ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis: MSc Econometrics and Management Science

Predicting COVID-19 Infections in France: Probabilistic State-Space SIR model

April 30, 2021

Author: Louis Perriens, 429200
Supervisor: dr. Paap, Richard
Second assessor: dr. Wang, Wendun

Abstract

This paper used a probabilistic state space model motivated by the deterministic SIR model to estimate and forecast the number of new COVID-19 infections in France, from July 1, 2020 until July 21, 2021. The effect of vaccination and governmental policies to control the spread of the virus have been integrated in the estimation and forecasting process. Through the artificial specification of different forecasted policies, we are able to assess the effectiveness of different strategies authorities could impose to successfully control the outbreak.

Keywords: COVID-19 ; SIR ; State-space model ; Bayesian Inference.

Contents

1 Introduction

The outbreak of the disease known as COVID-19 originated in Wuhan, the capital city of the Hubei province in China, and quickly rippled through the world, disrupting lives and economies alike. According to data published on John Hopkins University[1], by April 13, 2021, there were 137 million confirmed cases, and the pandemic had claimed 2.95 million lives, according to the data published on John Hopkins University[2].

In addition to the dramatic loss of lives, the pandemic presents a challenge to public health, food security and the world of work, according to a joint statement from Joint statement by the International Labour Organization (ILO), the Food and Agriculture Organization (FAO), the International Fund for Agricultural Development (IFAD), and the World Health Organisation. According to the World Health Organization et al. (2020), "tens of millions of people are at risk of falling into extreme poverty, while the number of undernourished people, currently estimated at nearly 690 million, could increase by up to 132 million by the end of the year. Nearly half of the world's 3.3 billion global workforce are at risk of losing their livelihoods. (...) Countries dealing with existing humanitarian crises or emergencies are particularly exposed to the effects of COVID-19. Responding swiftly to the pandemic, while ensuring that humanitarian and recovery assistance reaches those most in need, is critical."

Since the beginning of the outbreak, a lot of research has been carried out on the SARS-CoV2 and the disease it causes. Consequently, we now have a fairly good understanding of the disease, its transmission channels and policies effective in tackling it. In addition to vaccination, quarantining, medical isolation, hand-washing, mask-wearing, as well various degrees of community lockdowns have proved successful in slowing the spread of the virus Teslya et al. (2020). Though tremendous efforts have been put into the development of multiple vaccines (Moderna, AstraZeneca, Pfizer, Johnson & Johnson, Sputnik), this hope is overshadowed by the apparition of various mutations which threaten the efficacy of vaccines, and risk prolonging the duration of the pandemic[3]. Furthermore, development of the said vaccines has been fast-tracked. Their side effect profile has thus so far not been fully characterized, which leaves some uncertainty on their safety, and influences the willingness to receive them. Recently, doubts emerged about the safety of the AstraZeneca vaccine (Wise (2021)). As there is additionally considerable disparity in vaccines distribution throughout the world, favouring countries whose economies are stronger, their ability to control the pan-

[1]https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html/bda7594740fd40299423467b48e9ecf6
[2]https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html/bda7594740fd40299423467b48e9ecf6
[3]https://www.nytimes.com/2021/04/03/health/coronavirus-variants-vaccines.html

demic in the short term appears uncertain[4].

Tackling the pandemic effectively with swift responses by authorities to alleviate the burden on health care systems requires to have a good understanding how the disease will likely spread. Indeed, the public interest mainly lies on how the number of infected people will evolve, as well as on when the pandemic will converge to an end. This paper will tackle these questions through a prediction model using COVID-19 data from July 1, 2020 up to (and including) April 13, 2021 in France. We forecast on the future course of the pandemic in France until July 21, 2021, and model how different policies and human intervention to tackle COVID-19 determine the rate at which the disease propagates. To this end, we integrate a simple time-varying function into the model to describe the impact of human intervention on the rate of infection. This perspective was highlighted in Song et al. (2020), who introduced the concept of the transmission rate modifier within COVID-19 prediction framework. Forecasts on pandemic metrics are of extreme relevance to policy makers as these allow to assess which countermeasures will be effective in tackling the pandemic at its various stages, to efficiently allocate resources such as ventilators and medical staff, and overall allow for timely intervention in order to avoid snowball effects.

The epidemic process will be modelled through a variation of the Susceptible-Infected-Removed (SIR) model, which is a widely used model in epidemiology, originally developed by Kermack and McKendrick (1927). The SIR model is a compartmental model in which the population is partitioned into the three Susceptible-Infected-Removed respective groups. We integrate the SIR model into a probabilistic State-Space model, in which the observation equation consists of the measured time series of infections and removals (i.e. recoveries and deaths), and the state equation consists of the true underlying prevalence of infections and removals. The model is modified to allow for time-varying transmission modifiers, which will reflect government interventions imposed to slow down the spread of the COVID-19 virus. The main goal of this paper will be to forecast the evolution of the population within the three compartments of the SIR model in France, but we shall also determine some keys metrics regarding the pandemic evolution, such as potential end-points for the pandemic and the basic reproduction number (denoted R_0). This is defined as the average number of infections caused by a single case in a completely susceptible population. Additionally, we shall also do some policy evaluation to assess the efficacy of government intervention.

The outline of this paper will be as follows. In the next section, the current literature on

[4]https://www.theguardian.com/world/2021/mar/30/coronavirus-vaccine-distribution-global-disparity

similar models will be reviewed. In section 3, we present the French COVID-19 data used, and explain the scaling applied to it. In Section 4, the model will be presented in detail, along with the transmission rate modifier for different policies. Next, Section 5 presents the empirical results for France. Finally, there will be a general discussion of the findings and limitations of this paper.

2 Literature Review

Many variations of the SIR model have been proposed in the existing literature, for example through the incorporation of additional compartments within this framework. Such an example is the Susceptible-Exposed-Infected-Removed (SEIR) model, where individuals experience an incubation period prior to become infectious. These individuals are incorporated into the "Exposed" compartment. This variation was used by He, Peng, and Sun (2020) to model the COVID-19 outbreak within Hubei province in China. Another common extension to the SIR model is the Susceptible-Infected-Removed-Susceptible (SIRS) model, which does not add a new compartment to the model. This variation allows for re-infection after recovery due to imperfect immunity, hence we observe a cyclical flow of people through the three compartments.

A Hierarchical Bayesian SIRS model was proposed by Zhuang and Cressie (2014). Note however that many complications arose in the estimation for their specification. In all SIR models, it must hold that the sum of the proportions of all compartments must be one. This is relation is known as the balance equation. Zhuang and Cressie (2014) argued that pandemics are by nature random, hence the measurements of these proportions are subject to error, which leads to their sum not being one. In order to main the balance equation, they worked in terms of the log odds ratios of susceptible-over-removed and infectious-over-removed populations. This effectively changed the scale for their compartment from [0,1] to $(-\infty, \infty)$, and argued that on this scale, small errors would have no effect on the balance equation. While this did solve the problem of measurement errors on the counts for each compartment, the transformation made the results less interpretable.

Another variation the SIR embeds the model into a probabilistic State-Space framework, as was done by Shaman and Karspeck (2012). This approach exploits the forecasting power of compartmental SIR models while accounting for sources of uncertainty in a probabilistically consistent manner, specifically for measurement errors, as they assume no process error.

Shaman and Karspeck (2012) used this approach to model and forecast the influenza flu's disease transmission mechanisms within the SIRS framework, assuming normally distributed measurement errors and estimated the model parameters used a method called the ensemble adjustment Kalman filter (EAKF). A similar approach was implemented by Dukic, Lopes, and Polson (2012) who used State-State SEIR model, with a simulation-based approach including process errors. They use a particle filtering algorithm in combination with sequential Bayesian learning to track (not forecast) the flu's transmission in real-time and give uncertainty estimates each new surveillance data point. Both Dukic, Lopes, and Polson (2012) and Shaman and Karspeck (2012) used the normality assumption for both state and observation equations.

Although the models proposed by Dukic, Lopes, and Polson (2012) and Shaman and Karspeck (2012) allow for measurement errors to be accounted for, it is possible to make different (non-normal) distributional assumptions for the State-Space equations, as was done in Osthus et al. (2017). They choose to work within the SIR framework over more complex variations of the latter because of its parsimonious nature, and because of the ability to leverage analytical relations between the latent parameters of the SIR model and functions of the observed data when specifying priors. Indeed, without observing how people transition from being recovered to becoming susceptible again (on a macro scale), or without precise knowledge on the rate of loss immunity or the incubation periods, it is hard to identify and estimate the SEIR and SIRS models. Osthus et al. (2017) implemented a Dirichlet-Beta State-Space model (DBSSM) to make forecasts on the seasonal influenza in the USA. More specifically, they used data and metrics from previous flu seasons to make forecasts on new flu seasons. The choices of the Beta distribution for the observation equation and Dirichlet distribution for the state equation is natural, as these distributions match the supports of the quantities in questions: the vector y_t represents shares (in percentages) of the Infected and Removed compartments, hence these are bounded by [0,1], while the latent space vector θ_t was assumed to follow a Dirichlet distribution, which preserves the balance equation.

While the model proposed by Osthus et al. (2017) offers extremely relevant distributional choices for its State-Space equations, it requires data from previous flu seasons to estimate it, hence it is not directly applicable to the COVID-19 pandemic. Instead, Song et al. (2020) proposed a slightly simplified version of this model, still within the Dirichlet-Beta State-Space model framework, but requiring only data from the current outbreak. Additionally, they extend the SIR methodology through the introduction of a time-varying modifier to the rate of infection (briefly mentioned in Section 1), along with another extension that allows for

Quarantine compartment to be added to the existing three compartments of the SIR model.

A similar methodology was followed in Kobayashi et al. (2020), with respect to the implementation of the Dirichlet-Beta state-space model with the modifier on the rate of infection to predict future COVID-19 infections in Japan. Most notably, they offer a warning on how one should proceed when setting the rate of transmission modifier. They argue that in theory a reduction of mobility by 80% should be enough to terminate the pandemic, which is what the Japanese government was aiming for when they introduced the state of emergency on April 7, 2020. But they explain the actual reduction was more in the order of 60% to 70% because the state does not have the legal force to regulate all individual activities.

In our model, we consider a State-Space approach to the SIR over its deterministic SIR counterpart, as the deterministic version fails to explain the non-ignorable randomness in the epidemic process, while the State-Space approach allows us to account for multiple sources of uncertainty in the pandemic process. We opt for the classic SIR model because of its parsimonious nature and its interpretability. We follow in the steps of Osthus et al. (2017), making the same distributional choices for our observation and state equations. Additionally, this approach circumvents the need to transform the data, as was done in Zhuang and Cressie (2014).

As mentioned previously, we also integrate a modifier to the rate of transmission as proposed in Song et al. (2020), however we heed the warning by the Kobayashi et al. (2020) about how to specify this modifier, which will be set somewhat conservatively, as to reflect reality. Unlike Song et al. (2020) and Kobayashi et al. (2020), our observation period is very long, as we observe two waves in infections in our data, which should yield an interesting overview of how the pandemic evolved over time, and the effect policies had on the number of infections.

Lastly, unlike most papers, we choose to work with probable cases instead of confirmed cases, to tackle the issue of under-reporting. This allows us to get pragmatic results with respect to forecasting the pandemic.

3 Data

The data that is used is collected by the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE)[5]. It is composed of three daily cumulative time series: confirmed infected cases, recovered cases and deaths. We used data from July 1, 2020 until (and including) April 13, 2021. Because of the large range of the observation period, and the high variability in the daily reports, we decided to transform daily data into weekly data. Hence, we obtain an observation period of exactly 41 weeks.

In order to convert these cumulative time series to new incidences, we applied the first-difference operator to them. Upon further examination of the data, it became clear there were several caveats linked to under-reporting and the way the data was collected: we notice that on average, the time series for confirmed infections is larger than the series for confirmed recoveries, by a factor of approximately 18. This could be due to that recovery observations are collected majoritarily for patients who have been hospitalised. Additionally, it is known that the death rate for COVID-19 is around 2%, yet, in our sample we observed a death rate of 5.4%.

Under-reporting of the infected cases is a big problem when it comes to modelling COVID-19, and happens for a variety of reasons. Firstly, there is a large number of asymptomatic cases, which can go undetected and individuals continue to propagate the disease without being aware of it. Secondly, there is a disincentive for mildly affected individuals to get tested, because upon clinical confirmation of infection, they are required to follow strict self-quarantine measures. Instead, some people self-diagnose, and simply isolate themselves until they deem the symptoms to have passed. Thirdly, under-reporting is also due to imperfect testing accuracy, partially due to variability of the incubation period. Finally, the availability of testing infrastructure and reagents limits the ability to diagnose infections and carry out mass-scale testing. Since undetected cases are not quarantined, they are expected to transmit the disease at an accelerated rate (Deo and Grover (2020)).

To make a distinction between probable cases and confirmed cases, we use the definition from European Parliament's Policy Department for Economic, Scientific and Quality of Life Policies. A probable case is defined as: "Any person meeting the clinical criteria with an epidemiological link, or any person meeting the diagnostic criteria", and a confirmed case:

[5]https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

"Any person meeting the laboratory criteria". A study in the United States by Wu et al. (2020) suggests that by April 18, 2020, the number of probable cases could be anywhere between 3 to 20 larger than the number of confirmed infected cases.

In order to match the reality of the pandemic as closely as possible, we work in terms of probable cases in this paper. The data is up-scaled through the exploitation of some known relations between the time series: we upscale the confirmed cases time series by a factor of 3, such that the mean death rate for the new times infections time series is 1.8%, which is close the estimated empirical death rate for the virus. Next, we upscale the recovered time series by a factor of 50 (approximately 18x3), such that the new cumulative series of recovered cases is relatively close to the new cumulative series of infections. For this scaling, we used the ratio of cumulative infections to recoveries from data for Italy: on April 14, 2021, Italy accumulated 3.809.193 confirmed infected cases, against 3.178.976 recovered cases (ratio of infections to recoveries of 1.19). For the up-scaled French data, on April 13, 2021, we end up with 15.091.203 cumulative probable infected cases, against 13.981.200 cumulative probable recovered cases (ratio of 1.07).

Note that since we are dealing with weekly data for probable cases, the data points seem quite inflated. For the sake of comparison, two plots are presented below. Figure 1 presents the daily data of confirmed cases and recoveries without any scaling, and in Figure 2 we show the weekly data of probable cases, after scaling has been applied. In these plots, infections are plotted in red, while removals are plotted in blue.

