

QUANTITATIVE FINANCE  
RESEARCH CENTRE



UNIVERSITY OF  
TECHNOLOGY SYDNEY



## QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 409

April 2020

---

# Stochastic Modelling of the COVID-19 Epidemic

Eckhard Platen

---

ISSN 1441-8010

[www.qfrc.uts.edu.au](http://www.qfrc.uts.edu.au)

# Stochastic Modelling of the COVID-19 Epidemic

April 27, 2020

**Eckhard Platen**<sup>\* 1,2,3</sup>

**Abstract:** The need for the management of risks related to the COVID-19 epidemic in health, economics, finance and insurance became obvious after its outbreak. As a basis for respective quantitative methods, the paper models in a novel manner the dynamics of an epidemic via a four-dimensional stochastic differential equation. Crucial time dependent input parameters include the reproduction number, the average number of externally new infected and the average number of new vaccinations. The proposed model is driven by a single Brownian motion. When fitted to COVID-19 data it generates the typically observed features. In particular, it captures widely noticed fluctuations in the number of newly infected. Fundamental probabilistic properties of the dynamics of an epidemic can be deduced from the proposed model. These form a basis for managing successfully an epidemic and related economic and financial risks. As a general tool for quantitative studies a simulation algorithm is provided. A case study illustrates the model and discusses strategies for reopening the Australian economy during the COVID-19 epidemic.

*Key words and phrases:* stochastic epidemic model, stochastic differential equations, squared Bessel process, COVID-19 epidemic, simulation.

1991 *Mathematics Subject Classification:* 91G60, 60H10, 92D30

---

<sup>1</sup>University of Technology Sydney, School of Mathematical and Physical Sciences and Finance Discipline Group, PO Box 123, Broadway, NSW, 2007, Australia,  
Email: eckhard.platen@uts.edu.au, Phone: +61295147759.

<sup>2</sup>University of Cape Town, Department of Actuarial Science.

<sup>3</sup>College of Business and Economics, Australian National University, Canberra.

# 1 Introduction

The COVID-19 pandemic has demonstrated that there exist risks that have not been accounted for. There is an urgent need to assess quantitatively the risks related to an epidemic or pandemic with similar rigour as is common, e.g., for derivatives in quantitative finance. Given the way that viruses are known to mutate and move from animals to humans, it was not a question whether a pandemic could occur, it was only a question when this may happen. Also in future, new epidemics could emerge at any time and require appropriate long-term risk management. To provide a basis for accurate quantitative risk management for an epidemic or pandemic one needs an accurate understanding of its dynamics. This paper aims to provide such an understanding, suitable for risk management in areas such as health, economics, finance and insurance. It proposes a general stochastic model for the COVID-19 and similar epidemics. Fundamental probabilistic properties of the model are deduced and a simulation algorithm is provided.

A particular challenge in modelling epidemics emerges when the proportion of newly infected in the population is not too large, which is typically the case when an epidemic is managed. The challenge in such a situation is to keep a delicate balance between the number of newly infected and the imposed degrees of social distancing and travel restrictions, to allow finally the reopening of the economy. The understanding of the stochastic nature of widely observed fluctuations in the number of newly infected is critically important for decision making in an epidemic like COVID-19, which has relative high mortality and infection rates. Important questions that typically arise are: How much can one relax social distancing to keep with given probability the number of newly infected under a critical level? How important are severe travel restrictions? How many externally new infected are on average permitted to keep with given probability the number of newly infected under a critical threshold? How many susceptibles have to be vaccinated per day to reopen at a targeted date an economy? The proposed model provides in a unified and transparent manner a quantitative basis for answering these and other questions, which will be demonstrated in a case study.

The novel model has the ability to capture the stochastic dynamics of an epidemic in all stages of its evolution, in particular, when the number of newly infected is not too large and the number of newly infected fluctuates considerably. As will be explained in this paper, the key to the understanding of the stochastic nature of an epidemic is the insight that the evolution of the number of newly infected is captured by a generalized, time transformed squared Bessel process; see e.g. Revuz & Yor (1999), Platen & Heath (2010) and Feller (1971).

The proposed model characterizes the dynamics of an epidemic via a four-dimensional system of stochastic differential equations (SDEs) with time and state de-

pendent drift and diffusion coefficient functions. An introduction into the theory of SDEs and methods for their numerical solution can be found in Kloeden & Platen (1999). SDE solutions can be discretized with different time step sizes without changing parameters. A discrete time approximation for the solution of the system of SDEs describing the model is provided in the paper. This algorithm represents a flexible tool that can be generally applied for quantitative risk management related to epidemics using scenario simulation or Monte Carlo simulation.

The literature on epidemic modelling is very rich. Widely used and rather popular are variants of the SIR (Susceptible-Infectious-Recovered) model using deterministic ordinary differential equations; see Katriel (2010) and references therein. Other models employ Markov chains; see e.g. Chang, Harding, Zachreson, Cli & Prokopenko (2020) and references therein. Markov chain models for state variables of an epidemic can become extremely complex, in particular, when many geographical and network details are incorporated. As we show in this paper, a reasonably accurate model needs to consider about four evolving state variables. The complexity of Markov chain models, in particular the characterization of transition probabilities, makes it difficult to obtain on an aggregate level a deeper understanding of the modelled stochastic dynamics.

There exists some literature on the modelling of epidemics with SDEs, in some papers with time delay, where we may refer to Chunyan & Jiang (2014) and references therein, see also Küchler & Platen (2000). An advantage of SDE models is that these aggregate many unimportant minor details that average out when studying in continuous time evolving state variables, e.g. the number of infected. What remains are the core dynamics, characterized locally in time and space via the drift and diffusion coefficient functions. As in the case of the proposed model, SDE models can be often transformed into a standardized form that reveals well-understood fundamental probabilistic properties. These properties give access to accurate qualitative statements and valuations of quantities of interest that would not be possible without the deeper understanding of the underlying standardized dynamics. Beyond that, the modelling with SDEs is not only parsimonious, it provides also elegant access to the machinery of stochastic analysis, which has been the key to accurate quantitative methods in many areas, in particular in quantitative finance.

The paper is organized as follows: Section 2 describes a time discrete model. Section 3 provides the continuous time limit of the model. Fundamental properties of the proposed model are discussed in Section 4. Section 5 fits the model to Australian COVID-19 data and discusses in a case study managing an epidemic and assessing related risks.

## 2 Discrete Time Model for Simulation

This section presents a discrete time model for the evolution of the key state variables that characterize the dynamics of an epidemic. The discrete time model is designed in a way that permits the simulation of scenarios for studying the dynamic properties of an epidemic. The algorithm can also be employed to calculate flexibly via Monte-Carlo simulation quantities of interest.

One time unit is set to one day because most reported data during the outbreak of the COVID-19 epidemic have been provided on a daily basis. Let us introduce the equidistant time points  $t_i, i = 0, 1, 2, \dots$ , where  $t_{i+1} - t_i = 1$  such that  $t_i = i$  is counting the days from the initial time point  $t_0 = 0$ . We consider a population of size  $n_0 > 0$  at the initial time  $t_0 = 0$  that experiences an epidemic outbreak, similar to the COVID-19 epidemic, with time dependent *reproduction number*  $R_i$  at time  $i \geq 0$ . This number can be interpreted as the expected number of infections directly generated by one infected person during the full time the person is infectious if all individuals in the population are susceptible to infection. For the COVID-19 epidemic various sources estimated  $R_t \approx 2.25$  when no social distancing measures were implemented; see e.g. Roser & Ritchie (2020). Strong social distancing, with only essential services working and staying at home, achieves typically a reproduction number of about  $R_t \approx 0.5$ . For simplicity, we assume that a person can only get infected once and when recovered the infected person becomes immune. The parameter  $\sigma \geq 1$  characterizes the *average number of days during which a person infects* other people. Various studies show that for COVID-19 one finds most likely a value of about  $\sigma \approx 4.5$ ; see Roser & Ritchie (2020). The *infection variance*  $\nu \geq 0$  is proportional to the variance of the number of individuals that an infected infects when there are only susceptibles and the reproduction number equals 1.0. The case study at the end of this paper finds that  $\nu = 6$  generates the magnitude of fluctuations that one typically observes. When the entire population were susceptible, the average number of persons at time  $i$  that would become *externally infected* per day is denoted by  $\epsilon_i \geq 0$ .

