Sci. Numer. Simul.* 88: 105303–105304. DOI: 10.1016/j.cnsns.2020.105303.

[14] Caccavo, D. Chinese and Italian covid-19 outbreaks can be correctly described by a modified sir model. *medRxiv.* (2020).

[15] Colucci, R., Caraballo, T. (2017). A comparison between random and stochastic modeling for a sir model. *Commun. Pure Appl. Anal.* 16(1):151–162. 11

[16] Changguo, L., Yongzhen, P., Meixia, Z., Yue, D. (2018). Parameter estimation on a stochastic sir model with media coverage. *Discrete Dyn. Nat. Soc.* 2018:1–7. 06

[17] Plank, M. J., Binny, R. N., Hendy, S. C., Lustig, A., James, A., Steyn, N. A stochastic model for covid-19 spread and the effects of alert level 4 in aotearoa New Zealand. *medRxiv,* (2020).

[18] Platen, E. Stochastic modelling of the covid-19 epidemic. *Preprint. SSRN,* (2020).

[19] Yanev, N., Stoimenova, V., Atanasov, D. (2020). Stochastic modeling and estimation of covid-19 population dynamics. Preprint. Arxiv. 04.

[20] Gloter, A., Hoffmann, M. (2007). Estimation of the hurst parameter from discrete noisy data. *Ann. Statist.* 35(5):1947–1974.