COVID-19 Data Analysis Combined SIR Model with Actuarial Applications

Haomu Liu

B00780833

Supervised by Dr. Toby Kenney

Undergraduate Honor Research

Bachelor of Science Combined Honours Co-op in Statistics and Actuarial Science

Dalhousie University

Halifax, Nova Scotia

Date: 19 April 2021

Table of Contents

Abstract

Since the beginning of 2020, a global epidemic caused by Covid-19 has put the whole world in a panic. As the virus spreads by droplet and contract, many countries have been on lockdown and businesses have been closed temporarily. With such viral pneumonia's outbreak, it is no exaggeration to say that almost every industry has been affected, among which the insurance field has been challenged a lot with many companies adapting their products and financial arrangements to the pandemic.

This paper tries to establish a connection between this outbreak and an epidemiological model, SIR, and provides suggestions to the insurance industry through predicting the trend of the epidemic in Canada. In addition, the first part of this paper is aimed at building the SIR model in R and importing the Covid-19 data of Canada to analyze the incidence and infection of Covid-19 in Ontario; the second part mainly focuses on the impacts on the insurance industry, giving a subsequent analysis of insurance cost management, risk control, and the purchase of insurance products, and puts forward the adjustments that insurance companies should make.

Introduction of SIR Model

SIR Model Concept

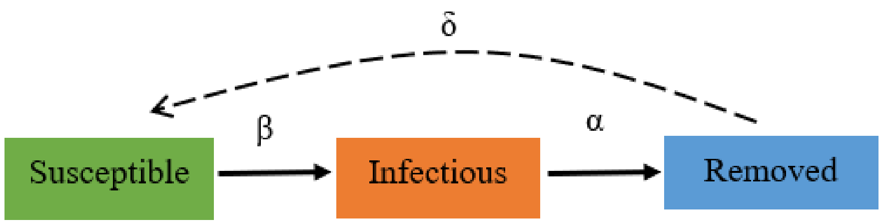
The SIR (Susceptible, Infectious, or Recovered) model not only is one of the most classic and basic of all epidemiological models, but also has become a mainstream model of infectious diseases analysis. It is on the basis of setting the rates of transition between each two groups and computing the proportion of the population of each group.

Origin of SIR Model

This model was first mentioned in Daniel Bernoulli's work on smallpox prevention in 1760 and was specifically introduced in the early 20th century based on the study of establishing mathematical models of infectious diseases developed by some scientists including Ronald Ross and William Hamer. It was not put forward until 1927 when A. G. McKendrick and W. O. Kermack studied the Black Death in London (Magal & Ruan, 2014).

Mathematical Model (Standard SIR Model)

In this model, we assume that the virus occurred in a short period of time, so we do not take into account the birth rate, mortality and mobility of the population, which means that the total number of individuals is constant. Here is the graph representing the standard SIR model:



From above, we know that SIR model is mainly for simulating the infection-remove situation of the population in a system. Group S is converted to I based on β, and I transfer to R based on α. When the immunity is likely to disappear over time, the R population may change back to the S based on δ. First of all, we need to define the variables:

S (Susceptible): all the people who could be infected.

I (Infectious): the people who are infected and could transmit the virus to group S.

R (Removed): the individuals who had been infected and have recovered or died.

β (effective contact rate): infection coefficient.

α (removal rate): remove coefficient.

δ (loss of immunity): immunity extinction coefficient.

N: total population in the model.

t: unit time in the model.

Then, we can write down the SIR differential equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \qquad \frac{dI}{dt} = \frac{\beta SI}{N} - \alpha I \qquad \frac{dR}{dt} = \alpha I$$

These equations represent the three rates of the population's conversion speed to other stages in all three states, from which we know that if the situation of loss of immunity is not considered, then the number of populations in group S will keep declining. Moreover, if $\frac{dI}{dt} > 0$, then the population in group I will show positive growth, so the pandemic would break out.

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \alpha I > 0 \rightarrow \frac{\beta SI}{N} > \alpha I \rightarrow \frac{\beta S}{\alpha N} > 1.$$

We now assume that at the beginning of the outbreak, everyone is at risk of being infected, i.e., S=N. Then the simplified equation is $\frac{\beta}{\alpha} > 1$. Let $R_0 = \frac{\beta}{\alpha}$, and is called the basic reproductive number or ratio. Therefore, when $R_0 > 1$, the infected population will keep increasing, which leads the virus will spread explosively; when $R_0 < 1$, the disease is coming under control and will die out eventually. Additionally, if we take the derivative of the differential equation of I, then we get $I(t) = e^{\left(\frac{\beta S}{N} - \alpha\right)t}$, so when $R_0 > 1$, the infected population will grow exponentially.

Methods of Controlling Infectious Diseases

1. Decrease the effective contact rate β. For example, wear a mask, and wash hands frequently.

2. Increase the recovery rate in order to rise the removal rate. For instance, use specific medicine.

3. In the above equation, we assume S=N. In real epidemics, people may become immune before they are infected by vaccination.

Application of SIR Model

One of the biggest concerns among the public about a pandemic is its eventual trend or size. In this section, we are going to estimate the ultimate number of infected individuals in Canada via formula derivations.

When $\frac{dI}{dt} = 0$, the value of I reaches its peak. So, $\frac{dI}{dt} = \frac{\beta SI}{N} - \alpha I = 0 \rightarrow S = \frac{\alpha N}{\beta}$

- Firstly, get the maximum of I, the largest number of people infected at a given time.

$\frac{dS}{dI} = \frac{-\frac{\beta SI}{N}}{\frac{\beta SI}{N} - \alpha I} = \frac{1}{\frac{\alpha N}{\beta S} - 1} \rightarrow dI = \left(\frac{\alpha N}{\beta S} - 1\right) ds$, integrate both sides →

$$\int dI = \int \left(\frac{\alpha N}{\beta S} - 1\right) ds \rightarrow I(t) - I(0) = \frac{\alpha N}{\beta}\log\big(S(t)\big) - S(t) - \frac{\alpha N}{\beta}\log\big(S(0)\big) + S(0) \qquad (*)$$

In the "Introduction of SIR Model" part, we assume S = N, so $S(0) = N, I(0) = 0$

$$(*) \rightarrow I_{max} = \frac{\alpha N}{\beta}\log\big(S(t)\big) - S(t) - \frac{\alpha N}{\beta}\log N + N \rightarrow \frac{I_{max}}{N} = \frac{\alpha}{\beta}\log\big(S(t)\big) - \frac{S(t)}{N} - \frac{\alpha}{\beta}\log N + 1 \ (**)$$

Since $S(t) = \frac{\alpha N}{\beta}$ and $R_0 = \frac{\beta}{\alpha}$, we have $S(t) = \frac{\alpha N}{\beta} = \frac{N}{R_0}$

$$\begin{aligned}(**) \rightarrow \frac{I_{max}}{N} &= \frac{1}{R_0}\log\left(\frac{N}{R_0}\right) - \frac{1}{R_0} - \frac{1}{R_0}\log N + 1 = \frac{1}{R_0}\log N - \frac{1}{R_0}\log R_0 - \frac{1}{R_0} - \frac{1}{R_0}\log N + 1 \\ &= 1 - \frac{1}{R_0}\log R_0 - \frac{1}{R_0} = 1 - \frac{\log R_0}{R_0} - \frac{1}{R_0}\end{aligned}$$

Therefore, the maximum proportion of people infected during the outbreak is related to R_0, and the formula is $\frac{I_{max}}{N} = 1 - \frac{\log R_0}{R_0} - \frac{1}{R_0}$.