Our main state variable  $Y_i$  denotes the number of persons that become at the  $i$ -th day *newly infected*. The second state variable  $X_i$  counts the number of *new deaths* at the  $i$ -th day. Our third state variable, denoted by  $Z_i$ , is capturing the number of *non-susceptibles* at time  $i$ . The latter includes the persons who had the disease and recovered, and also those who became immune through vaccination. Finally, the changing size  $n_i$  of the population represents our fourth state variable.

Let the number of newly infected individuals start at time  $t_0 = 0$  with  $0 \leq Y_0 \leq n_0$  and satisfy at time  $i + 1, i = 0, 1, 2, \dots$ , the relation

$$Y_{i+1} = Y_i + \left( \frac{Y_i}{\sigma} (R_i(1 - Z_i/n_i) - 1) + \epsilon_i(1 - Z_i/n_i) \right) + \sqrt{\frac{\nu Y_i}{\sigma} R_i(1 - Z_i/n_i)} U_i, \quad (2.1)$$

if the right hand side of (2.1) is nonnegative and less than  $n_i - Z_i$ . We set  $Y_{i+1} = 0$  if the right hand side of (2.1) is negative, and we set  $Y_{i+1} = n_i - Z_i$  if the right hand side of (2.1) is greater than  $n_i - Z_i$ . Here  $U_i, i = 0, 1, 2, \dots$ , are independent standard Gaussian distributed random variables.

The second summand (in large round brackets) on the right hand side of (2.1) models the trend of the number of newly infected. As long as no external infections occur in the population, the number of newly infected is on average exponentially increasing (decreasing) with *growth rate*

$$g_i = \frac{1}{\sigma}(R_i(1 - Z_i/n_i) - 1) \quad (2.2)$$

when  $R_i(1 - Z_i/n_i)$  is greater (smaller) than 1.0. Social distancing reduces the reproduction number  $R_i$  and, thus, the growth rate. An increased number of susceptibles  $n_i - Z_i$  also decreases the growth rate. There is an additional increase in the average number of newly infected if there are susceptibles and external infections, that is, when  $1 - Z_i/n_i > 0$  and  $\epsilon_i > 0$ . Travel restrictions between populations reduce  $\epsilon_i$ . When the proportion of susceptibles  $1 - Z_i/n_i$  has decreased so that  $R_i(1 - Z_i/n_i) < 1$ , then the average number of newly infected fluctuates around the *reference level*

$$\bar{Y}_i = \frac{\sigma\epsilon_i(1 - Z_i/n_i)}{1 - R_i(1 - Z_i/n_i)}. \quad (2.3)$$

This level is proportional to the average number of new externally infected and increases when the reproduction number increases. This means, relaxing travel restrictions and/or reducing social distancing raises the average number of newly infected.

So far, the average dynamics modelled reflects what typical deterministic SIR-type models capture. However, when the number of infected is not too large, as it is when an epidemic is managed, one observes in reality substantial fluctuations in the trajectory of the number of newly infected. These fluctuations are not primarily due to reporting errors. They represent an important stochastic feature of the dynamics of an epidemic. The proposed model captures this phenomenon. More precisely, the fluctuations of the number of newly infected  $Y_i$  are modelled in the third summand of relation (2.1). In particular, the *variance of the increment of the number of newly infected is proportional to the number of newly infected*.

To understand the reasoning behind this fundamental property of the fluctuations of the number of newly infected, assume that all infected individuals infect each an independent, identically distributed random number of individuals during a given day. Then due to the assumed independence, the variance of the increment of the total number of newly infected equals the sum of the variances of the newly infected caused by each individual that was infected at the beginning of the

day. Thus, the variance of the increment of the total number of newly infected is proportional to the number of individuals that were infected at the beginning of the day. This is a fundamental phenomenon that determines the feedback in the fluctuations of the number of newly infected in an epidemic.

The variance of the number of newly infected in (2.1) turns out to be proportional to the contact intensity, which is the product of the infection variance  $\nu$ , the reproduction number  $R_i$ , the proportion of susceptibles  $1 - Z_i/n_i$  and the inverse of the average number of days  $\sigma$  that an infected person infects others. This means, one observes larger fluctuations of the number of newly infected when the reproduction number is higher or the proportion of non-susceptibles is lower, and vice versa.

For simplicity, we do not consider the cases where individuals move between populations and do not account for births or for deaths that are not caused by an epidemic infection. We also do not consider the possibility that non-susceptibles can become susceptible again. However, immigration, emigration, births, deaths not caused by infection and the possibility that non-susceptibles can become susceptible again could be easily included in the model. Important is that we allow for the possibility that a vaccine becomes available that permits us to immunize susceptibles. We denote by  $\xi_i$  the per day *newly vaccinated* susceptibles at time  $i$ .

Unfortunately, an epidemic as that of COVID-19 causes a relatively high proportion of deaths among the infected. To model the number of deaths we introduce the *mortality rate*  $\lambda_i \geq 0$ , which captures the proportion of persons that pass away at time  $i$  as a result of an epidemic infection. We make the mortality rate time-dependent because this rate may change over time in a population, for instance, when more and more vulnerable become isolated from the majority of the population. In the case of COVID-19 these are elderly and those with prior health conditions. In some developed countries with a good health system a mortality rate of about  $\lambda \approx 0.01$  has been observed; see Roser & Ritchie (2020). However, this may vary considerably.

Let  $\psi > 0$  denote the average *lag time* (in days) between infection and death. A realistic number for this average time seems to be about  $\psi \approx 17$ ; see Roser & Ritchie (2020). It is reasonable to assume that when there was no infection in the population before the time  $t_0 = 0$  that the number  $X_i$  of *daily new deaths* in the population at time  $i = \psi, \psi + 1, \psi + 2, \dots$  takes the form

$$X_i = \lambda_i Y_{i-\psi}. \quad (2.4)$$

With these notations, the number  $Z_{i+1}$  of *non-susceptibles* at time  $i + 1$  equals

$$Z_{i+1} = Z_i + Y_i + \xi_i - X_i, \quad (2.5)$$

as long as the right hand side of (2.5) remains nonnegative and not greater than  $n_i$ . We set  $Z_{i+1} = n_i$  in case the right hand side of (2.5) becomes larger than  $n_i$ ,

and we set  $Z_{i+1} = 0$  should it become negative. We start with an initial value in the interval  $0 \leq Z_0 \leq n_0$ . Note that equation (2.5) is different to similar equations in the literature that uses SIR-type models; see e.g. Chunyan & Jiang (2014). This difference turns out to be crucial for modelling realistically the stochastic dynamics of an epidemic.

When non-susceptible individuals pass away at time  $i$ , caused by an epidemic infection, this reduces the total size  $n_i$  of the population to

$$n_{i+1} = n_i - X_i. \quad (2.6)$$

The four state variables  $Y_i$ ,  $X_i$ ,  $Z_i$  and  $n_i$  evolve jointly over time according to the system of equations (2.1), (2.4), (2.5) and (2.6).

Often reported quantities that can be easily derived from the model include the total number of deaths until time  $i$  :

$$V_i = \sum_{k=0}^i X_k. \quad (2.7)$$

The total number of persons that become infected until time  $i$  :

$$M_i = \sum_{k=0}^i Y_k. \quad (2.8)$$

The total number of recovered follows simply as the difference between the total number of those who became infected and the total number of deaths.

It is straightforward to simulate scenarios of the above modelled dynamics. These scenarios can help to compare the impact of alternative variants of strategies in the management of an epidemic. Furthermore, the above algorithm can be used to evaluate almost any quantity of interest via Monte-Carlo simulation; see e.g. Kloeden & Platen (1999). To gain a deeper understanding of the stochastic dynamics under the model it is extremely beneficial to study the probabilistic properties of the continuous time limit of the above dynamics, which emerges when letting the time step size tend to zero. The respective weak convergence can be secured by theorems given in Chapter 14 of Kloeden & Platen (1999).

### 3 Continuous Time Model

This section presents the continuous time model that follows by weak convergence from the above described discrete time model. The continuous time model is given in the form of a four-dimensional stochastic differential equation (SDE). We keep



the notations and interpretations of the previous section for the parameters of the continuous time model. From the mathematical perspective it is advantageous to normalize the number of newly infected and the number of non-susceptibles by the actual size of the population because the respective normalized quantities become uniformly bounded, which allows us to apply existence and uniqueness theorems for the solutions of the resulting system of SDEs. Therefore, our main state variable  $y_t \approx Y_t/n_t$  becomes the proportion of the population that is *newly infected* at time  $t$ . Our second state variable, denoted by  $z_t \approx Z_t/n_t$ , captures the proportion of the population that is *not susceptible* at time  $t \geq 0$ .