Figure 1: New daily confirmed cases, without scaling.

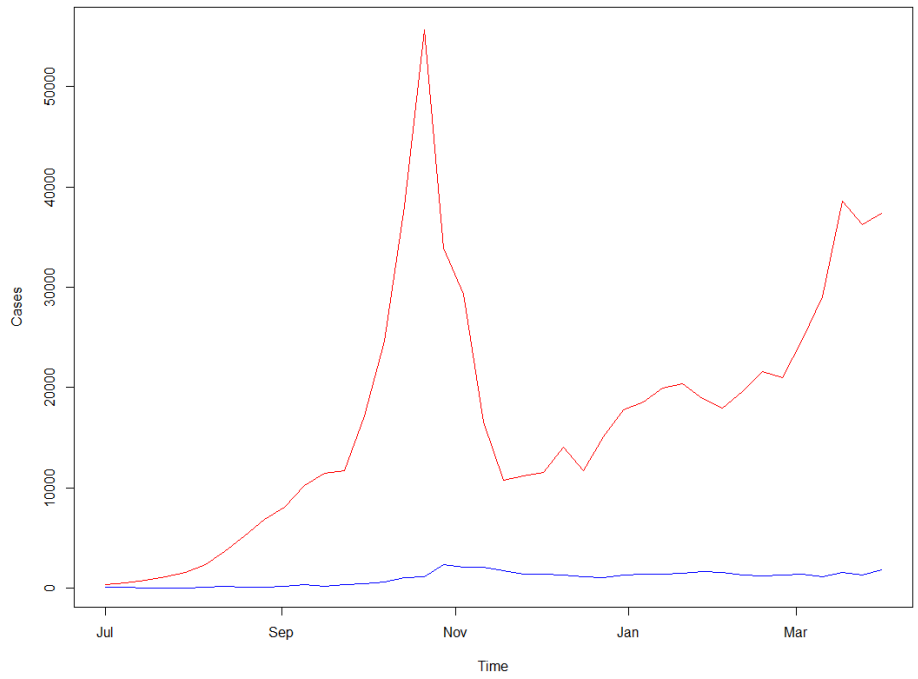
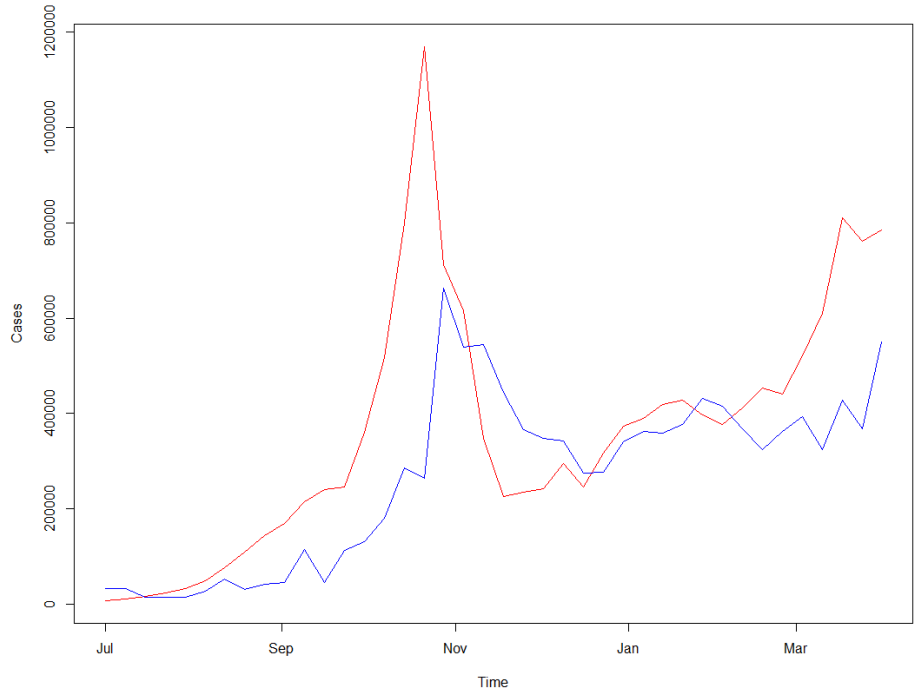


Figure 2: New weekly probable cases, after scaling up.



4 Model

To model the time series described in the previous section, we use the Susceptible-Infectious-Removed model, which we integrate into a Beta-Dirichlet State-Space model. The outline of this section is as follows: first we briefly describe the classic SIR model, and after that present the Beta-Dirichlet State-Space model in which the SIR model is incorporated. Then we present how we extend this model through the integration of government intervention in the estimation process. Lastly, details regarding posterior inference and specification prior distributions are described as well.

4.1 Susceptible-Infected-Removed (SIR) Model

The SIR was developed by Kermack and McKendrick (1927). It is a compartmental model which partitions a population into three compartments: Susceptible (S) to contract a disease, Infected (I), or Removed (R) which is a joint compartment for recoveries and deaths. The proportions of the population in these three respective categories are denoted S_t, I_t and R_t, such that at any given time t, any member of the population must be part of exactly one of these compartments, and the balance equation $S_t + I_t + R_t = 1$ holds at all times. The SIR model is accompanied by a number of assumptions:

I) The population is closed meaning that we observe no new births or exogenous deaths (deaths from another source than the disease), and there no migration in or out of the population.
II) The disease has a latent period of zero, meaning that individuals become infectious right after they contract the disease (i.e. no incubation period).
III) Recovering from the infection grants lifelong immunity.
IV) The population mixes homogeneously.

The SIR model describes the evolution of the pandemic over time, and the flow of the population within its three compartments is defined by the following a set of ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\beta S_t I_t, \tag{1}$$

$$\frac{dI}{dt} = \beta S_t I_t - \gamma I_t, \tag{2}$$

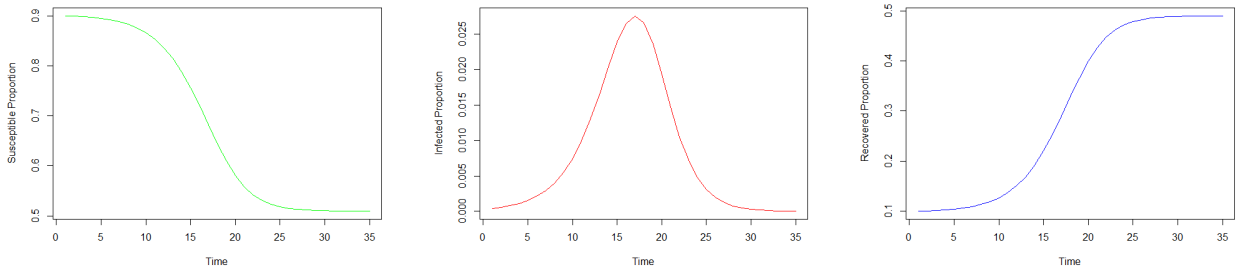
$$\frac{dR}{dt} = \gamma I_t, \tag{3}$$

where $\beta > 0$ is the rate of transmission and $\gamma > 0$ is the removal rate. Conceptually, as susceptible individuals contract the disease and become infectious, they move from the Susceptible to the Infectious compartment, and they eventually recover (or die), moving from Infectious to the Removed compartment.

The rate at which individuals circulate though these categories depend on the proportion of the population in the respective compartments at a given time, as well as the transmission and removals rates associated with the disease, which jointly (and deterministically) control the pandemic process. In the SIR model, the basic reproductive number R_0 can analytically be derived from the rates of infection and removal, such that $R_0 = \beta/\gamma$. To get a better understanding of the dynamics of the SIR model, simulated SIR curves for are shown in Figure 3, for $\beta = 2$, $\gamma = 1.4$, $S(0) = 0.9, I(0) = 0.0004, R(0) = 1 - S(0) - I(0)$:

Figure 3: Simulated SIR Curves

(a) Susceptible (b) Infectious (c) Removed



Note that the SIR keeps track of proportions in the three compartments, but in this paper, our main interest is the infections and removals compartments.

4.2 Dirichlet-Beta State-Space Model (DBSSM)

The SIR model described above is embedded in the Dirichlet-Beta State-Space model. The DBSSM replicates the dynamics of a deterministic SIR model while remaining more flexible than the traditional deterministic SIR model, as it allows for various sources of uncertainty to be accounted for. In particular, the model accounts for uncertainty in the SIR model parameters, uncertainty in the pandemic's underlying transmission dynamics (process error) and uncertainty in the data itself (measurement error). This is done through the simultaneous specification of the latent disease transmission mechanism and the observed data within

the state-space framework.

Let the two weekly time series Y_t^I and Y_t^R described in Section 3 represent the new probable infected cases, and probable removed cases respectively. We observe another time series for new deaths, but this information will be incorporated into Y_t^R, such that Y_t^R is the sum of the new probable recoveries and new confirmed deaths (the death time series is not scaled up). These series were made into proportions of whole population by dividing them by $N = 6.5387.859$, where N is the total population of France. Hence, we have $0 < Y_t^I, Y_t^R < 1$, with $Y_t^I + Y_t^R + Y_t^S = 1$, where Y_t^S is the time series for the proportions of the population that is susceptible to contract the disease.

Within the State-Space framework, the elements of $\boldsymbol{Y}_t = \{Y_t^I, Y_t^R\}$ constitute the observation equation and it is assumed that elements of $\boldsymbol{Y}_t$ follow Beta distributions. Beta distributions are a natural choice for modelling Y_t^I and Y_t^R as they are bound by [0,1], making them a good choice for modelling rates or proportions, and making certain that our forecasts will be feasible. Hence, we have:

$$Y_t^I|\boldsymbol{\theta}_t, \boldsymbol{\tau} \sim \text{Beta}(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I)) \tag{4}$$

$$Y_t^R|\boldsymbol{\theta}_t, \boldsymbol{\tau} \sim \text{Beta}(\lambda^R\theta_t^R, \lambda^R(1-\theta_t^R)), \tag{5}$$

where $\boldsymbol{\theta}_t = \{\theta_t^S, \theta_t^I, \theta_t^R\}$ are the true unobserved proportions of susceptible, infectious and removed individuals for the population at time t, and $\boldsymbol{\tau} = (\boldsymbol{\theta}_0, \beta, \gamma, \kappa, \boldsymbol{\lambda})$ are model parameters. More specifically, κ, $\boldsymbol{\lambda} = (\lambda^I, \lambda^R)$ are scalars, $\beta > 0$ is the rate of transmission, $\gamma > 0$ is the removal rate, and $\boldsymbol{\theta}_0$ is the initial state of $\boldsymbol{\theta}_t$. Parameters λ^I and λ^R play a role in controlling the conditional variance of $[\boldsymbol{Y}_t|\boldsymbol{\theta}_t, \boldsymbol{\tau}]$, but λ^I, λ^R do not affect the conditional expectation. As the values of λ^I and λ^R tend to infinity, the conditional variance of $\boldsymbol{Y}_t$ tends towards zero, reflecting we are certain of the values of $\boldsymbol{Y}_t$ (ie. no measurement error). This becomes clear when investigating moment properties for the Beta distribution. In particular, if $Y_t^I \sim \text{Beta}(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I))$ (analogous result for Y_t^R), then:

$$\begin{aligned} \text{E}(Y_t^I|\theta_t^I, \lambda^I) &= \theta_t^I, \\ \text{Var}(Y_t^I|\theta_t^I, \lambda^I) &= \frac{\theta_t^I(1-\theta_t^I)}{\lambda^I + 1}. \end{aligned} \tag{6}$$

The full derivation for this result can be found in the Appendix.

The time-indexed latent state vector $\boldsymbol{\theta}_t$ determines the dynamics of our model, and it is

assumed that $\boldsymbol{\theta}_{0:T} = (\boldsymbol{\theta}_0, ..., \boldsymbol{\theta}_T)$ follows a first-order Markov Process. This implies that $[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}] = [\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}]$ for all t. Additionally, for all $s \neq t$, Y_t and Y_s are independent conditionally on $\boldsymbol{\theta_t}$. We refer to this as the conditional independence assumption. The balance equation $\theta_t^S + \theta_t^I + \theta_t^R = 1$ must hold for all t with $0 < \theta_t^S, \theta_t^I, \theta_t^R < 1$. Hence, a natural choice for modeling $\boldsymbol{\theta}_t$ is the Dirichlet distribution, as its elements are constrained to be non-negative and sum to one. This allows us to proceed in the estimation without the need to transform the scale of the parameters, and this probabilistic specification allows for randomness in the evolution of the pandemic process. Within the State-space framework, the elements of $\boldsymbol{\theta_t}$ constitute the state equation. The distribution of $\boldsymbol{\theta_t}$ is specified as such:

$$\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau} \sim \text{Dirichlet}(\kappa f(\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau})), \tag{7}$$

where the function f determines the mean of the Dirichlet distribution (more detail in the next paragraph). Similarly to λ, the parameter κ does not affect the conditional expectation of $[\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}]$, but controls the conditional variance. As κ tends to infinity, the conditional variance tends to zero, reflecting confidence there is no process error. One can also view the parameter κ as controlling the randomness of the pandemic. For more details on the moment properties for the Dirichlet distribution, please see the Appendix.