- Secondly, verify the formula $\frac{I_{max}}{N} = 1 - \frac{\log R_0}{R_0} - \frac{1}{R_0}$.

As the basic reproduction number changes each day and is always reported based on provinces, so we verify the formula using related data in Ontario.

To simplify our calculation, we may assume that the given time is the entire outbreak, so I_{max} has the meaning of the total number of people infected during the epidemic. According to the report of mathematical modeling of Covid-19 in Ontario published on 24 March (Tuite, Fisman, & Greer, 2020), the estimated proportion of total infected people during the entire epidemic in Ontario is 42-63%, so we need to find the basic reproduction number R_0 in the most recent week including that day to compare the results. Based on the report, *All Ontario: Case numbers and spread* (Government of Ontario, 2021), we notice that the R_0 near 24 Mar 2020 is approximately 2.45, so the proportion of infected individuals would be $\frac{I_{max}}{N} = 1 - \frac{\log 2.45}{2.45} - \frac{1}{2.45} \approx 43.3\%$

which is in the range 42-63%. Therefore, the formula for the proportion of infected people over the epidemic is verified.

Another concern about a pandemic would be vaccination and herd immunity. We now focus on estimating the proportion of people who need to be vaccinated to control epidemics.

Vaccination is an artificial way to make the individuals who are in group S and get vaccinated produce antibodies enter group R directly. We set p as the percentage of people who are vaccinated. In order to control a pandemic, R_0 needs to be less than 1. So, $\mathrm{R}_0(1-p)<1 \rightarrow 1-p<\frac{1}{\mathrm{R}_0} \rightarrow p>1-\frac{1}{\mathrm{R}_0}$. If we use the data of Ontario near 24 Mar 2020, then $p>1-\frac{1}{\mathrm{R}_0}=1-\frac{1}{2.45} \approx 59.18\%$, which means that at least about 59.18% people in Ontario need to be vaccinated for controlling Covid-19. Fortunately, the trend of daily reported reproduction number has declined and tended to be stable around 0.9 less than 1. Therefore, based on the newest official data, we conclude that from today on, even if nobody is vaccinated, the epidemic will eventually end in Ontario.

Additionally, herd immunity works in a similar way as vaccination. People gradually acquire herd immunity along with the spread of the virus. As an increasing number of people get antibodies after being infected, the epidemic is gradually under control when the infected proportion tends to p that we calculate above. In order words, after about 59.18% population infected in Ontario, Covid-19 will die out in that province.

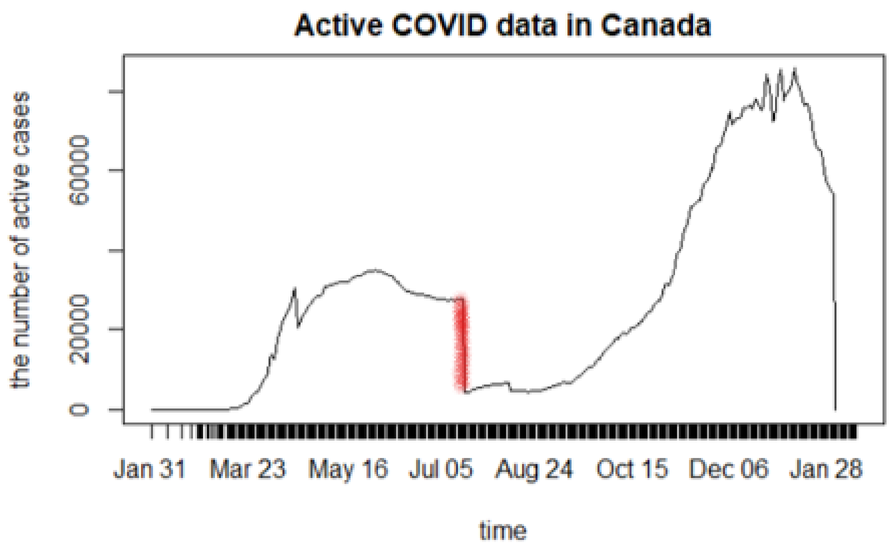
Real Data Analysis

Introduction

For the analysis following, the data that we will use is based on the real time csv file in the section "Current situation" (Government of Canada, 2021). The data file is converted to dynamic graphs and map in the official website for end users tracking the variation trend of covid-19 cases in each province in Canada.

The file includes 5142 rows and 39 columns, recording the situation of covid-19 cases from 31 Jan 2020 to 12 Feb 2021 in chronological order. Specifically, it mainly includes cumulative numbers of infections, numbers of active cases, and numbers of recoveries for each province on a daily basis. At the beginning the epidemic outbreak, the sampling intervals are relatively large since some days are missing, but with the pandemic going more seriously, there are no missing days after 11 March 2020.

Accuracy Analysis



prname	date	numconf	numprob	numdeaths	numtotal	numtested	numrecover	numtoday	numdeathstoday	numtestedtoday	numrecoveredtoday	numactive
Canada	7/12/2020	107579	11	8798	107590	3212879	71467	244	10	29363	201	27325
Canada	7/13/2020	108144	11	8805	108155	3257608	71841	565	7	44729	374	27509
Canada	7/14/2020	108475	11	8813	108486	3302483	72170	331	8	44875	329	27503
Canada	7/15/2020	108816	11	8830	108827	3341893	72485	341	17	39410	315	27512
Canada	7/16/2020	109253	11	8847	109264	3387755	72836	437	17	45862	351	27581
Canada	7/17/2020	109658	11	8859	109669	3438041	96689	405	12	50286	23853	4121
Canada	7/18/2020	109988	11	8868	109999	3481804	96914	330	9	43763	225	4217
Canada	7/19/2020	110327	11	8872	110338	3520542	97051	339	4	38738	137	4415
Canada	7/20/2020	111113	11	8878	111124	3573630	97474	786	6	53088	423	4772
Canada	7/21/2020	111684	13	8882	111697	3616728	97757	573	4	43098	283	5058

