Uncertainty Quantification Project 1

Group 3:
Jake Denton
Archie Boswell

December 10, 2021

Abstract

In this report we investigate the basic reproduction number R_0 of COVID-19 using data on the daily number of cases over a period of 98 days using non-linear regression techniques. We find that the best estimates for R_0 are near 3, and 95% confidence intervals indicate that the true value of R_0 is likely between 1.97 and 4.95. Given this, and the SAGE criterion of $R_0 > 1.3$, we are very confident that restrictions should not be relaxed.

1 Introduction

1.1 Structure and aims

The main goal of this report is to analyse uncertainty regarding the basic reproduction number (R_0) and using this, provide evidence that supports the tightening or relaxation of COVID-19 restrictions. To meet this aim, the SEIR epidemic model is fitted to data and it's parameters analysed using non-linear regression methods and subsequent asymptotic normality in the parameters.

This report begins with a short background on the SEIR model, it's parameters and the quantity of interest R_0. From here, the data used for the fitting is explored before details of the fitting are covered. Following this, the results from the model are discussed, including: a plot of the fit to the data, a histogram of the R_0 values, estimates for R_0 itself and multiple confidence intervals based on simulation and theory. Afterwards, we discuss the possible issues with fitting a model in this way and some further analysis based on initialising the fit with different starting parameters. Finally, we will conclude the report with our recommendation to SAGE.

1.2 Background

Epidemic modelling is a key tool for guiding public health measures, as it gives an idea of how a disease spreads through a population. Modelling of this kind helps scientists and politicians become more aware of the potential risks posed

by a pathogen so that they can actively make decisions that protect vulnerable members of society. Throughout this report, we use the SEIR compartmental epidemic model, which is governed by the following system:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} \\
\frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\
\frac{dI}{dt} &= \sigma E - \gamma I \\
\frac{dR}{dt} &= (1-p)\gamma I \\
\frac{dD}{dt} &= p\gamma I
\end{aligned}$$

Here, S, E, I, R and D represent the susceptible, exposed (in the incubation period, they are infected but not yet infectious), infectious, recovered and dead. N is the sum of the living members of the population i.e. $N = S + E + I + R$.Note that throughout this report, t refers to the time in days. There is also a sixth differential equation, which governs changes in the cumulative number of infected cases, given below:

$$\frac{dC}{dt} = \sigma E$$

The cumulative number of infected cases is particularly useful as it is very accessible from a data perspective. Indeed, in this report we use the number of daily cases, which is easy to source (using the results of testing) from which the cumulative number of cases can be found.

There are a number of parameters involved in the model. β denotes the infection rate, which can be thought of as the probability of contact between an infectious person and a susceptible multiplied by the number of contacts per day. $\frac{1}{\sigma}$ and $\frac{1}{\gamma}$ denote the average incubation and infectious period in days. Lastly, p denotes the death rate. Clearly, all of these parameters are non-negative, and we can confidently assume that γ and σ are less than 1 (so that the average incubation/infectious periods are longer than 1 day).

The idea behind the model is simple; a susceptible individual comes into contact with the pathogen and enters the incubation period for a time, until they become infectious. They can then pass the pathogen onto other susceptibles, and after a time either recover or die. The model makes a number of assumptions, which are detailed below:

- The population remains constant, with the exception of deaths caused by the virus (no immigration, emigration, births or deaths from other causes).
- Once an individual has recovered from the infection, they are permanently immune.
- All parameters are constant and do not change with time.
- Everyone who is susceptible will remain susceptible until they become infected.
- All susceptible individuals have equal probability of becoming infected.

The basic reproduction number, R_0, is the average number of secondary infections that result from a single infected individual. It is clear that if this number exceeds 1, then the size of the epidemic will grow over time as each individual passes the pathogen on to more than one individual (on average), and this larger number of individuals pass the pathogen on to even more people, and so on. The formula to calculate R_0 is given by $R_0 = \frac{\beta}{\gamma}$. This is intuitive, as β is the average number of individuals infected by one infected person each day, and $\frac{1}{\gamma}$ is the average number of days that a person is infected. The formula tells us that if we can estimate the parameters β and γ, then we can estimate R_0.