Under the proposed model the proportion of the currently infected population  $y_t$  satisfies the SDE

$$dy_t = \left( \frac{y_t}{\sigma} (R_t(1 - z_t) - 1) + \frac{\epsilon_t}{n_t} (1 - z_t) \right) dt + \sqrt{\frac{\nu y_t R_t}{\sigma n_t} (1 - z_t)} dW_t, \quad (3.9)$$

for  $t \geq 0$  with  $y_0 \geq 0$ . Here  $W = \{W_t, t \geq 0\}$  is the standard Brownian motion that models the uncertainty in the dynamics of the epidemic in  $t$ -time. This uncertainty is driving the fluctuations of the number of the newly infected in the population. Note that when the size of the population is large and the number of infected is so large that one is interested in studying the proportion of newly infected in the population, then the diffusion coefficient in (3.9) can be neglected and there is almost no randomness in the trajectory of the proportion of infected. However, in the case of a managed epidemic the number of newly infected is not too large. This number fluctuates in this case considerably in reality, as can be explained through the SDE (3.9) of the proposed model. Such fluctuations observed in COVID-19 data are often misinterpreted as observation errors. The proposed model gives a clear understanding of these fluctuations, which is important for proper decision making in an epidemic, e.g. when aiming at reopening an economy after a lockdown. For instance, it allows to calculate the probability for upward excursions of the number of newly infected reaching a critical level, as we demonstrate later on.

According to (2.4), the expected proportion of new deaths  $E(x_t) = \frac{E(X_t)}{n_t}$  (per day) at time  $t$  satisfies the differential equation with time delay

$$dE(x_t) = \lambda_t dE(y_{t-\psi}). \quad (3.10)$$

Note that with constant mortality rate  $\lambda_t = \lambda > 0$  the SDE for  $E(x_t)$  follows by (3.9) approximately in the form

$$dE(x_t) \approx \left( \frac{E(x_t)}{\sigma} (R_{t-\psi}(1 - E(z_{t-\psi})) - 1) + \lambda \frac{\epsilon_{t-\psi}}{n_{t-\psi}} (1 - E(z_{t-\psi})) \right) dt. \quad (3.11)$$

It reveals that the expected proportion of new deaths has, with some time delay  $\psi$ , the same growth rate as the proportion of newly infected. More precisely,

one can estimate the by the time  $\psi$  delayed growth rate of newly infected from the growth rate of the number of new deaths. This gives often more accurate estimates for the reproduction number than the reliance on the reported number of newly infected, which is often missing a large number of cases.

The differential equation for the proportion  $z_t = Z_t/n_t$  of non-susceptibles in the population starting with  $0 \leq z_0 \leq 1$  is given as

$$dz_t = (y_t + \frac{\xi_t}{n_t} - x_t)dt \quad (3.12)$$

for  $t \geq 0$ , as long as  $z_t$  stays in the interval  $[0, 1]$ . The proportion of non-susceptibles  $z_t$  is pulled back to the boundaries of this interval when the right hand side of (3.12) pushes  $z_t$  beyond these boundaries, analogous to the boundary behaviour in (2.5).

Finally, we get from (2.6) a differential equation for  $n_t$  in the form

$$dn_t = -x_t n_t dt \quad (3.13)$$

for  $t \geq 0$  with  $n_0 > 0$ .

The fluctuations of the proportion of currently infected  $y(t)$  are modelled in the SDE (3.9) via the diffusion coefficient. We emphasize the crucial modelling feature that the variance of the increments of the proportion of the infected in the population is proportional to the proportion of the infected in the population. This for the stochastic evolution of the number of newly infected crucial feature of the proposed model reflects the fundamental fact that we model the continuous time limit of a generalized birth process. Recall that we explained in Section 2 why the variance of the increments of the number of newly infected evolves proportionally to this number; see Feller (1971) for a similar property. Explanations for the form of the drift and diffusion coefficient functions of the four-dimensional system of SDEs (3.9), (3.10), (3.12) and (3.13) remain analogous to those given in Section 2 for the respective equations.

The four state variables at time  $t$  of the above model are the proportion  $y_t$  of the newly infected individuals in the population, the proportion of new deaths  $x_t$ , the proportion  $z_t$  of susceptible individuals in the population and the size  $n_t$  of the population. These quantities evolve jointly together according to the above four-dimensional system of SDEs. The existence and uniqueness of a strong solution of this system of SDEs can be secured by a combination of respective theorems in Ikeda & Watanabe (1989) and K uchler & Platen (2000). What makes this possible is that the key state variables are bounded, a Yamada condition for the diffusion coefficient of (3.9) is satisfied and the delayed proportion of deaths can be captured as an extra evolving component in a respective five-dimensional Markovian SDE, which satisfies the existence and uniqueness theorem given in Ikeda & Watanabe (1989).

## 4 Fundamental Properties and Strategies

The dynamics of  $y_t$  in (3.9) is that of a generalized, time transformed squared Bessel process; see e.g. Revuz & Yor (1999) and Section 8.4 in Platen & Heath (2010). To deduce fundamental probabilistic properties from its dynamics, we introduce its *intrinsic time*  $\tau_t$  with derivative

$$\frac{d\tau_t}{dt} = \frac{\nu R_t}{4\sigma n_t}(1 - z_t) \quad (4.14)$$

for  $t \geq 0$  with  $\tau_0 = 0$ . The intrinsic time runs faster when the reproduction number is higher and the proportion of non-susceptibles is smaller. It evolves slower for larger populations, which reveals a fundamental probabilistic property of epidemics. Furthermore, we have the crucially important *dimension*  $\delta_t$  of the underlying generalized squared Bessel process given by the formula

$$\delta_t = \frac{4\sigma\epsilon_t}{\nu R_t}. \quad (4.15)$$

Finally, we introduce the *intrinsic growth rate*  $\eta_t$  in the form

$$\eta_t = \frac{4n_t}{\nu} \left(1 - \frac{1}{R_t(1 - z_t)}\right). \quad (4.16)$$

These notations allow us to rewrite the SDE (3.9) as

$$dy_t = (\delta_t + \eta_t y_t)d\tau_t + 2\sqrt{y_t}d\tilde{W}_{\tau_t}. \quad (4.17)$$

Here  $\tilde{W}$  denotes a standard Brownian motion that evolves in the intrinsic time  $\tau_t$ . It aggregates in a canonical form the randomness driving the dynamics of an epidemic.

Due to the standardized form of the SDE (4.17), we can conveniently deduce below several fundamental probabilistic properties of the dynamics of the proportion of the infected population from well-studied properties of generalized squared Bessel processes to be found e.g. in Revuz & Yor (1999) and Section 8.4 in Platen & Heath (2010):

### Eradicating the Disease

When the dimension  $\delta_t$  is less than two, then  $y_t$  becomes *absorbed* at zero with probability one at some random future time. This is an important feature because it tells us due to (4.15) that the ratio  $\frac{\epsilon_t}{R_t}$  has to be less than  $\frac{1}{2\sigma}$  to *eradicate infections*. This quantifies the average number of new externally infected  $\epsilon_t$  that one can allow to emerge in the population and still eradicate infections at some

future time. As mentioned earlier, a realistic choice is  $\sigma \approx 4.5$ , which yields the inequality  $\epsilon_t < 0.11R_t$ . Without social distancing measures taken, that is with  $R_t \approx 2.25$ , this allows on average one externally infected person every four days. Anything more would not keep the epidemic on a path where its infections become eradicated if no new external infections occur.

Interestingly, the number of new externally infected allowed for eradicating infections becomes smaller when social distancing is in place. Say, for social distancing equivalent to a reproduction number of  $R_t \approx 0.5$ , one can only allow on average one externally infected every nine days to achieve eradication of infections. Unfortunately, these are extremely small numbers and they suggest to keep, in practice, the number of externally infected as close as possible to zero, which makes  $y$  a generalized squared Bessel process of dimension close to zero. If one does not respect this fundamental property of an epidemic and allows too many new external infections to occur, then one faces under the model the grim reality that the epidemic may randomly and forcefully break out again. Such outbreak is not too difficult to bring under control when strict social distancing measures are in place. However, when these are too early relaxed or not effective enough implemented, then one has a growing epidemic.