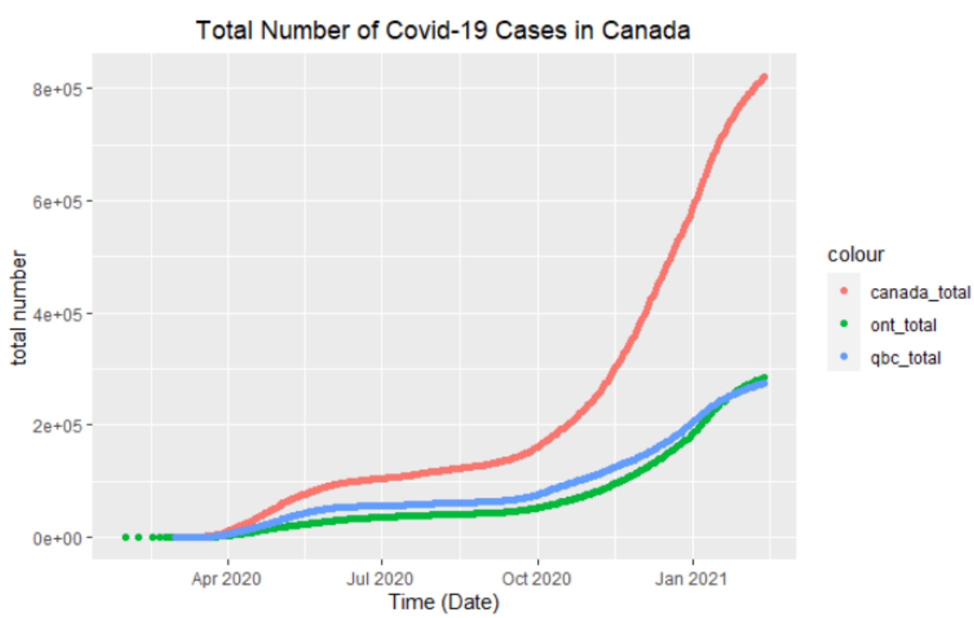
Chart 1

Although errors are easy to occur when recording data, there is an obvious artificial mistake of reported active cases after 5 July 2020. Through reading the downloaded .csv data file, we find that the exact date of this drop is 17 July 2020 where the number of active cases reduces by 27581-4121=23460 from the previous day, and the duration of the drop approximately lasts until early September in which the number starts to go up sharply (Chart 1). By glancing the data in other columns related to data recording around that date, we notice that on 17 July 2020, the statistics in the column 'numrecoveredtoday', 23853, is much larger than the rest statistics of the same column, but the variations of other statistics are slight. Therefore, it is highly possible that an artifact of recording recovered cases occurs on that day, which leads us to understand that the reported data is not absolutely accurate, and it should be considered when designing an insurance product in later sections.

SIR Model Prediction and Paraments Analysis

After knowing the basic knowledge of SIR model and the data we use, it is the time to build an SIR model in R and predict the overall trend of Covid-19 in Canada. Moreover, it should be emphasized that the model prediction is not reliable and can only show some trends in the data. The real situation still needs to be judged by oneself according to the facts.

First of all, we have a look at the overall trend of total number of Covid-19 cases in Canada and the first two provinces reporting the confirmed cases, Ontario and Quebec.



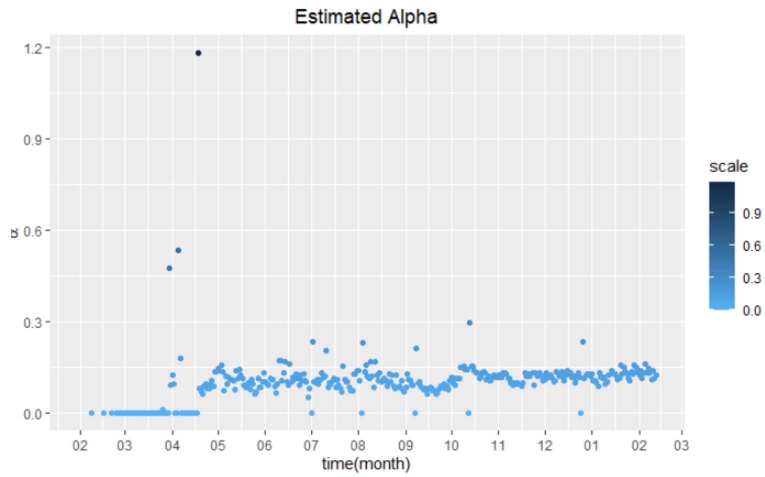
From this plot, we notice that the overall diagnosis in Ontario is lower than that in Quebec where the number of diagnoses rose exponentially in only 15 days after the record began, which is because the Ontario's government early attached great importance to the epidemic and enacted a ban about social distance.

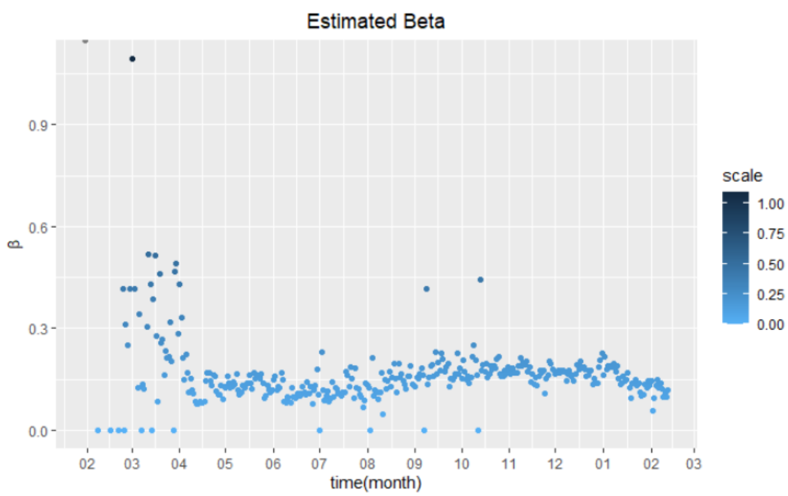
SIR Model Prediction in Ontario

As Ontario is the largest province and one of the first two provinces reporting the confirmed cases of Covid-19 in Canada, we are going to predict the epidemic trend of Ontario.

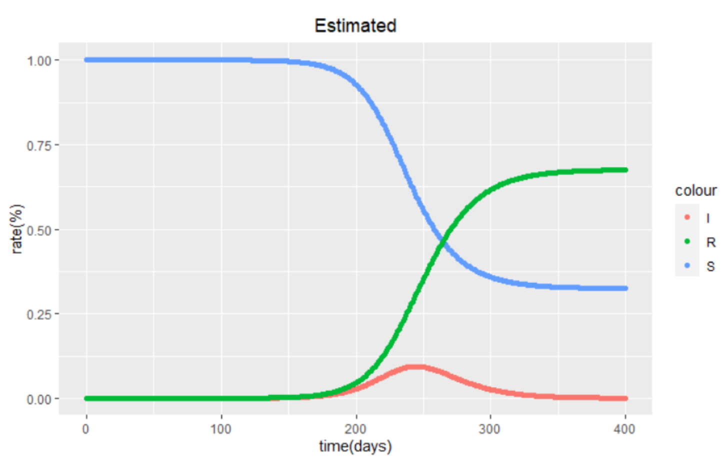
Ontario Covid-19 data were available from 31 January 2020 to 12 February 2021, for a total of 355 days. There are no statistics for the susceptible population in Ontario in the current data. Based on the general experience, this data is often difficult to obtain. What we assumed before in the part of Introduction of SIR is that all uninfected people are in the susceptible group, i.e., S=N, but in order to capture variation trends, we turn to look at their rates and make the following assumptions about the initial values: $\mathrm{S}(0) = 1 - \mathrm{I}(0)$ where $\mathrm{I}(0)$ is the ratio of the number of infected cases in the first day to the total population, and $\mathrm{R}(0) = 0$.