2 Data

We have information on the number of daily COVID infections over a 98-day period with a population of 1000 individuals. We are given that the epidemic starts from one infected, everyone else is susceptible and the parameter $\sigma = 0.25$. As mentioned in the previous section, in order to translate this data into something we can use to fit the model, we can calculate the cumulative number of cases. These values are plotted against the number of days below:

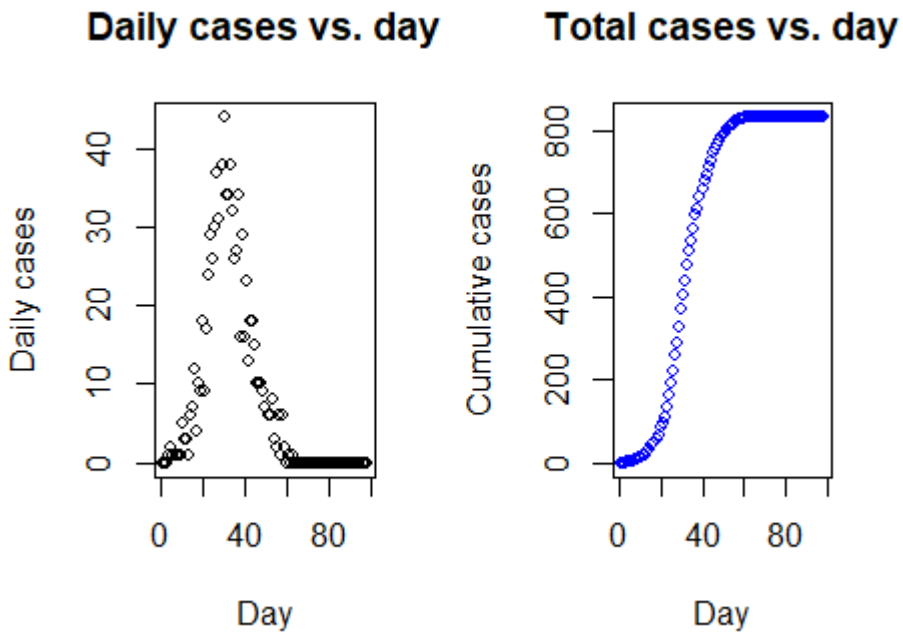

Figure 1: Plots of daily cases (data) and cumulative cases

Considering the left plot, we see that the number of daily cases explodes from about day 10, peaking at 44 cases on day 30. The number of daily cases falls off at about the same rate (very quickly), settling at 0 or 1 cases per day from day 60 onwards. The alarming rate at which infections occur is even more evident from the steepness of the right plot. The number of cases identified in total is 831, so 83.1 percent of this population became infected with the virus. It only takes 26 days for a quarter of the population to have been infected, all from one initial individual.

Now that we've had a look at the data, we can move on to describing how we will use it to fit the model.

3 Methods

3.1 Non-linear regression

One useful method to match the solutions of this system of ordinary differential equations to experimental data is non-linear regression. We have three parameters to fit: β,γ and p. In loose terms, regression methods aim to fit a model f to data y, which is assumed to be of the form $y = f(x, \theta) + \epsilon$, where ϵ is a vector of errors assumed to be independent with the same distribution, which has mean zero. Estimating f using a squared-error loss function, and assuming these errors are normally distributed, allows us to use the asymptotic properties of maximum likelihood estimators to quantify uncertainty about the parameter vector θ of f and the solutions f themselves. In our case, we can find distributions for the parameters β and γ then use the formula for R_0 and information from the two distributions to quantify uncertainty in R_0.

One way that non-linear regression methods can fit parameters is by using the Gauss-Newton algorithm. This is an iterative algorithm that takes an initial set of guesses for the parameter values before using the first-order Taylor expansion of f and residuals computed using the data to update the parameters. This process continues until convergence.

