Stockholms universitet

Stochastic modelling on epidemics and how they affect the workplace

Max Sjödin

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:17
Matematisk statistik
Juni 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2019:17**
http://www.math.su.se

Stochastic modelling on epidemics and how they affect the workplace

Max Sjödin*

June 2019

Abstract

In this thesis, we look at how diseases affect the workplace. We are mainly interested in the productivity of the workplace during the epidemic. We create a stochastic model which has a closed population and a two-level contact structure. Through simulation, we are able to analyze how the different parameters affect the productivity of the workplace. We derive the so-called basic reproduction number R_0 and discuss how it relates to the size and duration of the epidemic.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: max.lukas.sjodin@gmail.com. Supervisor: Pieter Trapman, Abhishek Pal Majumder.

Abstract

In this thesis, we look at how diseases affect the workplace. We are mainly interested in the productivity of the workplace during the epidemic. We create a stochastic model which has a closed population and a two-level contact structure. Through simulation, we are able to analyze how the different parameters affect the productivity of the workplace. We derive the so-called basic reproduction number R_0 and discuss how it relates to the size and duration of the epidemic.

Contents

Chapter 1

Introduction

The office is a place where virus and illness are common. It can result in multiple days of paid sick leave for every infected individual as well as lower productivity if the infected individuals show up to work while feeling a bit under the weather. It is important to look at different ways to approach this problem in terms of how it affects the employer from an economic perspective. To do this, we create an epidemic model and look at how the different parameters of the model affect the productivity of the workplace. The epidemic models have a contact structure on two levels to try and capture the work groups that are common in a workplace. We are looking at the so-called reproduction number R_0 and how this number is reflected in the simulations.

Chapter 2

Model

2.1 Stochastic model

There are two main types of models: stochastic and deterministic. The model discussed in this thesis is stochastic but it is important to understand the advantages and insufficiencies of both types of models. In deterministic models, there is no element of randomness. The main advantage of deterministic models is that since there is no randomness the results are easier to obtain and interpret. Stochastic models, although harder to interpret the results, do have some advantages. The element of randomness in stochastic models gives a more realistic way of modeling the spread of disease.

The stochastic model describes the probability of disease transmission between two individuals instead of stating with certainty when transmissions of the disease will happen. In the stochastic model, it is possible for models with a so-called reproduction number larger than 1 to end before an epidemic would emerge. In the deterministic model, an epidemic would always

occur if the reproduction number is larger than one and never occur if the reproduction number is smaller than one. This difference makes stochastic models more suitable for epidemics in smaller population or when there are few initially infectious individuals. This is because it captures the randomness of the final size of the epidemic. This means that some properties of the epidemic can only be analyzed in stochastic models.[1] A survey by Britton[2] looks more closely at deterministic and stochastic models. The survey also discusses when deterministic models are insufficient.

2.2 Model description

It is important to acknowledge that a mathematical model is always a simplification of reality. A model no matter how complex will not be able to accurately mirror reality. And with more complexity, it makes it harder to analyze how each factor influences the spreading of the disease. For this reason, we only include the parameters and characteristics that we want to analyze the influence of. In this way, we can derive some useful information from the analysis of the model.

In our model, we use discrete time intervals were all individuals contact each other at the same rate. We define the population as closed. This means that there are no outside effects that influence the population. It also means that no new individuals enter the population and that no individuals in the population leave. All individuals in the population are identical. This means that individuals in a certain state behave the same.

The model only considers weekdays which means that weekends are not

considered in any shape or form. The simplification of removing weekdays makes it so that the probabilities that individuals change state does not depend on the current time increment. Every time increment is defined as one whole day.

The day starts with every infectious individual deciding whether they stay home or go to work. Then every susceptible individual gets exposed to the disease in one way or another given that some infectious individuals are at work. Unrelated to the spread of the disease individuals that are at home resting either remain home another day or recover and return to work the next day and are immune.

To model the spread of disease in the workplace we divide the population into workgroups. Susceptible individuals in a specific workgroup can get infected on two different levels. The local level which means that the susceptible individual. The local level where a susceptible individual can contact the infectious individuals in the same workgroup, and the global level were the susceptible individual can contact any of the infectious individuals in the population.