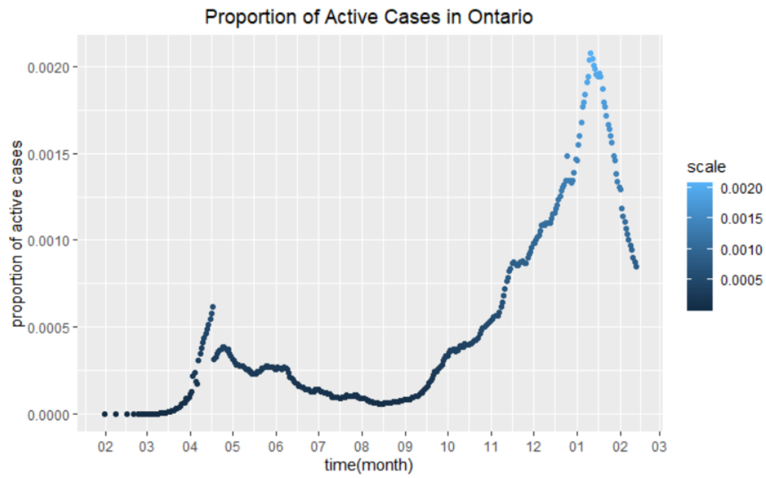
In the following, we fit the incidence of Covid-19 in Ontario using a proportional-based SIR. The values of alpha and beta need to be determined, but here is no more uniform standard for calculating alpha and beta since these two parameters are uncertain. However, we can estimate alpha and beta by referencing the parameters from previous epidemic models, such as SARS. We have known that beta is the infection coefficient and alpha is the remove coefficient, so the mathematical equations should be $\beta = \frac{\text{increased cases today}}{\text{active cases yesterday} * \mathrm{R}_0}$ and $\alpha = \frac{\text{recovered today}}{\text{active cases yesterday}}$. Referencing the existing pattern of the basic reproduction number R_0, we guess that the value of R_0 will be near 0.8 in a later stage (Government of Ontario, 2021). Therefore, in this research, we employ the following formula $\beta = \frac{\text{number today}}{(\text{number of active} - \text{number today}) * 0.8}$ and $\alpha = \frac{\text{number recovered today}}{\text{number of active} - \text{number today}}$ for alpha and beta prediction, respectively, and get the two plots following in which the estimated alpha plot indicates the majority points located in around 0.09 while the point accumulation area in the estimated beta plot occurs when beta equals to about 0.15. So, we set alpha and beta to be 0.09 and 0.15, respectively.





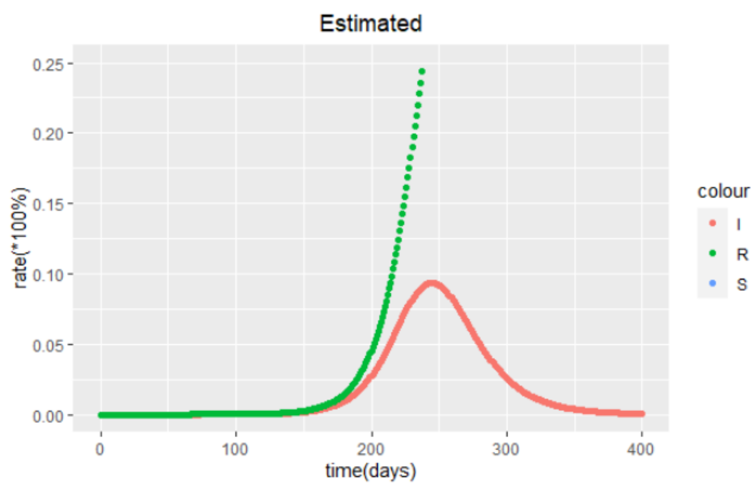
After predicting the value of these two parameters, we plug other related numbers into the SIR model built before and get an estimated plot, then using raw data to get a real plot, and make a comparison of these two.





The population in Ontario is around 14.73 million (Statistics Canada, 2021). As the Ontario's ban on social distance came early, the coefficient of the contact among the population in S (Susceptible) group is relatively small.

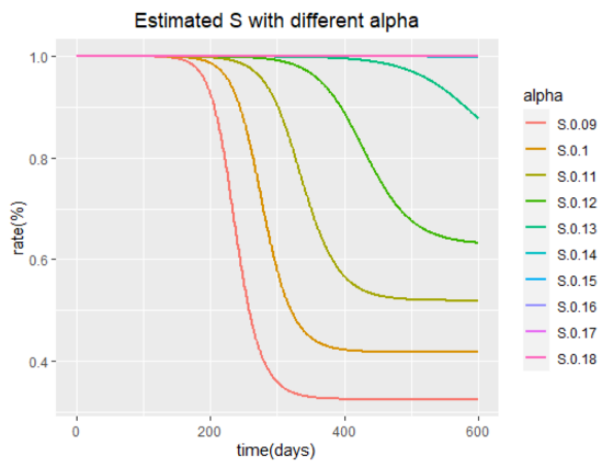
We now magnitude the estimated plot by setting the values of time axis from 100 to 300 and for rate axis, from 0 to 0.25.

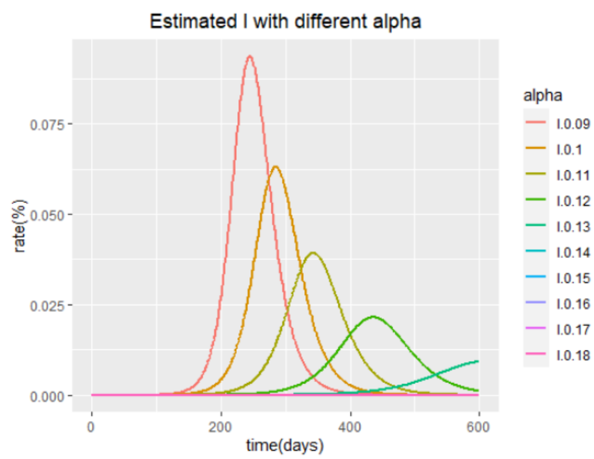


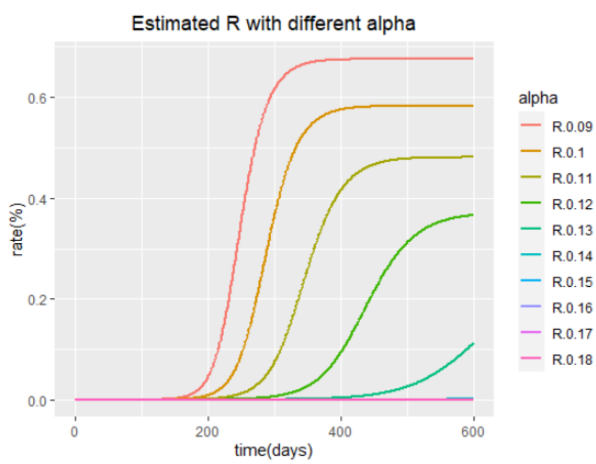
According to the actual data, until Aug 15, around the 200th day from the first recorded day, about 40565 people have been infected, which is around 0.2754% of the total population in Ontario. However, in the model, the number of infected people is about 2.5%, indicating that the predicted rate is more than 10 times higher than the real rate.

There are five possibilities of such things happening: 1) there is an error in the parameter; 2) the predictions are based on constant alpha and beta, but the actual values fluctuate; 3) the test sample is small and cannot reflect the overall diagnosis; 4) the development of the epidemic has been slowed down and suppressed by various measurements; 5) there is underreporting of cases. Among them, we focus on the first possibility in this research, because in SIR model the rate of growth is very sensitive to the values of alpha and beta. So, our next step is to analyze the trend of the three groups, S, I, and R, by varying the parameters separately. This idea can be achieved by fixing one parameter at a time and change the value of the other, which helps us observe the change in the fitting results of the SIR model.

Firstly, we assume that beta remains constant and alpha doubles the initial value in steps of 0.01 and get the trend plots of the number of susceptible, the number of infected, and the number of recovered people in the SIR model.