The initial estimates we use to fit the model are $(\beta, \gamma, p) = (0.5, 0.25, 0.05)$. For γ, recall that the reciprocal is the average length of the infectious period. Information online (e.g. NHS and gov.uk website) suggests that the average length of this period is between 3 and 10 days, meaning that gamma should be between $\frac{1}{10}$ and $\frac{1}{3}$. Secondly, for β, we know (from government press conferences say) that R_0 is normally between 0.5 and 3. Using the formula for R_0 and what we discussed regarding γ above, we pass $\beta = 0.5$ to the algorithm. Finally, the death rate p is bound to be small, since not many people (as a percentage of those infected) die of COVID. As a result, we pass the initial value p=0.05 to the algorithm. We also pass a lower bound to the algorithm, a very small positive number (1e-7) for all three parameters (since they are positive) and an upper bound on gamma at 1 (for reasons mentioned above).

Since we do not have the exact solution to the system available to us, we approximate the solution for each set of parameters using the Runge-Kutta method, which has a high order of accuracy (order 4). This means that we must pass initial values for the populations in the system. Luckily, we are given in the brief the initial values $(S(0), E(0), I(0), R(0), D(0), C(0)) = (999, 0, 1, 0, 0, 1)$.

Under the assumption of Gaussian errors, the MLE of the parameter vector θ is asymptotically normal if the number of data points is sufficiently large. Therefore, we have two approximate normal distributions for β and γ with which we can compute an estimate for and quantify uncertainty around the quantity of interest R_0. Methods for accomplishing this are detailed in the next section.

3.2 Ratio distribution and Confidence Intervals

The last section informs us that if we fit a non-linear regression model, we will receive parameters back that approximately follow normal distributions. The quantity of interest in this report is R_0, a ratio of β and γ. If these can be modelled by normal distributions (which are correlated), then how is R_0 distributed?

Firstly, what is the expectation of this distribution? We can use a multivariate Taylor expansion to approximate this. Note that to calculate the exact expectation, you would need to find $E(\frac{1}{\gamma})$. This moment does not exist (this is easily seen using the Law of the Unconscious Statistician and observing the factor $\frac{1}{x}$ in the integrand). This method, using Taylor expansions, is commonly referred to as the delta method. We expand around the means (μ_β, μ_γ) up to first order:

$$\frac{\beta}{\gamma} = \frac{\mu_\beta}{\mu_\gamma} + \frac{1}{\mu_\gamma}(\beta - \mu_\beta) - \frac{\mu_\beta}{\mu_\gamma^2}(\gamma - \mu_\gamma) + ...$$

Taking expectation gives:

$$E(\frac{\beta}{\gamma}) = \frac{\mu_\beta}{\mu_\gamma} + ...$$

The ellipses here refer to terms of higher order that we have ignored. Clearly one estimate for the expectation of R_0 is the ratio of the means. Using a second order Taylor expansion, a more accurate approximation is:

$$E(\frac{\beta}{\gamma}) \approx \frac{\mu_\beta}{\mu_\gamma} + \frac{\mu_\beta}{\mu_\gamma^3} Var(\gamma) - \frac{1}{\mu_\gamma^2} Cov(\gamma, \beta)$$

Okay, so we have approximations that we can use for the expectation of R_0, what about confidence intervals?

The distribution of the ratio of two non-central, correlated normal random variables was studied by Katz and Geary. Katz' approach, applied to our problem, is as follows:

$$R_0 \sim \frac{\mu_\beta + \mathcal{N}(0, \sigma_\beta^2)}{\mu_\gamma + \mathcal{N}(0, \sigma_\gamma^2)} = \frac{\mu_\beta + X}{\mu_\gamma + Y} = \frac{\mu_\beta}{\mu_\gamma} \frac{1 + \frac{X}{\mu_\beta}}{1 + \frac{Y}{\mu_\gamma}}$$

Here, β and γ are split up into sums of their means and zero-mean normal distributions X and Y, then there are some simple rearrangements. Taking the natural logarithm of both sides gives:

$$ln(R_0) = ln(\frac{\mu_\beta}{\mu_\gamma}) + ln(1 + \frac{X}{\mu_\beta}) - ln(1 + \frac{Y}{\mu_\gamma})$$