We separate the infectious period into two different states. In the first state the disease is not yet observed but the individual is still able to contact susceptible individuals. In the second state the disease is observed, and the individual decides to stays home from work the next day with a certain probability.

The purpose of this separation is to compare how different employers handle the diseases in the workplace. If an employer is very health conscious and mindful of the spread of diseases the idea is that the amount of time that

the disease is not yet observed would be shorter. The model also distinguishes between workdays were the individual is susceptible or infectious in terms of productivity since it is reasonable to assume that individuals are not as productive when they are sick. If illness would not affect productivity, there would not be any incentive to encourage or let employees stay home from work. Lastly an infectious individual cannot be sick for more than five days. Infectious individuals can however stay home and potentially recover in less than five days. This means that the total number of days spent sick and at home cannot exceed five. It is possible to recover from the disease while still at work.

2.3 Defining the model

The model has five different states and individuals move between them in discrete time. The five different states an individual can be in is: S if an individual is at work and susceptible to infection, I_1 if an individual is at work and infectious but the disease is yet to be detected, I_2 if an individual is at work and infectious, Q if an individual is at home recovering from previously being infected and R if an individual is recovered, back at work and immune to reinfection.

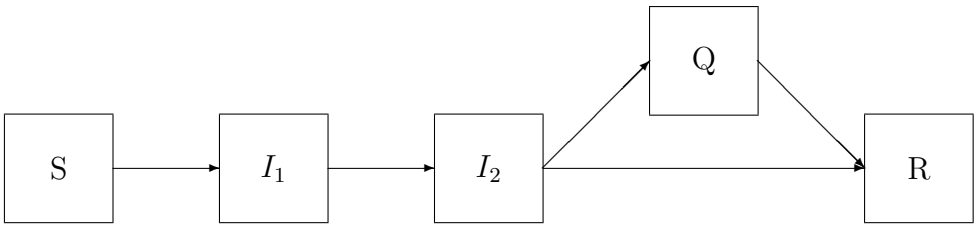

Figure 2.1: Different states of the model

Each day individuals move between the five states based on the outcome of a Bernoulli trial where the probability depends on their current state as well as the state of the population as a whole.

Infectious individuals contact a susceptible individual on the local level with probability $P(\Lambda_L)$ and on the global level with probability $P(\Lambda_G)$. If a susceptible individual gets infected they move to state $I_{1,t}$. Infectious individuals in state $I_{1,t}$ spread the disease but since the disease has not yet been observed the individual remains at work until they after a certain amount of time D move to state $I_{2,t}$. In our model, we have decided to let D be a fixed period of time. First when the infectious individual has moved to state $I_{2,t}$ they can decide to stay home from work and recover with probability γ which removes them from the population and places them in the state Q_t (quarantine). Individuals in state Q_t can then recover with probability δ and return to work the next day. The disease can last a maximum of $\rho = 5$ days until they inevitably recover, whether the individual goes home or not. This means that the individual can recover from the disease without any

days of paid sick leave. Individuals who have recovered and return to work are immune. An individual who has recovered can no longer switch state. This means that the epidemic inevitably ends if $S_t + R_t = N$. The model starts at time $t = 0$ with one infectious individual and the rest as susceptible individuals. More precisely

$$S_0 = N - 1, \qquad I_{1,0} = 1, \qquad I_{2,0} = 0, \qquad Q_0 = 0, \qquad R_0 = 0.$$

Since the population is closed it is always the case that

$$S_t + I_{1,t} + I_{2,t} + Q_t + R_t = N$$

for all $t \in \mathbb{N}$.

The population is divided into work groups $i = 1, ..., g$ which consist of h individuals. To differentiate between different work groups as well as time increments we create the following notation. $S_t^{(i)}$ denotes the number of susceptible individuals in group i at time t. $I_{1,t}^{(i)}$ and $I_{2,t}^{(i)}$ the number of infectious individuals in group i at time t. $Q_t^{(i)}$ the number of individuals in quarantine in group i at the time t. $R_t^{(i)}$ the number of recovered individuals in group i at the time t.