When the dimension  $\delta_t$  is greater than two, then the transmission will never be eradicated because  $y_t$  will never reach zero with probability one. The disease will continue to circulate in the population, supported from time to time by new external infections until there are no susceptibles in the population anymore. This also means that, finally, almost the entire population has the disease when assuming that no vaccine arrives. According to the mortality rate, a respective proportion of the population will in this scenario pass away as a result of the epidemic, which is, unfortunately, a rather large proportion in the case of COVID-19 and should be avoided if possible.

In summary, a real possibility to manage an epidemic successfully is to keep the dimension of the above generalized squared Bessel process extremely close to zero through strong travel restrictions. If one imposes additionally strong social distancing, this makes the growth rate of the newly infected negative, as we will see in the next subsection, and reduces dramatically the average time until eradication. When the eradication is achieved, one can even relax social distancing but has to continue to keep the number of new external infections at zero. Otherwise, a new outbreak emerges almost certainly from new externally infected. It needs to be emphasized that those populations that eradicate the disease have to remain isolated until a vaccine arrives or until all other populations they connect with have also eradicated the disease.

## Flattening the Curve

When the intrinsic growth rate  $\eta_t$  in (4.16) is positive and the impact of externally infected can be neglected, then the number of infected rises due to (4.16) and (4.14) on average exponentially with the *growth rate*  $g_t$  (with respect to  $t$ -time) according to the formula

$$g_t = \frac{1}{\sigma}(R_t(1 - z_t) - 1). \quad (4.18)$$

This growth rate is typically used to estimate the reproductive number from the smoothed version of a visibly exponentially growing trajectory of the number of newly infected. Of strategic importance in managing an epidemic at any of its stages is the ‘flattening of the curve’ of newly infected, which is typically achieved by two measures: First, by imposing strong travel restrictions so that these reduce  $\epsilon_t$  to such a low level that the term  $\frac{\epsilon_t}{n_t}(1 - z_t)$  in the drift of the SDE (3.9) can be neglected. Second, by social distancing that reduces the product of the reproductive number  $R_t$  and the proportion of the susceptibles  $1 - z_t$  to a level clearly below 1.0 so that the growth rate (4.18) becomes clearly negative.

Note that after a first outbreak of an epidemic the proportion of non-susceptibles,  $z_t$ , can be approximately set to zero. When the epidemic is managed reasonably well, this proportion can be also later on neglected until a vaccine is used to immunise the majority of the population. Thus, at the beginning of an epidemic, with social distancing in place that achieves a reproduction number  $R_t$  of about 0.5, one has for  $\sigma = 4.5$  a growth rate of about  $g_t \approx -0.11$ . With such a strategy one can expect the curve of the average number of newly infected exponentially to flatten according to this negative growth rate. The curve of newly infected will, by its stochastic nature, fluctuate visibly when the number of newly infected is in the typical range where epidemics are managed. On the other hand, when there would be no social distancing, the growth rate would be about  $g_t \approx 0.28$ , which yields the typically observed extremely fast exponential growth of a COVID-19 epidemic when not managed through social distancing.

To provide some information about the average number of newly infected, denote by  $\mu_t = E(Y_t)$  the expectation of the number of newly infected. Note that  $y_t$  fluctuates and moves rather independently from the proportion of non-susceptibles  $z_t$ , which is not fluctuating and is growing extremely slowly and almost deterministic. By (3.9) it follows approximately that

$$d\mu_t \approx \left( \frac{\mu_t}{\sigma}(R_t(1 - E(z_t)) - 1) + \epsilon_t(1 - E(z_t)) \right) dt \quad (4.19)$$

for  $t \geq 0$  with  $\mu_0 = Y_0$ . Similarly, we can calculate by using the Ito formula the variance of the number of newly infected, which satisfies approximately the

formula

$$E((Y_t - \mu_t)^2) \approx Y_0 \frac{4n_t}{\eta_t} (\exp(2 \int_0^t \eta_s d\tau_s) - \exp(\int_0^t \eta_s d\tau_s)) + \frac{2\delta_t n_t^2}{\eta_t^2} (1 - \exp(\int_0^t \eta_s d\tau_s))^2 \quad (4.20)$$

for  $t \geq 0$ . It turns out that both equations are very useful and provide valuable predictions for the average value and variance of  $Y_t$ , as can be confirmed via scenario simulation.

## Newly Infected in Local Equilibrium

When the growth rate in (4.18) becomes negative through social distancing and/or increased proportion of non-susceptibles, that is  $R_t(1 - z_t) < 1$ , then a local (in time) *equilibrium* emerges for the number of infected  $Y_t$ . As already mentioned in (2.3), this number fluctuates around the *reference level*

$$\bar{Y}_t = \frac{\sigma \epsilon_t (1 - z_t)}{1 - R_t(1 - z_t)}, \quad (4.21)$$

which does not depend on  $n_t$ . This reference level is lower when  $\epsilon_t$  is lower through stronger travel restrictions or when the reproduction number  $R_t$  is reduced through further social distancing. It is also lower when the proportion of non-susceptibles is higher. Note that when  $R_t(1 - z_t)$  approaches 1.0 from below, then the reference level of the local equilibrium goes to infinity.

The fluctuations around the above reference level occur by (4.20) with the variance

$$E((Y_t - \mu_t)^2) \approx \frac{\sigma \epsilon_t \nu}{2R_t} \left( \frac{R_t(1 - z_t)}{1 - R_t(1 - z_t)} \right)^2, \quad (4.22)$$

which is not depending on  $n_t$ . This variance increases proportionally to  $\epsilon_t$ ,  $\sigma$ ,  $\nu$  and approximately proportionally to  $R_t$  and the proportion of susceptibles  $1 - z_t$ . When the proportion of non-susceptibles reaches 1.0, then there are no longer any fluctuations. Note that when  $R_t(1 - z_t)$  approaches 1.0 from below, then major fluctuations of the number of newly infected are likely. This also means, that when one relaxes social distancing and comes close to 1.0 with  $R_t(1 - z_t)$ , one has a *warning sign when relaxing social distancing* measures: When the number of newly infected fluctuates significantly, then one is most likely rather close to ‘reversing the flattening of the curve’ and should tighten social distancing measures again to go safely back in a local equilibrium. In this way one can ride the curve. These and other properties of the model dynamics, when properly taken into account, can support the management of an epidemic.

In a local equilibrium the number of currently infected  $Y_t$  can be kept with some probability below a critical level  $C_t$ . This critical level may be chosen, for instance, to be proportional to the number of newly infected with severe disease that the

health system could just handle. To calculate the above mentioned probability, one can exploit another fundamental probabilistic property of the model, which is the shape of the stationary density for the number of newly infected. This density is in the discussed local equilibrium a gamma density with its mean given by (4.21) and its variance characterized by (4.22).

The respective gamma density is given by the formula

$$p_t(Y) = \frac{Y^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{Y}{\beta}\right), \quad (4.23)$$

with  $\alpha = \frac{\delta_t}{2}$  and

$$\beta = 0.5\nu\left(\frac{1}{R_t(1-z_t)} - 1\right)^{-1}, \quad (4.24)$$

where

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} ds \quad (4.25)$$

is the gamma function for  $\alpha > 0$ . Thus in a local equilibrium, the probability  $P(Y_t < C_t)$  to keep at time  $t$  the number of infected people  $Y_t$  below a critical level  $C_t$  is about

$$P(Y_t < C_t) \approx \int_0^{C_t} p_t(Y) dY. \quad (4.26)$$

The calculation of this probability can be very useful in assessing social distancing measures for controlling the risk of overloading the health system through random upward excursions of the number of newly infected. By relying on the above gamma distribution it is straightforward to calculate for a targeted probability  $P(Y_t < C_t)$  the respective critical level  $C_t$ , and vice versa. For a given proportion of non-susceptibles  $z_t$  and a number of on average per day externally infected  $\epsilon_t$ , one can via the above relations also identify the level permissible for the reproduction number  $R_t$ . This then translates into the required level of social distancing that has to be imposed to keep the fluctuations below the critical threshold  $C_t$ .