Finally, using the Taylor expansion of $ln(1 + \delta)$ (where δ assumed to be small, which it is for the above variables), we have:

$$ln(R_0) \approx ln(\frac{\mu_\beta}{\mu_\gamma}) + \frac{X}{\mu_\beta} - \frac{Y}{\mu_\gamma} \sim \mathcal{N}(ln(\frac{\mu_\beta}{\mu_\gamma}), \frac{\sigma_\beta^2}{\mu_\beta^2} + \frac{\sigma_\gamma^2}{\mu_\gamma^2})$$

This can be used to form a confidence interval around R_0, which we will call the Katz interval, with upper/lower bounds given by:

$$exp\left(ln(\frac{\mu_\beta}{\mu_\gamma}) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_\beta^2}{\mu_\beta^2} + \frac{\sigma_\gamma^2}{\mu_\gamma^2}} \right)$$

where α is the significance level and $z_{\frac{\alpha}{2}}$ is the appropriate quantile from the standard normal distribution.

A second way to find a confidence interval is by using the so-called Geary-Hinkley transformation (see [1] for reading on this). In our context, this is given as follows:

$$\frac{\mu_\gamma R_0 - \mu_\beta}{\sqrt{\sigma_\gamma^2 R_0^2 - 2\sigma_{\beta\gamma}\sigma_\beta\sigma_\gamma R_0 + \sigma_\beta^2}} \sim \mathcal{N}(0, 1)$$

Again, we can clearly make a confidence interval from this. This is a bit trickier as the denominator depends on R_0 as well as the numerator. Going through this calculation gives a quadratic which can be solved and analysed in order to satisfy the resulting inequality, this can be done by hand but is lengthy, so we omit the details.

We can now move on to actually fitting some non-linear regression models, analysing the distributions of the parameters outputted and quantifying uncertainty around R_0.

4 Results

A non-linear regression model was fitted to the cumulative cases data using the initial values and starting guesses for the parameters given in section 3.1. From this fitted model, the cumulative cases were extracted and unzipped into daily cases, so that a plot of the model with the original data can be seen. This is given below:

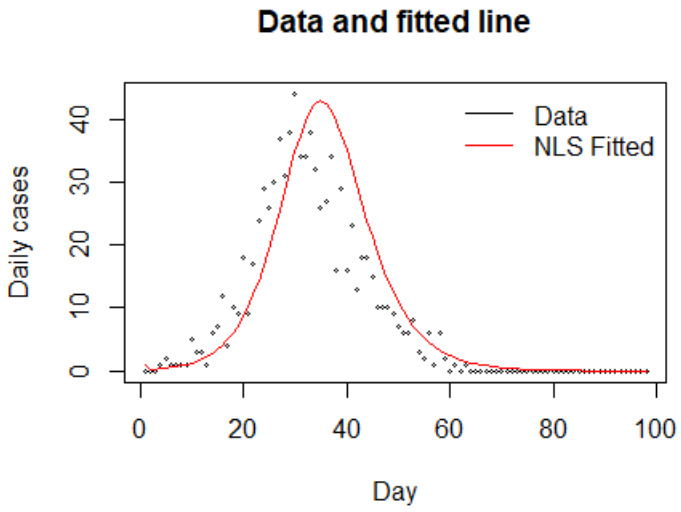

Figure 2: Plot showing fitted model and data

The fitted curve models the rise in cases between 0 and 30 days very well. The peak is below the maximum daily number of cases given in the data and occurs just after, at about the 40th day. As a result, the daily number of cases after this point overestimates those given in the data until approximately the 55th day.