It is always the case that

$$S_t = \sum_{i=1}^{g} S_t^{(i)}, \qquad I_{1,t} = \sum_{i=1}^{g} I_{1,t}^{(i)}, \qquad I_{2,t} = \sum_{i=1}^{g} I_{2,t}^{(i)}$$

$$Q_t = \sum_{i=1}^{g} Q_t^{(i)}, \qquad R_t = \sum_{i=1}^{g} R_t^{(i)}.$$

2.4 Disease spread in the workplace

In this section, we are looking more closely at the probability of getting infected. In this model, we have a household structure with infectious individuals making contact on two levels. The local level, the probability that a susceptible individual gets infected by an infectious individual in the same work group. The global level, the probability that a susceptible individual gets infected by any infectious individual in the population. This means that an infectious individual can infect individuals in their own work group on the global as well as the local level. It is important to note that the two levels of infection are independent of each other. To calculate the probability that a certain individual gets infected we need to look at the probability that the individual gets infected on any level. To do this we first need to look at the probability of getting infected on each level separately.

Let λ_L denote the probability that an infectious individual contact a certain susceptible individual on the local level. The probability of getting infected on the local level is

$$P(\Lambda_L)_i = 1 - (1 - \lambda_L)^{I^{(i)}}.$$

Since λ_L is the probability of making contact with a single infectious individual the probability $(1 - \lambda_L)$ is the probability of not making contact with that specific infectious individual. The probability of not making contact with any of the infectious individuals in the same group is $(1 - \lambda_L)^{I^{(i)}}$. The probability $1-(1-\lambda_L)^{I^{(i)}}$ is the complementary event to not getting infected on the local level. With the same reasoning

$$P(\Lambda_G) = 1 - (1 - \lambda_G)^I$$

is the probability of getting infected on the global level. The probability of getting infected on any level can be derived using basic probability theory for two independent events. The probability that one of these events occur is

$$P(\Lambda_L \cup \Lambda_G)_i = P(\Lambda_L)_i + P(\Lambda_G) - P(\Lambda_L \cap \Lambda_G)_i$$

and since $P(\Lambda_L)_i$ and $P(\Lambda_G)$ is independent it follows that the joint distribution $P(\Lambda_L \cap \Lambda_G)_i = P(\Lambda_L)_i \cdot P(\Lambda_G)$.[3]

Putting this together we get

$$P(\Lambda_L \cup \Lambda_G)_i = P(\Lambda_L)_i + P(\Lambda_G) - P(\Lambda_L)_i \cdot P(\Lambda_G).$$

The probability of getting infected varies between the different work-groups. In order to differentiate between them, we let $P(\Lambda_L \cup \Lambda_G)_i$ be the probability of getting infected on any level for a susceptible individual in group i.

2.5 Summary of the model

At $t = 0$ there is $N-1$ susceptible individuals and $m = 1$ infectious individual in state $I_{1,t}$. Susceptible individuals can then get infected with a group-specific probability $P(\Lambda_L \cup \Lambda_G)_i$. Ones infected the susceptible individual moves to state $I_{1,t}$ were they stay for a predetermined number of days D and then move to state $I_{2,t}$. Once in state $I_{2,t}$ the infectious individual can then stay home from work the next day with probability γ and move to state Q_t. Individuals in state Q_t can then recover with probability δ and move to state R_t were they have returned to work and are immune to infection. If an individual has been sick for a total of ρ days the individual recovers, moving from state $I_{2,t}$ or Q_t to state R_t.

2.6 Reproduction number

When looking at an epidemic there are generally two end scenarios. Either the disease dies out quickly with only a few infected individuals or a large

part of the population gets infected. The reproduction number R_0 plays a central part when looking at the final size of the epidemic. In general R_0 is defined as the expected number of individuals infected by an infectious individual in a mostly susceptible population. In our model, we have a two-level contact structure. Because of this, we are interested in the average number of global infections as well as the local sub-epidemics within groups. The average number of global infections can be derived using branching process approximation. We assume that the number of work groups $g \to \infty$ and denote the average infectious period of an infectious individual as τ. Let λ_G^*/n denote the rate at which an infectious individual contacts a given individual in the population on the global level. The initially infectious individual on average contacts $\lambda_G^*\tau$ individuals who all with high probability belong to different work groups. All these newly infected individuals will then start sub-epidemics in their respective household. The size of these sub-epidemic result in on average $M_L = \sum_{j=1}^{h} jP_{1,j}$ were $P_{i,j}$ is the probability that the epidemic results in j infected individuals. We are interested in the average number of groups as well as the sub-epidemics that they generate. We define the basic reproduction number as $R_0 = M_L\lambda_G^*\tau$.[1, p. 56] Branching process approximation of epidemics with two levels of mixing is looked at more closely by Ball, Mollison and Scalia-Tomba.[4] In a research report by Pellis, Ball and Trapman[5] they discuss another way of calculating R_0. The research report also gives an intuitive argument on how R_0 is constructed. For our model M_L and, by extension R_0, is hard to calculate numerically. For this reason, we decided to simulate R_0. The simulation of R_0 is done in chapter 3.1.