## Achieving Herd Immunity

When managing the epidemic by keeping the number of infected in a local equilibrium with a large probability  $P(Y_t < C_t)$  that avoids overloading the health system, it takes a long time to get the, so called, ‘herd immunity’ to a level that allows removing social distancing and travel restrictions. This is the case when the number of newly infected reaches a local equilibrium where  $R_t(1-z_t)$  is clearly less than 1.0. This means with  $R_t \approx 2.25$  for no social distancing, one needs a proportion of about 60% of non-susceptibles to remove all social distancing measures and still have a local equilibrium. Note that this theoretical proportion is a *critical borderline* and we will refer to it several times later on. Even if a

population has such a high level of non-susceptibles reached, it has, according to formula (4.21), to be very cautious about relaxing travel restrictions because these increase  $\epsilon_t$ . In particular, as pointed out, the average number of newly infected is proportional to the number  $\epsilon_t$  of per day new externally infected. Similarly, by formula (4.22) also the variance of the number of infected is proportional to  $\epsilon_t$ . As described above, one can calculate the probability  $P(Y_t < C_t)$  of staying with  $Y_t$  below a critical level  $C_t$  and enforce travel restrictions and social distancing measures that keep this probability on a targeted level.

To reach the critical border line of 60% of non-susceptibles through immunization by infection and recovery, one has to let a proportion of more than 60% of the population to become infected. This means, for a population of about 25 Million and a mortality rate of 1%, a toll of about 150,000 deaths would be the consequence of that strategy, which would be tragic. Even if the health system could handle about  $C_t = 100,000$  newly infected per day (a large number of infected may show only mild symptoms but at the same time about 1000 infected would pass away each day), then it would take about five months to reach the proportion of non-susceptibles necessary to remove social distancing measures. This would be a horror scenario for the population. Since social distancing would have to be rather strict, parts of the economy would suffer dramatically and large parts of the population would become unemployed. One can clearly see the risks for the vulnerable, the economy and its financial market. The entire society, in the way as it has evolved and as it has been functioning, is here at risk.

If one opens the economy too early, then there would be an explosion of new cases and the epidemic would force its way through the entire population until a high level of non-susceptibles is reached. After that the epidemic would slow down, come into some local equilibrium again and would finally stop the spread of newly infected because there would be finally almost no susceptibles anymore. Many infected would in such scenario pass away without a chance for obtaining reasonable medical care because the health system would not have enough capacity. As a consequence, the mortality rate for infected would be much higher due to the lack of medical care.

## Delaying Flattening the Curve

As mentioned earlier, to get the economy going again without a vaccine and with a low death toll, the leadership of a population has to act ‘hard and fast’ from the very beginning of the epidemic by imposing as soon as possible strict travel restrictions and social distancing measures, aiming at a low number of infections and a fast eradication of the disease. This strategy is realistic and permits to achieve over a rather short period of time the eradication of the disease within an isolated population. However, it requires living conditions, leadership and discipline of the population in implementing social distancing that keeps for a sufficiently long time the reproduction number clearly below 1.0 and excludes



new external infections. In case one cannot avoid a few new external infections, the above strategy does not lead to eradication but keeps the number of newly infected extremely low, which allows to isolate emerging clusters of new infections to keep the overall reproduction number sufficiently low.

If a population does not act as ‘fast and hard’ as above indicated, then it allows the number of newly infected exponentially to grow and many avoidable deaths. Still, even when acting late, through the exponential decline of the number of newly infected under strong social distancing the population can manage and even eradicate the disease. Unfortunately, this takes longer and the total number of deaths may be rather high due to the late response to the arrival of the epidemic. If after controlling the number of newly infected by ‘flattening the curve’ the social distancing measures would be relaxed too early, then another avoidable death toll would arise caused by the then exponentially growing number of newly infected. One would have then to impose stronger social distancing again to control the epidemic again or give this up and let the epidemic force its way through the population, which would be catastrophic.

In a population, as the one of Australia when no new externally infected are assumed, respective calculations indicate that eradication of the disease could be achieved in about two to three weeks after the peak of new deaths with strict social distancing measures in place, equivalent to about  $R_t \approx 0.5$ , and about currently 7 new deaths per day. If one would have instead currently about 700 new deaths, then it would take about two to three months to achieve the eradication of the disease. Important is to note that the second scenario would cause more than hundred times the number of deaths than the first scenario. It should be emphasized that the social distancing could have been practically removed after less than half the time, which would make the economic damage much lower. It should be emphasized that this is not the end state of an epidemic. Even when the disease seems to be eradicated in a population, the vulnerable should be very cautious because the epidemic could still have been unobserved circulated and unexpected new externally infected may appear. Wide, almost complete regular testing and new case tracking would be essential to safeguard the vulnerable and monitor the epidemic. From the above discussion one can clearly see how responsible leadership based on a scientific, quantitatively supported understanding of the epidemic can avoid large numbers of deaths and also significant economic damage.

## **No Possibility for Sufficient Social Distancing**

For populations that are living under conditions that do not allow social distancing at a level where the reproduction number is low enough for a sufficiently long period of time, the arrival of an epidemic like COVID-19 can be catastrophic. When the epidemic begins it may not look so serious because social distancing and travel bans are often implemented, which reduces the growth rate of the num-

ber of newly infected but may not sustainably ‘flatten the curve’. For instance, in many countries a large part of the population needs to continue to work and socialise when performing economic activities needed to survive. Additionally, the individual living conditions for the just mentioned populations do even often not allow sufficient social distancing. For such a population it is very difficult to manage an epidemic successfully because it does not have the main weapon to fight the epidemic, which is social distancing for bringing the reproduction number sufficiently down. The insufficient level of social distancing, when implemented, would only slow the epidemic down, which may give infected with severe disease access to medical care that may reduce the mortality rate. Otherwise, it lengthens the time that the epidemic is evolving.

One possibility for such a population may be to isolate almost totally the vulnerable and let the epidemic take its course with some or without social distancing by providing medical care as much as possible. There is then not much need for travel restrictions from the perspective of this population. The number of non-susceptibles will exponentially rise through the large number of infected and then recovered. The dynamics continues until the proportion of non-susceptibles is so high that there is only a small chance that vulnerable individuals would become infected. This strategy, which may not even have to enforce any social distancing or travel restrictions, could become very painful for the population if the vulnerable would be not well enough protected so that a high death toll would arise.

In the case of COVID-19 the mortality rate may be not so high for on average young populations living in a clean environment. If a young population cannot afford to implement sufficiently long strong social distancing, it can still protect well its vulnerable. This gives it a chance to get through the COVID-19 epidemic fast via ‘herd immunization’ without too many deaths and without major economic damage. Furthermore, no vaccine is needed for this scenario. In this case, a long period of social distancing would most likely harm the economy and could even create a famine or some social uprising.

It is extremely unfortunate that those populations that cannot implement social distancing on a sufficient level and for a sufficiently long period of time, have not many options in the absence of a vaccine.

## Reopening with Vaccination

Let us finally discuss the possibility that a vaccine becomes available that provides immunization towards infections. We denoted by  $\xi_t$  the number of vaccinations per day at time  $t$ . Then the SDE (3.12) captures the evolution of the proportion of non-susceptibles. At the beginning of the epidemic one has almost no non-susceptibles. When managing the epidemic with a low number of new deaths,

then there is for a long time only a rather small proportion of non-susceptibles in the population and one can almost neglect this proportion at the time when a vaccine becomes available. The number of vaccinations per day determines when the critical borderline of about 60% of non-susceptibles will be reached, where another major outbreak is unlikely because the number of newly infected evolves in a local equilibrium even without social distancing. One could at this point slowly start to remove some of the travel restrictions because one would be in a local equilibrium. However, new externally infected would increase the number of newly infected and the number of new externally infected should remain rather small compared to the per day vaccinated.

Assume for the moment again a population of 25 Million. To reach the critical borderline of a level of non-susceptibles of 60% in about one month, one needs to immunize approximately 500,000 susceptibles per day. As the proportion of non-susceptibles increases during the vaccination campaign, the reproduction number can be increased accordingly so that the number of newly infected remains still in equilibrium on a low level. This means, during the vaccination campaign one can step by step relax social distancing. When aiming during the vaccination campaign for a product of reproduction number and proportion of non-susceptibles close to  $\Phi < 1.0$ , one can relax more and more the social distancing because of the then possible higher reproduction number given by the formula

$$R_t = \frac{\Phi}{1 - z_t}. \quad (4.27)$$

In summary, the most likely mistakes that the leadership of a population could make in managing an epidemic like COVID-19 are as follows:

- (i) Not acting ‘hard and fast’ through delay in implementing social distancing and delay in travel restrictions because one wants to keep the economy as long as possible open.
- (ii) Relaxing too early social distancing because one wants to reopen the economy as soon as possible.
- (iii) Underestimating the number of externally infected that may emerge in the population because one needs external workers or has to let some travellers in.
- (iv) Not protecting enough the vulnerables because it seems expensive or is considered to be too restrictive.