The parameters we obtain from the model (the means of the approximate normal distributions for each variable) can be summarised as follows:

β	γ	p
0.8991097	0.3004804	0.0000001

The mean of the parameter β is almost three times that of γ, indicating that R_0 is likely to exceed 1.3. The death rate p is extremely small, in fact this is actually the minimum bound that was passed to the model fitter. The variance-covariance matrix for the parameters is given below:

Covariance Matrix			
Parameter	β	γ	p
β	0.011941325	0.006158393	-0.010062125
γ	0.006158393	0.003212934	-0.005076886
p	-0.010062125	-0.005076886	0.019949790

We can square root the variances for the standard deviations, which for β and p exceed 0.1. This means that, for example, β may be between 0.7 and 1.1 and p

is between 0 and 0.3 (a large interval, but since our subsequent analysis relates just to β and γ, this can be ignored).

Using the formula for R_0 given in the introduction, we can estimate its expected value using the delta method which was discussed in the previous section. The estimate based off the first-order expansion is 2.992241, and the more accurate estimate based off the second-order expansion is 3.030511. Both of these estimates greatly exceed the tolerance 1.3 by which we'd recommend tightening measures.

A sample of 100000 was taken from each of the normal distributions of β and γ with parameters given above and the value of R_0 calculated for each sample. The histogram below shows the result of this sampling, and gives an idea of the shape of the distribution of R_0.

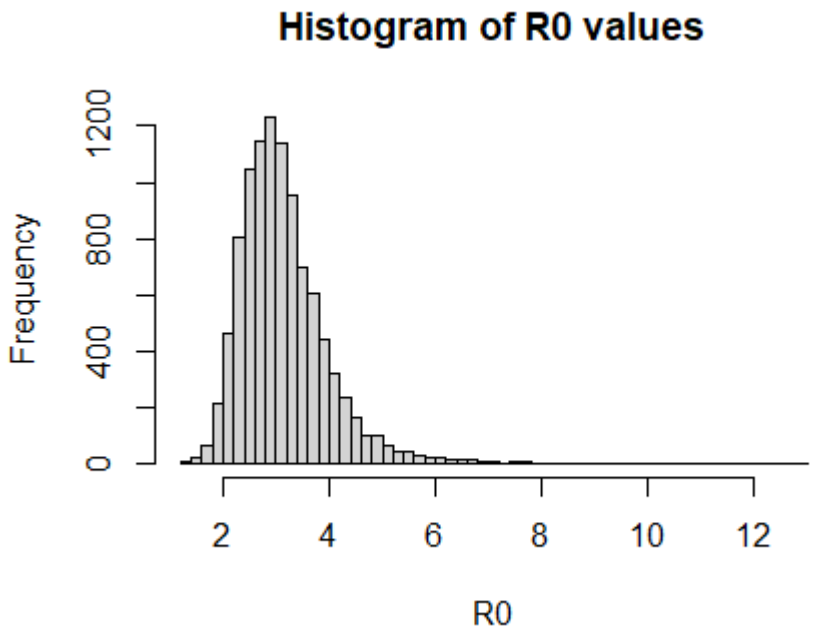

Figure 3: Histogram of R_0 values

The histogram shows that most values of R_0 lie between 2 and 4. There is a long tail showing that some large values of R_0 were observed. The minimum value observed in this sample of 100000 was 1.042911, which is below the threshold value 1.3 with which we have to base our recommendation, however in this sample just 7 values below 1.3 were actually observed. As a result, it seems that the true value of R_0 will exceed this threshold.

Using this large sample of values, we can estimate the 95% confidence interval for R_0. Sorting the sample in order of magnitude from smallest to largest and pulling the 2500th and 97500th values gives (1.972596,4.950001). Note that this interval does not contain the value 1.3.

In the previous section we introduced some theoretical distributions for R_0 based off the asymptotic normal distributions of β and γ. We calculated the 95% intervals using the Katz and Geary methods. The Katz method gives the interval (1.927455,4.645249) whilst the Geary method gives (1.974968,4.953367). These intervals match up closely with the interval obtained through simulation, and provide further evidence that the true value of R_0 exceeds 1.3.

In the following section, we discuss some of the flaws with the non-linear regression method and how this affects our results as well as the flaws of the SEIR model itself and assumptions we made about the parameters involved.