The function $f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau}) \in \mathbb{R}^3$ governs the SIR flow through the determination of the mean values for $\boldsymbol{\theta_t}$. Its role is to propagate the latent state vector $\boldsymbol{\theta}_{t-1}$ one step into the future, hence it is the key to the infection dynamics. The set of ordinary differential equations (ODEs) defining the SIR flow within the state-space model are:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I \tag{8}$$

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I \tag{9}$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I, \tag{10}$$

and the function $f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau})$ corresponds the solution of the above set of ODEs. As there is no exact solution for the above system, we approximate $\hat{f}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau})$ using the fourth order Runge-Kutta (RK4) approximation to solve this system of ODEs (Chauhan and Srivastava

(2019)). Explicitly, we have :

$$\hat{f}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau}) = \begin{pmatrix} \theta^S_{t-1} + 1/6[k^{S_1}_{t-1} + 2k^{S_2}_{t-1} + 2k^{S_3}_{t-1} + k^{S_4}_{t-1}] \\ \theta^I_{t-1} + 1/6[k^{I_1}_{t-1} + 2k^{I_2}_{t-1} + 2k^{I_3}_{t-1} + k^{I_4}_{t-1}] \\ \theta^R_{t-1} + 1/6[k^{R_1}_{t-1} + 2k^{R_2}_{t-1} + 2k^{R_3}_{t-1} + k^{R_4}_{t-1}] \end{pmatrix} := \begin{pmatrix} \alpha_{1(t-1)} \\ \alpha_{2(t-1)} \\ \alpha_{3(t-1)} \end{pmatrix} \tag{11}$$

where

$$\begin{aligned} k^{S_1}_t &= -\beta\theta^S_t\theta^I_t \\ k^{S_2}_t &= -\beta[\theta^S_t + 1/2k^{S_1}_t][\theta^I_t + 1/2k^{I_1}_t] \\ k^{S_3}_t &= -\beta[\theta^S_t + 1/2k^{S_2}_t][\theta^I_t + 1/2k^{I_2}_t] \\ k^{S_4}_t &= -\beta[\theta^S_t + k^{S_3}_t][\theta^I_t + k^{I_3}_t], \end{aligned}$$

$$\begin{aligned} k^{I_1}_t &= \beta\theta^S_t\theta^I_t - \gamma\theta^I_t \\ k^{I_2}_t &= \beta[\theta^S_t + 1/2k^{S_1}_t][\theta^I_t + 1/2k^{I_1}_t] - \gamma[\theta^I_t + 1/2k^{I_1}_t] \\ k^{I_3}_t &= \beta[\theta^S_t + 1/2k^{S_2}_t][\theta^I_t + 1/2k^{I_2}_t] - \gamma[\theta^I_t + 1/2k^{I_2}_t] \\ k^{I_4}_t &= \beta[\theta^S_t + k^{S_3}_t][\theta^I_t + k^{I_3}_t] - \gamma[\theta^I_t + k^{I_3}_t], \end{aligned}$$

and

$$\begin{aligned} k^{R_1}_t &= \gamma\theta^I_t \\ k^{R_2}_t &= \gamma[\theta^I_t + 1/2k^{I_1}_t] \\ k^{R_3}_t &= \gamma[\theta^I_t + 1/2k^{I_2}_t] \\ k^{R_4}_t &= \gamma[\theta^I_t + k^{I_3}_t]. \end{aligned}$$

4.3 Integrating Policy

4.3.1 Estimation

We implement time-varying modifier to the transmission rate β, dependant on the strictness of measures against COVID-19 in place, in an attempt to obtain better estimates, and ultimately better forecasts.

To this end, we introduce the function π, with $0 < \pi(t) < 1, t > 0$, which reflects the effects of the policy/intervention in place at time t, and we scale the rate of infection β by this factor.

Hence the new rate of transmission becomes $\beta\pi(t)$. Intuitively, one can think of $\pi(t)$ as the probability that a susceptible person encounters an infectious person or vice versa, given current regulations and quarantine measures in place, or how successful the implemented policy is at reducing mobility of individuals. We set values for $\pi(t)$ deterministically, with $\pi(t) = 1$ being equivalent to no measures at all. Implementing this new rate boils down to adjusting the values of β in the Runge-Kutta approximation for $f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau})$ above. This results in the following new SIR flow:

$$\frac{d\theta_t^S}{dt} = -\beta\pi(t)\theta_t^S\theta_t^I \tag{12}$$

$$\frac{d\theta_t^I}{dt} = \beta\pi(t)\theta_t^S\theta_t^I - \gamma\theta_t^I \tag{13}$$

$$\frac{d\theta_t^R}{dt} = \gamma\theta_t^I, \tag{14}$$

There are many factors affecting the rate of infection of the disease, such as the use of protective face mask, as well as community-level lockdown, or simply a continuous increase in awareness of the population. In order to model the effects of such actions, we follow the guideline used in Song et al. (2020) and Kobayashi et al. (2020), who propose the following values for $\pi(t)$: $\pi(t) = 1$ reflects that human mobility has returned to its level before the intervention (no concrete quarantine protocols), $\pi(t) = 0.8$ reflects alertness in public awareness of the disease, where individuals voluntarily avoid outdoors and practice social distancing. Next, $\pi(t) = 0.5$ would correspond to the closing of all non-essential businesses, and lastly, $\pi(t) = 0.2$ represents an enhanced quarantine. Note that these values are somewhat subjective, thus prone to error. As mentioned in the literature review, Kobayashi et al. (2020) warn that these values should be set conservatively, as it is difficult for governments to monitor individual activity, leading to a slight reduction in the efficacy of these measures. Examples of this failure include gatherings with more participants than legally permitted, people breaking curfew, as well as illegal parties. Such an example is the illegal rave that went on for 36 hours on New Year's Eve in a warehouse in Bretagne with over 2500 participants [6], which defied these three measures simultaneously.

In Table 1 we give a brief outline of the policies that have been implemented by the French government [7] that are the most relevant to this study, and we quantify them according to the guidelines described above. Our sample begins on the 1^{st} of July 2020, but it is important to note that first lockdown was progressively lifted from May 11, 2020 to June 2, 2020.

[6] https://www.washingtonpost.com/world/2021/01/02/france-party-new-years-eve-coronavirus

[7] https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement

Table 1: Summary of the policies and their quantification

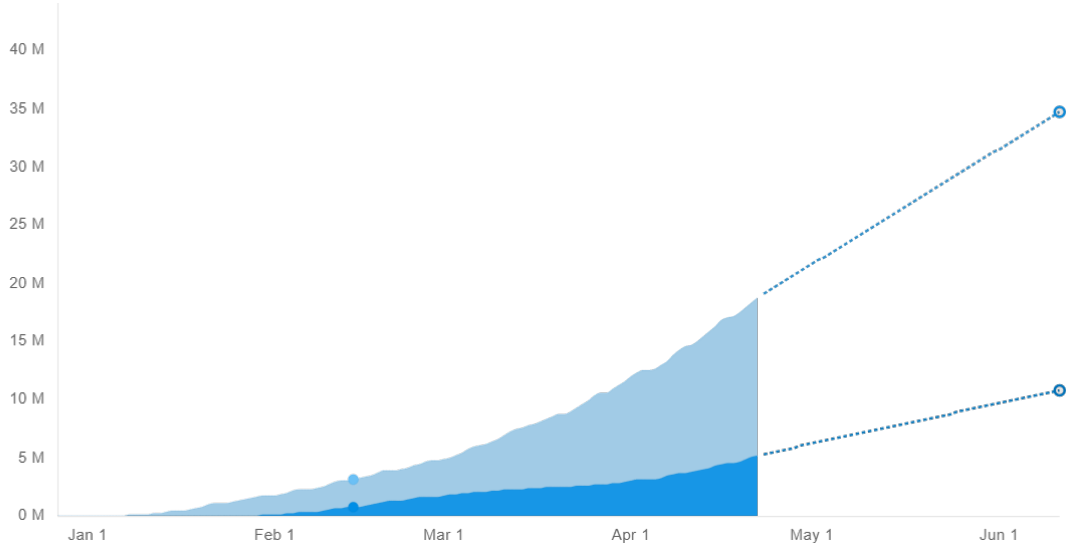
Date	Policy Description	Quantification
1rst July 2020	Beginning of the observation period	0.8
10th July 2020	End of the sanitary state of emergency	0.85
20th July 2020	Introduction of a decree imposing the wearing of masks in all indoor spaces, and in the street in some major cities.	0.75
17th October 2020	Introduction of the curfew at 20h in some French regions.	0.50
22th October 2020	Generalisation of the curfew to all French regions.	0.50
30th October 2020	Beginning of the second lockdown, while curfew remains.	0.30
15th December 2020	Gradual deconfinement, while curfew at 20h remains active.	0.55
2nd January 2021	Stricter curfew at 18h in some French regions	0.45
16th January 2021	Generalisation of the curfew at 18h to all French regions.	0.40
18th March 2021	Reinforced sanitary measures (i.e. lockdown) for 16 departments (cumulating about 60% of population)	0.35
31th March 2021	Generalisation of the lockdown to all French regions.	0.30

Another very important aspect in the control of the pandemic is the vaccination roll-out, which we seek to integrate into the policy function $\pi(t)$. According to the Covid Vaccine Tracker [8] by the April 13, 2021, 11.430.203 had received a single vaccine shot (17.5% of the population), while 4.008.284 had received the second shot (6.1% of the population), with a vast majority of people receiving the Pfizer vaccine. As the SIR model assumes that the population mixes homogeneously, we do not account for the fact that vaccines were first given to people at high risk, before being distributed to the general population.

Figure 4 is a plot from CovidTracker.fr[8] showing the cumulative number of persons having received single and double doses of a vaccine, with the dotted lines representing the projections for these quantities (data from the Ministry of Health).
In a study in the United-States about the necessity to take two doses of Pfizer or Moderna, Livingston (2021) state "When the vaccines were first tested, a relatively weak immune reaction was found within a few weeks after people received the first dose of vaccine, followed by a strong reaction when a second dose was given. (...) It is possible that people who get only 1 dose will have only partial immunity to COVID-19 infection, [but] there is no evidence that people who get only 1 dose have adequate long-term protection against COVID-19 infection". For this reason, they insist that the second dose is necessary to confer adequate immunity. Hence, the information on vaccines that we incorporate into the $\pi(t)$ function focuses primary on the proportion of people having received a second dose. We can make this assumption without the loss of generality as people who receive a first shot will take a second dose on average 4 weeks after receiving the first one[9]. To reflect a reduction in the infection pattern,

[8]https://covidtracker.fr/vaccintracker/
[9]https://www.cdc.gov/coronavirus/2019-ncov/vaccines/second-shot.html

Figure 4: Cumulative vaccine shots: light blue for single dose, dark blue for double dose



the values set for $\pi(t)$ are slightly and progressively deflated, from 0 on February 1, 2021 to -0.06 (=6%) on April 13, 2021. Figure 5 shows the evolution of the policies over time, with the adjustments made to account for vaccination roll-out.

We choose to work with a step function over a continuous function for $\pi(t)$, at it is a lot more reflective of how the measures were imposed and the changes in behaviour these changes in policies induce for the population.

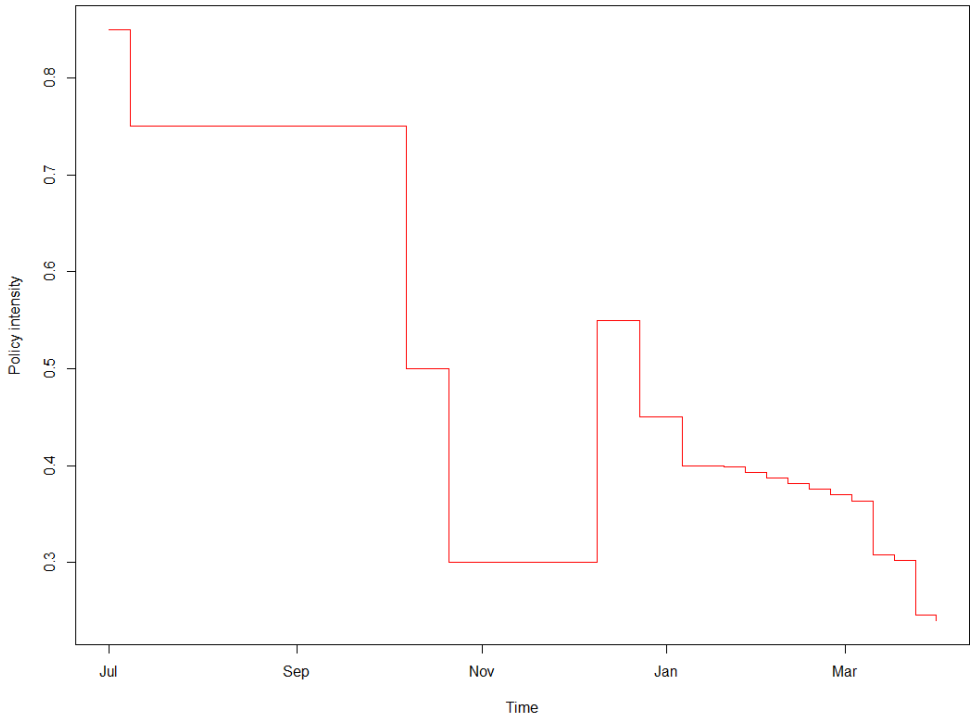
4.3.2 Forecasting

The main goal of this paper is to obtain forecasts on the number of new COVID-19 infections. Integrating policy in the model allows us to predict the number of cases that would occur for various predicted policies (i.e. policy evaluation). France entered its third lockdown on March 31, 2021 which was been extended to last until May 3, 2021, according to the latest news [10]. As for the plan afterwards, the country is set to gradually open up, in the hope that the vaccination roll-out will help counteract the rise of new cases. Authorities project that the lockdown measures will be fully lifted by the end of June.