## Aggregation of Epidemics in a Pandemic

After closing borders one observes in the COVID-19 pandemic epidemics evolving independently with their own fluctuations in the number of newly infected in different countries. One may ask the question: What is the nature of the dynamics

for the entire pandemic? The somehow surprising and for our understanding important result will be derived below that *the stochastic dynamic of a pandemic is of the same nature as each of its epidemics*.

Let for the  $k$ -th population  $y_t^k$  denote the proportion of newly infected,  $z_t^k$  the proportion of non-susceptibles,  $n_t^k$  its size,  $R_t^k$  its reproduction number and  $W_t^k$  its independent driving Brownian motion at time  $t \geq 0$ ,  $k = 1, 2, \dots, d$  with  $d > 1$ . For simplicity, assume that the proportion of new externally infected per day in each population can be neglected because most travel between populations is severely restricted in a pandemic. Then for the total number  $\hat{Y}_t$  of the at time  $t \geq 0$  newly infected in the  $d$  independently evolving epidemics we obtain from the SDE (3.9) the SDE

$$d\hat{Y}_t = \sum_{k=1}^d \left( \left( \frac{n_t^k y_t^k}{\sigma} (R_t^k (1 - z_t^k) - 1) \right) dt + \sqrt{\frac{n_t^k y_t^k}{\sigma} R_t^k (1 - z_t^k)} dW_t^k \right), \quad (4.28)$$

with

$$\hat{Y}_t = \sum_{k=1}^d n_t^k y_t^k. \quad (4.29)$$

Thus, with average reproduction number

$$\hat{R}_t = \sum_{k=1}^d \frac{n_t^k y_t^k}{\hat{Y}_t} R_t^k \quad (4.30)$$

and average proportion of non-susceptibles

$$\hat{z}_t = \sum_{k=1}^d \frac{n_t^k y_t^k R_t^k}{\hat{Y}_t \hat{R}_t} z_t^k \quad (4.31)$$

we obtain for the proportion of newly infected  $\hat{y}_t = \hat{Y}_t / \hat{n}_t$  the SDE

$$d\hat{y}_t = \frac{\hat{y}_t}{\sigma} (\hat{R}_t (1 - \hat{z}_t) - 1) dt + \sqrt{\frac{\nu \hat{y}_t}{\hat{n}_t \sigma} \hat{R}_t (1 - \hat{z}_t)} d\hat{W}_t, \quad (4.32)$$

with total population size

$$\hat{n}_t = \sum_{k=1}^d n_t^k. \quad (4.33)$$

Here  $\hat{W}$  denotes a new standard Brownian motion with respect to  $t$ -time. One notes that this type of dynamics is the same as that we had before obtained for the epidemic in a single population. Thus, it turns out that one can aggregate the numbers of infected of different populations and still remain in *the same stochastic continuous time model class*. This is a remarkable fundamental fact for the aggregation of epidemics. The dynamics of the newly infected of an entire pandemic is, therefore, again that of a generalized, time transformed squared

Bessel process. This squared Bessel process is of dimension zero if there are no new externally infected. Thus, we can eradicate the pandemic, if we do not get new externally infected through, e.g., transmission from animals.

Interestingly, it follows from (4.14) that the *intrinsic time for the pandemic runs slower* than those of most of the epidemics that form part of it. In particular, its derivative  $\frac{d\hat{\tau}_t}{dt}$  follows by (4.14) in the form

$$\frac{d\hat{\tau}_t}{dt} = \frac{\hat{R}_t}{4\sigma\hat{n}_t}(1 - \hat{z}_t). \quad (4.34)$$

This means that the COVID-19 pandemic, as it forces its way through the world population, evolves much slower than each epidemic in a country. When the dimension  $\hat{\delta}_t$  of the squared Bessel process  $\hat{y}$  is zero, the disease will be eradicated with probability one in a finite random time. Unfortunately, this random time period can be extremely long because the size of the world population is extremely large and the respective time scale in which the pandemic evolves is proportional to the inverse of this size. If hundred years ago a pandemic raged in about two years through the world population and was not much managed, then a similar pandemic would take today about *eight* years. Therefore, a better strategy to manage the pandemic is to let each population eradicate its epidemic before it opens itself up to other populations that have done the same. Smaller self-sustained island populations have an advantage because they can better isolate themselves than populations that are close to neighbouring populations with many natural ties. This predicts that small islands will be the first to eradicate the epidemic. Large populations will take more time to eradicate the disease. They may consider to subdivide their population into smaller parts that can be isolated and eradicate in each part the disease first before opening these up. This would make the process of eradication faster.

A possibility that we may have to face in future is that after a while the gained immunity gets lost. One can easily capture this feature by an obvious extension of the model, which leads to some SDE with an extra time delay.

## 5 A Case Study

It is critical to fit the proposed model to observed data. There exist various sources capturing the evolution of the COVID-19 epidemic in different countries. We rely on the web site of Roser & Ritchie (2020) for the employed Australian data and check whether the evolution of the observed data makes sense under the proposed model.

The reported number of newly infected at the beginning of an epidemic is rather unreliable and underestimates often significantly the number of newly infected. A more informative number is the number of new deaths, which also may not

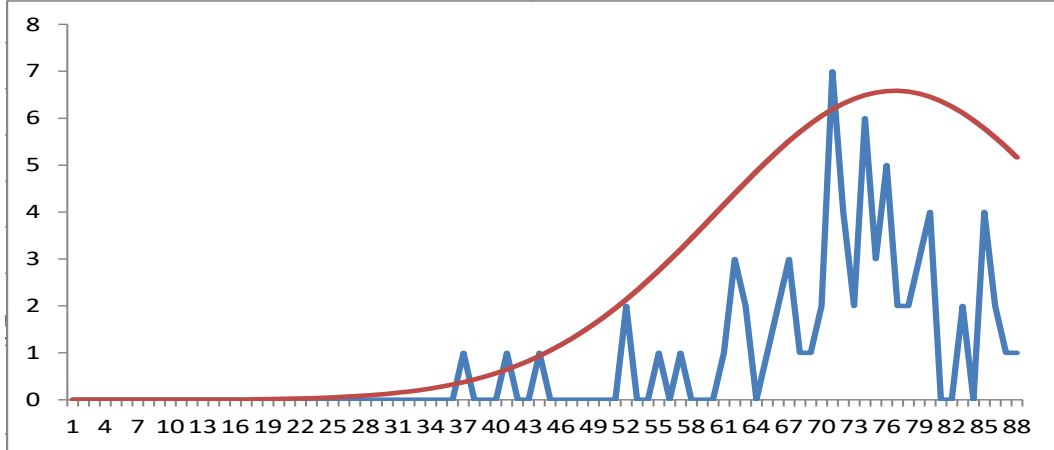


Figure 5.1: Number of new deaths in Australia (blue) and its expectation under the model (red).

be perfectly accurate. Therefore, we display in Figure 5.1 (in blue) the number of new deaths reported in Australia, where we set  $t_0$  to 0 at 25 January 2020, the day when the first case of a COVID-19 infection was reported in Australia. By using the equations (3.10) and (4.19) and the average time delay between infection and death of  $\psi = 17$  days, as well as, a mortality rate of  $\lambda = 0.01$ , we calculate the expected number of deaths under the model, also shown in Figure 5.1 (in red), with the parameters chosen as explained below.

A key parameter is the average number  $\epsilon_t$  of per day new externally infected. Since Australia detected externally infected over several months and still after the ‘curve became flattened’ around the beginning of April, we assume, for simplicity, this parameter to be constant with one external infection every ten days. This seems to be an extremely small number. However, it is sufficient to generate the number of infected observed in Australia. Moreover, the model shows that any number much higher than  $\epsilon_t = 0.1$  seems not to allow fitting the observed data well. This means, it was extremely important to close the borders of Australia soon after the outbreak to avoid larger numbers of external new infections.

Australia has a population with a high proportion of elderly and a respective high mortality rate. Therefore, it aimed at avoiding the high death toll of the ‘herd immunization’ strategy by managing the epidemic with a low number of infections. Fortunately, Australia has a developed health system with significant capacity, can afford to isolate the vulnerable and also an economy that can hopefully sustain a longer partial lockdown. Thus, when managing the epidemic under these circumstances, it comes down to the timing of travel restrictions and social distancing measures and their removal.