5 Discussion

One of the largest flaws in our method is the effect that different starting guesses for the parameters has on the fitted model. Changes to these initial guesses can lead to different outputs, which is why it's essential that many different guesses are tried and the fits analysed.

As a result, we passed different initial guesses for β and γ to the model fitter. The values varied between 0.1 and 1 in increments of 0.1. This gave back a lot of results where beta hits the upper bound at 1. Ignoring these, we plotted some of the different curves obtained to the daily case data. This is shown below:

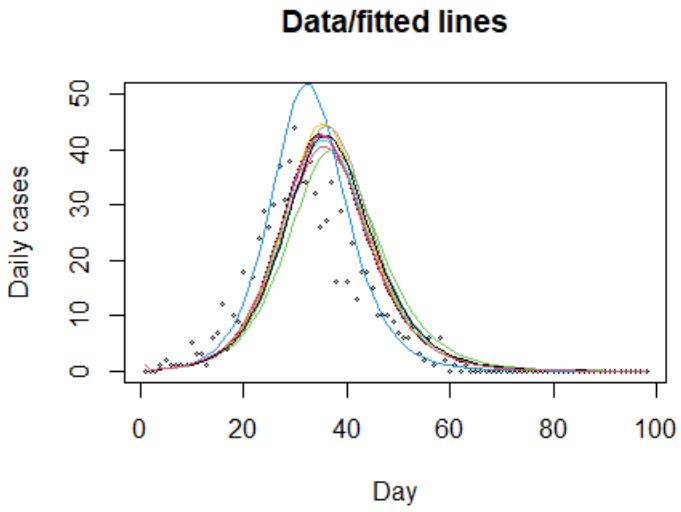

Figure 4: Plot showing models fitted with different initial guesses for parameters

The fitted curves (with the exception of the light blue curve) are all very similar to the curve we had in figure 2, which is a good sign in terms of our results being reliable. We decided to look at the light-blue curve more closely.

This curve is a result of the initial guesses (β,γ)=(0.5,0.1) being passed to the model. The fitted parameters and their variances are:

μ_β	μ_γ	σ_β^2	σ_γ^2
0.73907583	0.16379526	0.04290897	0.01530403

It is worth noting that the variances for the parameters for this model are about 4x larger for β and 5x larger for γ. As a result, it is very possible when sampling from the distribution of γ to get negative values, which is not acceptable with this model. The possibility of negative values makes simulating and approximating confidence intervals difficult and unreliable. For example, the Katz interval becomes much wider (between 0.93 and 21.88) as it relies on the assumption that $\frac{Y}{\mu_\gamma}$ is small, where $Y \sim \mathcal{N}(0, \sigma_\gamma^2)$. This is not true in general for this set of parameters, as μ_γ and σ_γ are of the same order of magnitude. Similarly, the Geary-Hinkley transformation is accurate only when γ is unlikely to take negative values, which is not the case here. Solving the quadratic equations from this theory gives only a lower bound for R_0 using this method at 1.359. As a result, we can be reasonably confident since all the other fits are close to what we had that our information on R_0 is reasonable.

Besides the fitting of the model, there are a few issues with the SEIR model itself. One challenge is that all infected individuals are assumed to be able

to spread the disease as effectively as each other. This does not account for differences between asymptomatic, mild and severe cases. On top of this, we know that sometimes COVID is passed on at scale by so-called superspreader events. It is obvious that not all infected individuals behave in the same way, and this is not accounted for in the model. Lastly, it is worth noting that we have not been told where our data has come from. As a result, we cannot confirm whether or not this sample of 1000 people is representative of the population as a whole. Without this information it is hard to provide a recommendation. Nevertheless, based on the criterion that restrictions should be relaxed if $R_0 < 1.3$, we provide our recommendation in the next section.

6 Recommendation

Drawing all of our results together, we can be very confident that the value of R_0 derived from the provided dataset exceeds the threshold value 1.3.

