Hainaut, Donatien

Article

An actuarial approach for modeling pandemic risk

Risks

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Hainaut, Donatien (2021) : An actuarial approach for modeling pandemic risk, Risks, ISSN 2227-9091, MDPI, Basel, Vol. 9, Iss. 1, pp. 1-28, https://doi.org/10.3390/risks9010003

This Version is available at:
https://hdl.handle.net/10419/258093

risks

MDPI

Article

An Actuarial Approach for Modeling Pandemic Risk

Donatien Hainaut †

LIDAM Institute of Statistics, Biostatistics and Actuarial Sciences, Université Catholique de Louvain, 5032 Isnes, Belgium; donatien.hainaut@uclouvain.be
† Current address: 20 voie du Roman Pays, 1348 Louvain-La-Neuve, Belgium.

Abstract: In this article, a model for pandemic risk and two stochastic extensions is proposed. It is designed for actuarial valuation of insurance plans providing healthcare and death benefits. The core of our approach relies on a deterministic model that is an efficient alternative to the susceptible-infected-recovered (SIR) method. This model explains the evolution of the first waves of COVID-19 in Belgium, Germany, Italy and Spain. Furthermore, it is analytically tractable for fair pure premium calculation. In a first extension, we replace the time by a gamma stochastic clock. This approach randomizes the timing of the epidemic peak. A second extension consists of adding a Brownian noise and a jump process to explain the erratic evolution of the population of confirmed cases. The jump component allows for local resurgences of the epidemic.

Keywords: SIR; epidemic risk; COVID-19; jump diffusion

1. Introduction

The recent outbreak of COVID-19 reminds us that epidemics raise not only sanitary but also financial issues. There is a clear and growing need for covering the epidemic risk and for developing analytically tractable models. In this article, we propose a new deterministic model in which the contagion rate is inversely proportional to time instead of to the susceptible population. This model presents a great analytical tractability and replicates the first wave of COVID-19 in Belgium, Germany, Italy and Spain. In this framework, we infer a closed-form expression for the fair premium rate of an insurance plan covering healthcare expenses of infected persons and providing a lump sum capital payment in case of death.

The deterministic model is next extended in two different directions. In the first one, the time scale is random and ruled by a process called subordinator. Using a gamma process as stochastic clock preserves the analytical tractability and randomizes the dynamic of the pandemic. In the second extension, the evolution of the infectious population is noised by a Brownian diffusion and a jump process.

Multiple contributions are presented in this article. The literature on mathematical modeling of epidemic is abundant, but most of the existing solutions, as compartment models, do not provide any analytical expression for the evolution of populations in each compartment. The valuation of an epidemic-linked insurance requires one to calculate integrals of infected and susceptible population sizes and is therefore computationally intensive. The three models proposed in this article do suffer from this problem and have a high level of analytical tractability for actuarial applications. Furthermore, estimating their parameters does not pose any problem, and their empirical explanatory power is comparable to the one of the susceptible-infected-recovered (SIR) approach. Finally, the jump diffusion extension allows one to simulate realistic random epidemic scenarios and to value reinsurance treaties.

We first present an overview of previous research. This is followed by the introduction of the deterministic model that is compared to the susceptible-infected-recovered (SIR) approach. Section 4 studies the valuation of an insurance plan with healthcare and death

Citation: Hainaut, Donatien. 2021. An Actuarial Approach for Modeling Pandemic Risk. *Risks* 9: 3. https://dx.doi.org/10.3390/risks9010003

Received: 23 November 2020
Accepted: 11 December 2020
Published: 23 December 2020

benefits. The model is fitted to COVID-19 datasets for Belgium, Germany, Italy and Spain in the next section. Section 6 introduces the first stochastic variant of this model, based on a random gamma clock. The valuation of reinsurance treaties is developed in the next paragraph and the time-changed model is estimated in Section 8. Sections 9 and 10 explore the features of the jump diffusion model. Finally, we propose an estimation method and fit the model to COVID-19 datasets.

2. Overview of Previous Research

Communicable diseases have always been an important part of human history as underlined by Smith (2017), who reviews their nature and proposes a brief history of pandemics. He also introduces the mathematical models used to evaluate the risk that pandemics pose to human populations. Another detailed survey of these quantitative models, including future perspective, is available in Brauer (2017).

The reference model in epidemiology is the susceptible-infected-recovered (SIR) model proposed by Kermack and McKendrick (1927). Various disease outbreaks, including the SARS epidemic of 2002–2003, the concern about a possible H5N1 influenza epidemic in 2005, the H1N1 influenza pandemic of 2009 and the Ebola outbreak of 2014 have reignited interest in epidemic models, beginning with the reformulation of the Kermack–McKendrick model by Diekmann et al. (1995). In the SIR, the population is homogeneous, whereas in reality, an epidemic spreads differently according to factors like the age or susceptibility to infection. It is then necessary to follow the secondary infections in the subpopulations separately; this is done through the next-generation matrix as explained by Diekmann et al. (1990) and Van den Driessche and Watmough (2002). Anderson and May (1979) extended the SIR model by considering the host population as a dynamic variable rather than a constant. Tchuenche et al. (2007) studied the stability of a SIR model with a time delay in the contagion dynamic. Under the assumption that all individuals are susceptible, they showed that the endemic equilibrium is stable. Similarly, Zhang and Wang (2013) studied a nonautonomous SIRS epidemic model with time delay.

The SIR model has been extended to multiple compartments with labels such as M, S, E, I and R that are often used for the epidemiological classes. The class M contains infants with passive immunity inherited at birth. After that, the infant moves to the susceptible class S. When a susceptible individual has adequate contact with an infected individual such that transmission occurs, then the susceptible individual enters the exposed class E of those in the latent period, who are infected but not yet infectious. After the latent period ends, the individual enters the class I of infectious. We refer the reader to Hethcote (2000) for a comparison of MSEIR and SEIR models for various diseases, including measles in Niger and pertussis in the United States.

Drawing conclusions from mathematical models raises the question of the origin of data and of methodological best practices. Walters et al. (2018) reviewed the literature, highlighting common approaches and good practice and identifying research gaps. They extracted information from 78 records and found that most epidemiological data come from published journal articles, population data come from a wide range of sources and travel data mainly come from statistics or surveys. Rhodes et al. (2020) traced how models can be investigated as matters of correspondence and enactment in relation to their social and policy contexts.

At the beginning of a disease outbreak, there is a small number of infectious individuals and the transmission of infection is a stochastic event depending on the pattern of contacts between members of the population. The Watson and Galton (1874) process was one of the first approaches to successfully described this pattern. An alternative consists of introducing random noise into differential equations defining each compartment. Zhang and Wang (2013) explored this alternative and studied the asymptotic behavior of SIR model with Brownian noise and a jump process. The stochastic model containing a standard Brownian motion was studied by Caraballo and Colucci (2017). Caraball and Keraani (2018) explored the features of a stochastic SIR model with a frac-

tional Brownian motion. The book by Daley and Gani (1999) contains an account of some of the more recent extensions.

In actuarial sciences, the literature on epidemics model is rather scarce. Jia and Tsui (2005) proposed and estimated a compartment model for severe acute respiratory syndrome (SARS) data. Chen and Cox (2009) employed the theory of real options and considered a regime-switching process for modeling the number of infected individuals. Feng and Garrido (2011) quantified the risk of infection with a classical epidemiological compartment model. They formulated financial arrangements between an insurer and insured using actuarial methodology and applied their framework to the SARS epidemic in 2003. Gathy and Lefèvre (2009) and Lefevre and Utev (1999) proposed extensions of deterministic compartment models and provided additional tools to account for randomness in epidemiological dynamics. Based on a Markov chain formulation of the susceptible-infected-recovered (SIR model), Lefèvre et al. (2017) developed a recursive method to calculate the cost components and the corresponding premium levels. More recently, Clara-Rahola (2020) proposed two distinct exponential models for the infection rate before and after lockdown. The fit to data from China, Spain, South Korea and Italy revealed that a crossover point between pre- and postlockdown infection rates is found one week after lockdown, which, in turn, is the average COVID-19 incubation period.

3. A Deterministic Epidemic Model

We consider a population of size S_0 hit by an epidemic disease at time 0. We propose to model the number of infectious persons at time $t \geq 0$, denoted by I_t, by the following relation:

$$I_t = S_0 e^{-(\alpha+\mu)t}(\beta t)^{\gamma} \ , \ t \geq t_0, \tag{1}$$

where $\alpha, \beta, \mu, \gamma \in \mathbb{R}^+$. This function is the product of two terms. The first one is an exponential decreasing function, $e^{-(\alpha+\mu)t}$, whereas the second one is an increasing function. The number of infectious decreases exponentially with a rate $(\alpha + \mu)$. The parameter α is the recovery rate from the disease, whereas μ is assumed to be the death rate of infected persons. The average duration before recovery is then $1/\alpha$. At time 0, the initial number of infected individuals is equal to $I_0 = 0$. In order to understand the role played by the parameter β in the dynamics of infectious population, we differentiate Equation (1):

$$\begin{aligned} dI_t &= -(\alpha+\mu)I_t dt + S_0 e^{-(\alpha+\mu)t}\beta^{\gamma}\left(\gamma\, t^{\gamma-1}\right)dt \\ &= -(\alpha+\mu)I_t dt + I_t\frac{\gamma}{t}dt\,. \end{aligned} \tag{2}$$

This differential equation reveals that the initial contagion rate per capita is equal to $\frac{\gamma}{t}$. This is a decreasing function to 0 when $t \to \infty$. Empirical tests in following sections emphasizes that Equation (2) explains the evolution of the first wave of COVID-19 in Belgium, Germany, Spain and Italy.

This model slightly differs from the susceptible-infected-recovered (SIR) model developed by Kermack and McKendrick (1927), which is a standard in the literature. In the SIR model, the evolution of the epidemics is described by the ordinary differential equations (ODEs):

$$\begin{aligned} dI_t^{SIR} &= -(\alpha_{SIR}+\mu_{SIR})I_t^{SIR}dt + \frac{\beta_{SIR}S_t^{SIR}}{S_0}I_t^{SIR}dt\,, \\ dS_t^{SIR} &= -\frac{\beta_{SIR}S_t^{SIR}}{S_0}I_t^{SIR}dt\,, \end{aligned} \tag{3}$$

where $\beta_{SIR} \in \mathbb{R}^+$ and S_t^{SIR} is the number of persons that are susceptible to be infected at time t. As in our model, α_{SIR} and μ_{SIR} are the recovery and mortality rates, respectively. The contagion rate per capita in the SIR is proportional to the population of susceptible,

$\frac{\beta_{SIR} S_t^{SIR}}{S_0}$, whereas it is a function of time, $\frac{\gamma}{t}$, in the approach proposed in this article. This assumption allows us to obtain the closed form expression (2) for the infectious populations. This is the greatest benefit of this model compared to the SIR that does not admit analytical solutions for the system (3). In actuarial applications developed in following sections, we have to integrate I_t. Therefore, having an analytical formula allows us to avoid numerical integration of an ODE numerical solution and propagation of numerical errors. Furthermore, the model (2) is easily extended to stochastic frameworks, as detailed in following sections.

The basic reproduction number, R_0, is defined as the average number of secondary cases arising from a typical primary case. Under the assumption that the population of susceptible is large, we have $\frac{S_t^{SIR}}{S_0} \approx 1$ and the basic reproduction number of the SIR model is $R_0 = \frac{\beta}{\alpha+\mu}$. In our model, the reproduction number is instead a function of time equal to $R_0(t) = \frac{\gamma}{t(\alpha+\mu)}$. The time-varying R_0 allows us to take into account the impact of preventive measures to curb the epidemic, such as a lockdown or the wearing of masks. The empirical analysis of the next section confirms that this approach offers a better fit than the SIR model. Our model presents other interesting features. In particular, the peak of the epidemic is known and obtained by canceling the first order derivative of Equation (1):

$$t_{max} = \frac{\gamma}{\alpha + \mu}\,. \tag{4}$$

Combining Equations (1) and (4) allows us to evaluate the size of the infected population when the peak is reached:

$$I_{t_{max}} = S_0 e^{-\gamma} \left(\frac{\gamma}{\alpha + \mu} \right)^{\gamma} \beta^{\gamma}\,. \tag{5}$$

Since the population of infectious persons may not exceed S_0, we infer the necessary conditions $\beta^{\gamma} \leq \frac{e^{\gamma}}{(t_{max})^{\gamma}}$. As μ is the mortality rate, the total number of deaths up to time t is a function, denoted by D_t, that is solution of the ordinary differential equation:

$$dD_t \quad = \quad \mu I_t dt\,. \tag{6}$$

Under the assumptions that recovery does not not provide a protective immunity and that there is no entry into or departure from the population, the size of the population of susceptible, denoted by S_t, is then solution of the following ODE:

$$dS_t \quad = \quad \alpha I_t dt - I_t \frac{\gamma}{t} dt \quad , \; t > t_0. \tag{7}$$

By construction, the population of susceptible growths when infected individuals recover from the disease and is decreased by the number of new contaminated. As we do not consider new entrants in the population, the sum of the number of susceptible individuals, infected individuals and deaths remains constant and equal to S_0:

$$S_t + I_t + D_t = S_0 \quad , \; t \geq 0. \tag{8}$$

4. Actuarial Valuation of an Insurance Plan

In the same spirit as Feng and Garrido (2011), we consider an infectious disease insurance plan that collects premiums in the form of continuous annuities from susceptibles, as long as they are healthy. The premium rate is assumed constant and noted p. Collected premiums cover medical expenses which are continuously reimbursed for each infected policyholder during the period of treatment. The benefit rate is noted b. The plan terminates when the individual recovers or dies from the disease. In case of death, a lump sum benefit, c, is paid.