After controlling the new external infections through travel restrictions, Australia needed to bring the exponential growth of newly infected rapidly under control. The model clearly shows, to achieve this Australia had to reduce the reproduction

number  $R_t$  below 1.0. Recall that reducing the reproduction number is the primary tool to manage successfully an epidemic. It governs the exponential increase or decrease of the number of newly infected  $Y_t$ .

The latter is determining through the mortality rate  $\lambda$  the number of new deaths  $X_t$ , which we exhibit in Figure 5.1. Under the model the expected number of new deaths  $E(X_t)$  is by (3.10) proportional to the mortality rate  $\lambda_t = 0.01$  and to the by  $\psi$  delayed expected number of newly infected  $E(Y_{t-\psi})$ . This allows us to deduce from the expected number of new deaths the expected number of newly infected

$$E(Y_t) \approx E(X_{t+\psi})/\lambda_t. \quad (5.1)$$

When using an average time span of  $\sigma = 4.5$  days that an infected infects others,

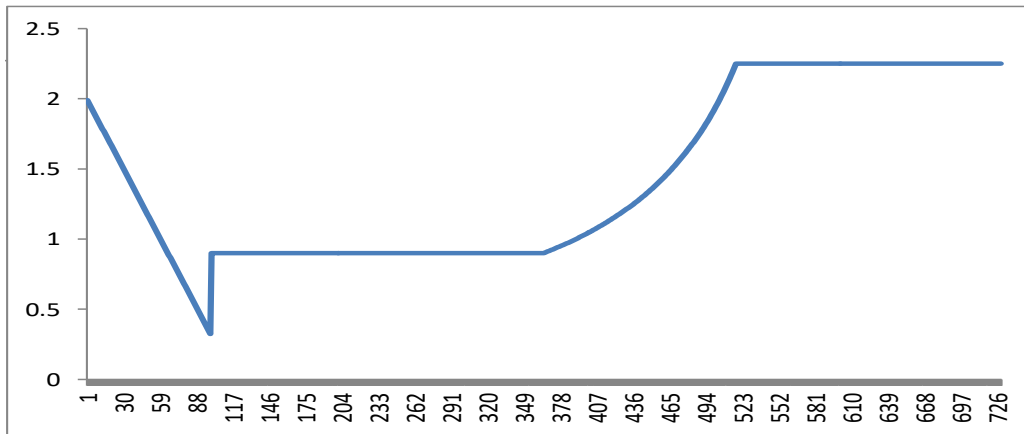


Figure 5.2: Reproduction number for Australia.

the data fitted to the number of observed daily new deaths suggest an approximate evolution of the reproduction number  $R_t$  as displayed in Figure 5.2 in the first part until about  $t = 89$ . The reproduction number starts from a high level of about 2.0, which is only slightly below the level of 2.25 where no social distancing measures would be in place. At that end of January already some social distancing had been implemented in Australia. From that time onward, with every day, social distancing became more and more implemented and practiced. The simplified, steady decrease in the reproduction number during the first period shown in Figure 5.2 is most likely a reasonable reflection of the impact of social distancing on the reproduction number in Australia for March and April in 2020, that is until about  $t = 100$ . It is interesting to note that even a minor increase or decrease in the general level of the fitted declining line in Figure 5.2 for the observed period until about 21 April 2020, that is  $t = 89$ , would change substantially the expected number of new deaths predicted under the model, which is also shown in Figure 5.1. It should be emphasized that in this case study the reported number of new deaths has been used to fit the time dependent reproduction number in the first part of Figure 5.2.

The model allows us to calculate via the differential equation (4.19) the expected number of newly infected  $\mu_t$ . The trajectory of  $\mu_t$  and the reported number of newly infected in Australia are shown in Figure 5.3. One notes that the number

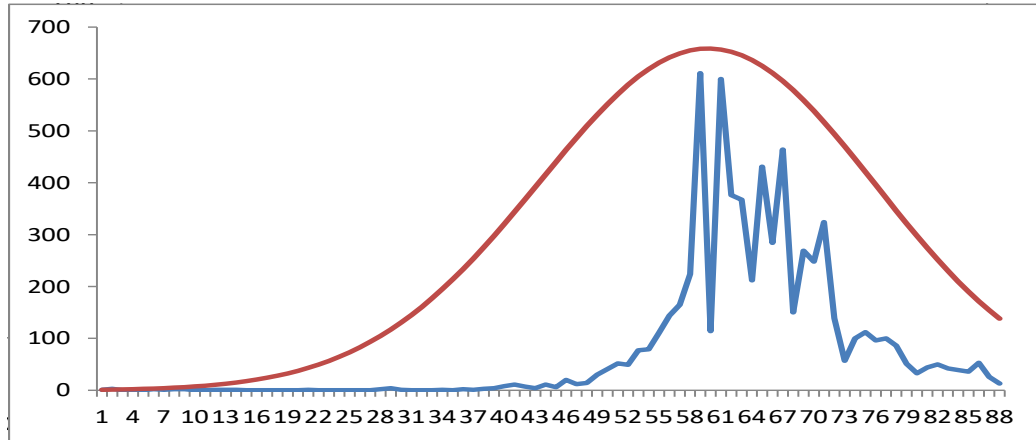


Figure 5.3: Number of newly infected (blue) and its expectation (red).

of reported newly infected (in blue) and its theoretical expected number (in red) evolve approximately similarly in shape what concerns the right hand part from the peak. However, the reported number of newly infected evolves on a lower level. Before its peak the reported number of newly infected appears to be even significantly smaller than what the model would have expected. This is plausible because many newly infected have only mild or no symptoms and were not reported. After some initial time around the time of the peak in Figure 5.3, the public awareness towards COVID-19 became with every day stronger so that the reporting became more accurate but may still not fully reflect the true number of newly infected. Testing became also more widely used, which provided over time a more and more accurate number of newly infected.

Let us now position ourselves at the end of the observed data set on 21 April 2020, that is  $t = 89$ . By using its properties, the model allows us to describe probabilistically the future evolution of the epidemic under the model using an assumed parameter set. Due to the small number of newly infected the numbers of new deaths and newly infected fluctuate strongly. As we explained, this is a property of the underlying stochastic dynamics and causes no problem because we can use our theoretical understanding of the evolution of the epidemic to predict probabilistically its path under given assumptions.

As indicated in the previous section, a recommendable strategy would be to leave the increased implementation of social distancing in place until, say, about 2 May 2020, that is  $t = 100$ , which is three weeks after the peak of the new deaths has passed. An on average further decreasing but fluctuating number of new deaths and daily new infections can be expected during the time interval  $[89,100]$ . Around 3 May of 2020, that is  $t = 101$ , it should then be safe to move



into the by the model predicted equilibrium regime.

It is clear from our discussions in the previous section, to secure the equilibrium regime the product  $R_t(1 - z_t)$  of the reproduction number and the proportion of susceptibles has to remain clearly below 1.0. Therefore, it should be possible around 3 May 2020 to ease the social distancing rules so that they still ensure a reproduction number of, say, about  $R_t = 0.9 < 1.0$ . This number can be confirmed as a suitable number through scenario simulation for the given set of parameters, which includes the parameter  $\nu = 6$ . We assume in Figure 5.2 that a reproduction number of 0.9 has been achieved for the period from  $t = 101$  until  $t = 365$ . The easing of social distancing after  $t = 101$  should be slowly and carefully executed to make sure that one does not reach a reproduction number too close to 1.0 or even above. As we pointed out, a *warning sign* would be a *strongly fluctuating number of newly infected*. This should raise an alarm and tougher social distancing would have to be implemented again.

As discussed in the previous section, what would be a severe mistake is the easing of travel restrictions with countries that have still infections circulating. One can see from equation (4.21), the average number of newly infected rises proportionally to the average number of per day new externally infected. This means, even when keeping social distancing on a level where one remains in a local equilibrium, the number of newly infected can become rather large when removing travel restrictions.

In a local equilibrium it is important to know accurately how large the number of newly infected may become when it goes through an upward excursion. The gamma distribution, as stationary distribution for the equilibrium dynamic, is given in (4.26) and allows us to characterize the probability for the number of newly infected to be found below a given threshold  $C_t$  at a time  $t$ . For the parameters we assume for the equilibrium regime after  $t = 101$  and until  $t = 365$ , the respective gamma distribution is shown in Figure 5.4. We note that the probability to have at a day less than 55 newly infected is above 0.99. Just as a side note, the average number of daily new infected  $\bar{Y}_t$  is here by (4.21) around  $\bar{Y}_t \approx 4.5$  and its deviation is by (4.22) about  $\sqrt{E((Y_t - \mu_t)^2)} \approx 11$ . Recall that  $\bar{Y}_t$  is the average that the number of newly infected approaches asymptotically, when we still allow one new externally infected every ten days. In case we would have no new externally infected, the number of newly infected would soon drop to zero and the disease would be eradicated for the moment in Australia.