In terms of forecasting, we look at three different scenarios. Note that these scenarios will only diverge from May 3, 2021 onward when the current lockdown ends. Until then, we impose the same restrictions as the ones imposed by the government. In the first scenario,

[10] https://www.linternaute.com/actualite/politique/2498313-discours-de-macron-des-annonces-dimanche-2-mai-sur-le-covid-et-le-deconfinement/

Figure 5: Graph of the evolution of imposed policies



the current policies in place will be prolonged until the end the forecasting period: we impose a continued lockdown for a period of 15 weeks, alongside the curfew that is currently in place at 19h. This corresponds to setting the baseline value for $\pi_1(t)$ at 0.3 for the weeks $t \in$ [April 14, 2021; July 21, 2021]. In the second scenario, we relax the quarantine measures in place, however in order to keep this scenario realistic, we assume that the public awareness of the disease remains high, and that individuals will continue to practice some social distancing, as well as engage in simple preventive behaviour such as mask-wearing and hand-washing. Hence, we set the baseline for $\pi_2(t)$ at 0.8 until the end of the forecasting period. Finally, we follow a scenario parallel to the measures announced by the government, which involves a gradual deconfinement after May 3, 2021. After this date, the gradual re-opening is modelled in baseline policy function $\pi_3(t)$ as a monotonically increasing function:

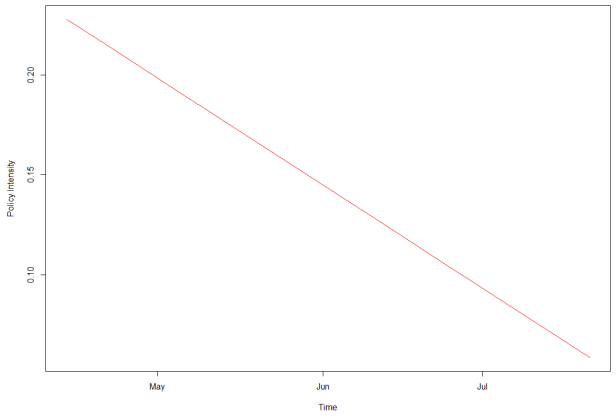
$$\pi_3(t) = \pi(13^{th}\text{April}) + \exp(0.037 \times t) - 1,$$

where the value for $\pi(13^{th}\text{April})$ includes the effect of vaccination mentioned in the previous section. Note that we refer to the baseline values of the function $\pi(t)$ throughout this section. Indeed, after setting baseline values for $\pi(t)$ in the three different scenario, all three

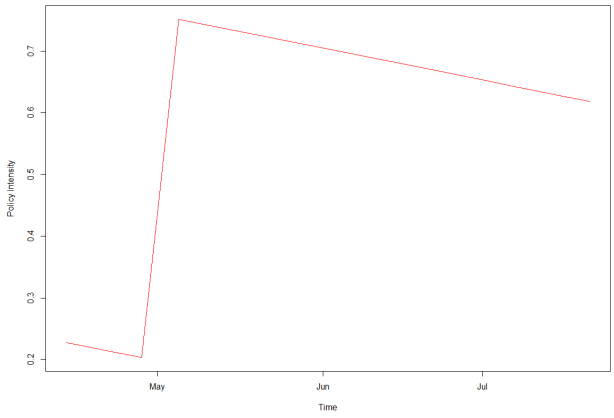
specifications of this function are modified identically to integrate the effect of vaccination roll-out in the policy function, as was done for the estimation process. According to Covid Vaccine Tracker[8], 11.392.352 people are expected to have received their second vaccination shot by the 9^{th} of June 2021 (17.4% of the population), versus the 4.276.086 second doses administered by April 14^{th} corresponding to an increase of 10.88% between these two dates. Hence, values for the functions $\pi_i(t), i = 1, 2, 3$ are linearly decreased by that amount over the course of that period. For dates beyond June 9, 2021, we assume the linear relation continues to hold, and values for $\pi(t)$ are interpolated using the same increments. Different specifications of the functions $\pi_i(t), i = 1, 2, 3$ are plotted in Figure 6 below:

Figure 6: Policy curves, including vaccination roll-out

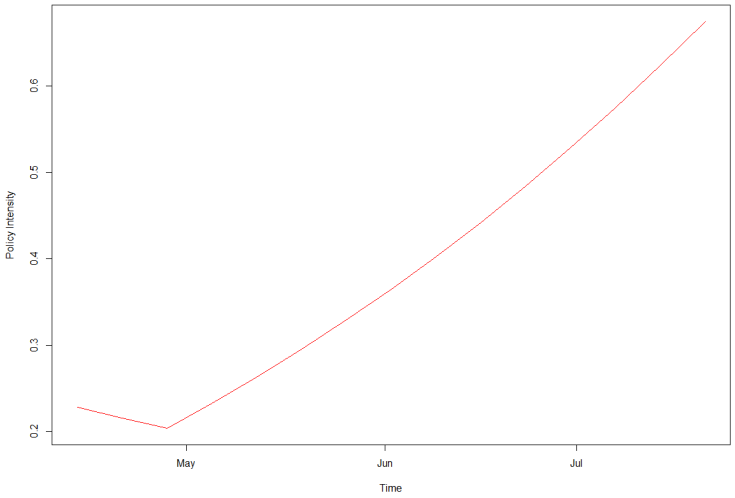
(a) Strict policy $\pi_1(t)$



(b) Loose policy $\pi_2(t)$



(c) Gradual opening policy $\pi_3(t)$



4.4 Posterior Inference

4.4.1 Prior Specification

As Bayesian research can be somewhat subjective, it is important to define defensible priors in practice. To this end, we follow the hyper-parameter prior specifications made in Song et al. (2020), in terms of distributional choices. However, we do make some changes to the location and scale parameters of these distributions due to new information becoming available, and differences between the outbreak in Wuhan and France.

As mentioned previously, high values for the controls κ, λ^I and λ^R reduce the conditional variances for $\boldsymbol{\theta_t}$, Y_t^I and Y_t^R respectively. These values also define how tight our 95% credible intervals for our estimates will be. Hence, we increase the location parameter for κ to 20, instead of 2 as was done in Song et al. (2020), to show we have some confidence in how well the model can describe the pandemic process. We only increase this value slightly, because for very high values of κ we do not observe randomness in the evolution of the pandemic. Indeed, the values for κ tends to infinity, the SIR flow collapses to the deterministic SIR flow. However, we do not modify the values for λ^I and λ^R because of the uncertainty linked to the under-reporting and the up-scaling of the data. All in all, these remain relatively flat prior distributions, reflection we know little about these three parameters.

Next, we modify the prior of the basic reproductive number R_0, as in France, R_0 was estimated to be 1.02[11], whereas Song et al. (2020) estimated a R_0 of 3. In our prior, we scale down the variance of R_0 to reflect the fact the basic reproductive number does not fluctuate excessively.

In fact, prior distributions for γ, β, R_0 are set jointly according to analytical relations of the SIR model. More specifically, we set $\gamma = 0.16$ according to the specification of Kobayashi et al. (2020), then set $\beta = 0.20$, such that such that $R_0 = \beta/\gamma = 1.25$. Next we impose $\text{Var}(R_0 = 0.2^2)$ and $\text{Var}(\gamma = 0.1^2)$. The parameters γ and R_0 are then transformed such that they follow log-normal distributions which satisfy the above constraints. Lastly, the prior for the initial state $\boldsymbol{\theta}_0$ is set using the empirical values for the proportions Y[1] and R[1], and

[11] https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-15-avril-2021

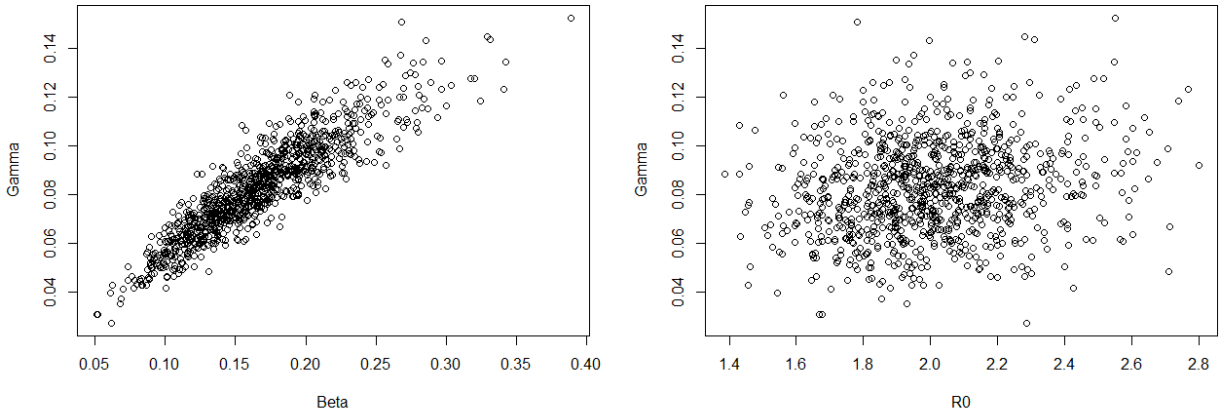
the initial proportion of susceptible people is derived from these. Then, we have:

$$
\begin{aligned}
\theta_0 &\sim \text{Dir}(1 - Y[1] - R[1], Y[1], R[1]), \\
R_0 &\sim \text{LogN}(0.211, 0.0253), \\
\gamma &\sim \text{LogN}(-1.998, 0.330), \\
\beta &\sim R_0\gamma, \\
\kappa &\sim \text{Gamma}(20, 0.0001), \\
\lambda^I &\sim \text{Gamma}(2, 0.0001), \\
\lambda^R &\sim \text{Gamma}(2, 0.0001).
\end{aligned}
$$

Note that we do not set a distribution for the prior of β, but we calculate this value as $\beta = R_0\gamma$. We do not sample β and γ simultaneously, because these two parameters are highly correlated. Instead, we sample γ and R_0 which are less correlated and derive β from these quantities. This approach was recommended by Osthus et al. (2017), who explain that the Gibbs sampler can have difficulties sampling from the DBSSM due to highly correlated structures. To illustrate this point, Figure 7 shows the correlation plots for R_0, β and γ, based on the first 1000 posteriors samples for these.

Figure 7: Correlations plots

(a) Posterior samples of β and γ (b) Posterior samples of R_0 and γ



4.4.2 Posterior Sampling

We use Bayesian methods for inference and forecasts within the state-space framework. Let $t' \in [1, 2, ..T]$ denote the time index for all values within our observed sample. We seek to get posterior knowledge on the model parameters $\boldsymbol{\tau}$ as well as on the latent space vector $\boldsymbol{\theta}_t$ given the observed data in the series Y_t^I and Y_t^R. Let $[\boldsymbol{\tau}]$ denote the prior distributions for elements in $\boldsymbol{\tau}$. Then,

$$[\boldsymbol{\theta}_{1:t'}, \boldsymbol{\tau}|Y_{1:t'}] \propto [\boldsymbol{\tau}][Y_{1:t'}, \boldsymbol{\theta}_{1:t'}|\tau] = [\boldsymbol{\tau}] \prod_{t=1}^{t'} [y_{1:t}|\theta_t, \boldsymbol{\tau}][\theta_t|\theta_{t-1}, \boldsymbol{\tau}]. \tag{15}$$

The above equation holds because of the model's conditional independence assumption described in section 3. The above form does not have a closed form posterior, thus we use MCMC methods, specifically a Gibbs sampler to collect draws from the posterior distributions. This is done using the package **rjags** (Plummer et al., 2019), a tool within the *R* programming language (R Core Team (2020)). The package calls the software JAGS which stands for Just Another Gibbs Sampler (Plummer (2003)) from which the simulation and sampling is done. In terms of how sampler selection, the JAGS user manual (Plummer (2012)) states that "The user has no direct control over the process of choosing Samplers", as samplers are automatically chosen for each parameter by JAGS during the initialisation of the model. However, we do run a long adaptative phase after the burn-in to ensure that the samplers reach optimal efficiency.

For posterior inference, we focus on values θ_t^I and θ_t^R, as it shown in equation 6 that on average the prevalence of infection and removal was equal to the proportions θ_t^I and θ_t^R respectively. Samples from JAGS are collected through the procedure described in the pseudo-algorithm below:

Algorithm 1 Posterior Sampling JAGS

for $m = 1, 2, \ldots, M$ **do**
  Draw $\kappa^{(m)}$ from $[\kappa]$
  Draw $\lambda^{I,(m)}$ from $[\lambda^I]$
  Draw $\lambda^{R,(m)}$ from $[\lambda^R]$
  Draw $\boldsymbol{\theta}_0^{(m)}$ from $[\boldsymbol{\theta}_0]$
  Draw $\gamma^{(m)}$ from $[\gamma]$
  Draw $R_0^{(m)}$ from $[R_0]$
  Calculate $\beta^{(m)}$ as $R_0^{(m)}\gamma^{(m)}$
  Then $\boldsymbol{\tau}^{(m)} = (\kappa^{(m)}, \lambda^{I,(m)}, \lambda^{R,(m)}, \boldsymbol{\theta}_0^{(m)}, \gamma^{(m)}, R_0^{(m)}, \beta^{(m)})$ constitutes draw from $[\boldsymbol{\tau}]$
  for $t = 1, 2, \ldots, T$ **do**
    Calculate $f(\boldsymbol{\theta}_t^{(m)}, \boldsymbol{\tau}^{(m)})$
    Draw $\boldsymbol{\theta}_t^{(m)}$ from $[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(m)}, \boldsymbol{\tau}^{(m)}]$
    Draw $\boldsymbol{Y}_t^{(m)}$ from $[\boldsymbol{Y}_t|\boldsymbol{\theta}_t^{(m)}, \boldsymbol{\tau}^{(m)}]$
  end for
  Then $\boldsymbol{\theta}_{1:T}^{(m)} = (\boldsymbol{\theta}_1^{(m)}, \boldsymbol{\theta}_2^{(m)}, \ldots, \boldsymbol{\theta}_T^{(m)})$ constitutes a draw from $[\boldsymbol{\theta}_{1:T}]$
  And $\boldsymbol{Y}_{1:T}^{(m)} = (\boldsymbol{Y}_1^{(m)}, \boldsymbol{Y}_2^{(m)}, \ldots, \boldsymbol{Y}_T^{(m)})$ constitutes a draw from $[\boldsymbol{Y}_{1:T}]$
end for

4.4.3 Inference and Forecast

From the posterior, we are able to make forecasts on the number of probable infections and removals for COVID-19. We make forecasts 15 weeks into the future, until the week of July 21, 2021. Let $t^* \in [T+1, .., T_{end}]$ We seek to obtain forecasts for $Y_{t^*}^I$, $Y_{t^*}^R$ and $\boldsymbol{\theta_{t^*}}$, which we calculate as:

$$[\boldsymbol{Y}_{(T+1):T_{end}}, \boldsymbol{\theta}_{(T+1):T_{end}}|\boldsymbol{\theta}_T, \tau] = \prod_{t=T+1}^{T_{end}} [\boldsymbol{Y}_t|\boldsymbol{\theta_t}, \tau][\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \tau], \tag{16}$$

where the above holds true given the conditional independence assumption, and where $[\boldsymbol{Y}_t|\boldsymbol{\theta_t}, \boldsymbol{\tau}]$ and $[\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \tau]$ are defined as in equations 4 & 5 and 7 respectively.

Predicted policies are incorporated into these forecasts, and we can vary the intensity of the imposed protocols to see how this would affect new incidences of COVID. For example, we can compare the predictions of new cases under a continued strict lockdown versus a gradually attenuated lockdown, or the outcome with no quarantining protocols at all.

5 Empirical Application

In this section, we present the results from the Dirichlet-Beta state-space SIR model, using weekly probable cases as data. We start by presenting posterior results obtained from our simulation within JAGS. Then, we present the forecasts obtained based on the posterior results. We consider forecasts for different policy scenarios, which serves as policy evaluation.

5.1 Posterior Results

In the JAGS run, we run four parallel chains whose seeds are generated through RNG. In JAGS, the burn-in is the period between the initialisation of the model and the creation of the first monitoring point (i.e. we do not run the model and then discard these observations). We ran 62500 iterations, with 12500 iterations as burn-in, and used a thinning of 10. This was followed by an adaptative phase, which we ended only when the samplers reached optimal efficiency. This was done through six recursive calls of the *adapt* for 50000 iterations. In total, we obtain $(62500/10) \times 4 = 25000$ draws from JAGS.