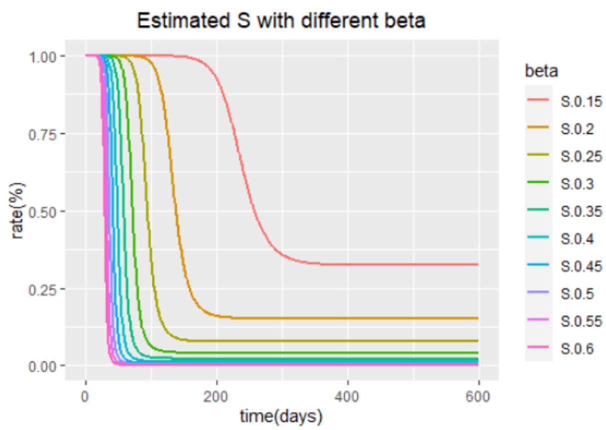


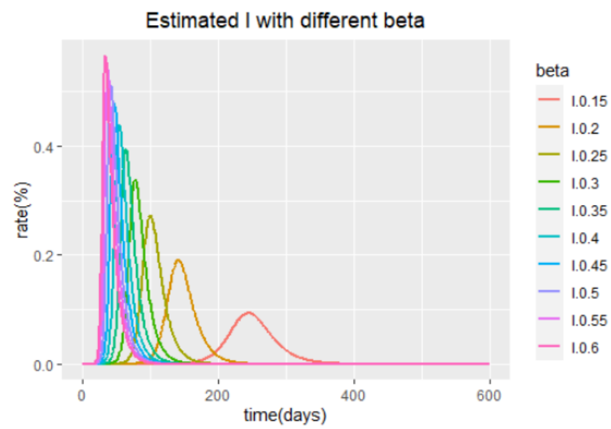


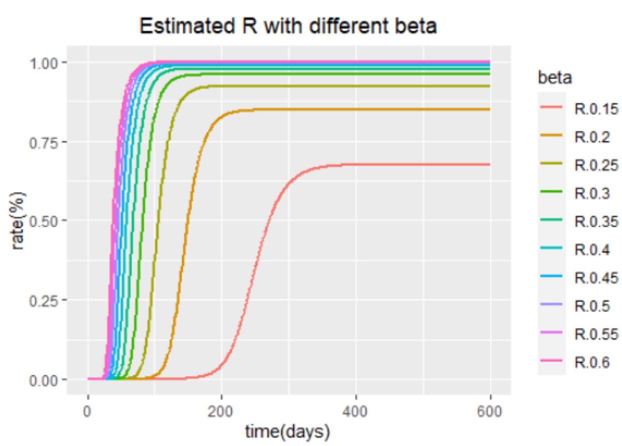


From the plots, we notice that with the value of alpha increasing, the proportion of susceptible people fitted by the SIR model decreases more and more slowly before stabilizing at a relatively high proportion interval. In other words, when the mortality rate of an epidemic is relatively high, the speed and size of the infected population decrease. Conversely, when an epidemic's mortality rate is low, its intensity and infection rate is relatively more significant. As for the proportion of infected people, when alpha is increasing, the speed and size of the number of infections increasing both decrease, which is inversely consistent with the trend in the susceptible population. In contrast, when comparing the removed (recovered) population, an increase in alpha makes recovery time increase accordingly, decreasing the size of recovery populations at the same time, which can be explained that epidemics with higher mortality rates are more difficult to be treated.

Secondly, we assume that alpha remains constant and gradually increases beta, in steps of 0.05, by three times. The SIR model was fitted separately for each combination of parameters. The trends in the number of susceptible, infected, and recovered people in the SIR model were obtained as shown below.

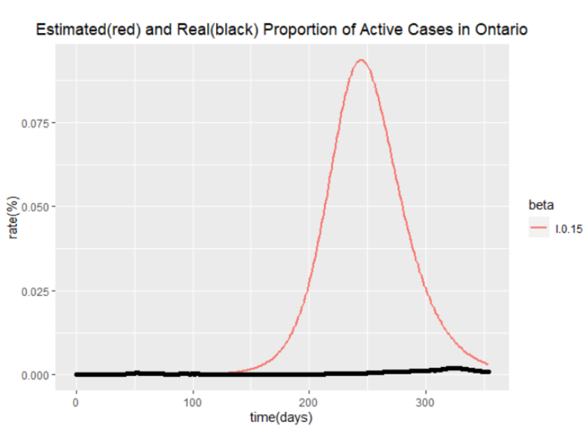


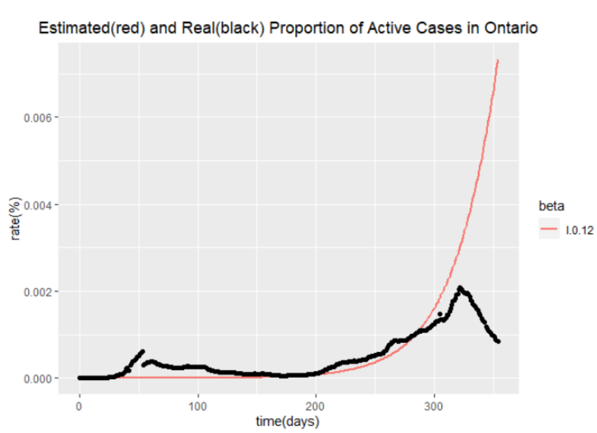




The distribution of the susceptible population shows that as beta increases, the declining speed of the susceptible population increases, and the declining size increases significantly. This result is easier to be explained that as the probability of infection rises, the susceptible population rapidly becomes infected. With beta growing, the infected group increases more rapidly and reaches a peak earlier. For the removed or recovery population, the recovery time and speed become shorter and faster with the increasing beta, and the proportion of recovery increases. We also notice that, unlike alpha, as beta increases to a relatively large value, it will have less impact on the SIR model fitting results.

After analyzing how alpha and beta values affect the SIR model, we reference the two patterns to find the most proper combination of these two parameters for obtaining a better fit model. Firstly, we compare the SIR model fitted with initial values of alpha = 0.09 and beta = 0.15 with the real distribution of infections in Ontario. As the left-handed side graph shown below, the current SIR fit is not ideal, and the actual distribution of infections is much flatter than predicted. Combined with the analysis above, the current alpha is small, and beta is large. However, considering that the fitting curves vary smaller when changing beta than when changing alpha, our optimization strategy of choosing parameters is to keep alpha unchanged and reduce beta. Through several experiments, we find that the fitted results of the SIR model are closest to the results of actual data when beta is taken as 0.12. In this new model, we notice that the fitted result is good in general before the first 350 days, but afterwards, the predicted number of infections is much higher than the actual number. More specifically, the growth rate is predicted to increase continuously while the actual data shows a significant decline in later stages with a turning point, which indicates that the spread of the epidemic is effectively controlled probably due to many valid measurements taken by the government.





Application Analysis for the Insurance Industry

Epidemics pose a significant challenge to the effectiveness of insurance industry. Good insurance products and benefits strategies can help people survive a pandemic, while a lack of consideration for the occurrence of epidemics can expose insurance products to high benefits risks, resulting in huge losses for both insurers and policyholders.

We analyze the total cost of insurance and premiums under the conditions of an outbreak of Covid-19 in Ontario, through an insurance theoretical analysis based on the construction of the epidemic model. The cost of insurance, or the amount of benefits, is affected by the epidemic in various ways. It is assumed that each susceptible person is insured and would pay premiums until they become infected and begin to receive treatment in a hospital. During the period of treatment, insurance benefits occur. In this study, we will only consider the annuity benefit and the lump-sum benefit and assume that the premium is paid annually.