The risk-free rate is constant and denoted by r. If the insurance plan covers the whole population, premium, benefit rates and lump sum death capital must ensure the financial equilibrium of the plan. Under the assumption that the plan starts at time 0 and finishes at time T, discounted premiums have to cover all discounted benefits:

$$p \int_0^T e^{-rs} S_s \, ds \quad = \quad b \int_0^T e^{-rs} I_s \, ds + c \int_0^T e^{-rs} \, dD_s \, . \tag{9}$$

As stated in the previous section, the discounted integral of I_t admits a closed-form expression.

Proposition 1. *For $r \geq 0$, we have that:*

$$\int_0^T e^{-rs} I_s ds = \frac{S_0 \beta^\gamma}{\theta^{\gamma+1}} \Gamma_l(\gamma + 1, T\theta) \, , \tag{10}$$

where $\theta = r + \alpha + \mu$ and $\Gamma_l(\gamma + 1, x) = \int_0^x u^\gamma e^{-u} du$ is the gamma lower incomplete function.

Proof. Using Equation (1), we develop the integral as follows:

$$\int_0^T e^{-rs} I_s ds \quad = \quad S_0 \beta^\gamma \int_0^T e^{-(r+\alpha+\mu)s} s^\gamma ds \, ,$$

and we perform a change of variable $u = \theta s$ in order to rewrite the integral:

$$\int_0^T e^{-rs} I_s ds \quad = \quad S_0 \beta^\gamma \theta^{-\gamma-1} \int_0^{\theta T} e^{-u} u^\gamma du \, .$$

From the definition of the gamma incomplete function, we directly obtain Equation (10). □

From this last proposition, we immediately infer the expressions of D_t, $\int_0^T e^{-rs} \, dD_s$ and S_t.

Corollary 1. *The cumulated number of deceases caused by the epidemic at time $t \geq 0$ is equal to*

$$D_t \quad = \quad \mu S_0 \beta^\gamma (\alpha + \mu)^{-\gamma-1} \Gamma_l(\gamma + 1, t(\alpha + \mu)) \, . \tag{11}$$

If $\theta = r + \alpha + \mu$, the second term on the right-hand side of Equation (9) is:

$$\int_0^T e^{-rs} \, dD_s \quad = \quad \mu S_0 \beta^\gamma \theta^{-\gamma-1} \Gamma_l(\gamma + 1, T\theta) \, . \tag{12}$$

The size of the population of susceptible at time $t \geq 0$ is deduced from the relation $S_t + I_t + D_t = S_0$:

$$S_t = S_0 - S_0 e^{-(\alpha+\mu)t} (\beta t)^\gamma - \mu S_0 \beta^\gamma (\alpha + \mu)^{-\gamma-1} \Gamma_l(\gamma + 1, t(\alpha + \mu)) \, . \tag{13}$$

The next proposition reports the closed-form expression for the premium rate solution of Equation (9).

Proposition 2. *For the benefit rate (b, c), the fair premium rate that ensures the actuarial equilibrium of the plan is given by*

$$p \quad = \quad \frac{(b + c\mu) S_0 \beta^\gamma \theta^{-\gamma-1} \Gamma_l(\gamma + 1, T\theta)}{\int_0^T e^{-rs} S_s \, ds} \, , \tag{14}$$

where the denominator is equal to:

$$\begin{aligned}\int_0^T e^{-rs} S_s \, ds = \frac{S_0}{r}\left(1 - e^{-rT}\right) - \frac{S_0 \beta^\gamma}{\theta^{\gamma+1}} \Gamma_l(\gamma+1, T\theta)\left(1 + \frac{\mu}{r}\right) \\ + \frac{\mu S_0 \beta^\gamma}{r(\alpha+\mu)^{\gamma+1}} e^{-rT} \Gamma_l(\gamma+1, T(\alpha+\mu)) \,.\end{aligned} \tag{15}$$

Proof. From Equation (13), we infer that

$$\int_0^T e^{-rs} S_s \, ds = S_0 \int_0^T e^{-rs} \, ds - \int_0^T e^{-rs} I_s \, ds - \int_0^T e^{-rs} D_s \, ds$$

The first integral on the right-hand side is equal to $\frac{S_0}{r}\left(1 - e^{-rT}\right)$ whereas $\int_0^T e^{-rs} I_s \, ds$ is provided by Equation (10). From Equation (11), we infer that the integral of the discounted number of deaths is:

$$\int_0^T e^{-rs} D_s \, ds = \frac{\mu S_0 \beta^\gamma}{(\alpha+\mu)^{\gamma+1}} \int_0^T e^{-rs} \Gamma_l(\gamma+1, s(\alpha+\mu)) \, ds.$$

Integrating by parts leads to the following result:

$$\begin{aligned}\int_0^T \frac{e^{-rs}}{(\alpha+\mu)^{\gamma+1}} \Gamma_l(\gamma+1, s(\alpha+\mu)) \, ds = \left[-\frac{e^{-rs}}{r\,(\alpha+\mu)^{\gamma+1}} \Gamma_l(\gamma+1, s(\alpha+\mu)) \right]_{s=0}^{s=T} \\ - \int_0^T \frac{e^{-rs}}{r} \frac{d}{ds}\left(\int_0^s e^{-(\alpha+\mu)v} v^\gamma dv \right) ds \,.\end{aligned}$$

The integral in this last expression may also be reformulated in terms of an incomplete lower gamma function:

$$\begin{aligned}\int_0^T \frac{e^{-rs}}{r} \frac{d}{ds}\left(\int_0^s e^{-(\alpha+\mu)v} v^\gamma dv \right) ds &= \frac{1}{r} \int_0^T e^{-(r+\alpha+\mu)s} s^\gamma \, ds \\ &= \frac{1}{r\theta^{\gamma+1}} \Gamma_l(\gamma+1, T\theta) \,.\end{aligned}$$

Combining these results leads to Equation (15) and the fair premium comes from the actuarial equilibrium equation. □

5. Empirical Illustration

We fit the model to data about the COVID-19 outbreak in Belgium, Germany, Italy and Spain. The first three countries are selected because they have reported the highest death rates in Europe during the 2020 first wave of COVID-19. In comparison, Germany has better managed the spread of the virus but the distribution of infected individuals over time has decreased at a lower pace than other countries considered in this study. We use the datasets from the library "coronavirus[1]" in R which provides daily time-series of the number of deaths and detected cases of COVID from the beginning of 2020 up to the end of July. We choose as starting date the day when the number of confirmed cases passes above a threshold set to 0.005% of the total population of the country. As the model is designed for modeling a single epidemic wave, the ending date is set to the 15th of June 2020, which corresponds to the end of the lockdown period in the considered countries. Figure 1 shows these time series and Table 1 reports some statistics of the datasets. For Spain, the number of confirmed cases or deaths is negative for a few days. This is due to retrospective corrections.

[1] See, https://github.com/RamiKrispin/coronavirus, package developed by Rami Krispin.

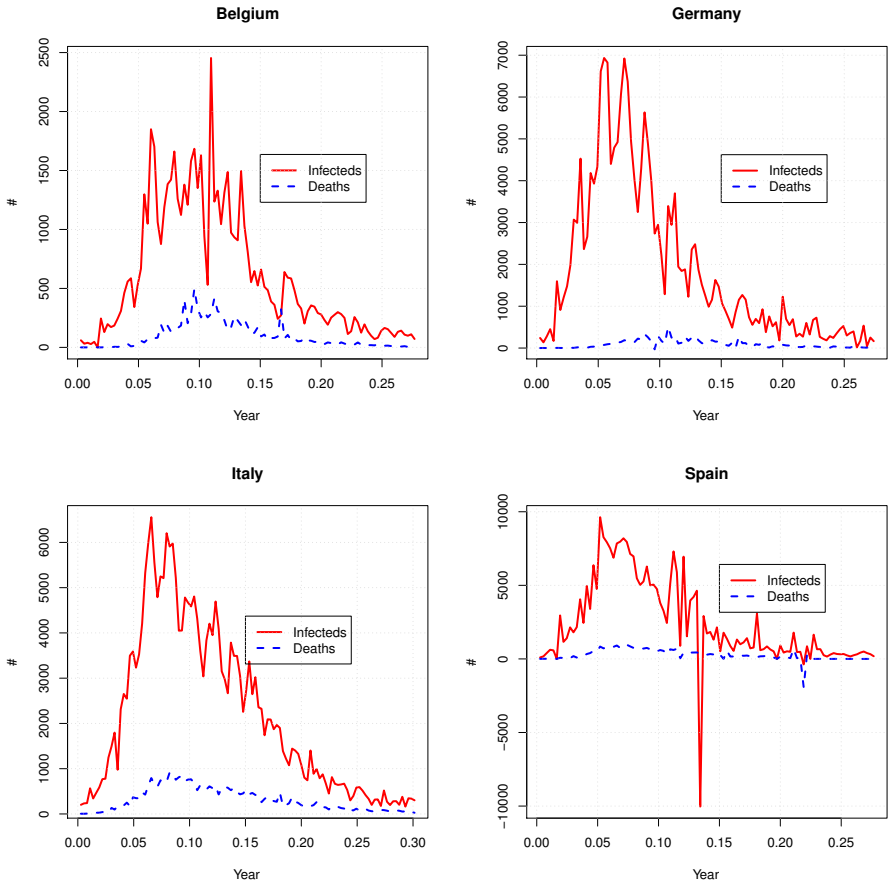

Figure 1. Graphs of the number of COVID-19 reported cases and deaths from March to June, for Belgium, Germany, Italy and Spain.

Table 1. Statistics about the time series of deaths and confirmed COVID cases.

Country	Starting Date	Number of Days	Number of Infected Individuals	Number of Deaths	Population Size, S_0
Belgium	7/3/2020	101	59,991	9661	11,589,623
Germany	8/3/2020	100	186,883	8807	83,770,952
Italy	27/2/2020	110	236,837	34,359	60,461,826
Spain	7/3/2020	101	243,709	27,131	46,934,632

Both the SIR and our model aim to describe the evolution of the number of infected persons, I_t. As the data sets only report new confirmed cases, we assume that contaminated individuals remain infected, on average, for 12 days, which is slightly less than the duration of the quarantine imposed, e.g., in Belgium (14 days) after being in touch with a contaminated person. If $\left(I_t^{obs}\right)_{t=1,\dots,n_{obs}}$ is the time series of observations and n_{obs} is the number of days, parameters are obtained by a weighted least-square minimization:

$$(\widehat{\alpha+\mu},\widehat{\gamma},\widehat{\beta}) \quad = \quad \arg\min \frac{\sum_{k=1}^{n_{obs}} \omega_k \left(I_k - I_k^{obs}\right)^2}{\sum_{k=1}^{n_{obs}} \omega_k}, \tag{16}$$

where I_k is the value of I_t at time t_k. As the impacts of α and μ on I_t are indistinguishable, we first estimate their sum. The annualized mortality rate is estimated as the ratio of the total number of deaths on the cumulated number of infecteds forecast by the model, multiplied by 365. Given that the COVID testing was far from being generalized in March,

it is likely that the number of infected cases was higher than the one reported. In order to take this into account in the estimation procedure, more importance is granted to most recent daily observations as follows:

$$\begin{cases} \omega_k = 0.05 & k \in \{1, ..., 25\}, \\ \omega_k = 0.5 & k \in \{36, ..., 73\}, \\ \omega_k = 1.5 & k \geq 74. \end{cases} \tag{17}$$

These weights are chosen in order to obtain the best compromise fitting both the tail and the peaks of the I_t curve. The results of the calibration procedure are reported in Table 2.

Table 2. Parameter estimates for the model ruled by Equation (1) (unit of time t: year). The SSE is the value of the optimization criterion in Equation (16).

	$\widehat{\alpha}$	$\widehat{\gamma}$	$\widehat{\beta}$	$\widehat{\mu}$	t_{max}(days)	I_{max}	SSE
Belgium	40.718	4.74	6.606	4.457	38	17,829	535,988
Germany	40.633	3.124	3.693	1.239	27	65,665	12,513,688
Italy	30.878	3.382	3.709	3.931	35	65,103	3,093,842
Spain	46.631	3.937	6.979	2.966	29	89,479	24,203,834

We benchmark the capacity of Equation (1) to model the evolution of the infected population with the SIR model, fitted by least-square minimization. Our numerical experiments reveals that the SIR model fails to replicate the curve of I_t. The only way to fit this model consists to consider that S_0 is also a parameter. Parameter estimates are reported in Table 3. Figure 2 compares observed to forecast I_t with our and SIR models. The SIR offers at a first sight an excellent fit but considering that S_0 is adjustable is hard to justify. Furthermore, the adjusted S_0 are considerably smaller than the real size of considered populations. This confirms that our approach is a reliable alternative compared to the SIR model.