Key statistics for all the model parameters and the true underlying proportions of S_t, I_t and $R_t, t \in [1, ..T]$ are summarized in the Table 2 below.

Table 2: Summary of key statistics

		2.5%	25%	50%	75%	97.5%		
Parameter	Description	Quantile					Mean	SD
β	Rate of infection	0.0537	0.0897	0.114	0.144	0.216	0.119	0.0418
γ	Rate of removal	0.0427	0.0642	0.0773	0.0911	0.120	0.0783	0.0198
R_0	Basic reproduction number	1.049	1.315	1.482	1.674	2.130	1.510	0.277
λ^I	Y_t^I variance scaling	2899	9723	15504	23389	48022	18052	11812
λ^R	Y_t^R variance scaling	6300	11280	15192	20779	39412	17100	8508
κ	$\boldsymbol{\theta}_t$ variance scaling	4591	5888	6781	7947	13228	7363	3205
θ_t^S	Proportion susceptible	0.978	0.980	0.980	0.981	0.983	0.980	0.000663
θ_t^I	Proportion infected	0.00950	0.0111	0.0117	0.0122	0.0132	0.0116	0.000526
θ_t^R	Proportion removed	0.00673	0.00760	0.00804	0.00846	0.00929	0.00803	0.000386

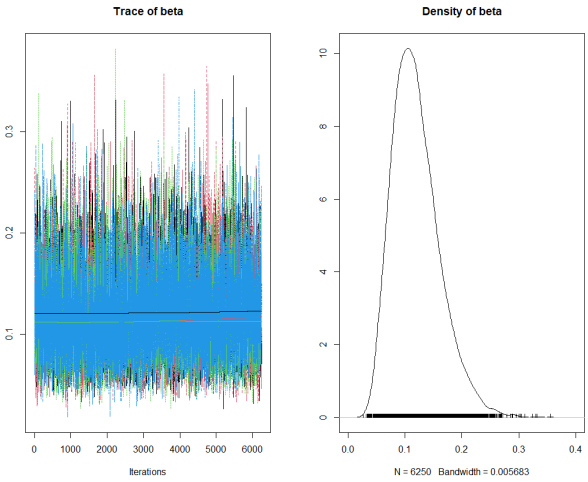
The estimates for θ_t^S, θ_t^I and θ_t^R are first constructed by averaging out the 25.000 draws, then statistics are calculated for each time t, and finally averaged out across t. The estimates for λ^I, λ^R and κ allow us to examine the scale of process and measurement errors. The value κ serves as a control for the randomness in the pandemic process, and its mean value across the simulation returns lower than the prior we had set for it, with a very lower variance. The lower value of $\mathrm{E}(\kappa)$ implies that the pandemic is more random than anticipated and also indicates that a deterministic SIR model (equivalent to $\kappa = \infty$) would fail to capture the pandemic process. As for values λ^I and λ^R, these are relatively large, but the magnitude of

the measurement errors is hardly interpretable.

In Figure 8 below, we present the densities of posterior samples for the main model parameters: the rate of infection β, the removals rate γ and the reproductive number R_0. The combination of the trace plots and the density plots helps to visually assess how the draws fluctuated throughout the sampling process. The different colours in the trace plots correspond to parallel simulation chains. Similar plots for λ^I, λ^R and κ can be found in the Appendix.

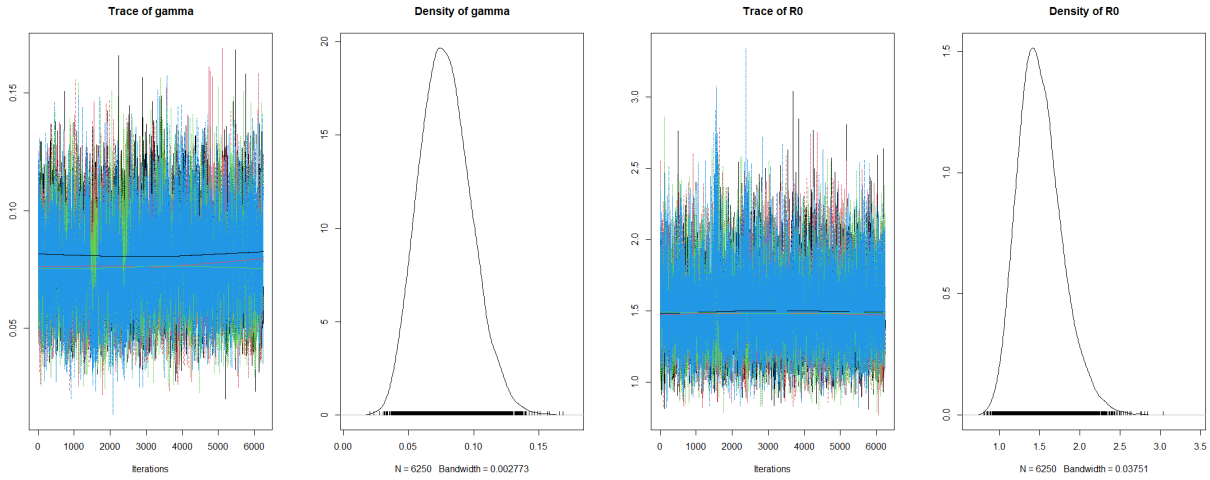
Figure 8: SIR Model Parameters

(a) Rate of infection



(b) Removal rate

(c) Reproductive number



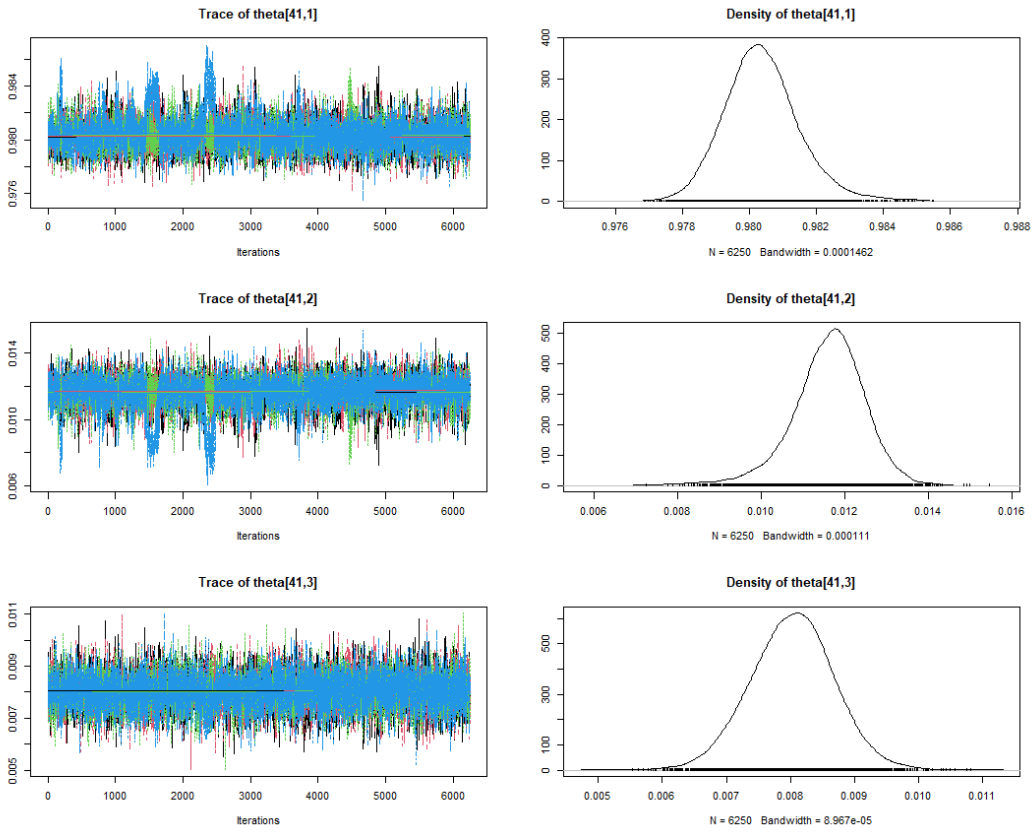
Note that we will tend to overestimate $R_0 = \beta/\gamma$, as our samples for β are slightly inflated

due to the fact they are multiplied by $0 < \pi(t) < 1$. In this study, we estimate an R_0 of 1.51, close to the estimate from Kobayashi et al. (2020) of 1.44 from Japan (95% CI for $R_0 \in (1.22, 1.64)$), but lower than estimates from Song et al. (2020) who consistently estimated R_0's of the order of 4, under various quarantining measures.

In reality, the basic reproductive number was estimated to be 0,95 in France, according to Public Health Ministry[12] on the 17^{th}April 2021. Once again, note that this is no surprise, since this model explicitly incorporates interventions, such that R_0 is the adjusted basic reproduction number that would occur if all quarantine protocols would be removed. In contrast, the estimate of $R_0 = 0.95$ incorporates the effect of government intervention into the transmission rate, leading to a reduced estimate due to the contribution of intervention.

In Figure 9 we present the posterior densities for the latent space vector $\boldsymbol{\theta}_{T+1}$ (T= week of April 14, 2021), corresponding to the expected proportions of individuals in the Susceptible (θ^S_{T+1}), Infected (θ^I_{T+1}) and Removed (θ^R_{T+1}) compartments, respectively.

Figure 9: Densities of latent space vector samples



[12]COVID-19 : point épidémiologique du 22 avril 2021

Lastly, we present the evolution of the posterior samples for θ_t^I and $\theta_t^R, t \in [1, ..T]$. These are plotted below, in Figures 10 and 11. The estimated quantities for θ_t^I and θ_t^R are plotted in colour (infections in red, removals in blue), while the measured values of Y_t^I and Y_t^R are plotted in black, for comparison. On average, values for Y_t^I and Y_t^R are very close estimates for θ_t^I and θ_t^R, as expected from the moment conditions from equation 6. This remains a good indicator of the performance of the model.

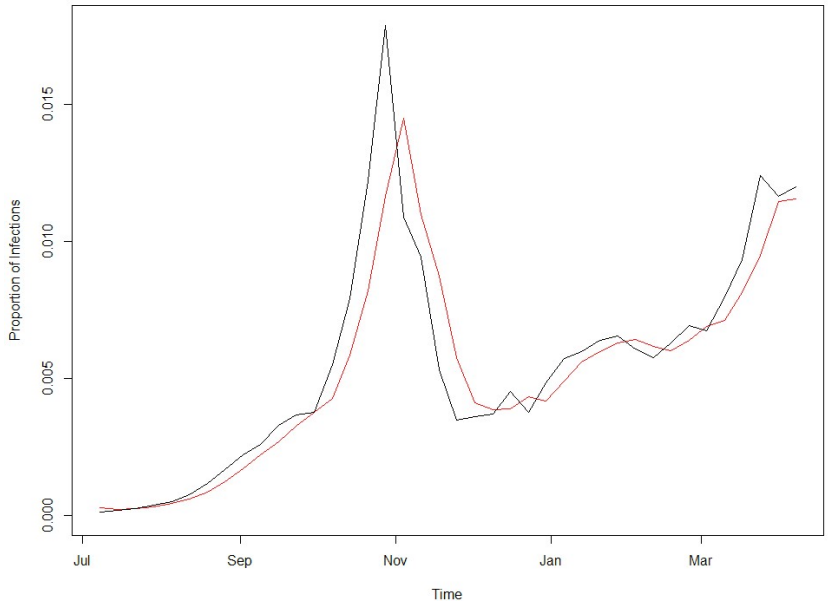


Figure 10: Measured infections versus true estimated proportions

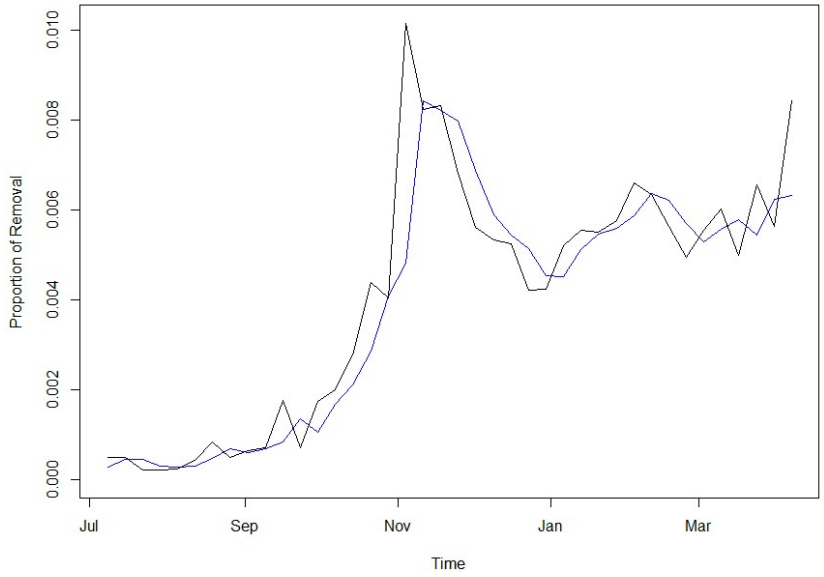
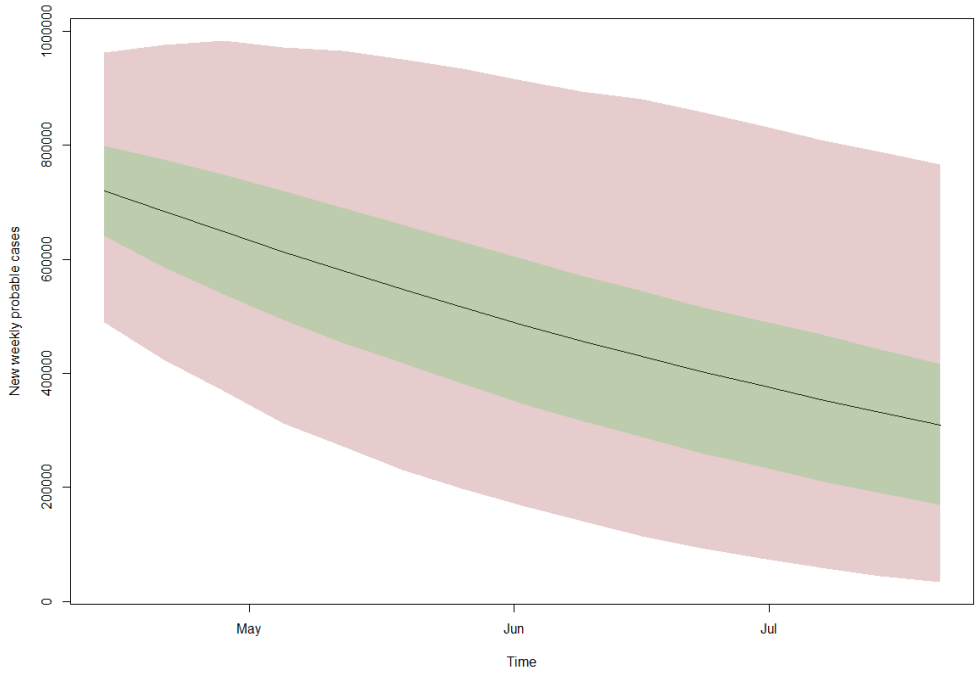


Figure 11: Measured removals versus true estimated proportions

5.2 Forecasts and Policy Evaluation

In this section, we present the forecasts obtained, under the three different scenarios described in the Section 4.3.2. Here, we focus on the predicted quantities for $Y^I_{t*}, t^* \in [T+1, .., T_{end}]$. We multiply these proportions of infections by N, the total population for France, to make them into new weekly probable infections. The forecasts for the new incidences are plotted with their 50% and 95% credible intervals, where the 50% confidence intervals are shaded in green, while the 95% confidence intervals are shaded in red. Below are the plots for the two first scenarios: the continued strict lockdown scenario is shown in Figure 12, and the scenario where no concrete measures are imposed (with public awareness of the disease) in Figure 13:

Figure 12: Predicted infections: continuous lockdown



From Figure 12, we note that the cases continuously and monotonically go down if a continued lockdown is applied, in combination with vaccines being distributed. Though these are weekly probable predicted cases, meaning these are higher than the typical daily confirmed cases, it is interesting to note that despite a very strict lockdown, the number of cases does not converge to 0 before the end of the sample, on July 21, 2021. This fact underlines the importance of imposing measures for a sufficiently long time in order to mitigate the propagation of the COVID-19 virus.