Furthermore, we calculate the EPV (Expected Present Value) of the annuity premium, annuity benefit and lump-sum benefit using the following formulae (Feng & Garrido, 2012):

t-year annuity premium: $\bar{a}^{s}_{t\rceil} \triangleq \int_0^t e^{-\delta x} s(x)dx$;

t-year annuity benefit: $\bar{a}^{i}_{t\rceil} \triangleq \int_0^t e^{-\delta x} i(x)dx$;

lump-sum benefit: $\bar{A}^{i}_{\infty\rceil} \triangleq \beta \int_0^\infty \mathrm{e}^{-\delta \mathrm{x}} \mathrm{s(t)i(t)dt}$;

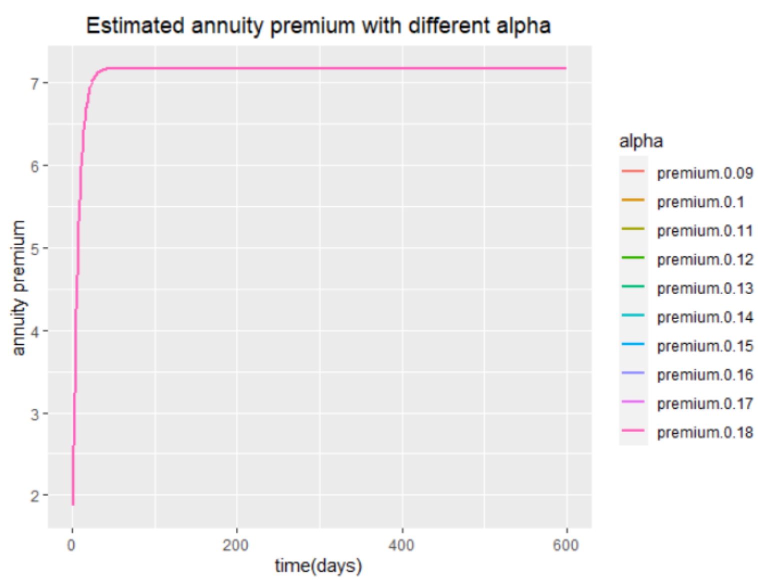
where $s(x)$ and $i(x)$ are the susceptible population ratio and infected population ratio, respectively, and $\delta = \ln{(1+i)}$ is the discounting force of interest. In the following calculation, we choose $i = 0.16\%$ as the Canadian interest rate (Bank of Canada, 2021).

When we fit the SIR model to an epidemic, we are able to obtain the corresponding $s(x)$ and $i(x)$, which allows us to calculate the theoretical estimates of premiums and costs. From the *SIR Model Prediction and Paraments Analysis* section, we have already known that the two main parameters affecting the fitting are alpha and beta. Since premiums and benefits are determined by the number of the infected and susceptible populations obtained from the fitted model, we should also focus on the influence of alpha and beta on the results when analyzing the variation features of insurance costs and premiums. Ideally, we should treat alpha and beta as two random variables together, but after considering that this method is more difficult to programming, we turn to fix one of these two parameters and find the relationship between the annuity premium or benefit and the other variable.

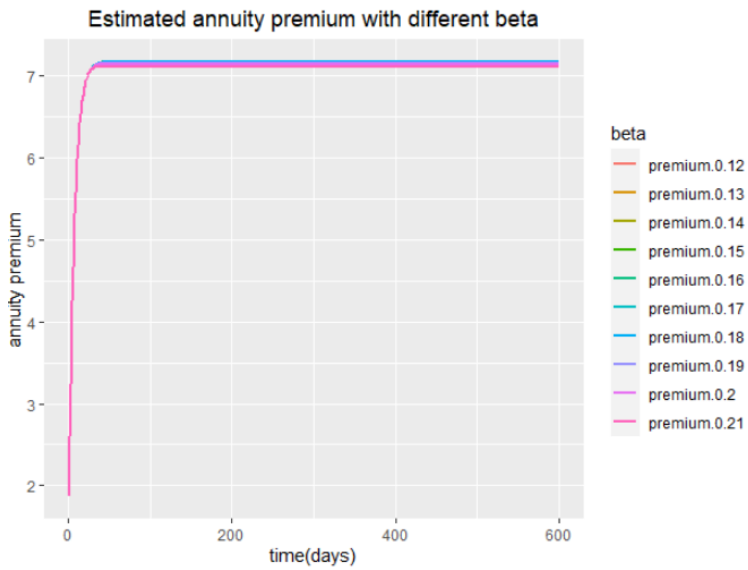
Annuity Premium

As the previous section where we studied the effect of alpha and beta on the SIR model of Covid-19, this section also considers the impact of the two parameters on the annuity premium which should be paid in the susceptible state using the same idea of controlling one parameter.

First of all, we select the initial parameters and step sizes associated with alpha and hold beta constant to obtain the effect of alpha on annuity premiums. The result is shown below where we find that for different values of alpha, the curves of annuity premiums overlap together. In other words, the effect of alpha on annuity premium is relatively low. We explain this phenomenon by combining $s(x)$ and the formula of annuity premium: $s(x)$ is unaffected by alpha in early stages, but afterwards, it starts to decline at different rates as alpha varies; and under discrete distribution conditions, the annuity premium is cumulative and increases by $e^{-\frac{\delta}{365}}$ compared to the previous day where $\delta = \ln{(1 + 0.0016)}$ is a yearly rate as we mentioned before. Hence, during the mid and late stages, the change of alpha has minimal effect on the annuity premium.

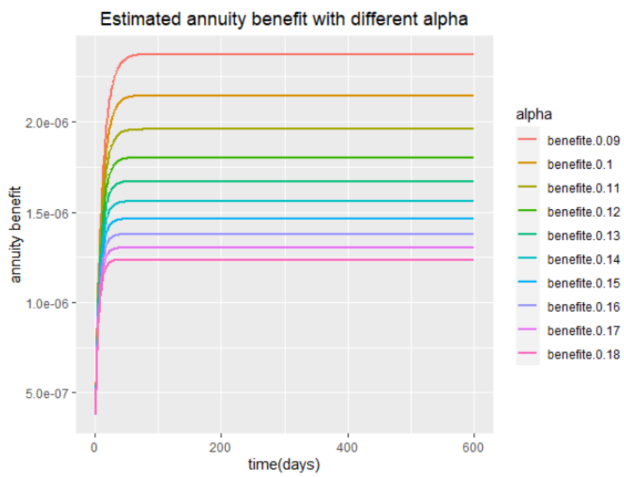


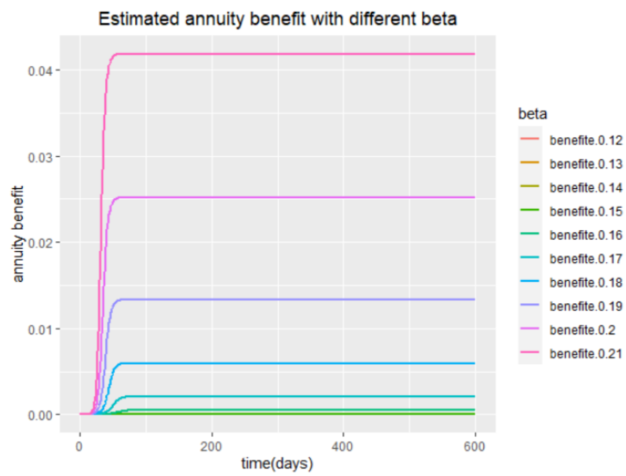
Secondly, we keep alpha constant and increase beta in the same way and get the plot below. In general, these two plots of estimated annuity premium and the time from the start of the epidemic are roughly present a Pareto distribution. What is more, we notice that the trends of the annuity premium are almost identical but do not precisely overlap. Hence, the effect of the variation of beta on $s(x)$ is more significant compared to alpha.