Table 3. Parameter estimates of the SIR model.

	Adjusted $\widehat{S}_0$	$\widehat{\alpha}_{SIR}$	$\widehat{\beta}_{SIR}$	$\widehat{\mu}$	SSE
Belgium	38,203	12.407	107.76	4.457	16,109
Germany	128,250	16.816	131.649	1.239	1,372,666
Italy	118,250	10.616	113.425	3.931	1,175,335
Spain	171,750	15.914	146.516	2.966	2,743,222

Next, we use parameter estimates in order to evaluate the fair premium rates of an insurance plan, such as described in Section 4. Two cases are considered. In the first one, collected premiums cover exclusively medical expenses. An allowance of 1000 EUR per day is paid during the treatment (c = 365,000 EUR on a yearly basis). The second plan covers exclusively the death risk: a lump sum capital of 200,000 EUR is paid at the decease of an infected patient. The duration of both plans is six months and the risk-free rate is set to 2%. Tables 4 and 5 report the fair premium rates calculated with our approach and the SIR model, respectively, for Belgium, Germany, Italy and Spain. We also test the sensitivity of these rates to variations of parameters. Per country, premium rates computed with our approach or the SIR model are similar. The premium rates for both benefits (62.38 EUR and 42.35 EUR per year) for Germany are the lowest due to the low number of confirmed cases and deaths reported by this country. The death coverage is the most expensive for Belgium (338 EUR/year), whereas Italy and Spain are in the same range (230.52 EUR and 232.57 EUR). The healthcare benefit is most expensive in Spain and Belgium (143.1 EUR/year and 138.54 EUR/year).

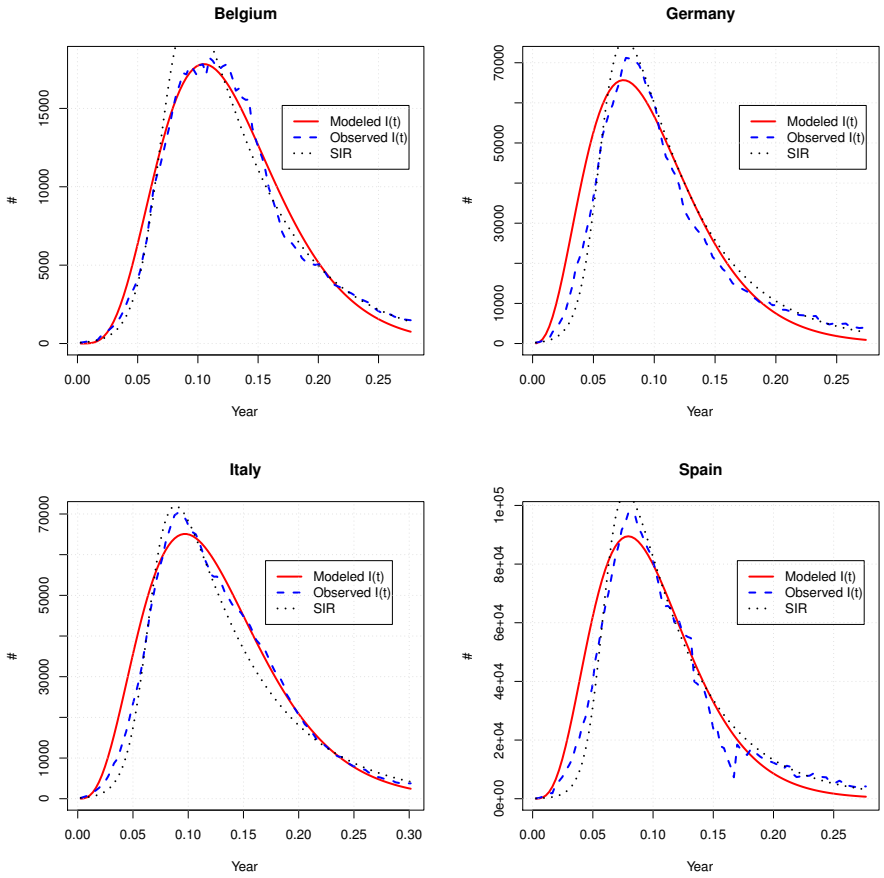

Figure 2. Comparison of fitted I_t with observed ones and those computed with the SIR model.

Table 4. Fair premium rates with the model in (1). Duration $T = 0.5$ year and $r = 2\%$.

	b	c	Fair p	$\beta+$ 1%	$\beta-$ 1%	$\alpha+$ 1%	$\alpha-$ 1%	$\gamma+$ 1%	$\gamma-$ 1%
Be	365,000	0	138.54	145.23	132.09	131.58	145.93	136.83	140.32
	0	200,000	338.35	354.7	322.59	321.36	356.41	334.19	342.71
Ge	365,000	0	62.38	64.35	60.45	59.95	64.94	60.21	64.65
	0	200,000	42.35	43.69	41.04	40.70	44.09	40.88	43.89
It	365,000	0	107.02	110.68	103.44	102.95	111.28	103.89	110.27
	0	200,000	230.52	238.42	222.81	221.77	239.7	223.78	237.53
Sp	365,000	0	143.1	148.82	137.54	136.64	149.94	140.5	145.8
	0	200,000	232.57	241.87	223.54	222.07	243.68	228.35	236.95

Table 5. Fair premium rates with the model (1). Duration $T = 0.5$ year and $r = 2\%$.

	b	c	Fair p, SIR	$\beta_{SIR}+1\%$	$\beta_{SIR}-1\%$	$\alpha_{SIR}+1\%$	$\alpha_{SIR}-1\%$
Be	365,000	0	142.43	142.45	142.4	141.38	143.49
	0	200,000	347.84	347.89	347.79	345.29	350.43
Ge	365,000	0	61.86	61.86	61.85	61.28	62.44
	0	200,000	42.00	42.00	41.99	41.61	42.39
It	365,000	0	97.95	97.95	97.94	97.24	98.66
	0	200,000	210.98	211.00	210.97	209.47	212.52
Sp	365,000	0	141.51	141.52	141.51	140.33	142.72
	0	200,000	229.99	230	229.98	228.07	231.95

6. Time-Changed Extension

The model introduced in Section 3 is fully deterministic. In practice, we observe random fluctuations of the number of infected persons. In order to replicate such random variations, we propose two stochastic extensions of Equation (1). The first one developed in this section consists of replacing the time t by a stochastic clock, also called a subordinator. This clock is an increasing positive process denoted by $(\tau_t)_{t\geq 0}$, defined on a probability space Ω endowed with a probability measure P and its natural filtration $(\mathcal{F}_t)_{t\geq 0}$. We consider that τ_t is a gamma process, i.e., τ_t is gamma-distributed with expectation and variance equal to λt. The probability density function of τ_t is given by

$$f_{\tau_t}(x) \quad = \quad \mathbf{1}_{x>0}\frac{x^{\lambda t-1}e^{-x}}{\Gamma(\lambda t)}\,, \tag{18}$$

where $\Gamma(a)=\int_0^\infty x^{a-1}e^{-x}dx$ is the standard gamma function such that $\Gamma(a+1)=a\Gamma(a)$ for $a>0$. A straightforward calculation shows that the characteristic function for the gamma process is given by

$$\begin{aligned}\mathbb{E}\left(e^{iu\tau_t}\right) &= (1-iu)^{-\lambda t}\\ &= e^{-t\psi(u)}\,,\end{aligned}$$

where $\psi(u)=\lambda\ln(1-iu)$ is the characteristic exponent and $u\in\mathbb{R}$. This Lévy–Khinchine representation of the characteristic function reveals that τ_t is also a Lévy process. Indeed, $\psi(u)$ may be rewritten as the following integral:

$$\psi(u) \quad = \quad \int_0^\infty \lambda x^{-1}e^{-x}(1-e^{iux})dx\,,$$

from which we infer that the Lévy measure of τ_t is $\nu(dx)=\lambda x^{-1}e^{-x}1_{(x\geq 0)}dx$. τ_t is a process with finite variations. Therefore, for any function $f(t,\tau_t)$ of time and of the subordinator, Itô's lemma for semimartingales states that

$$df(t,\tau_t)=\frac{\partial f(t,\tau_t)}{\partial t}dt+f(t,\tau_{t-}+\Delta\tau_t)-f(\tau_{t-}))\,.$$

Whereas the $\mathcal{F}_t-$expectation of its infinitesimal variation is

$$\mathbb{E}(df(t,\tau_t)|\mathcal{F}_t)=\frac{\partial f(t,\tau_t)}{\partial t}dt+\int_0^\infty (f(t\,,\,\tau_{t-}+x)-f(\tau_{t-}))\,\lambda\,x^{-1}e^{-x}dx\,dt\,.$$

The time-changed version of the deterministic model is obtained by replacing t with the chronometer τ_t. The dynamics of the population of infectious individuals is then:

$$I_t=S_0e^{-(\alpha+\mu)\tau_t}(\beta\tau_t)^\gamma\;t\geq 0, \tag{19}$$

where $\alpha,\beta,\mu,\gamma\in\mathbb{R}^+$. Using Itô's lemma and first-order Taylor developments, we infer that:

$$\begin{aligned}dI_t &= S_0e^{-(\alpha+\mu)(\tau_{t-}+\Delta\tau_t)}(\beta(\tau_{t-}+\Delta\tau_t))^\gamma-S_0e^{-(\alpha+\mu)\tau_{t-}}(\beta\tau_{t-})^\gamma\\ &= I_{t-}\left(-(\alpha+\mu)\Delta\tau_t+\tfrac{\gamma}{\tau_{t-}}\Delta\tau_t\right)+\mathcal{O}\left((\Delta\tau_t)^2\right).\end{aligned} \tag{20}$$

This first-order approximation emphasizes the strong relation between the deterministic and the time-changed dynamics. Parameters α and μ may still be interpreted as recovery and death rates, but over a random time interval of size Δ_t. The contagion rate per capita at time t is equal to $\frac{\gamma}{\tau_{t-}}$ for a period Δ_t. The next proposition gives the first two moments of $(I_t)_{t\geq 0}$.

Proposition 3. *The expected number of infected persons at time t is equal to:*

$$\mathbb{E}(I_t \,|\, \mathcal{F}_0) \;=\; \frac{S_0 \beta^\gamma \Gamma(\gamma + \lambda t)}{(\alpha + \mu + 1)^{\gamma + \lambda t} \Gamma(\lambda t)}\,, \tag{21}$$

whereas its variance is given by the following relation:

$$\mathbb{V}(I_t \,|\, \mathcal{F}_0) \;=\; S_0^2 \beta^{2\gamma} \left(\frac{\Gamma(2\gamma + \lambda t)}{(2\alpha + 2\mu + 1)^{2\gamma + \lambda t} \Gamma(\lambda t)} - \left(\frac{\Gamma(\gamma + \lambda t)}{(\alpha + \mu + 1)^{\gamma + \lambda t} \Gamma(\lambda t)} \right)^2 \right). \tag{22}$$

Proof. The expectation of I_t is rewritten in its integral form:

$$\begin{aligned} \mathbb{E}(I_t \,|\, \mathcal{F}_0) &= S_0 \int_0^\infty \frac{u^{\lambda t - 1} e^{-u}}{\Gamma(\lambda t)} e^{-(\alpha+\mu)u} \, (\beta u)^\gamma \, du \\ &= \frac{S_0 \beta^\gamma}{\Gamma(\lambda t)} \int_0^\infty u^{\gamma + \lambda t - 1} e^{-(\alpha + \mu + 1)u} \, du\,. \end{aligned}$$

Next, we do a change of variable $v = (\alpha + \mu + 1)u$ in order to obtain Equation (21). We obtain the moments of second order in a similar manner:

$$\begin{aligned} \mathbb{E}\left(I_t^2 \,|\, \mathcal{F}_0\right) &= S_0^2 \int_0^\infty \frac{u^{\lambda t - 1} e^{-u}}{\Gamma(\lambda t)} e^{-2(\alpha+\mu)u} \, (\beta u)^{2\gamma} \, du \\ &= \frac{S_0^2 \beta^{2\gamma}}{\Gamma(\lambda t)} \int_0^\infty u^{2\gamma + \lambda t - 1} e^{-(2\alpha + 2\mu + 1)u} \, du \\ &= \frac{S_0^2 \beta^{2\gamma} \Gamma(2\gamma + \lambda t)}{(2\alpha + 2\mu + 1)^{2\gamma + \lambda t} \Gamma(\lambda t)}\,. \end{aligned}$$

Equation (22) is the difference of this second moment and of the square of the expectation. □

Notice that the maximum of the epidemics is reached at t_{max} with $\tau_{t_{max}} = \frac{\gamma}{\alpha + u}$. The cumulated number of deaths is the time-changed version of Equation (11):

$$\begin{aligned} D_t &= \mu S_0 \beta^\gamma (\alpha + \mu)^{-\gamma - 1} \Gamma_l(\gamma + 1, \tau_t(\alpha + \mu)) \\ &= \mu S_0 \beta^\gamma (\alpha + \mu)^{-\gamma - 1} \int_0^{\tau_t} (\alpha + \mu) u^\gamma e^{-u} du \\ &= \mu S_0 \beta^\gamma \int_0^{\tau_t} v^\gamma e^{-(\alpha + \mu)v} dv\,. \end{aligned} \tag{23}$$