Figure 13: Predicted infections: no concrete measures

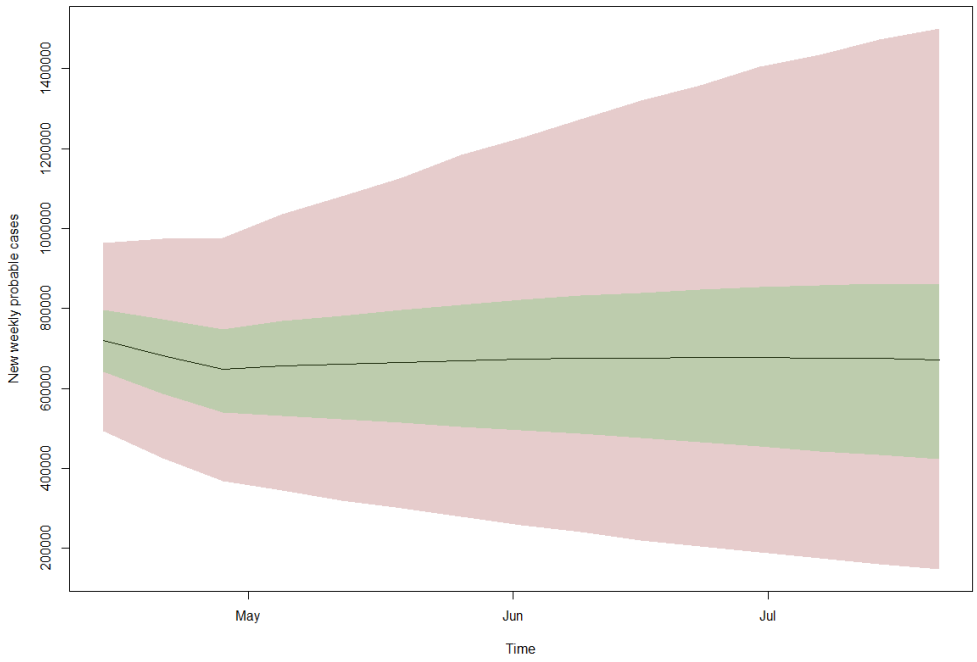
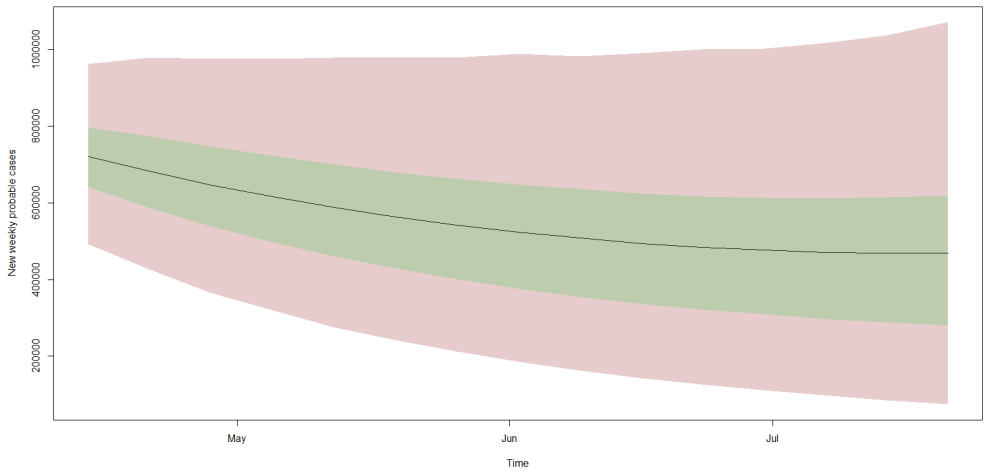


Figure 13 shows what would happen if the lockdown was lifted on May 3, 2021, with no additional measures besides mask-wearing and basic preventive behaviour of individuals. After May 3, cases break the downwards trend initially observed, and begin to rise again. What is particularly interesting to note is that though the slope is positive after May 3, it gradually flattens and after June 30, 2021, the number of cases begins to slope downwards. In the absence of any concrete policy, this decrease in the number of new infections can solely be attributed to the vaccination efforts, which is a very optimistic result. Indeed, this implies that if a sufficiently large proportion of the population gets vaccinated, the number of cases should be able to converge to a very low number.

In Figure 14 we show the predictions made under the last specification $\pi_3(t)$, which models the gradual re-opening of the economy from May 3, 2021, paralleled with vaccinations. We observe a monotonically decrease in number of cases for all the predicted data points, except the last point in the week of July 21, 2021, though this increase is barely noticeable, and invisible graphically. Nevertheless, this estimate could serve as a warning against relaxing too many measures too quickly. To policy makers, this would imply that a gradual relaxation can be done without an increase in the number of cases in the medium term, however an enhanced relaxation of measures could cause the number of cases to rise again in the long-

term.

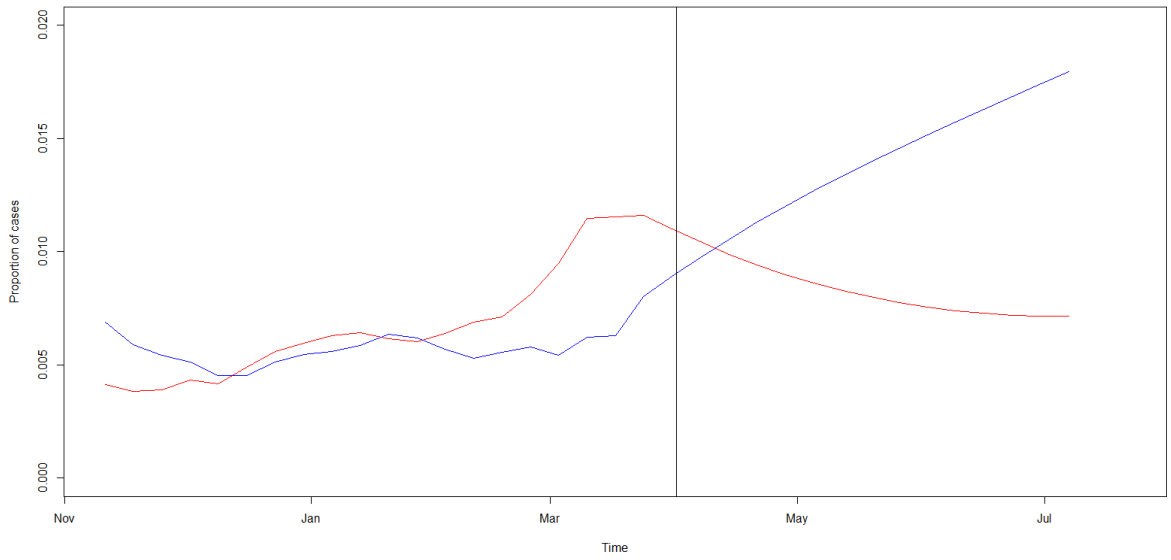
Figure 14: Predicted infections: Gradually lifting of the lockdown



In summary, the three policy specifications offer the following insights: Figure 12 the number of cases will only converge to a truly low number if the impactful measures (such as lockdowns, curfew and working from home) are kept in place for a sufficiently long period. Figures 13 shows that the roll-out of vaccines will also be a significant part in reducing the infections, and that if a sufficiently high proportion of the population gets vaccinated, the pandemic should become more manageable. Lastly, Figure 14 offers a warning for the relaxation of too many measures in a too short amount of time. Indeed, it would be best to open up the economy at a slowed down rates, in parallel with a continued effort to get as many people vaccinated as possible.

Finally, we present how the proportions of infections and removals θ_t^I and θ_t^R evolve from the second half of observed sample. The predicted quantities are based on the third scenario which follows the predicted government policies for the gradual lifting of the lockdown. These are shown in plot 15 below, where the vertical black line is the date France entered its third lockdown (March 31, 2021). Visually, it is noticeable that imposing a third lockdown helped significantly in controlling the proportions of new weekly probable infections and removals. In particular, the proportions of removals is projected to reach an all-time-high for the entire observation period, and will continue to increase.

Figure 15: Proportions of new weekly infections in Red, removals in Blue.



6 Discussion

In this paper, we modelled the COVID-19 outbreak in France using the Dirichlet-Beta state-space model inspired by the deterministic SIR model. We extended the classic SIR by incorporating a time-varying function to account for quarantining protocols imposed by authorities, which translated to changing the rate of infection over time. Understanding the effects of quarantining measures is crucial to crafting effective preventive policies to control the spread of the COVID-19 virus, hence forecasts were carried out with various degrees of government intervention. The quality of the data is limited due to under-reporting, which motivated up-scaling the observations. Thus, we worked in terms probable cases rather than confirmed cases, which is a distinct perspective of this paper. The use of MCMC methods within the state-space framework allowed us to integrate uncertainty linked to under-reporting and the virus's complex dynamics in the prediction process.

We found that provided adjustment to the reported data, the model accurately predicted the recent course of the pandemic in France. The model predicts that, at current coverage of vaccination and its likely further slow roll-out, the lockdowns should be lifted more gradually than the government in France is currently envisaging. Another finding was that at least for the short term, the impact of social behaviour on transmission is greater than that of vaccination because it is possible to restrict social behaviour to a certain scale, whereas the same is impossible with vaccines. As attaining population vaccine coverage will remain a

huge challenge, and as maintaining social distancing is not feasible long term, the government will need to find a way to mitigate the impact of the epidemic through other means. This might require increased investment in health infrastructure to deal with a potential additional waves of infections following measures which might be too liberal, and the development of complication-blocking treatment for COVID infections.

The DBSSM model proposed in this paper can be further extended to accommodate for the incubation period individual experience prior to being exposed to the COVID-19 virus, or to account for imperfect immunity subsequent to recovery. Both of these extensions be be achieved through the addition of a new compartments in the SIR model, such that we obtain SEIR and SIRS models respectively. However, these specifications add complexity to the model, and require adequate surveillance data to be implemented. As more research is carried out on the effect of policies on controlling the outbreak, numerical quantification of quarantine measures should also improve. A few drawbacks in methodology of this paper are the fact the SIR model assumes we observe a closed population. Yet, migration plays a strong role in how the virus propagates, as in France 82.80% of new COVID-19 cases were due to the UK-variant, and 4.2% due to the South-African variant. However, the strongest limitation to this paper and all COVID-19 related literature is the quality of the data. Indeed, it is very difficult, if not impossible, to acquire data on infected individuals who self-quarantined or asymptomatic cases, hence the true scale of under-reporting remains unknown, and inestimable in practice.

References

Chauhan, Vijeyata and PK Srivastava (2019). "Computational techniques based on Runge-Kutta method of various order and type for solving differential equations". In: *International Journal of Mathematical, Engineering and Management Sciences* 4.2, pp. 375–386.

Deo, Vishal and Gurprit Grover (2020). "A New Extension of State-Space SIR Model to Account for Underreporting-An Application to the COVID-19 transmission in California and Florida". In: *medRxiv*.

Dukic, Vanja, Hedibert F Lopes, and Nicholas G Polson (2012). "Tracking epidemics with Google flu trends data and a state-space SEIR model". In: *Journal of the American Statistical Association* 107.500, pp. 1410–1426.

He, Shaobo, Yuexi Peng, and Kehui Sun (2020). "SEIR modeling of the COVID-19 and its dynamics". In: *Nonlinear Dynamics* 101.3, pp. 1667–1680.

Kermack, William Ogilvy and Anderson G McKendrick (1927). "A contribution to the mathematical theory of epidemics". In: *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772, pp. 700–721.

Kobayashi, Genya et al. (2020). "Predicting intervention effect for COVID-19 in Japan: state space modeling approach". In: *BioScience Trends*.

Livingston, Edward H (2021). "Necessity of 2 doses of the Pfizer and Moderna COVID-19 vaccines". In: *JAMA* 325.9, pp. 898–898.

Organization, World Health et al. (2020). "Impact of COVID-19 on people's livelihoods, their health and our food systems". In: *World Health Organization. https://www. who. int/news/item/13-10-2020-impact-of-covid-19-onpeople's-livelihoods-their-health-and-our-food-systems*.

Osthus, Dave et al. (2017). "Forecasting seasonal influenza with a state-space SIR model". In: *The annals of applied statistics* 11.1, p. 202.

Plummer, Martyn (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.

— (2012). "JAGS Version 3.3. 0 user manual". In: *International Agency for Research on Cancer, Lyon, France*.