Annuity Benefit

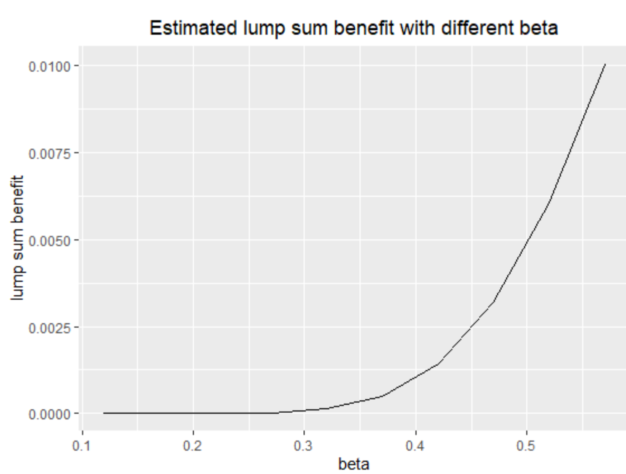
The distribution of annuity benefits that is payable in the infected state is also influenced by alpha and beta. The trend of annuity benefits is shown in the left-hand graph below where we fix beta and gradually increase alpha, and the right-hand graph below whose alpha is fixed and beta rises gradually. The results indicate that the annuity benefit decreases as alpha increases, while an increase in beta leads to an increase in the annuity benefit. This is because the larger alpha is, the slower the peak of the number of infected populations will reach, so for a short period, a smaller value of alpha results in a more rapid increase in the number of infected people, which makes the annuity benefit increase as well. Moreover, the increase in beta also results in a rapid increase in the infected population, so that the annuity benefit rises as beta growth.





<u>Lump-sum Benefit</u>

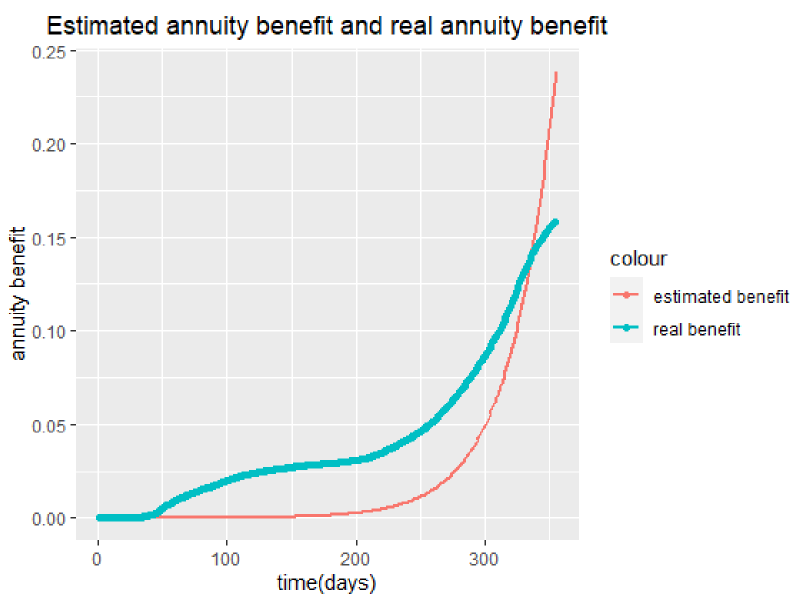
The formula of lump-sum benefit, $\bar{A}^{i}_{\infty\rceil} \triangleq \beta \int_0^{\infty} e^{-\delta x} s(t) i(t) dt$, tells us that the value is mainly determined by beta, so we fix alpha as 0.09 and gradually increase the value of beta. From the trend plot of lump-sum benefit below, we observe that with beta growth, the lump-sum benefit shows an increasing trend. Moreover, the shape of the curve indicates that the lump-sum benefits and the value of beta are probably inverse exponential distributed. When beta is greater than 0.45, the benefit tends to increase slowly; when beta becomes large enough and reach a high level, the susceptible population declines to a stable status, while the infected population peaks and the lump sum benefit would surge. If an insurance company adopts this claim method, it needs to make detailed analysis of the transmission characteristics of the epidemic, assigning reasonable rates to avoid risks and ensuring that the cost and premium are roughly in balance.



Difference between Estimated Benefit Amount and Real Benefit Amount

The number of people infected in Ontario is currently known, but that of the susceptible population is not available, so we are able to calculate the real benefit amount using the real active cases in Ontario. After that, we should compare it to the estimated benefit amount from the number of infections predicted by the SIR model, for analyzing the difference in benefit amounts between the real epidemic outbreak and the ideal epidemic model.

As for the parameters of SIR model, we choose the optimal combination obtained in *SIR Model Prediction and Paraments Analysis* section where alpha equals to 0.09 and beta equals to 0.12. From the plot below, we know that the real annuity benefit amount is higher than the estimated amount in the first 330 days. Compared the prediction that we did in the previous section, the real proportion of infections is higher than the estimated before entering the late stage where the estimated proportion is much higher, and since the impact of the number of infections at a late stage on the annuity benefit amount is relatively small, we can conclude that in general, the estimated benefit amount is smaller than the actual benefit amount in the first year of the outbreak. Therefore, insurance companies should be cautious when estimating benefit amounts to prevent bankruptcy.



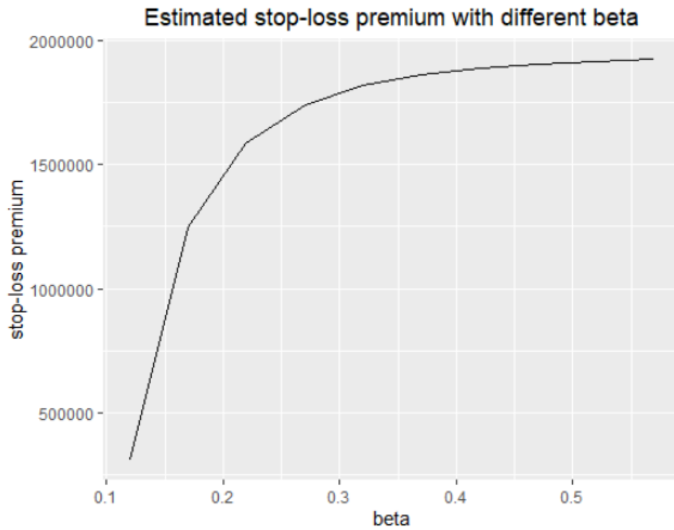
Stop-Loss Insurance

A well-structured insurance policy can prevent catastrophic losses, and the stop-loss insurance is a typical representative. The purpose of stop-loss insurances is mainly to provide protection against unpredictable huge losses, and therefore, setting a proper deductible is important as it provides a buffer between given minimum losses and truly catastrophic losses. It also helps insurance companies ensure the financial stability and reduce the risk of dishonest conducts by insureds (Moorcraft, 2020).