We use Itô's lemma and first-order Taylor developments to check that the dynamics of D_t is compliant with the one of I_t:

$$\begin{aligned} dD_t &= \mu S_0 \beta^\gamma \int_{\tau_{t-}}^{\tau_{t-} + \Delta\tau_t} v^\gamma e^{-(\alpha+\mu)v} dv \\ &\approx \mu \underbrace{S_0 (\beta\, \tau_{t-})^\gamma e^{-(\alpha+\mu)\tau_{t-}}}_{I_{t-}} \Delta\tau_t + \mathcal{O}\Big((\Delta\tau_t)^2\Big)\,. \end{aligned} \tag{24}$$

which is the stochastic equivalent equation to (6). This equation confirms that the infinitesimal variation of D_t is a fraction $\mu\Delta\tau_t$ of the population of infecteds. Unfortunately, the

expectation and variance of the number of deaths only admit a semiclosed form expression and their valuation requires numerical integration:

$$\begin{aligned}\mathbb{E}(D_t \mid \mathcal{F}_0) &= \mu S_0 \beta^\gamma \int_0^\infty \frac{x^{\lambda t-1} e^{-x}}{\Gamma(\lambda t)} \int_0^x v^\gamma e^{-(\alpha+\mu)v} dv\, dx \\ &= \frac{\mu S_0 \beta^\gamma}{\Gamma(\lambda t)(\alpha+\mu)^{\gamma+1}} \int_0^\infty x^{\lambda t-1} e^{-x} \int_0^{(\alpha+\mu)x} u^\gamma e^{-u} du\, dx \\ &= \frac{\mu S_0 \beta^\gamma}{\Gamma(\lambda t)(\alpha+\mu)^{\gamma+1}} \int_0^\infty x^{\lambda t-1} e^{-x} \Gamma_l(\gamma+1, x(\alpha+\mu))\, dx\,.\end{aligned} \tag{25}$$

and

$$\begin{aligned}\mathbb{E}(D_t^2 \mid \mathcal{F}_0) &= (\mu S_0 \beta^\gamma)^2 \int_0^\infty \frac{x^{\lambda t-1} e^{-x}}{\Gamma(\lambda t)} \left(\int_0^x v^\gamma e^{-(\alpha+\mu)v} dv \right)^2 dx \\ &= \frac{(\mu S_0 \beta^\gamma)^2}{\Gamma(\lambda t)(\alpha+\mu)^{2\gamma+2}} \int_0^\infty x^{\lambda t-1} e^{-x} \left(\int_0^{(\alpha+\mu)x} u^\gamma e^{-u} du \right)^2 dx \\ &= \frac{(\mu S_0 \beta^\gamma)^2}{\Gamma(\lambda t)(\alpha+\mu)^{2\gamma+2}} \int_0^\infty x^{\lambda t-1} e^{-x} (\Gamma_l(\gamma+1, x(\alpha+\mu)))^2\, dx\,.\end{aligned} \tag{26}$$

The variance of the cumulated number of deaths up to time t is therefore:

$$\begin{aligned}\mathbb{V}(D_t \mid \mathcal{F}_0) = \frac{(\mu S_0 \beta^\gamma)^2}{\Gamma(\lambda t)(\alpha+\mu)^{2\gamma+2}} \Bigg[&\int_0^\infty x^{\lambda t-1} e^{-x} (\Gamma_l(\gamma+1, x(\alpha+\mu)))^2\, dx \\ &- \left(\int_0^\infty x^{\lambda t-1} e^{-x} \Gamma_l(\gamma+1, x(\alpha+\mu))\, dx \right)^2 \Bigg]\,.\end{aligned} \tag{27}$$

The size of the population of susceptible at time $t \geq 0$ is deduced from the relation $S_t + I_t + D_t = S_0$ and is given by the following expression:

$$\begin{aligned}S_t = S_0 - S_0 e^{-(\alpha+\mu)\tau_t} (\beta\, \tau_t)^\gamma - \\ \mu S_0 \beta^\gamma (\alpha+\mu)^{-\gamma-1} \Gamma_l(\gamma+1, \tau_t\, (\alpha+\mu))\,.\end{aligned}$$

The expected size of the population of susceptible is simply equal to $\mathbb{E}(S_t \mid \mathcal{F}_0) = S_0 - \mathbb{E}(D_t \mid \mathcal{F}_0) - \mathbb{E}(I_t \mid \mathcal{F}_0)$ where $\mathbb{E}(I_t \mid \mathcal{F}_0)$ and $\mathbb{E}(D_t \mid \mathcal{F}_0)$ are respectively provided by Equations (21) and (25).

If we consider the insurance plan introduced in Section 4, the fair premium rate that finances expected benefits is such that

$$p = \frac{b \int_0^T e^{-rs} \mathbb{E}(I_s \mid \mathcal{F}_0)\, ds + c \int_0^T e^{-rs}\, \mathbb{E}(dD_s \mid \mathcal{F}_0)}{\int_0^T e^{-rs} \mathbb{E}(S_s \mid \mathcal{F}_0)\; ds}\,. \tag{28}$$

Contrary to the deterministic model, the integrals present in this last expression do not admit a closed-form expression but can easily be approached by a sum over a partition of the interval $[0, T]$. If we consider a partition $\{s_0 = 0, s_1, ..., s_m = T\}$ of equispaced times and if we note by Δ_m the length of interarrival times, the integrals are computed as:

$$\int_0^T e^{-rs}\mathbb{E}(S_s \mid \mathcal{F}_0)\ ds \approx \sum_{i=1}^{m} e^{-rs_i}\mathbb{E}(S_{s_i}|\mathcal{F}_0)\Delta_m\,,$$
$$\int_0^T e^{-rs}\,\mathbb{E}(dD_s \mid \mathcal{F}_0) \approx \sum_{i=1}^{m} e^{-rs_i}\big(\mathbb{E}(D_{s_i}|\mathcal{F}_0) - \mathbb{E}\big(D_{s_{i-1}}|\mathcal{F}_0\big)\big)\,,$$
$$\int_0^T e^{-rs}\mathbb{E}(I_s \mid \mathcal{F}_0)\ ds \approx \sum_{i=1}^{m} e^{-rs_i}\mathbb{E}(I_{s_i}|\mathcal{F}_0)\Delta_m\,.$$

In the next Section, we present a different stochastic extension of our deterministic model that leads to analytical expressions for these integrals.

7. Reinsurance in the Time-Changed Model

The introduction of randomness in the dynamic of the epidemic allows us to price various reinsurance coverages. As illustration, we consider excess-of-loss contracts providing a compensation if the number of infected or the number of deaths exceeds a certain threshold. The following proposition provides an analytical expression for a reinsurance treaty that plans the payment of an amount $C\,(I_t - K)$ at time t, where $C, K \in \mathbb{R}^+$, if $I_t > K$. This contract is similar to a financial option with the underlying being the size of the infected populations. The risk-free rate is noted r and the model parameters for I_t and τ_t are assumed the same under the pricing and real measures[2].

Proposition 4. *The value of an excess-of-loss reinsurance covering an excessive number of infectious is equal to:*

$$\begin{aligned} &Ce^{-rt}\mathbb{E}\big((I_t - K)_+ \mid \mathcal{F}_0\big) = \\ &\qquad Ce^{-rt}\left(\frac{S_0\beta^\gamma\Gamma_u(\lambda t+\gamma\,,\,(\alpha+\mu+1)u_K)}{\Gamma(\lambda t)(\alpha+\mu+1)^{\lambda t+\gamma}} - K\frac{\Gamma_u(\lambda t, u_K)}{\Gamma(\lambda t)}\right), \end{aligned} \tag{29}$$

where u_K is the positive solution of the following equation

$$\ln S_0 - (\alpha+\mu)u + \gamma\ln\beta + \gamma\ln u = \ln K\,, \tag{30}$$

and $\Gamma_u(\gamma+1, x) = \int_x^\infty u^\gamma e^{-u}du$ is the upper gamma incomplete function.

Proof. We insert the definition of I_t and rewrite the expectation as an integral with respect to the density of τ_t:

$$Ce^{-rt}\mathbb{E}\big((I_t - K)_+ \mid \mathcal{F}_0\big) \quad = \quad Ce^{-rt}\int_0^\infty \frac{u^{\lambda t-1}e^{-u}}{\Gamma(\lambda t)}\Big(S_0e^{-(\alpha+\mu)u}(\beta u)^\gamma - K\Big)_+ du\,.$$

The integrand is positive if and only if u is above u_K, such as it is defined in Equation (29). This allows us to rewrite the integral as a difference

$$\begin{aligned} &\int_0^\infty \frac{u^{\lambda t-1}e^{-u}}{\Gamma(\lambda t)}\Big(S_0e^{-(\alpha+\mu)u}(\beta u)^\gamma - K\Big)_+ du = \\ &\qquad \frac{S_0\beta^\gamma}{\Gamma(\lambda t)}\int_{u_K}^\infty u^{\lambda t+\gamma-1}e^{-(\alpha+\mu+1)u}\,du - K\int_{u_K}^\infty \frac{u^{\lambda t-1}e^{-u}}{\Gamma(\lambda t)}\,du\,, \end{aligned}$$

[2] In practice, the reinsurer would use a set of parameters more conservative than the estimated one in order to include a safety margin.

where the second term is the ratio $\frac{K\Gamma_u(\lambda t,u_K)}{\Gamma(\lambda t)}$. We replace the integrating variable by v that is $(\alpha+\mu+1)u = v$ and infer that the first term is:

$$\begin{aligned}\int_{u_K}^{\infty} u^{\lambda t+\gamma-1}e^{-(\alpha+\mu+1)u}\,du &= \frac{\int_{(\alpha+\mu+1)u_K}^{\infty} v^{\lambda t+\gamma-1}e^{-v}\,dv}{(\alpha+\mu+1)^{\lambda t+\gamma}}\\ &= \frac{\Gamma_u(\lambda t+\gamma\,,\,(\alpha+\mu+1)u_K)}{(\alpha+\mu+1)^{\lambda t+\gamma}}\,,\end{aligned}$$

and we recover Equation (29). □

Notice that, in practice, the reinsurer will price reinsurance treaties with a set of parameters more conservative than the ones fitted to data in order to include a safety loading. The time-changed model can also be used for the evaluation of contracts covering an excess of mortality caused by the epidemic. To illustrate this, we consider a treaty that plans the payment of an amount $C\,(D_t - K)$ at time t, where $C, K \in \mathbb{R}^+$, if $D_t > K$. The value of this contract is similar to an option written on D_t. If r is the risk-free rate, u_K is the positive solution of the equation:

$$\Gamma_l(\gamma+1, u(\alpha+\mu)) = \frac{K}{\mu S_0\beta^{\gamma}(\alpha+\mu)^{-\gamma-1}}\,,$$

then the value of the reinsurance treaty is:

$$\begin{aligned}&Ce^{-rt}\mathbb{E}\big((D_t-K)_+\,|\,\mathcal{F}_0\big) =\\ &Ce^{-rt}\left(\frac{\mu S_0\beta^{\gamma}}{(\alpha+\mu)^{1+\gamma}}\int_{u_K}^{\infty}\frac{u^{\lambda t-1}e^{-u}}{\Gamma(\lambda t)}\Gamma_l(\gamma+1,u(\alpha+\mu))\,du - K\frac{K\Gamma_u(\lambda t,u_K)}{\Gamma(\lambda t)}\right).\end{aligned} \tag{31}$$

Unfortunately, the integral in this last equation does not admit a simple analytical form and must be computed numerically.

8. Estimation and Illustration

In order to illustrate the ability of the time-changed model to explain the evolution of a pandemic, we fit it to COVID-19 data sets for Belgium, Germany, Italy and Spain. As in Section 5, parameter estimates are found by a weighted least-square minimization between expected and observed sizes of infected population. We use the same weights as those in Equation (17). Given that the remission and mortality rates have the same impact on I_t, we estimate their sum. The force of mortality is next found by considering the ratio deaths to the number of infected persons adjusted by a time coefficient. More precisely, given that τ_t has increments that are gamma-distributed, $\mathbb{E}(\Delta\tau_t) = \lambda\,dt$. From Equation (24), the expectation is $\mathbb{E}(dD_t|\mathcal{F}_0) \approx \mu\mathbb{E}(I_{t-}|\mathcal{F}_0)\lambda\,dt$. Using a moment-matching approach, we then estimate the mortality rate as follows:

$$\widehat{\mu} = \frac{\sum_{k=1}^{n_{obs}} dD_k^{obs}}{\widehat{\lambda}\left(\sum_{k=1}^{n_{obs}} I_k\right)dt}\,.$$

Parameter estimates are presented in Table 6 and Figure 3 compares the expected number of infectious obtained with the time-changed model and its deterministic counterpart. Globally, we do not observe any similarities between parameters of time-changed and deterministic models. The goodness of fit, measured by the SSE, is also worse for the time-changed version than for the deterministic one. The time change does not fit the right tail of the I_t curve better and overestimates, on average, the number of infected individuals at the beginning of the pandemic.