Some equilibrium regime, as the one described, is likely to become for several months the new normal until a vaccine becomes available, unless the Australian population succeeds through strong social distancing and isolation of new cases to eradicate the disease. This scenario seems to be preferable but requires strong social distancing and tracking and isolation of new cases. Once the disease would be eradicated many social and economic activities could go back to ‘normal’. Still it would be wise for the vulnerable to remain very cautious until a vaccine arrives and they would become vaccinated.

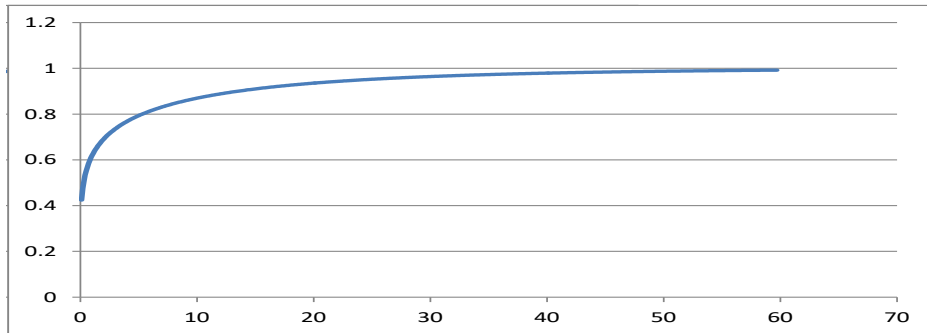


Figure 5.4: Stationary distribution of newly infected.

Assume that a vaccine could be rolled out to the population about one year after the first infected was reported in Australia, which we set hypothetically for our case study to be on 22 January 2021, that is  $t = 363$ . Then it depends on how fast the immunization could be achieved. Until that date the proportion of non-susceptibles would be still almost negligible.

As indicate in the previous section, when vaccinating fast one can take advantage of the increasing proportion of non-susceptibles. When aiming during the vaccination campaign for a product of reproduction number and proportion of non-susceptibles close to 0.9, one can relax more and more social distancing equivalent to a level of

$$R_t = \frac{0.9}{1 - z_t} \quad (5.2)$$

for the reproduction number. For instance, when starting vaccinating on 24 January 2021 ( $t = 365$ ) with a daily number of immunizations of about  $\xi = 100,000$ , Figure 5.2 shows the increase of the respective reproduction number after  $t = 365$ . It takes then about 156 days until the level of 2.25 for the reproduction number is reached. This means that no more social distancing is required after  $t = 521$ . The vaccination campaign would take about 5 months, however, it could end much earlier if the number of vaccinations per day would be much higher. Still, only about 60% of the population would be non-susceptible at that time  $t = 521$  and it would take several more months to vaccinate the entire population. If the vulnerables would be vaccinated first, then one would have fewer potential deaths.

To round up the case study for the Australian COVID-19 epidemic, we use scenario simulation to generate a path of an approximate solution of the SDE (3.9) by using the algorithm described in equation (2.1) and related equations. Figure 5.5 shows a path of the simulated number of per day newly infected (in blue) and its expectation (in red). We see the similarity between both paths and also some upward excursions of the newly infected that randomly occur but do not lead

far away from the expected number. The probability for the number of newly infected to remain during the period from  $t = 101$  until  $t = 365$  below a certain level  $C$  is captured in Figure 5.4. An interesting feature shown by formula (4.22) is that after  $t = 101$  the variance of the number of newly infected decreases more and more as the proportion of non-susceptibles comes closer to 1.0. The scenario simulation confirms that at the time  $t = 628$ , when the proportion of non-susceptibles reaches 1.0, the variance of newly infected declines to zero, as predicted by formula (4.22).

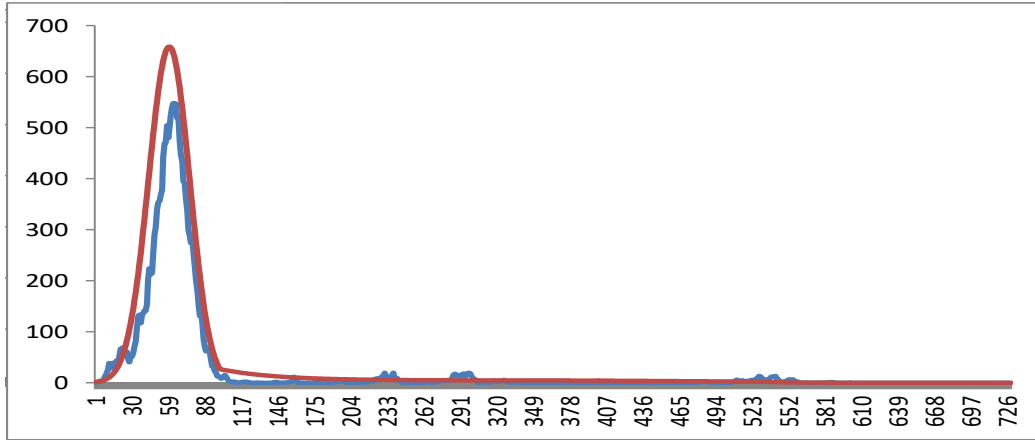


Figure 5.5: Simulated scenario of newly infected (blue) and its expectation (red).

## Conclusions

The proposed model provides an accurate and efficient tool for managing, in a quantitative manner, an epidemic but also related risks. The provided algorithm for scenario simulation is widely applicable. Furthermore, the revealed fundamental probabilistic properties of the dynamics of the model give important insights into the quantitative stochastic behaviour of an epidemic. In particular, they allow the calculation of quantities that are critical for strategies that aim to manage an epidemic. The model covers realistically the dynamics of small numbers of newly infected, which fluctuate in reality. It is the power of the exponential growth that makes the epidemic a deadly enemy. However, the exponential growth is also the only strategic weapon that can be used, when there is no vaccine available. An epidemic can be brought under control by making the growth rate negative through social distancing. It is shown that under severe travel restrictions and with strong social distancing one can bring the number of newly infected to an extremely low level or even eradicate the disease in a relatively short time. It is the responsibility of the leadership of a population to harness this exponential power to beat an epidemic. The quantitative relationships that this paper outlines generalize widely used results, known for popular SIR-type

models, to a new level of reliability. They capture realistically, in a unified manner the different possible regimes of the stochastic dynamics of an epidemic. The proposed model reveals through its probabilistic properties a deeper understanding of an epidemic. These properties can be accurately quantified and exploited to support the managing of an epidemic or perform quantitative risk management in health, finance, economics or insurance. Forthcoming studies of the COVID-19 epidemic for different populations will demonstrate that the model reflects extremely well reality and can provide crucial support for decision making to save lives and economic value.

## References

- Chang, C. C., N. Harding, C. Zachreson, O. M. Cli, & M. Prokopenko (2020). Modelling transmission and control of the COVID-19 pandemic in Australia. *arXiv:2003.10218v2*, 27.
- Chunyan, J. & D. Jiang (2014). Threshold behaviour of a stochastic SIR model. *Applied Mathematical Modelling* **38**(21-22), 5067–5079.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed.), Volume 2. Wiley, New York.
- Ikeda, N. & S. Watanabe (1989). *Stochastic Differential Equations and Diffusion Processes* (2nd ed.). North-Holland. (first edition (1981)).
- Katriel, G. (2010). Epidemics with partial immunity to reinfection. *Mathematical Bioscience* **228**, 153–159.
- Kloeden, P. E. & E. Platen (1999). *Numerical Solution of Stochastic Differential Equations*, Volume 23 of *Appl. Math.* Springer. Third printing, (first edition (1992)).
- Küchler, U. & E. Platen (2000). Strong discrete time approximation of stochastic differential equations with time delay. *Math. Comput. Simulation* **54**, 189–205.
- Platen, E. & D. Heath (2010). *A Benchmark Approach to Quantitative Finance*. Springer Finance. Springer.
- Revuz, D. & M. Yor (1999). *Continuous Martingales and Brownian Motion* (3rd ed.). Springer.
- Roser, M. & H. Ritchie (2020). Coronavirus disease (COVID-19) the data. Retrieved from: <https://ourworldindata.org/coronavirus>.