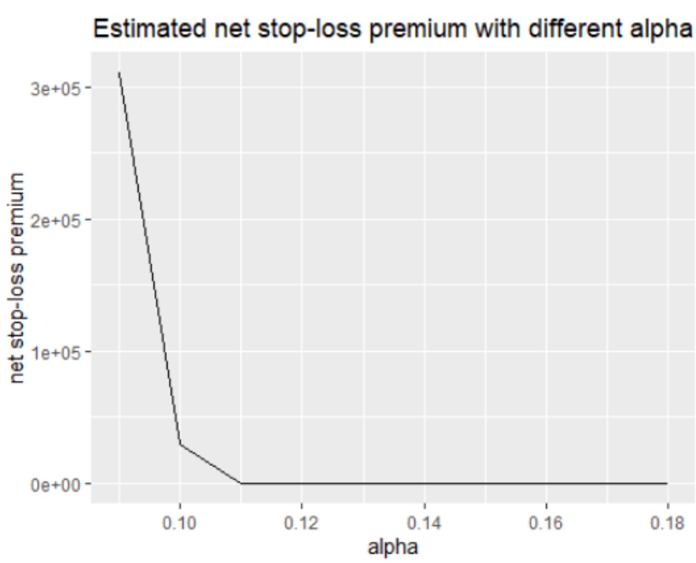
In the reality, the values of alpha and beta are fluctuating throughout the epidemic. Since they are same for every policyholder, there exist non-diversifiable risks which cannot be reduced by selling enough policies. Thus, if the premium is priced based on a value of beta that is too small, then every policy will lose money, even after including a contingency loading, and selling more policies will just lead to more losses. To avoid significant risks brought by underestimating premiums and losses, insurance companies need to set up diversified insurance policies, such as the stop-loss reinsurance.

In this section, we employ the discrete function $E[(S-d)_+] = \sum_{x>d}(x-d) * f_S(x)$ to calculate the expected cost of stop-loss insurance, i.e., net stop-loss premium, where $f_S(x)$ is the distribution of the aggregate cost for all policies and d is the deductible. Specifically, we use the optimal combination of alpha (0.09) and beta (0.12) obtained before, setting one of them fixed and changing the other, for exploring the expression of $f_S(x)$ for each parameter. Moreover, since the raw data is recorded in days, we set the deductible to $5 for each day but to calculate the total aggregate losses.

Firstly, we let alpha stay the same and select the initial parameters and step sizes associated with beta, getting the plot below. It shows that with beta rising, the premium grows more and more slowly, but generally beta and net stop-loss premium have a positive correlation. In other words, the net stop-loss premium rises as the infection rate increases.



Secondly, we select the initial parameters and step sizes associated with alpha and hold beta constant to obtain the trend of estimated net stop-loss premium with alpha changing. As the result shown below, we notice that this distribution is heavy tailed. For the value of alpha smaller or equal to 0.11, there is a negative relationship between alpha and net stop-loss premium in general. What is more, a turning point occurs when alpha equals to 0.10, before which the declining speed of the net stop-loss premium is much more dramatic than after. Hence, when the beta is fixed at 0.12, the net stop-loss premium decreases as the removal rate gradually increases until 0.11. Moreover, since the 95% quantile of the aggregate loss premium, around 0.6698, is much higher than the cost of losses not covered by the reinsurance, around 0.0219, we conclude that insurance companies need to set an unreasonably high premium to avoid bankruptcy.



Impact of the epidemic on the Insurance Industry

Individuals' risk awareness and insurance awareness have been generally improved and greatly enhanced during the outbreak. For example, the number of people taking the initiative to consult and purchase life and health insurance has increased significantly. Moreover, it can be predicted that in the post-epidemic era, various kinds of health insurance products will be attractive to the public, which reflects that purchasing health insurances has become one of the most effective means to resist risks and protect human lives.

Apart from that, a lot of policyholders will prefer profit-sharing insurance, which serves as a mean of protection and wealth inheritance, as well as a tool of risk aversion and steady appreciation. In the post-epidemic era, low interest rates may remain in place for a long time to stimulate economic recovery, so the whole life insurance whose price is sensitive to interests will be less attractive. Although a long-term low interest rate in profit-sharing insurance may lead the interest rate spread shown, insurance companies can adjust their investment portfolio to offset this spread. In addition, the benefits of profit-sharing insurance are relatively stable and reliable, and more importantly, the annual rate of dividend is never lower than zero, and as long as policyholders receive a payment, its cumulative total value never goes down. What is more, many insurance companies maintain its dividend rate or slightly adjust the rate around 0.25% during Covid-19. For instance, Equitable Life of Canada have announced that they will remain their dividend scale interest rate at 6.2% unchanged until 20 June 2021 at least, and Sun Life maintained their dividend ratio of 6.25% from 2017 to 1 April 2021 on which they low the ratio to 6.00%.

Future Work

During this research, we fixed the values of alpha and beta and the predictions are based on constant alpha or beta or both, but the actual values can change at any time throughout the real epidemic. Thus, it is well worth considering how fluctuations of these two values affect our predictions, such as building dynamic simulation models, and to make reasonable assumptions about the distributions of the related parameters. After that, more work can be conducted about deciding whether or when companies should sell insurance policies.

In addition, another interesting direction for insurance companies dealing with non-diversifiable risks is to choose a participating life insurance, where companies charge a premium with a large loading and then give the benefit to policyholders via cash dividends, increased benefits, etc. It is probably appropriate for health insurances during a pandemic, but insurers need to consider a problem that a high-profit margin would deter customers from buying such products.

Reference

Bank of Canada. (2021). Interest Rates. Retrieved from https://www.bankofcanada.ca/rates/interest-rates/

Equitable Life of Canada. (2021). No change to dividend scale for 2020/2021. Retrieved from https://www.equitable.ca/en/already-a-client/dividend-information/

Feng, Runhuan., & Garrido, Jose. (2012, December 27). Actuarial Applications of Epidemiological Models. Retrieved from https://www.researchgate.net/publication/44165089_Actuarial_Applications_of_Epidemiological_Models

Government of Canada. (2021, February 14). Coronavirus disease 2019 (COVID-19): Epidemiology update. Retrieved from https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

Government of Ontario. (2021 February 15). All Ontario: Case numbers and spread. Retrieved from https://covid-19.ontario.ca/data

Magal, Pierre., & Ruan, Shigui. (2014). Susceptible-infectious-recovered models revisited: From the individual level to the population level. Retrieved from https://www.math.miami.edu/~ruan/MyPapers/MagalRuan-MathBiosci2014.pdf

Moorcraft, Bethan. (2020, Mar 20). What is stop-loss insurance? Retrieved from https://www.insurancebusinessmag.com/us/news/healthcare/what-is-stoploss-insurance-217479.aspx

Statistics Canada. (2021, February 20). Population estimates, quarterly. Retrieved from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901

Sun Life. (2021, April 1). The Sun Life dividend scale is changing on April 1, 2021. Retrieved from https://www.sunnet.sunlife.com/files/advisor/english/PDF/2021_2022_Sun_Life_Dividend_Scale_Announcement.pdf

Tuite, Ashleigh R., Fisman, David N., & Greer, Amy L. (2020, March 24). Mathematical modeling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. Retrieved from https://www.medrxiv.org/content/10.1101/2020.03.24.20042705v1