We make this conclusion supported by the histogram, estimates and confidence intervals discussed earlier in this report. The histogram and sampling showed that the vast majority of values sampled were within the range (1.97,4.95), which was supported by the theoretical confidence intervals. The value 1.3 is comfortably below the lower bound of this interval.

Fitting many non-linear regression models and analysing the fitted parameters repeatedly returned an estimate for R_0 of about 3. A final argument we haven't yet mentioned is that the number of daily cases given as data were likely under-reported due to limits on testing. Accounting for this, it's likely even more people were infected which would result in higher values for R_0.

Despite all this evidence, more work could be done to check the reliability of our results. We were not told where this data has come from and if it is a representative sample of the population. Other techniques such as bootstrapping and Monte Carlo could help get a better estimate for the true value of R_0, and other more sophisticated epidemic models could be investigated.

A Appendix

Selections of the R code we used to complete the work presented in this report can be found in this section. The first code snippet gives an idea of how the model was fitted.

```
seir.model<-function(t,x,params){
  beta=params[1]; gamma=params[2]; p=params[3]
  N<-x[1]+x[2]+x[3]+x[4] #keeps N up-to-date so dead removed from population
  dS<-((-1)*beta*x[1]*x[3])/N #change in susceptibles
  dE<-(beta*x[1]*x[3])/N-0.25*x[2] #change in exposed
  dI<-0.25*x[2]-gamma*x[3] #change in infected
  dR<-(1-p)*gamma*x[3] #change in recovered
```

```
  dD<-p*gamma*x[3] #change in dead
  dC<-0.25*x[2] #change in cumulative infected
  list(c(dS,dE,dI,dR,dD,dC)) #return list of all these
}

initial_values=c(999,0,1,0,0,1) #list of initial values
param.start=list(beta=0.5,gamma=0.25,p=0.005) #initial guesses for parameters
fills<-rep(0,98) #make a zero vector to fill covidData with
covidData<-data.frame(DailyCases$day,fills,fills,fills,
fills,fills,DailyCases$CumulativeCases)
colnames(covidData)=c("day","S","E","I","R","D","C") #give appropriate names
nls.fit=nlsLM(C~rk(initial_values,day,seir.model,
c(beta=beta,gamma=gamma,p=p))[,7], start=param.start,data=covidData,
lower=c(0.0000001,0.0000001,0.0000001), upper=c(1,1,1))
coef(nls.fit) #coefficients of the model
vcov(nls.fit) #variance of these coefficients
```

The next snippet simulates the distributions of the variables so that we can plot a histogram and finds confidence intervals.

```
set.seed(31415) #set seed for reproducibility
betaSamples<-rnorm(100000,mean=coef(nls.fit)[1],sd=sqrt(vcov(nls.fit)[1,1]))
gammaSamples<-rnorm(100000,mean=coef(nls.fit)[2],sd=sqrt(vcov(nls.fit)[2,2]))
R0Samples<-rep(0,100000) #initialise vector for R0 values
for (i in 1:100000)
{
  R0Samples[i]<-betaSamples[i]/gammaSamples[i]
}

R0Samples<-sort(R0Samples)
Interval<-c(R0Samples[2500],R0Samples[97500])

#Katz method for confidence interval (nicely lines up with what we have)
logmean<-log(coef(nls.fit)[1]/coef(nls.fit)[2])
logvariance<-vcov(nls.fit)[1,1]/coef(nls.fit)[1]^2+
vcov(nls.fit)[2,2]/coef(nls.fit)[2]^2
lowerKatz<-exp(logmean-qnorm(0.975)*sqrt(logvariance))
upperKatz<-exp(logmean+qnorm(0.975)*sqrt(logvariance))
intervalKatz<-c(lowerKatz,upperKatz)

#Geary-Hinkley transformation for confidence interval
a<-coef(nls.fit)[2]^2-qnorm(0.975)^2*vcov(nls.fit)[2,2]
b<-2*(vcov(nls.fit)[1,2]*sqrt(vcov(nls.fit)[1,1]*vcov(nls.fit)[2,2])
*qnorm(0.975)^2-coef(nls.fit)[1]*coef(nls.fit)[2])
c<-coef(nls.fit)[1]^2-qnorm(0.975)^2*vcov(nls.fit)[1,1]
TSolutionUpper<-(-b+sqrt(b^2-4*a*c))/(2*a)
TSolutionLower<-(-b-sqrt(b^2-4*a*c))/(2*a)
```

```
a2<-coef(nls.fit)[2]^2+qnorm(0.975)^2*vcov(nls.fit)[2,2]
b2<--2*(vcov(nls.fit)[1,2]*sqrt(vcov(nls.fit)[1,1]*vcov(nls.fit)[2,2
])*qnorm(0.975)^2+coef(nls.fit)[1]*coef(nls.fit)[2])
c2<-coef(nls.fit)[1]^2+qnorm(0.975)^2*vcov(nls.fit)[1,1]
determinant<-b2^2-4*a2*c2 #There is no real solution on this side
c2 #This is positive, can prove by completing the square
intervalGeary<-c(TSolutionLower,TSolutionUpper)
```