Table 7 shows the fair premium rates for the healthcare and death insurances valued in Section 5. Since the time-changed model predicts on average higher healthcare bene-

fits during the growing phase of the outbreak, the healthcare insurance is slightly more expensive than the one valued with the deterministic model. The death benefits and death insurance premium rates are comparable in both models.

Table 6. Parameter estimates for the model ruled by Equation (19) (unit of time t: year). The SSE is the value of the optimization criterion in Equation (16).

	$\widehat{\alpha}$	$\widehat{\gamma}$	$\widehat{\beta}$	$\widehat{\mu}$	$\widehat{\lambda}$	**SSE**
Belgium	44.806	105.021	1.113	0.149	27.776	1,005,094
Germany	17.208	29.866	1.295	0.041	29.353	18,592,728
Italy	12.407	25.088	1.09	0.149	25.307	5,906,960
Spain	67.311	81.602	2.133	0.112	23.427	40,570,084

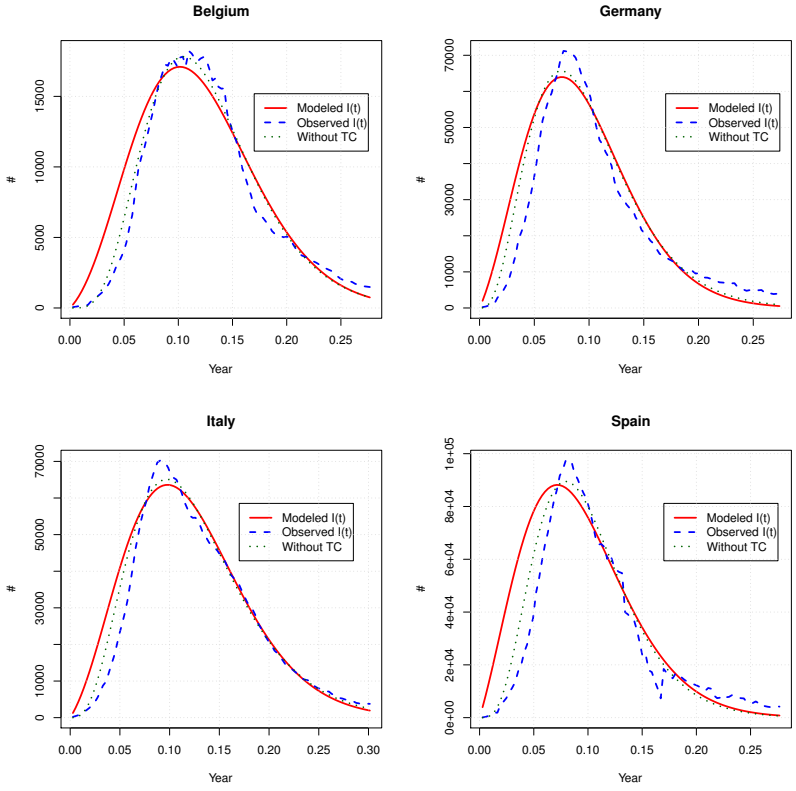

Figure 3. Comparison of $\mathbb{E}(I_t|\mathcal{F}_0)$ with the observed values and those computed with the deterministic model of Section 6.

Table 7. Fair premium rates with the model in (1). Duration $T = 0.5$ year and $r = 2\%$.

	b	c	**Fair** p	**Model 1, Fair** p	$\lambda + 1\%$	$\lambda - 1\%$
Be	365,000	0	148.76	138.54	147.29	150.26
	0	200,000	336.76	338.35	336.77	336.76
Ge	365,000	0	63.98	62.38	63.34	64.62
	0	200,000	41.9	42.35	41.9	41.9
It	365 000	0	110.99	107.02	109.9	112.11
	0	200 000	228.32	230.52	228.32	228.31
Sp	365,000	0	161.6	143.1	160	163.23
	0	200,000	227.73	232.57	227.73	227.73

If we limit our analysis to a comparison of the expected number of infected individuals and premium rates, we have the impression that both deterministic and time-changed

models are quite similar. However, this is far from being the case, given that the second one is stochastic. In order to emphasize the different behavior of this model, we simulate 2000 samples paths with parameter estimates obtained for Italy. Figure 4 shows three of these paths and the average over the 2000 simulations. We see that the time-changed model generates curves of I_t with a different shape than the average sample path. The simulated peaks of the epidemic may be far above the observed one and the timing of this peak displays a high variance. The sample paths are also much more discontinuous than the real evolution of I_t.

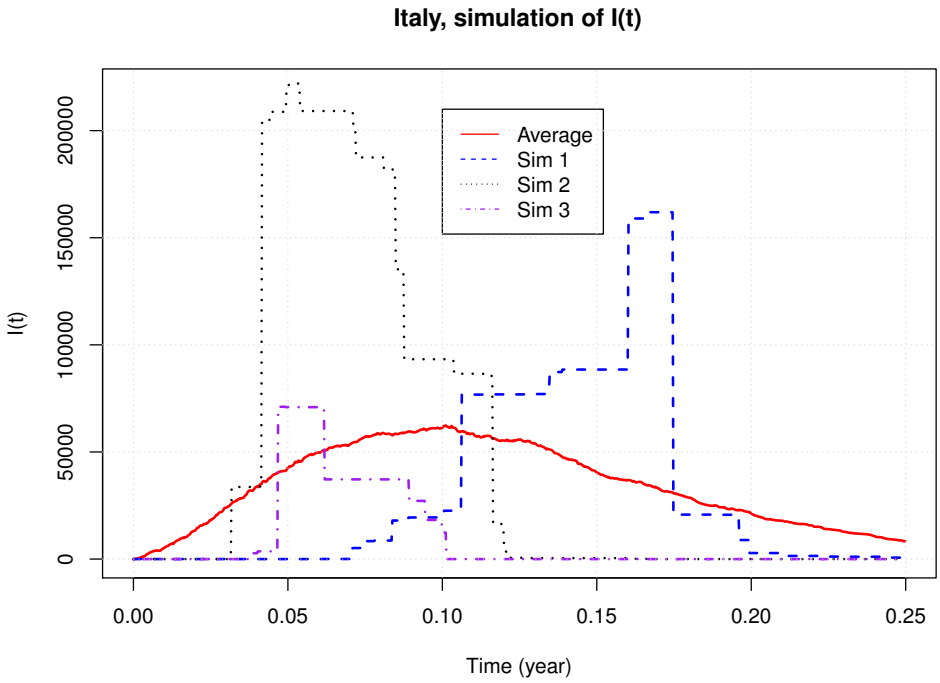

Figure 4. Simulation of three sample paths for I_t and comparison with the simulated average of 1000 sample paths for Italy.

It is also interesting to look at the distribution of the cumulated number of deaths. Figure 5 presents the histograms of D_t for 2000 simulations at time $t = 0.1$ and $t = 0.25$. For $t = 0.1$, we have a bimodal distribution, whereas the number of cumulated deaths after a quarter is nearly deterministic. This is explained by the type of randomness driving the model. The stochastic clock either delays or advances the time of the epidemic peak, but it does not modify the ultimate total number of deaths caused by the pandemic. Table 8 reports the expectations, standard deviations, 5% and 95% percentiles of D_t, computed by simulations for the other countries. We draw from those figures the same conclusions. The discontinuities of I_t and the bimodal behavior of D_t being unlikely in practice (at least for COVID-19), we investigate in the next section, an alternative stochastic extension of the deterministic model.

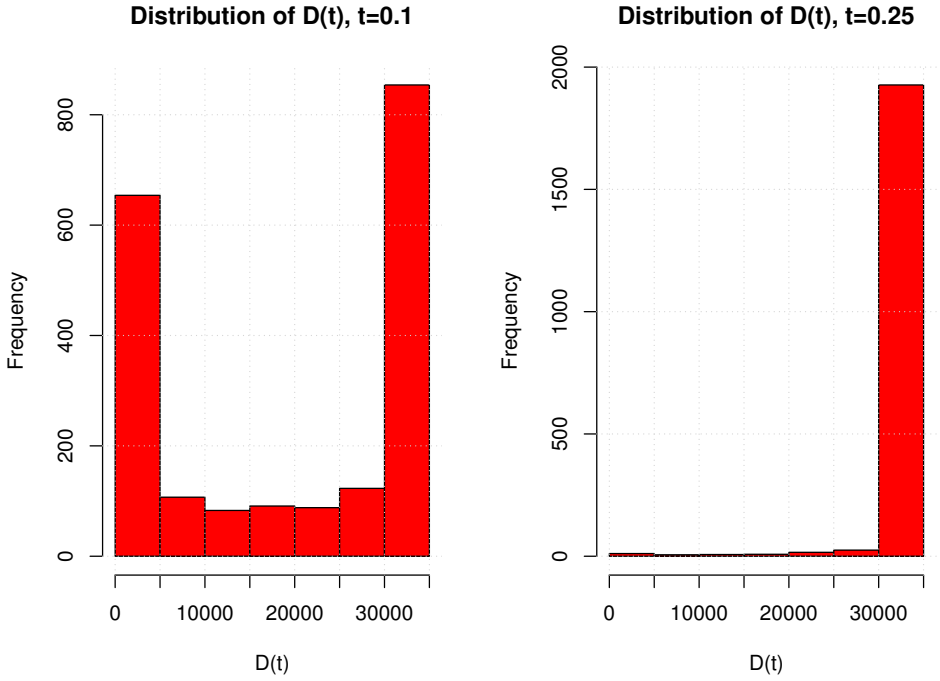

Figure 5. Histograms of the cumulated number of deaths at time $t = 0.1$ and $t = 0.25$ for the Italian population. 2000 simulations.

Table 8. Expectations, standard deviations, 5% and 95% percentiles of simulated D_t. 2000 simulations.

	$\mathbb{E}(D_{0.25})$	$\sqrt{\mathbb{V}(D_{0.25})}$	$\mathbb{E}(D_{0.10})$	$\sqrt{\mathbb{V}(D_{0.10})}$	$D_{0.10}$: 5% Percentile	$D_{0.10}$: 95% Percentile
Belgium	9574.027	1005.182	5138.069	4504.648	0	9718.298
Germany	8708.296	432.437	6288.87	3403.712	0	8741.932
Italy	33,735.473	3497.944	19071.41	14,798.839	0	34,382.855
Spain	26,533.64	1116.695	19,959.171	10,804.449	0	26,602.104

9. A Jump Diffusion Model

This section presents an alternative stochastic model for the dynamic of the infected population that includes random noise and local resurgence of the epidemic. This model also has an excellent analytical tractability and is estimated by a peak over threshold approach. We consider a probability space $(\Omega, \mathcal{F}, P)$ on which a Brownian motion $(W_t)_{t\geq 0}$ and a compound jump process $(L_t)_{t\geq 0}$ are defined. We denote by $(N_t)_{t\geq 0}$ a Poisson process with intensity $\lambda \in \mathbb{R}^+$, and by $(J_k)_{k\in\mathbb{N}} \sim J$ i.i.d. random variables defined on $\mathbb{R}^+$ with a probability density function denoted by $f_J(.)$. The expectation and variance of a jump are denoted by $\mu_J = \mathbb{E}(J)$ and $\mathbb{V}(J) = \sigma_J^2$, respectively. The compound Poisson process $(L_t)_{t\geq 0}$ is defined as the sum of jumps J_k up to time t:

$$L_t = \sum_{k=1}^{N_t} J_k \,. \tag{32}$$

We assume that the dynamics of the population of infected individuals is ruled by the following geometric jump diffusion:

$$dI_t = -(\alpha + \mu) I_t dt + I_t \frac{\gamma}{t} dt + \sigma I_t dW_t + I_t \, dL_t \,, \tag{33}$$

where α, μ, $\frac{\gamma}{t}$ are the recovery, mortality and contagion rates, respectively. The term $\sigma I_t dW_t$, with $\sigma \in \mathbb{R}^+$, is a Gaussian noise, whereas $I_t\, dL_t$ introduces random discontinuities caused by the discovery of clusters of infection. The next proposition presents the solution of Equation (32). Under the assumption that I_t is ruled by Equation (32), the size of the population of infected people is equal to:

$$I_t = S_0 e^{-\left(\alpha+\mu+\frac{\sigma^2}{2}\right)t+\sigma W_t}(\beta t)^\gamma \prod_{k=1}^{N_t}(1+J_k)\,. \tag{34}$$

This result is a direct consequence of Itô's lemma for jump diffusion. In order to evaluate an insurance plan, we need to adopt a dynamic for the cumulated number of deaths, D_t. As in previous models, we assume that the instantaneous number of death at time t is a proportion μdt of the population of sick persons. The differential equation ruling S_t, the population of susceptible individuals, guarantees that the total size of the population remains equal to S_0:

$$\begin{cases} dD_t &= \mu I_t dt\,, \\ dS_t &= I_t\left(\alpha - \frac{\gamma}{t}\right)dt - \sigma I_t dW_t - I_{t-}\, dL_t\,. \end{cases}$$

The number of deaths is $D_t = \mu \int_0^t I_s ds$, whereas $S_t = S_0 - I_t - D_t$. Of course, we could include a random noise in the dynamics of D_t, but this would not fundamentally modify the results developed in the remainder of this section. The next proposition presents the first two moments of I_t.