Plummer, Martyn et al. (2019). "Package 'rjags'". In:

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Shaman, Jeffrey and Alicia Karspeck (2012). "Forecasting seasonal outbreaks of influenza". In: *Proceedings of the National Academy of Sciences* 109.50, pp. 20425–20430.

Song, Peter X et al. (2020). "An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China". In: *MedRxiv*.

Teslya, Alexandra et al. (2020). "Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study". In: *PLoS medicine* 17.7, e1003166.

Wise, Jacqui (2021). *Covid-19: European countries suspend use of Oxford-AstraZeneca vaccine after reports of blood clots*.

Wu, Sean L et al. (2020). "Substantial underestimation of SARS-CoV-2 infection in the United States". In: *Nature communications* 11.1, pp. 1–10.

Zhuang, Lili and Noel Cressie (2014). "Bayesian hierarchical statistical SIRS models". In: *Statistical Methods & Applications* 23.4, pp. 601–646.

Appendices

A Moment Condition Beta Distribution

If $Y_t^I \sim \text{Beta}(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I))$, then:

$$\begin{aligned}
\text{E}(Y_t^I|\theta_t^I, \lambda^I) &= \frac{\lambda^I\theta_t^I}{\lambda^I\theta_t^I + \lambda^I(1-\theta_t^I)} = \theta_t^I, \\
\text{Var}(Y_t^I|\theta_t^I, \lambda^I) &= \frac{\lambda^I\theta_t^I\lambda^I(1-\theta_t^I)}{(\lambda^I\theta_t^I + \lambda^I(1-\theta_t^I))^2(\lambda^I\theta_t^I + \lambda^I(1-\theta_t^I)+1)} = \frac{\theta_t^I(1-\theta_t^I)}{\lambda^I+1}.
\end{aligned}$$

B Moment Condition Dirichlet Distribution

If $\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau} \sim \text{Dirichlet}(\kappa f(\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}))$, then:

$$\begin{aligned}
\text{E}(\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}) &= f(\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}) \\
\text{Var}(\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}) &= \frac{1}{1+\kappa}\Sigma_\alpha,
\end{aligned}$$

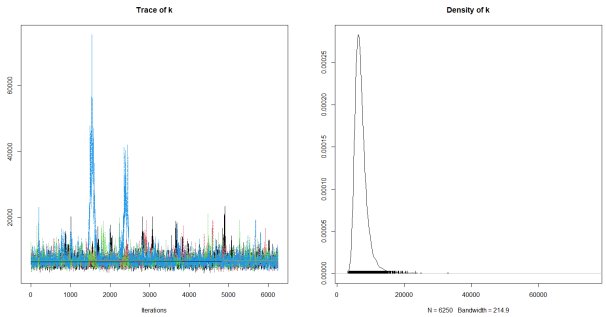
where $f(\boldsymbol{\theta_{t-1}}, \boldsymbol{\tau}) = \begin{pmatrix} \alpha_{1(t-1)} \\ \alpha_{2(t-1)} \\ \alpha_{3(t-1)} \end{pmatrix}$, and $\Sigma_\alpha = \begin{bmatrix} \alpha_1(1-\alpha_1) & -\alpha_1\alpha_2 & -\alpha_1\alpha_3 \\ -\alpha_1\alpha_2 & \alpha_2(1-\alpha_2) & -\alpha_2\alpha_3 \\ -\alpha_1\alpha_3 & -\alpha_2\alpha_3 & \alpha_3(1-\alpha_3) \end{bmatrix}$.

C Trace Plots for κ, λ^I and λ^R

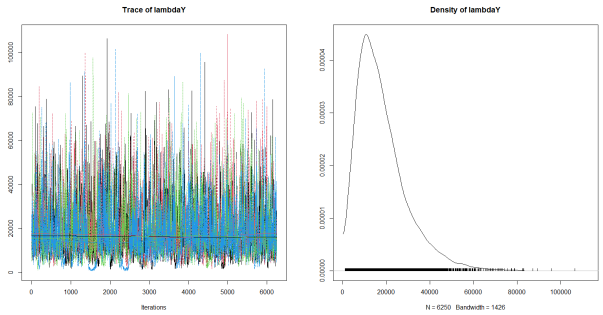
Figure 16 below show the trace plots and densities for κ, λ^I and λ^R obtained from the JAGS simulation:

Figure 16: Additional SIR Model Parameters

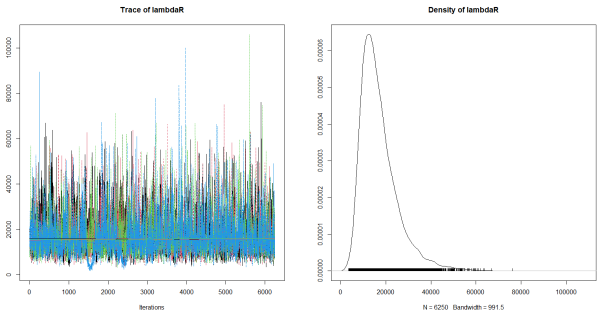
(a) Trace and density for κ samples



(b) Trace and density for λ^I samples



(c) Trace and density for λ^R samples



D Full script

```
#package library
library(gridExtra)
library(dplyr)
library(utils)
library(ggplot2)
options(stringsAsFactors = FALSE)
options(scipen=999)
library(readxl)
library(stats)
library(lognorm)
library(emdbook)
library(rjags)
library(runjags)
library(R2OpenBUGS)
library(coda)
library(nimble)
library(Formula)
library(DirichletReg)
```

```
19
20
21 #Very basic Simulated SIR plots
22 beta=2
23 gamma=1.4
24 S_sim=rep(0,35)
25 I_sim=rep(0,35)
26 R_sim=rep(0,35)
27 S_sim[1]=0.9
28 I_sim[1]=0.0004
29 R_sim[1]=1-S_sim[1]-I_sim[1]
30
31 for (i in 2:35)
32 {
33   S_sim[i]=S_sim[i-1]-beta*S_sim[i-1]*I_sim[i-1]
34   I_sim[i]=I_sim[i-1]+beta*S_sim[i-1]*I_sim[i-1]-gamma*I[i-1]
35   R_sim[i]=R_sim[i-1]+gamma*I_sim[i-1]
36
37 }
38 plot(S_sim,type="l",col="green",xlab="Time",ylab="Susceptible␣Proportion")
39 plot(I_sim,col="red",type="l",xlab="Time",ylab="Infected␣Proportion")
40 plot(R_sim,col="blue",type="l",xlab="Time",ylab="Recovered␣Proportion")
41
42 #Data (clean)
43 FR <- read_excel("Cours␣Uni␣+␣important␣documents/Masters␣Econometrics/
44 Thesis␣Active/Code/LatestDataFR.xlsx")
45 #to correct negative recovery rates, take data from two previous weeks
46 rec <- read_excel("Cours␣Uni␣+␣important␣documents/Masters␣Econometrics/
47 Thesis␣Active/Code/rec.xlsx", col_names = FALSE)
48 datesdaily<-seq(as.Date("2020/07/01"), by = "day", length.out = 347)
49 datesweek<-seq(as.Date("2020/07/01"), by = "week", length.out = 52+4)
50
51
52 N<-65387859
53 Infected<-as.numeric(FR[1,1:287])/N
54 Death<-as.numeric(FR[2,1:287])/N
55 Recovered<-as.numeric(FR[3,1:287])/N
56 Removed<-Recovered+Death
57 BigPolicy<-as.numeric(FR[4,])
58
59
60 VaccinesModPre<-rep(0,71)
61 single<-(-0.06/71)
62 VaccinesModPre[1]<-single
```

```
for (i in 2:71)
{
  VaccinesModPre[i]<- VaccinesModPre[i-1]+single
}
#plot(VaccinesModPre)
for (i in 1:71)
{
  BigPolicy[i+216]<-BigPolicy[i+216]+VaccinesModPre[i]
}


n_days<-as.numeric(length(Infected))
n_weeks<-n_days/7
T_fin=n_weeks+15 #=364 days, approx a year.

#Integrate vaccine in the Pi function:


weeklyInfected<-rep(0,n_weeks)
weeklyDeath<-rep(0,n_weeks)
weeklyRecovered<-rep(0,n_weeks)
weeklyBigPolicy<-rep(0,T_fin)

for (i in 1:n_weeks)
{
  weeklyInfected[i]<-Infected[i*7]
  weeklyDeath[i]<-Death[i*7]
  weeklyRecovered[i]<-Recovered[i*7]
}

for (i in 1:T_fin)
{
  weeklyBigPolicy[i]<-BigPolicy[i*7]
}

weeklyPolicy<-weeklyBigPolicy[2:n_weeks]
weeklyPredPolicy<-weeklyBigPolicy[(n_weeks+1):T_fin]

#Fix negative values due to first difference operator based on rates of
recoveries from previous weeks
weeklybb<-rep(0,43)
for (i in 1:43)
{
  weeklybb[i]<-as.numeric(rec[i*7])/N
```

```
}
weeklybb<-diff(weeklybb*50)
weeklybb[3:4]<-weeklybb[2]
weeklybb[6]<-weeklybb[5]
weeklybb<-weeklybb[3:42]

nonscaledRecovery<-(weeklybb/50)+diff(weeklyDeath)


plot(datesweek[1:n_weeks-1],diff(weeklyInfected)*N/7,type="l",xlab="Time"
,ylab="Cases",col="red")
lines(datesweek[1:n_weeks-1],nonscaledRecovery*N/7,type="l",col="blue")


#Upscale the data & and correct the negative values in Recovered based on the
data from previous weeks
Y <- diff(weeklyInfected*3)
R <- weeklybb+diff(weeklyDeath)
#plot(datesweek[1:(n_weeks-1)],diff(weeklyDeath))

#Visualisation
plot(datesweek[2:n_weeks],Y*N,type="l",xlab="Time",ylab="Cases",col="red")
lines(datesweek[2:n_weeks],R*N,type="l",col="blue")
plot(datesweek[2:n_weeks],weeklyPolicy, type="s",lwd=1,xlab="Time",
ylab="Policy␣intensity",col="red")
#plot(datesweek[1:(n_weeks-1)],diff(weeklyDeath))


#Lognormal parameter maker
lognorm.parm <- function(mu0,var0){
  var <- log(var0 / mu0^2 + 1)
  mu <- log(mu0) - var / 2
  list(mu = mu, var = var)}

#initial conditions / hyperparameters
beta0=0.20
gamma0=0.16
gamma0_sd=0.1
gamma_var <- gamma0_sd^2
R0=beta0/gamma0 #returns about 1.214
R0_sd=0.2
R0_var <- R0_sd^2
lognorm_gamma <- lognorm.parm(gamma0,gamma_var)
gam0<-c(lognorm_gamma$mu,lognorm_gamma$var)
```

```
lognorm_R0 <- lognorm.parm(R0,R0_var)
initR0<-c(lognorm_R0$mu,lognorm_R0$var)
state0=c(1-Y[1]-R[1],Y[1],R[1])

#MCMC  parameters
M=62500
thn=10
nchain=4
nadapt=50000
nburnin=12500


#MCMC to JAGS
model.string <-paste0("
model{

 #Hyperpriors
  theta[1,1:3] ~ ddirch(state0)
  gamma ~ dlnorm(gam0[1], 1/gam0[2])
  R0 ~ dlnorm(initR0[1],1/initR0[2])
  beta <- R0*gamma
  k ~ dgamma(20,0.0001)
  lambdaY ~ dgamma(2,0.0001)
  lambdaR ~ dgamma(2,0.0001)

  for(t in 2:(n_weeks)){

    KtS[t-1,1] <- -beta*weeklyPolicy[t-1]*theta[t-1,1]*theta[t-1,2]
    KtI[t-1,1] <- beta*weeklyPolicy[t-1]*theta[t-1,1]*theta[t-1,2]
    -gamma*theta[t-1,2]
    KtR[t-1,1] <- gamma*theta[t-1,2]

    KtS[t-1,2] <- -beta*weeklyPolicy[t-1]*(theta[t-1,1]+0.5*KtS[t-1,1])*
    (theta[t-1,2]+0.5*KtI[t-1,1])
    KtI[t-1,2] <- beta*weeklyPolicy[t-1]*(theta[t-1,1]+0.5*KtS[t-1,1])*
    (theta[t-1,2]+0.5*KtI[t-1,1]) - gamma*(theta[t-1,2]+0.5*KtI[t-1,1])
    KtR[t-1,2] <- gamma*(theta[t-1,2]+0.5*KtI[t-1,1])

    KtS[t-1,3] <- -beta*weeklyPolicy[t-1]*(theta[t-1,1]+0.5*KtS[t-1,2])*
    (theta[t-1,2]+0.5*KtI[t-1,2])
    KtI[t-1,3] <- beta*weeklyPolicy[t-1]*(theta[t-1,1]+0.5*KtS[t-1,2])*
    (theta[t-1,2]+0.5*KtI[t-1,2]) - gamma*(theta[t-1,2]+0.5*KtI[t-1,2])
    KtR[t-1,3] <- gamma*(theta[t-1,2]+0.5*KtI[t-1,2])