Multiple models were fitted in order to produce the plot in the discussions section. Part of the code for finding the parameters with different starting guesses is given below:

```
#Do a grid search, pass different initial guesses to model fit
initialbeta<-seq(0.1,1,by=0.1)
initialgamma<-seq(0.1,1,by=0.1)
searchInitials=matrix(0,nrow=100,ncol=6)
rownumber<-1
for (i in 1:10)
{
  for (j in 1:10)
  {
    initialcoefs<-c(initialbeta[i],initialgamma[j])
    searchInitials[rownumber,1:2]<-initialcoefs
    param.start1=list(beta=initialbeta[i],gamma=initialgamma[j],p=0.005)
    nls.fit1=nlsLM(C~rk(initial_values,day,seir.model,
    c(beta=beta,gamma=gamma,p=p))[,7],start=param.start1,
    data=covidData,lower=c(0.0000001,0.0000001,0.0000001),
    upper=c(1,1,1)) #fit non-linear regression model
    finalcoefs<-coef(nls.fit1)[1:2]
    searchInitials[rownumber,3:4]<-finalcoefs
    finalvars<-c(vcov(nls.fit1)[1,1],vcov(nls.fit1)[2,2])
    searchInitials[rownumber,5:6]<-finalvars
    rownumber<-rownumber+1
  }
}
#Now look at light blue curve
sensibleRows<-which(searchInitials[,3]<1)
dailyFits=matrix(0,nrow=length(which(searchInitials[,3]<1)),ncol=98)
for (i in 1:length(which(searchInitials[,3]<1)))
{
  row<-sensibleRows[i]
  param.start_fitted=c(beta=searchInitials[row,3],
  gamma=searchInitials[row,4],p=0.005)
  solution=as.data.frame(rk(initial_values,t,seir.model,param.start_fitted))
  cumulative<-solution$'6'
  daily_fitted<-rep(0,98) #make a vector to fill with daily cases
  for (j in 0:97)
```

```
  {
    if (j==97)
    {
      daily_fitted[1]<-cumulative[1]
    }
    else
    {
      daily_fitted[98-j]<-cumulative[98-j]-cumulative[97-j] #find difference
    }
  }
  dailyFits[i,]<-daily_fitted
}
#Find Katz interval
meanAndVariances<-searchInitials[sensibleRows[3],3:6]
logmean2<-log(meanAndVariances[1]/meanAndVariances[2])
logvariance2<-meanAndVariances[3]/(meanAndVariances[1]^2)+
meanAndVariances[4]/(meanAndVariances[2]^2)
lowerKatz2<-exp(logmean2-qnorm(0.975)*sqrt(logvariance2))
upperKatz2<-exp(logmean2+qnorm(0.975)*sqrt(logvariance2))
intervalKatz2<-c(lowerKatz2,upperKatz2) #Katz interval much wider
```

The full R script used is available upon request.

References

[1] R.C.Geary. *The Frequency Distribution of the Quotient of Two Normal Variates*. Journal of the Royal Statistical Society. 93 (3): 442–446.