Proposition 5. *The expectation and variance of the number of infected persons at time t are respectively equal to*

$$\mathbb{E}(I_t|\mathcal{F}_0) \;=\; S_0 e^{\left(\lambda\mu_J-(\alpha+\mu)\right)t}(\beta t)^\gamma\,, \tag{35}$$

and

$$\mathbb{V}(I_t|\mathcal{F}_0) \;=\; S_0^2 e^{-2(\alpha+\mu)t}(\beta t)^{2\gamma} e^{2\lambda\mu_J t}\left(e^{\left(\sigma^2+\lambda\mathbb{E}(J^2)\right)t}-1\right). \tag{36}$$

Proof. Given that W_t is independent from N_t, the expectation of I_t is equal to a product of expectations:

$$\mathbb{E}(I_t|\mathcal{F}_0) \;=\; S_0 e^{-(\alpha+\mu)t}(\beta t)^\gamma \mathbb{E}\left(e^{-\frac{\sigma^2}{2}t+\sigma W_t}|\mathcal{F}_0\right)\mathbb{E}\left(\prod_{k=1}^{N_t}(1+J_k)|\mathcal{F}_0\right).$$

As $\sigma W_t \sim \mathcal{N}(0,\sigma^2 t)$ is normal, $\mathbb{E}\left(e^{-\frac{\sigma^2}{2}t+\sigma W_t}|\mathcal{F}_0\right) = 1$. The moment-generating function of N_t is $\mathbb{E}\left(e^{\omega N_t}\right) = e^{\lambda t(e^\omega-1)}$ and N_t is independent for $(J_k)_{k=1\ldots,N_t}$. Therefore, we conclude that the expectation of the product of jumps is:

$$\begin{aligned} \mathbb{E}\left(\prod_{k=1}^{N_t}(1+J_k)|\mathcal{F}_0\right) &= \mathbb{E}\left(\prod_{k=1}^{N_t}(1+\mu_J)|\mathcal{F}_0\right) \\ &= \mathbb{E}\left(e^{N_t\ln\left(1+\mu_J\right)}|\mathcal{F}_0\right) \\ &= e^{\lambda\mu_J t}\,. \end{aligned}$$

Combining these elements leads to Equation (35). In a similar manner, we calculate

the second order moment of I_t—that is,

$$\mathbb{E}\left(I_t^2|\mathcal{F}_0\right) = S_0^2 e^{-2(\alpha+\mu)t}(\beta t)^{2\gamma}\mathbb{E}\left(e^{-\sigma^2 t+2\sigma W_t}|\mathcal{F}_0\right)\mathbb{E}\left(\left(\prod_{k=1}^{N_t}(1+J_k)\right)^2|\mathcal{F}_0\right).$$

Since $e^{2\sigma W_t}$ is log-normal, $\mathbb{E}\left(e^{2\sigma W_t}|\mathcal{F}_0\right) = e^{2\sigma^2 t}$ and $\mathbb{E}\left(e^{-\sigma^2 t+2\sigma W_t}|\mathcal{F}_0\right) = e^{\sigma^2 t}$. Furthermore, the expectation of the square of the product of jumps is equal to:

$$\begin{aligned}
\mathbb{E}\left(\left(\prod_{k=1}^{N_t}(1+J_k)\right)^2|\mathcal{F}_0\right) &= \mathbb{E}\left(\mathbb{E}\left(\left(\prod_{k=1}^{N_t}(1+J_k)\right)^2|\mathcal{F}_0\vee N_t\right)|\mathcal{F}_0\right)\\
&= \mathbb{E}\left(\prod_{k=1}^{N_t}\mathbb{E}\left((1+J)^2\right)|\mathcal{F}_0\right)\\
&= \mathbb{E}\left(e^{N_t\ln(1+2\mu_j+\mathbb{E}(J^2))}|\mathcal{F}_0\right)\\
&= e^{\lambda\left(2\mu_J+\mathbb{E}(J^2)\right)t}.
\end{aligned}$$

We obtain, then, the second moment of I_t and the variance in (36). □

The next proposition presents an analytical formula to evaluate expected healthcare benefits paid up to time t. This result is similar to the one for the deterministic model, except that it takes into account the frequency and the average size of jumps.

Proposition 6. *Let us consider a discount rate $r \in \mathbb{R}^+$. The integral of expected discounted numbers of infected is equal to:*

$$\int_0^t e^{-rs}\mathbb{E}(I_s|\mathcal{F}_0)ds = \frac{S_0\beta^\gamma}{\theta_J^{\gamma+1}}\Gamma_l\left(\gamma+1,\theta_J t\right), \tag{37}$$

where $\theta_J = r+\alpha+\mu-\lambda\mu_J$ and $\Gamma_l(\gamma+1,x)$ is the gamma lower incomplete function.

The proof is similar to the one of Proposition 1. We insert the expression (34) of $\mathbb{E}(I_s|\mathcal{F}_0)$ in the integral and obtain Equation (37). Contrary to the time-changed model, the expected number of deaths and its discounted integral admit closed-form expressions in terms of incomplete lower gamma functions.

Corollary 2. *The expected cumulated number of deaths caused by the epidemic at time $t \geq 0$ is equal to*

$$\mathbb{E}(D_t\,|\,\mathcal{F}_0) = \frac{\mu S_0\beta^\gamma}{\left(\alpha+\mu-\lambda\mu_J\right)^{\gamma+1}}\Gamma_l\left(\gamma+1,t\left(\alpha+\mu-\lambda\mu_J\right)\right). \tag{38}$$

If $\theta_J = (r+\alpha+\mu)-\lambda\mu_J$, the expectation of the integral of discounted variation of D_t is equal to:

$$\mathbb{E}\left(\int_0^T e^{-rs}\,dD_s\,|\,\mathcal{F}_0\right) = \frac{\mu S_0\beta^\gamma}{\theta_J^{\gamma+1}}\Gamma_l\left(\gamma+1,\theta_J t\right). \tag{39}$$

Let us again consider the insurance plan introduced in Equation (4) with a maturity T. The premium rate is still denoted as p. The rate of healthcare expenses is b and the lump sum benefit in case of death is c. The next proposition presents the fair premium rate that ensures the equilibrium of this plan under the assumption that the pricing and real measures are identical.

Proposition 7. *For the benefit rates* (b, c)*, the fair premium rate that guarantees the actuarial equilibrium of the plan is given by:*

$$p = \frac{(b+c\mu)S_0\beta^\gamma\theta_J^{-\gamma-1}\Gamma_l(\gamma+1,\theta_J T)}{\int_0^T e^{-rs}\mathbb{E}(S_s \mid \mathcal{F}_0)\, ds}, \tag{40}$$

where the denominator is equal to:

$$\begin{aligned}\int_0^T e^{-rs}\mathbb{E}(S_s \mid \mathcal{F}_0)\, ds &= \frac{S_0}{r}\left(1-e^{-rT}\right) - \frac{S_0\beta^\gamma}{\theta_J^{\gamma+1}}\Gamma_l(\gamma+1,\theta_J T)\left(1+\frac{\mu}{r}\right) \\ &\quad + \frac{\mu S_0\beta^\gamma e^{-rT}\Gamma_l\left(\gamma+1\,,\,(\alpha+\mu-\lambda\mu_J)T\right)}{r\left(\alpha+\mu-\lambda\mu_J\right)^{\gamma+1}}.\end{aligned} \tag{41}$$

Proof. The fair premium rate in a stochastic framework is given by Equation (28). The integral $\int_0^T e^{-rs}\mathbb{E}(I_s \mid \mathcal{F}_0)\, ds$ is provided by Equation (37). As $S_t = S_0 - I_t - D_t$, we infer that

$$\begin{aligned}\int_0^T e^{-rs}\mathbb{E}(S_s \mid \mathcal{F}_0)\, ds &= \frac{S_0}{r}\left(1-e^{-rT}\right) - \frac{S_0\beta^\gamma}{\theta_J^{\gamma+1}}\Gamma_l(\gamma+1,\theta_J T) \\ &\quad - \int_0^T e^{-rs}\mathbb{E}(D_s \mid \mathcal{F}_0)\, ds\,,\end{aligned} \tag{42}$$

where $\int_0^T e^{-rs}\mathbb{E}(D_s \mid \mathcal{F}_0)\, ds$ admits the following integral representation:

$$\begin{aligned}\int_0^T e^{-rs}\mathbb{E}(D_s \mid \mathcal{F}_0)\, ds &= \mu\int_0^T e^{-rs}\int_0^s \mathbb{E}(I_u \mid \mathcal{F}_0)\, du\, ds \\ &= \mu S_0\beta^\gamma \int_0^T e^{-rs}\int_0^s e^{\left(\lambda\mu_J-(\alpha+\mu)\right)u}u^\gamma\, du\, ds\,.\end{aligned}$$

Given that $\int_0^s e^{\left(\lambda\mu_J-(\alpha+\mu)\right)u}u^\gamma\, du = \frac{\Gamma_l\left(\gamma+1,\left(\alpha+\mu-\lambda\mu_J\right)s\right)}{\left(\left(\alpha+\mu-\lambda\mu_J\right)\right)^{\gamma+1}}$, the double integral in this last expression becomes:

$$\begin{aligned}\int_0^T e^{-rs}\int_0^s e^{\left(\lambda\mu_J-(\alpha+\mu)\right)u}u^\gamma\, du\, ds &= \left[-\frac{1}{r}\frac{e^{-rs}\Gamma_l\left(\gamma+1\,,\,(\alpha+\mu-\lambda\mu_J)s\right)}{\left(\alpha+\mu-\lambda\mu_J\right)^{\gamma+1}}\right]_{s=0}^{s=T} \\ &\quad + \frac{1}{r}\int_0^T e^{\left(\lambda\mu_J-(r+\alpha+\mu)\right)s}s^\gamma ds \\ &= -\frac{e^{-rT}\Gamma_l\left(\gamma+1\,,\,(\alpha+\mu-\lambda\mu_J)T\right)}{r\left(\alpha+\mu-\lambda\mu_J\right)^{\gamma+1}} \\ &\quad + \frac{\Gamma_l(\gamma+1\,,\theta_J T)}{r\theta_J^{\gamma+1}}\end{aligned}$$

Therefore, the integral of the discounted expected number of deaths is equal to:

$$\begin{aligned}\int_0^T e^{-rs}\mathbb{E}(D_s \mid \mathcal{F}_0)\, ds &= -\frac{\mu S_0\beta^\gamma e^{-rT}\Gamma_l\left(\gamma+1\,,\,(\alpha+\mu-\lambda\mu_J)T\right)}{r\left(\alpha+\mu-\lambda\mu_J\right)^{\gamma+1}} \\ &\quad + \frac{\mu S_0\beta^\gamma\Gamma_l(\gamma+1\,,\theta_J T)}{r\theta_J^{\gamma+1}}.\end{aligned} \tag{43}$$

Combining expressions (42) and (43) leads to Equation (41). □

10. Reinsurance in the Jump Diffusion Model

As in the time-changed model, we can evaluate various reinsurance coverages by Monte Carlo simulations. When jumps are constant and equal to $J = \kappa \in \mathbb{R}^+$, reinsurance treaties with a payoff dependent on I_t can be valued with a closed-form expression. In this case, conditionally to $N_t = n$ jumps, I_t is log-normal:

$$(I_t \,|\, N_t = n) = S_0(\beta t)^\gamma e^{X_t^{(n)}},$$

where $X_t^{(n)} \sim \mathcal{N}\big(\mu_X(t,n)\,;\,\sigma_X^2(t,n)\big)$. The mean and variance of $X_t^{(n)}$ are respectively equal to

$$\begin{aligned}\mu_X(t,n) &= n\ln(1+\kappa) - \left(\alpha + \mu + \frac{\sigma^2}{2}\right)t\,,\\ \sigma_X^2(t,n) &= \sigma^2 t\,.\end{aligned}$$

The following proposition provides an analytical expression for a reinsurance treaty that plans the payment of an amount $C\,(I_t - K)$ at time t, where $C, K \in \mathbb{R}^+$, if $I_t > K$. Model parameters are assumed to be the same under the pricing and real measures.