```

```
    KtS[t-1,4] <- -beta*weeklyPolicy[t-1]*(theta[t-1,1]+KtS[t-1,3])*
    (theta[t-1,2]+KtI[t-1,3])
    KtI[t-1,4] <- beta*weeklyPolicy[t-1]*(theta[t-1,1]+KtS[t-1,3])*
    (theta[t-1,2]+KtI[t-1,3]) - gamma*(theta[t-1,2]+KtI[t-1,3])
    KtR[t-1,4] <- gamma*(theta[t-1,2]+KtI[t-1,3])

    alpha[t-1,1] <- theta[t-1,1]+(KtS[t-1,1]+2*KtS[t-1,2]+2*KtS[t-1,3]+KtS[t-1,4])/6
    alpha[t-1,2] <- theta[t-1,2]+(KtI[t-1,1]+2*KtI[t-1,2]+2*KtI[t-1,3]+KtI[t-1,4])/6
    alpha[t-1,3] <- theta[t-1,3]+(KtR[t-1,1]+2*KtR[t-1,2]+2*KtR[t-1,3]+KtR[t-1,4])/6

    theta[t,1:3] ~ ddirch(k*alpha[t-1,1:3])
    Y[t-1] ~ dbeta(lambdaY*theta[t,2],lambdaY*(1-theta[t,2]))
    R[t-1] ~ dbeta(lambdaR*theta[t,3],lambdaR*(1-theta[t,3]))
  }
}
")

model.spec<-textConnection(model.string)
posterior <- jags.model(model.spec,
data = list ('Y'=Y, 'R'=R, 'n_weeks'= n_weeks,'weeklyPolicy'=weeklyPolicy,'state0'=sta
'gam0'=gam0,'initR0'=initR0),  n.chains = nchain,  n.adapt = nadapt)

adapt(posterior, nadapt,end.adaptation=FALSE)
adapt(posterior, nadapt,end.adaptation=FALSE)
adapt(posterior, nadapt,end.adaptation=FALSE)
adapt(posterior, nadapt,end.adaptation=FALSE)
adapt(posterior, nadapt,end.adaptation=FALSE)
adapt(posterior, nadapt,end.adaptation=FALSE)

update(posterior,nburnin)
jags_sample <- jags.samples(posterior,c('theta','gamma','R0','beta','Y','R',
'lambdaY','lambdaR','k'), n.iter = M, thin = thn)

#clean up
remove(rec)
remove(weeklybb)
remove(Infected)
remove(Recovered)
remove(Death)
remove(BigPolicy)
remove(FR)
remove(n_days)
remove(i)
remove(weeklyInfected)
```

```
remove(weeklyRecovered)
remove(model.spec)
remove(model.string)
remove(beta0)
remove(gamma0)
remove(gamma_var)
remove(gamma0_sd)
remove(gam0)
remove(R0)
remove(R0_sd)
remove(R0_var)
remove(initR0)
remove(state0)
remove(nadapt)
remove(nburnin)
remove(nchain)
remove(M)
remove(thn)
remove(lognorm_gamma)
remove(lognorm_R0)
remove(posterior)
remove(lognorm.parm)

#get some plots
#posterior true probabilities, instead of Y and R.
plot(as.mcmc.list(jags_sample$theta)[,(1:3)*(n_weeks)])
plot(as.mcmc.list(jags_sample$R0))
plot(as.mcmc.list(jags_sample$gamma))
plot(as.mcmc.list(jags_sample$beta))
plot(as.mcmc.list(jags_sample$lambdaY))
plot(as.mcmc.list(jags_sample$lambdaR))
plot(as.mcmc.list(jags_sample$k))


#make array and get some statistics
R0_p <- unlist(as.mcmc.list(jags_sample$R0))
gamma_p <- unlist(as.mcmc.list(jags_sample$gamma))
beta_p <- unlist(as.mcmc.list(jags_sample$beta))
lambdaY_p <- unlist(as.mcmc.list(jags_sample$lambdaY))
lambdaR_p <- unlist(as.mcmc.list(jags_sample$lambdaR))
k_p <- unlist(as.mcmc.list(jags_sample$k))

drawsize=length(R0_p)

```

```
#stats on these posteriors (confidence intervals)
theta_p <- array(Reduce(rbind,as.mcmc.list(jags_sample$theta)),
dim=c(drawsize,n_weeks,3))
theta_p_ci <- as.vector(apply(theta_p[,n_weeks,],2,quantile,
c(0.025,0.25,0.5,0.75,0.975)))
theta_p_mean <- apply(theta_p[,n_weeks,],2,mean)

#theta mean across all simulation, for each time t: (beautiful)
theta_p_tS_mean <- apply(theta_p[,,1],2,mean)
theta_p_tS_SD <- mean(apply(theta_p[,,1],2,sd))


theta_p_tI_mean <- apply(theta_p[,,2],2,mean)
theta_p_tI_SD <- mean(apply(theta_p[,,2],2,sd))

theta_p_tR_mean <- apply(theta_p[,,3],2,mean)
theta_p_tR_SD <- mean(apply(theta_p[,,3],2,sd))

#Mean values at each time t for all the posterior draws
plot(datesweek[1:n_weeks],theta_p_tI_mean,type="l",xlab="Time",
ylab="Proportion␣of␣cases",col="red")
lines(datesweek[1:n_weeks],theta_p_tR_mean,col="blue")


#plot(datesweek[1:n_weeks],theta_p_tR_mean,type="l",xlab="Time",
ylab="Proportion␣of␣Removal",col="blue")


#We can compare how the obtained Thetas fare against the actual measured data
plot(datesweek[2:n_weeks],theta_p_tR_mean[1:40],type="l",xlab="Time",
ylab="Proportion␣of␣Removal",col="blue",ylim = c(0, max(R)))
lines(datesweek[2:n_weeks],R)

plot(datesweek[2:n_weeks],theta_p_tI_mean[1:40],type="l",xlab="Time",
ylab="Proportion␣of␣Infections",col="red",ylim = c(0, max(Y)))
lines(datesweek[2:n_weeks],Y)


#Addional statistics
R0_p_mean <- mean(R0_p)
R0_p_ci <- quantile(R0_p,c(0.025,0.25,0.5,0.75,0.975))
R0_p_sd<-sd(R0_p)
gamma_p_mean <- mean(gamma_p)
gamma_p_ci <- quantile(gamma_p,c(0.025,0.25,0.5,0.75,0.975))
```

```
gamma_p_sd<-sd(gamma_p)
beta_p_mean <- mean(beta_p)
beta_p_ci <- quantile(beta_p,c(0.025,0.25,0.5,0.75,0.975))
beta_p_sd<-sd(beta_p)

#Addtional results for appendix.
lambdaY_p_mean <- mean(lambdaY_p)
lambdaY_p_ci <- quantile(lambdaY_p,c(0.025,0.25,0.5,0.75,0.975))
lambdaY_p_sd<- sd(lambdaY_p)

lambdaR_p_mean <- mean(lambdaR_p)
lambdaR_p_ci <- quantile(lambdaR_p,c(0.025,0.25,0.5,0.75,0.975))
lambdaR_p_sd<- sd(lambdaR_p)

k_p_mean <- mean(k_p)
k_p_ci <- quantile(k_p,c(0.025,0.25,0.5,0.75,0.975))
k_p_sd <- sd(k_p)

#FORECAST#
theta_f <- array(0,dim=c(drawsize,(T_fin-n_weeks),3))
Y_f <- matrix(NA,nrow=drawsize,ncol=(T_fin-n_weeks))
R_f <- matrix(NA,nrow=drawsize,ncol=(T_fin-n_weeks))
weeklyPredNoPolicy<-rep(1,T_fin-n_weeks)
eps<-1e-10

percentagediff<-(11392352-4276086)/N
single<-(-percentagediff)/9
#nine periods between 14th April and 9th June,
then we interpolate vaccine roll-out for future period

#Vax Mod Prediction
VaccinesModPost<-rep(0,15)
VaccinesModPost[1]<-single
for (i in 2:15)
{
  VaccinesModPost[i]<- VaccinesModPost[i-1]+single
}
#continue lockdown for three first periods
PolCont<-rep(weeklyPolicy[n_weeks-1],15)  of prediction sample (-+ 28th April)
StrictPol<-PolCont
NoPol<-rep(0.8,15)
NoPol[1:3]<-weeklyPolicy[n_weeks-1]

for (i in 1:12)
```

```
{
  PolCont[i+3]<-exp(0.04*i)-1+weeklyPolicy[n_weeks-1]
}
plot(PolCont)
#Integrate vaccine effects in all Policy scenarios
for (i in 1:15)
{
  PolCont[i]<-PolCont[i]+VaccinesModPost[i]
  StrictPol[i]<-StrictPol[i]+VaccinesModPost[i]
  NoPol[i]<-NoPol[i]+VaccinesModPost[i]
}

plot(datesweek[(n_weeks+1):T_fin],PolCont,xlab="Time",
ylab="Policy␣Intensity",col="red",type="l")

plot(datesweek[(n_weeks+1):T_fin],StrictPol,xlab="Time",
ylab="Policy␣Intensity",col="red",type="l")

plot(datesweek[(n_weeks+1):T_fin],NoPol,xlab="Time",
ylab="Policy␣Intensity",col="red",type="l")



for(i in 1:drawsize){
  thetalt1 <- theta_p[i,(n_weeks),1]
  thetalt2 <- theta_p[i,(n_weeks),2]
  thetalt3 <- theta_p[i,(n_weeks),3]
  beta_i <- c(beta_p)[i]
  gamma_i <- c(gamma_p)[i]
  k_i <- c(k_p)[i]
  lambdaY_i <- c(lambdaY_p)[i]
  lambdaR_i <- c(lambdaR_p)[i]
  if(beta_i<0 |gamma_i<0 |thetalt1<0 |thetalt2<0 |thetalt3<0) next

  for(t in 1:(T_fin-n_weeks)){
    KtS<-NULL
    KtI<-NULL
    KtR<-NULL
    alpha_f<-NULL
    Pol<-PolCont[t]

    KtS[1] <- -beta_i*Pol*thetalt1*thetalt2
    KtI[1] <-  beta_i*Pol*thetalt1*thetalt2 - gamma_i*thetalt2
    KtR[1] <-  gamma_i*thetalt2
```

```
415
416     KtS[2] <- -beta_i*Pol*(thetalt1+0.5*KtS[1])*(thetalt2+0.5*KtI[1])
417     KtI[2] <-  beta_i*Pol*(thetalt1+0.5*KtS[1])*(thetalt2+0.5*KtI[1])
418     -gamma_i*(thetalt2+0.5*KtI[1])
419     KtR[2] <-  gamma_i*(thetalt2+0.5*KtI[1])
420
421     KtS[3] <- -beta_i*Pol*(thetalt1+0.5*KtS[2])*(thetalt2+0.5*KtI[2])
422     KtI[3] <-  beta_i*Pol*(thetalt1+0.5*KtS[2])*(thetalt2+0.5*KtI[2])
423     -gamma_i*(thetalt2+0.5*KtI[2])
424     KtR[3] <-  gamma_i*(thetalt2+0.5*KtI[2])
425
426     KtS[4] <- -beta_i*Pol*(thetalt1+KtS[3])*(thetalt2+KtI[3])
427     KtI[4] <-  beta_i*Pol*(thetalt1+KtS[3])*(thetalt2+KtI[3])
428     -gamma_i*(thetalt2+KtI[3])
429     KtR[4] <-  gamma_i*(thetalt2+KtI[3])
430
431     alpha_f[1] <- thetalt1+(KtS[1]+2*KtS[2]+2*KtS[3]+KtS[4])/6
432     alpha_f[2] <- thetalt2+(KtI[1]+2*KtI[2]+2*KtI[3]+KtI[4])/6
433     alpha_f[3] <- thetalt3+(KtR[1]+2*KtR[2]+2*KtR[3]+KtR[4])/6
434
435     alpha_f[2]<-pmax(alpha_f[2],eps)
436
437     thetalt_tmp <- rdirichlet(1,k_i*c(alpha_f))
438     thetalt1<-theta_f[i,t,1] <- thetalt_tmp[1]
439     thetalt2<-theta_f[i,t,2] <- thetalt_tmp[2]
440     thetalt3<-theta_f[i,t,3] <- thetalt_tmp[3]
441
442     Y_f[i,t] <- rbeta(1,lambdaY_i*thetalt2,lambdaY_i*(1-thetalt2))
443     R_f[i,t] <- rbeta(1,lambdaR_i*thetalt3,lambdaR_i*(1-thetalt3))
444   }
445 }
446
447 remove(thetalt1)
448 remove(thetalt2)
449 remove(thetalt3)
450 remove(beta_i)
451 remove(gamma_i)
452 remove(i)
453 remove(k_i)
454 remove(lambdaY_i)
455 remove(lambdaR_i)
456
457
458 Y_fmean <- c(colMeans(Y_f))
```

```
R_fmean <- c(colMeans(R_f))

Y_fCI95 <- data.frame(t(apply(Y_f,2,quantile,probs=c(0.025,0.975))))
Y_fCI50 <- data.frame(t(apply(Y_f,2,quantile,probs=c(0.25,0.75))))

R_fCI95 <- data.frame(t(apply(R_f,2,quantile,probs=c(0.025,0.975))))
R_fCI50 <- data.frame(t(apply(R_f,2,quantile,probs=c(0.25,0.75))))



plot(datesweek[(n_weeks+1):T_fin],Y_fmean*N,type="l",col="black",lwd=1,
ylim = c(min(Y_fCI95[,1])*N, max(Y_fCI95[,2])*N),xlab="Time",
ylab="New␣weekly␣probable␣cases")

#plot(datesweek[(n_weeks+1):T_fin],Y_fmean*N,type="l",col="black",lwd=1,
ylim = c(min(Y_fmean*N), max(Y_fmean)*N),xlab="Time",
ylab="New␣weekly␣probable␣cases")

polygon(c(datesweek[(n_weeks+1):T_fin],rev(datesweek[(n_weeks+1):T_fin])),
c(Y_fCI95[,1]*N,rev(Y_fCI95[,2]*N)),col = rgb(0.5, 0, 0,0.2), border = NA)

polygon(c(datesweek[(n_weeks+1):T_fin],rev(datesweek[(n_weeks+1):T_fin])),
c(Y_fCI50[,1]*N,rev(Y_fCI50[,2]*N)),col = rgb(0.1, 0.8, 0.2,0.2), border = NA)



thetaI_fmean <- c(colMeans(theta_f[,,2]))
#thetaI_fCI <- data.frame(t(apply(theta_p[,-1,2],2,quantile,probs=c(0.025,0.975),
na.rm=T)))
#thetaR_fmean <- c(colMeans(theta_f[,,3]))
#thetaS_fmean <- c(colMeans(theta_f[,,1]))

#acquisition of key statistics (turning points, first derivatives) put mean from simul
thetaS_mean <- c(colMeans(theta_p[,-1,1],na.rm = T),colMeans(theta_f[,,1],
na.rm = T))
thetaI_mean <- c(colMeans(theta_p[,-1,2],na.rm = T),colMeans(theta_f[,,2],
na.rm = T))
thetaR_mean <- c(colMeans(theta_p[,-1,3],na.rm = T),colMeans(theta_f[,,3],
na.rm = T))

plot(datesweek[1:T_fin],thetaI_mean[1:T_fin+1],type="l",xlab="Time",
ylab="Proportion␣of␣cases",col="red",ylim = c(0, 0.02))

#theta_p_tR_mean <- apply(theta_p[,,3],2,mean)
```

```
#plot(datesweek[1:n_weeks],theta_p_tR_mean,type="l",xlab="Time",
ylab="Proportion of Removal",col="blue")

lines(datesweek[1:T_fin],thetaR_mean[1:T_fin+1],col="blue")
abline(v=as.Date("2021-04-01"))
```