Proposition 8. *Let us define*

$$d_2(t,n) = \frac{1}{\sigma_X(t,n)}\left(\ln\left(\frac{K}{S_0(\beta t)^\gamma}\right) - \mu_X(t,n)\right), \tag{44}$$

$$d_1(t,n) = d_2(t,n) - \sigma_X(t,n)\,. \tag{45}$$

The value of an excess-of-loss reinsurance covering an excessive number of infected is equal to:

$$\begin{aligned}&Ce^{-rt}\mathbb{E}\big((I_t - K)_+ \,|\, \mathcal{F}_0\big) =\\ &\quad Ce^{-rt}\sum_{n=0}^{\infty} P(N_t = n)\left(S_0(\beta t)^\gamma e^{\mu_X + \frac{\sigma_X^2}{2}}\Phi(-d_1(t,n)) - K\Phi(-d_2(t,n))\right),\end{aligned} \tag{46}$$

where $P(N_t = n) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$ *and* $\Phi(.)$ *is the cumulative probability distribution of a standard normal random variable.*

Proof. We can rewrite the price of this treaty as a sum of conditional expectations with respect to N_t:

$$\begin{aligned}&Ce^{-rt}\mathbb{E}\big((I_t - K)_+ \,|\, \mathcal{F}_0\big) =\\ &\quad Ce^{-rt}\sum_{n=0}^{\infty} P(N_t = n)\mathbb{E}\left(\left(S_0(\beta t)^\gamma e^{X_t^{(n)}} - K\right)_+ \,|\, \mathcal{F}_0\right).\end{aligned}$$

The conditional expectations may be developed as the difference of two integrals:

$$\begin{aligned}\mathbb{E}\left(\left(S_0(\beta t)^\gamma e^{X_t^{(n)}} - K\right)_+ \,|\, \mathcal{F}_0\right) &= S_0(\beta t)^\gamma \int_{e^{\mu_X + \sigma_X u} \geq \frac{K}{S_0(\beta t)^\gamma}} e^{\mu_X + \sigma_X u}\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du\\ &\quad - K\int_{e^{\mu_X + \sigma_X u} \geq \frac{K}{S_0(\beta t)^\gamma}} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du\,.\end{aligned} \tag{47}$$

The second term, after a change of variable, is equal to:

$$\begin{aligned}K\int_{e^{\mu_X + \sigma_X u} \geq \frac{K}{S_0(\beta t)^\gamma}} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du &= K\int_{u \geq \frac{1}{\sigma_X}\left(\ln\left(\frac{K}{S_0(\beta t)^\gamma}\right) - \mu_X\right)} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du\\ &= K\Phi(-d_2(t,n))\,.\end{aligned}$$

where $d_2(t,n) = \frac{1}{\sigma_X(t,n)}\left(\ln\left(\frac{K}{S_0(\beta t)^\gamma}\right) - \mu_X(t,n)\right)$. In the same manner, the first term of Equation (47) becomes:

$$\begin{aligned}\int_{e^{\mu_X+\sigma_X u}\geq \frac{K}{S_0(\beta t)^\gamma}}^{\infty} e^{\mu_X+\sigma_X u}\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du &= e^{\mu_X+\frac{\sigma_X^2}{2}}\int_{u\geq \frac{1}{\sigma_X}\left(\ln\left(\frac{K}{S_0(\beta t)^\gamma}\right)-\mu_X\right)}^{\infty}\frac{e^{-\frac{u^2-2\sigma_X u+\sigma_X^2}{2}}}{\sqrt{2\pi}}du\\ &= e^{\mu_X+\frac{\sigma_X^2}{2}}\int_{s\geq \frac{1}{\sigma_X}\left(\ln\left(\frac{K}{S_0(\beta t)^\gamma}\right)-\mu_X\right)-\sigma_X}^{\infty}\frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}}ds\\ &= e^{\mu_X+\frac{\sigma_X^2}{2}}\Phi(-d_1(t,n))\,,\end{aligned}$$

and we retrieve the result. □

By construction, the cumulated number of deaths up to time t is proportional to the integral of I_t. Unfortunately, the statistical distribution of D_t is unknown and, therefore, reinsurance treaties covering excess of mortality must be valued by simulations.

11. Estimation of the Jump Diffusion Model

The jump diffusion model is fitted in two steps. Let us recall that $\mathbb{E}(I_k|\mathcal{F}_0) = S_0 e^{\eta t_k}(\beta t_k)^\gamma$ where $\eta = \lambda\mu_J - (\alpha+\mu)$ and t_k are the observation times for $k = 1,...,n_{obs}$. The first step consists in estimating β, γ and η by minimizing the weighted sum of squares between the expected and observed numbers of infected persons:

$$(\widehat{\eta},\widehat{\gamma},\widehat{\beta}) = \arg\min \frac{\sum_{k=1}^{n_{obs}}\omega_k\left(\mathbb{E}(I_k|\mathcal{F}_0) - I_k^{obs}\right)^2}{\sum_{k=1}^{n_{obs}}\omega_k}\,. \tag{48}$$

Given that the expectation of I_t in the jump diffusion approach coincides with the deterministic model, we obtain the same estimates $\widehat{\beta}$, $\widehat{\gamma}$ as those in Table 2. In the second stage, we fit the jump process by the peak over threshold method. From Equations (33) and (35), we define Y_t as the ratio of the process I_t on its expectation:

$$Y_t := \frac{I_t}{\mathbb{E}(I_t|\mathcal{F}_0)} = e^{\left(-\lambda\mu_J-\frac{\sigma^2}{2}\right)t+\sigma W_t}\prod_{k=1}^{N_t}(1+J_k)\,.$$

Using Itô's lemma, we infer that Y_t is driven by the following infinitesimal dynamics:

$$\frac{dY_t}{Y_t} = +\sigma dW_t + dL_t - \lambda\mu_J dt\,. \tag{49}$$

The value of Y_t at time t_k is noted Y_k and we define $Z_k = \frac{Y_k - Y_{k-1}}{Y_k}$ the discrete approximation of $\frac{dY_t}{Y_t}$. If the time lag of one day between two successive observations is noted Δ, according to Equation (49), Z_k is approximately the sum

$$Z_k \approx \epsilon_k + \Delta L_k - \lambda\mu_J\Delta\,, \tag{50}$$

where $\epsilon_k \sim \mathcal{N}\left(0,\sigma\sqrt{\Delta}\right)$ and $\Delta L_k = L_{t_k} - L_{t_{k-1}}$. A jump is believed to occur when Z_k is above a threshold, noted by $g(q)$, where q is a confidence level. To define the threshold, we fit a pure Gaussian process to the time series of $Z_k \sim \mu_z\Delta + \sigma_z W_\Delta$. The unbiased estimators of μ_z and σ_z are:

$$\widehat{\mu}_z = \frac{1}{n_{obs}\Delta}\sum_{j=1}^{n_{obs}} Z_j \qquad \widehat{\sigma}_z^2 = \frac{1}{(n_{obs}-1)\Delta}\sum_{j=1}^{n_{obs}}(Z_j - \widehat{\mu}_z)^2\,.$$

If $\Phi(.)$ denotes the cumulative distribution function of a standard normal, $g(q)$ is set to the q percentile of the Gaussian process: $g(q) = \widehat{\mu}_z\Delta + \widehat{\sigma}_z\sqrt{\Delta}\Phi^{-1}(q)$. The time s_k of the k^{th} jump is therefore:

$$s_k \quad = \quad \min\{t_j \in \{t_1, ..., t_n\} \,|\, z_j \geq g(q)\,,\, j \geq k\}\,.$$

and the sample path of $(N_t)_{t\geq 0}$ is approached by the following time series:

$$N(t_k) = \max\{j \in \mathbb{N} \,|\, s_j \leq t_k\}\,.$$

Under the assumption that the diffusion is negligible with respect to jumps, we assimilate $z_k - \mu_z\Delta$ to J_k when a jump is detected at time t_k.

We have applied this estimation procedure to COVID-19 data sets for Belgium, Germany, Italy and Spain. The estimates $\widehat{\beta}$ and $\widehat{\gamma}$ are found by minimization of weighted least squares as shown in Equation (48). We use the same weights as those used to fit previous models (see Equation (17)). The other parameters are estimated by the peak over threshold method and are reported in Table 9. Given that the number of confirmed cases in March is underestimated due to the penury of tests, we mainly focus on the period from mid-April to mid-June to calculate the time series of Z_k. The threshold confidence level is set to 90%. Figure 6 shows those series and the threshold level. For Spain, we have not taken two abnormally high and low values into account due to retrospective adjustments of the number of confirmed cases. For Belgium and Spain, the Z_k values have a mean close to zero and the variance seems more or less constant for different time windows. This confirms that the postulated dynamic in Equation (50) for Z_k is acceptable for those countries. For Italy, the variance of Z_k seems to raise during the month of June, but their mean is close to zero. For Germany, the Z_k values have a residual linear increasing trend and their variance seems to increase as it does for Italy. The consequence is that the peak-over-threshold method tends to overestimate the size and frequency of jumps. Nevertheless, this is a conservative approach from the insurer point of view and, therefore, we accept these parameter estimates. For the same reason, we do not consider negative jumps.

Table 9. Parameter estimates for the model ruled by Equation (33) (unit of time t: year). The SSE is the value of the optimization criterion in Equation (16).

	$\widehat{\alpha}$	$\widehat{\gamma}$	$\widehat{\beta}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$	$\widehat{\mu_J}$	$\widehat{\sigma_J}$	**SSE**
Belgium	41.944	4.740	6.606	4.457	0.589	20.278	0.06	0.007	535,988
Germany	43.758	3.125	3.693	1.239	0.727	30.845	0.101	0.022	12,513,688
Italy	31.88	3.382	3.709	3.931	0.375	13.519	0.074	0.024	3,093,842
Spain	51.465	3.937	6.979	2.966	1.512	25.347	0.191	0.026	24,203,834

By construction, the expectation of the jump diffusion model corresponds to the deterministic model of Section 3. Therefore, the fair premium rate of an insurance plan covering healthcare expenses and death benefits are the same with both approaches. We refer the reader to the previous Section 5 for numerical examples. However, the jump diffusion model can generate a wide variety of sample paths for I_t. This point is illustrated in Figure 7 that shows 1000 simulated paths and the expectation of I_t over a quarter. Notice that these simulations are performed with the assumption that jumps J are constant and equal to μ_J.

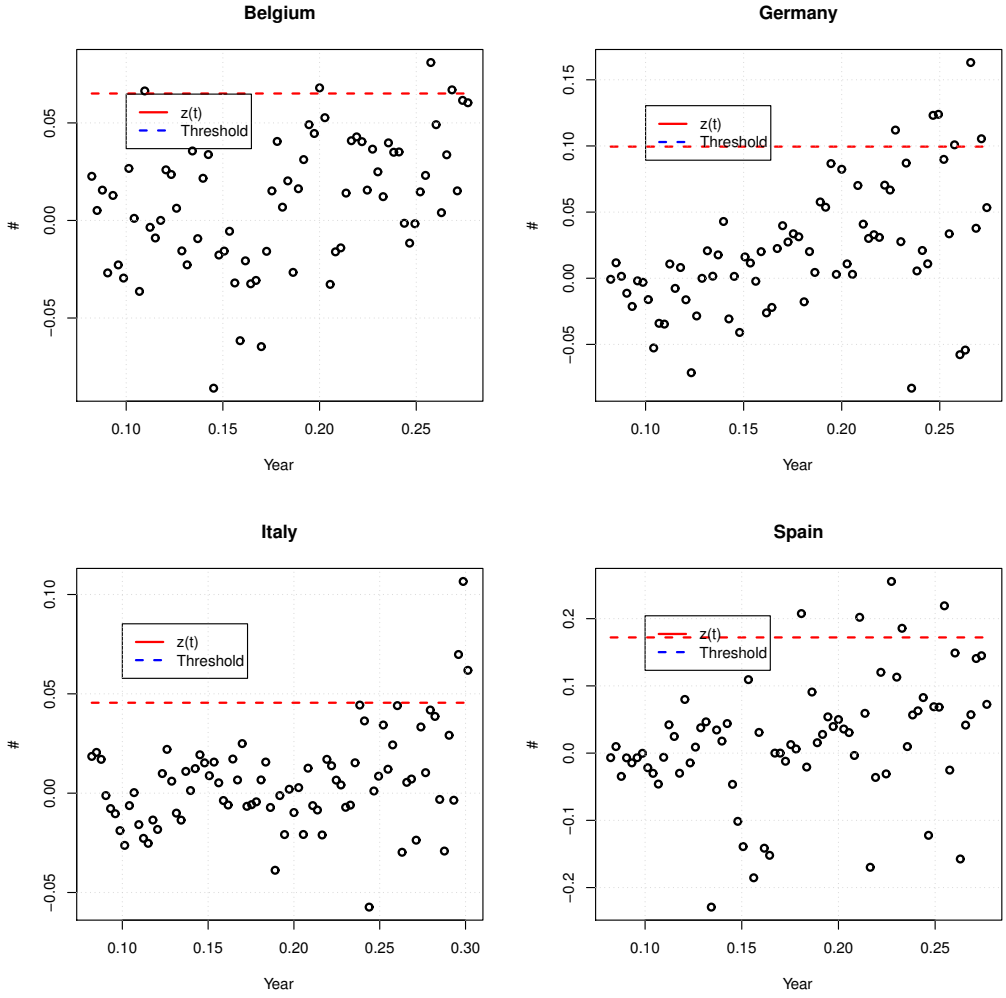

Figure 6. Plot of Z_k time series for Belgium, Germany, Italy and Spain. The threshold is calculated with $q = 90\%$.

Contrary to the time-changed approach, the jump diffusion model generates smoother sample paths which all appear as likely scenarios for the evolution of the infected population. These graphs also display the expected and the observed number of infected cases for each country. It reveals that the real sample path of I_t for each country is a likely realization of the jump diffusion model, at least for $t \geq 0.07$. At the beginning of the pandemic, the real sample path of I_t bounds from below simulated trajectories. This is a direct consequence of choices made in Section 5 for calibrating the average trend of the model. At the start of the epidemic, testing polices were in the process of being deployed and the number of infected cases was probably underestimated. This motivates our choice of underweighting observations collected in the early ascending phase. We also remark that the peak of infected cases is significantly higher than the observed one in some scenarios. To illustrate this, Table 10 reports the 90% and 95% percentiles of the simulated maximum of infected cases. The ratios of these 95% percentiles on real numbers of infected cases at the epidemic peak range from 121.23% for Italy up to 202.49% for Spain.

Table 11 reports statistics about the simulated number of deceases over one quarter. Globally, the mean and standard deviation of $D_{0.25}$ seems credible and the distribution of D_t does not display any bimodality as in the time-changed model.

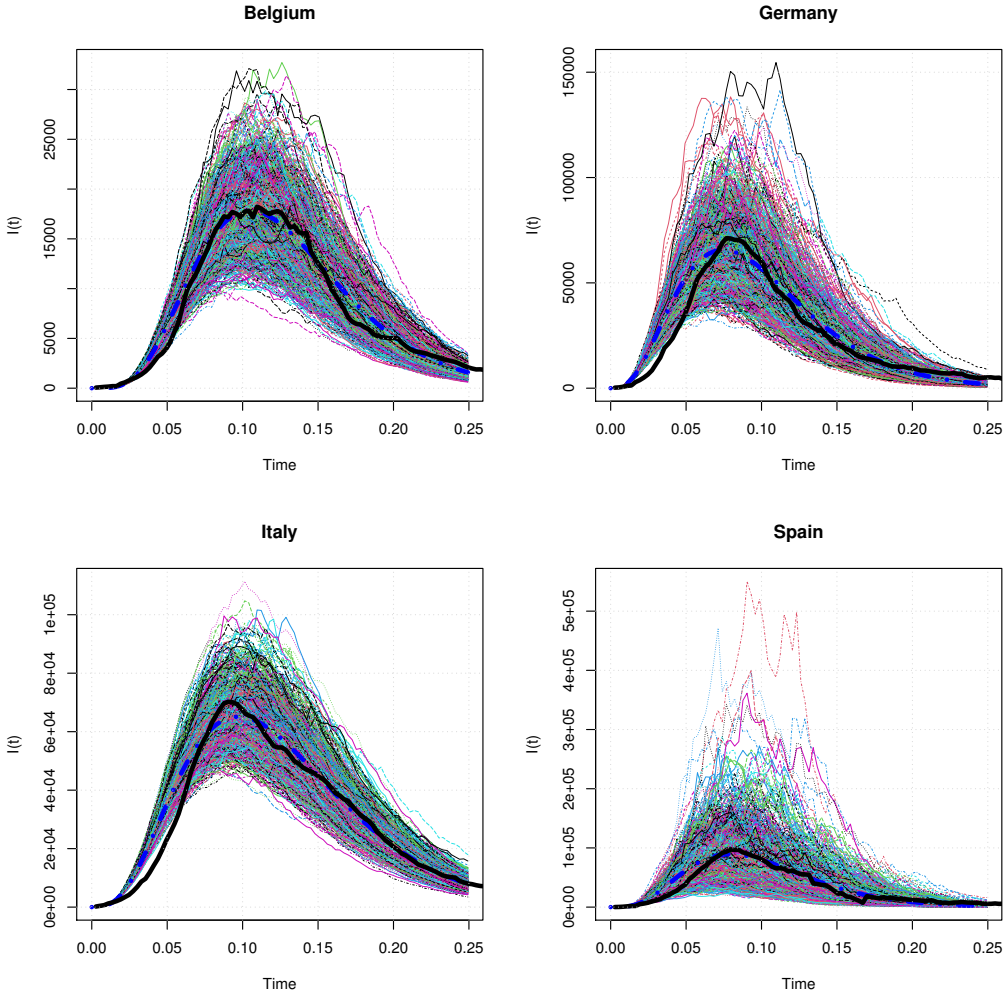

Figure 7. Simulations of 1000 sample paths of I_t, with parameter estimates in Table 9 and constant J. The thick dotted blue line is the expectation of I_t, whereas the thick black line is the observed number of infected cases.

Table 10. Statistics about the observed and simulated epidemic peaks (1000 simulations).

	Maximum Number of Infected	90% Percentile Max. Number of Infected	95% Percentile Max. Number of Infected	Ratio 95% Percentile on Max. Cases
Belgium	18,225	23,702	25,260	138.6%
Germany	71,219	93,506	100,764	141.49%
Italy	70,233	79,827	85,141	121.23%
Spain	97,400	164707	197,224	202.49%

Table 11. Expectation, standard deviation and 5%–95% percentiles of simulated cumulated number of deaths, D_t. 1000 simulations.

$D_{0.25}$	Expected	Standard Deviation	5% Percentile	95% Percentile
Belgium	9379.404	1850.385	6664.908	12,717.609
Germany	8576.788	2079.589	5671.554	12,163.033
Italy	33,166.37	4704.36	26,483.31	41,795.193
Spain	24,948.895	12,595.321	10,641.107	46,904.273

Table 12 presents the prices of a few excess-of-loss reinsurances with I_t and D_t as

underlying risks. The treaties on I_t have a threshold K equal to one-fourth of the total number of reported cases over the considered period of time (see Table 1). The time horizon of the contract and the capital by unit in excess are set to $t = 0.1$ year and $C = 1$, respectively. Prices are calculated with the closed-form expression (46). We use Monte Carlo simulations for the valuation of reinsurance contracts covering an excessive mortality. The threshold K is set to the observed number of deaths observed over the considered period (see Table 1). The time horizon and the capital by unit in excess are respectively set to $t = 0.1$ year and $C = 1$.

Table 12. Prices of excess of loss reinsurance treaties. $r = 2\%$.

	K	$e^{-r\,0.10}\mathbb{E}\Big((I_{0.10} - K)_+\Big)$	K	$e^{-r\,0.25}\mathbb{E}\Big((D_{0.25} - K)_+\Big)$
ine Belgium	14,998	3154.536	9661	610.114
Germany	46,721	11,411.364	8807	705.235
Italy	59,209	7088.726	34,359	1362.358
Spain	60,927	25,778.164	27,131	3739.249

12. Conclusions

The valuation of actuarial commitments requires us to integrate over time the size of infectious and susceptible populations. Since existing compartment models do not admit closed-form expressions for these quantities, actuarial calculations are in this framework computationally intensive and subject to numerical errors. The three models proposed in this article remedy to this issue and present a high degree of analytical tractability.

The first model is purely deterministic. The basic reproduction number, R_0, decays with time in order to replicate the impact of preventive measures to curb the epidemic. Contrary to the SIR, the empirical tests performed on COVID-19 data confirm that the model explains the first wave of such an epidemic. Furthermore, the insurance premium rate admits a closed-form expression within this framework. The main disadvantage of this approach is the absence of random effects that prevents to evaluate the incurred extreme costs.

The second model is a time-changed extension of the deterministic one. The time of the pandemic peak is randomized by observing the process on a stochastic time scale. The main advantage of this approach is that it preserves the main features of the deterministic model and leads to comparable premium rates. Nevertheless, the simulation study reveals that simulated scenarios display a different trend from what is observed for the COVID-19 outbreak. Furthermore, the stochastic clock modulates the speed at which the pandemic evolves but do not modify the sizes of infected and susceptible populations.

This article proposes a second stochastic extension of the deterministic model based on a jump diffusion process. In this approach, the rate at which patients cease to be considered as infected is noised by a Brownian motion. This allows us to randomize the duration of illness. The apparition of local clusters of infected causing a sudden increase of the number of contagious cases is replicated by the jump component. This model presents several interesting features. As it behaves, on average, as the deterministic approach does, it keeps a high analytical tractability for actuarial applications. On the other hand, the model is able to generate realistic noised sample paths of infected cases. This feature allows us to price reinsurance contracts, such as excess of loss treaties, that cannot be valued in a deterministic framework. Last but not least, it is remarkably easy to estimate its parameters with the proposed "peak-over-threshold" method.

Notice that, by construction, the contagion rate per capita decreases as $\frac{1}{t}$, the size of the infected population, converges to zero after having reached the epidemic peak. In a similar way to the SIR model, the solutions proposed in this article are then designed for explaining a single epidemic wave with, eventually, random recovery duration and discovery of local clusters of infected individuals.

This observations opens the way to further research. Instead of a deterministic starting date for the epidemic, we can replace it by the jump time of a self-exciting point process, e.g., Hainaut and Moraux (2019). In its simplest version, the intensity of this process is persistent and suddenly increases as soon as a jump occurs. Within this approach, the starting date of the pandemic becomes random and the probability of observing a new epidemic wave raises after the first one but decay exponentially to its baseline level. Other possible extensions consist to randomize the mortality rates or to develop a compartmental version with subpopulations of infected individuals.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author is grateful to the "Fonds de la Recherche Scientifique—FNRS" for financial support under Grant number 33658713.

Conflicts of Interest: The authors declare no conflict of interest.

References

Anderson, Roy M., and Robert M. May. 2008. Population biology of infectious diseases. *Nature* 280: 361–67. [CrossRef] [PubMed]
Brauer, Fred. 2017. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* 2: 113–27. [CrossRef] [PubMed]
Caraballo, Tomas, and Renato Colucci. 2017. A comparison between random and stochastic modeling for a SIR model. *Communications on Pure and Applied Analysis* 16: 151–62. [CrossRef]
Caraballo, Tomas, and Sami Keraani. 2018. Analysis of a stochastic SIR model with fractional Brownian motion. *Stochastic Analysis and Applications* 36: 895–908. [CrossRef]
Chen, Hua, and Samuel H. Cox. 2009. An option-based operational risk management model for pandemics. *North American Actuarial Journal* 13: 54–76. [CrossRef]
Clara-Rahola, Joaquim. 2020. An empirical model for the spread and reduction of the COVID19 pandemic. *Estudios de Economia Aplicada* 38. [CrossRef]
Daley, David J., and J. Gani. 1999. *Epidemic Models: An Introduction*. Cambridge Studies in Mathematical Biology 15. Cambridge: Cambridge University Press.
Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1995. The legacy of Kermack and McKendrick; In *Epidemic Models: Their Structure and Relation to Data*. Edited by D. Mollison. Cambridge: Cambridge University Press. pp. 95–115.
Diekmann, Odo, J. A. P. Heesterbeek, and Johan A. J. Metz. 1990. On the definition and the computation of the basic reproductive ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* 28: 365–82. [CrossRef]
Feng, Runhuan, and Jose Garrido. 2011. Actuarial applications of epidemiological models. *North American Actuarial Journal* 15: 112–27. [CrossRef]
Gathy, Maude, and Claude Lefèvre. 2009. From damage models to SIR epidemics and cascading failures. *Advances in Applied Probability* 41: 247–69. [CrossRef]
Hainaut, Donatien, and Franck Moraux. 2019. A switching self-exciting jump diffusion process for stock prices. *Annals of Finance Volume* 15: 267–306. [CrossRef]
Hethcote, Herbert W. 2000. The Mathematics of Infectious Diseases. *SIAM Review* 42: 599–653. [CrossRef]
Jia, Na, and Lawrence Tsui. 2005. Epidemic Modelling Using SARS as a Case Study. *North American Actuarial Journal* 9: 28–42. [CrossRef]
Kermack, William Ogilvy, and A. G. McKendrick. 1927. Contributions to the mathematical theory of epidemics—Part I. *Proceedings of the Royal Society of London. Series A* 115: 700–21.
Lefèvre, Claude, and Sergey Utev. 1999. Branching Approximation for the Collective Epidemic Model. *Methodology and Computing in Applied Probability Volume* 1: 211–28. [CrossRef]
Lefèvre, Claude, Picard Philippe, and Matthieu Simon. 2017. Epidemic risk and insurance coverage. *Journal of Applied Probabilit* 54: 286–303. [CrossRef]
Rhodes, Tim, Kari Lancaster, Shelley Lees, and Melissa Parker. 2020. Modelling the pandemic: attuning models to their contexts. *BMJ Global Health* 5: e002914. [CrossRef]
Smith, Dominic. 2017. Pandemic Risk Modelling. In *The Palgrave Handbook of Unconventional Risk Transfer*. Edited by M. Pompella and N. A. Scordis. Cham: Palgrave Macmillan, pp. 463–495.
Tchuenche, Jean M., Alexander Nwagwo and Richard Levins. 2007. Global behaviour of an SIR epidemic model with time delay. *Mathematical Methods in the Applied Sciences* 30: 733–49. [CrossRef]

Van den Driessche, P., and James Watmough. 2020. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences* 180: 29–48. [CrossRef]

Walters, Caroline E., Margaux M.I. Meslé and Ian M. Hall. 2018. Modelling the global spread of diseases: A review of current practice and capability. *Epidemics* 25: 1–8. [CrossRef]

Watson, H. W., and Francis Galton. 1874. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland* 4: 138–44. [CrossRef]

Zhang, Xianghua, and Ke Wang. 2013. Stochastic SIR model with jumps.*Applied Mathematics Letters* 26: 867–74. [CrossRef]