Chapter 3

Simulations and results

3.1 Reproduction number

The basic reproduction number R_0 is approximated by simulating the average number of global contacts $\lambda_G^*\tau$ and the average size of the sub-epidemics $M_L = \sum_{j=1}^{h} jP_{1,j}$ separately. From these two simulations we can approximate $R_0 = M_L\lambda_G^*\tau$. We simulated R_0 for the following parameter values: $\lambda_L = 0.1$, $\lambda_G^* = 1$, $\rho = 5$, $D = 1$, $\gamma = 0.75$ and group size $h = 10$. We approximate $R_0 \approx 9.7$. This means that there is a positive probability that the epidemic results in a major outbreak for these parameter values.

3.2 Results

We are interested in how the epidemic affets the productivity of the workplace as a whole. In order to take into account that infectious individuals are less productive we value each day that a sick individual work as θ. θ is a value between 0 and 1 which we have decided to be fixed. It is important to point out that θ is arbitrary.

We want to look at how each parameter impacts the epidemic. To do this we look at all parameters individually. We simulate the model where we let a single parameter take all possible values in a given range. Lastly, we plot the parameters different values against the lost productivity of the workplace. When we analyze how a specific parameter affects the loss of productivity, all other parameters remain constant. From Figure 3.1 we can tell that λ_G has quite a large impact on the loss of productivity. This is true for values lower or equal to 0.04. For values larger than 0.04 there is no visible effect in terms of lost productivity. This can be because once $\lambda_G \geq 0.04$ the whole population gets infected quickly.

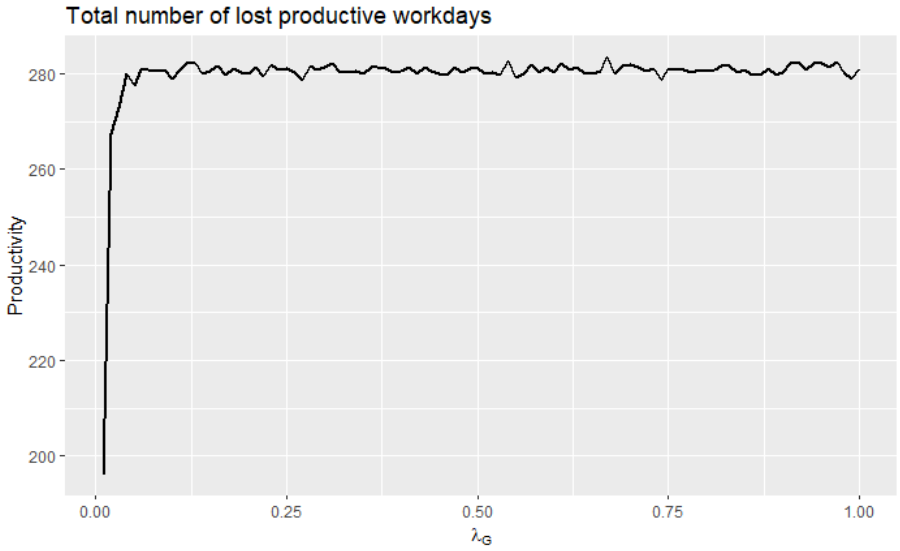

Figure 3.1: 100 simulations have been run for each value of the parameter λ_G. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_L = 0.1$, $D = 1$, $\gamma = 0.75$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$.

Figure 3.2 shows a similar effect of λ_L as that of λ_G in Figure 3.1. λ_L has quite a large impact on the loss of productivity. This is true for values lower or equal to 0.3. Once $\lambda_L \geq 0.3$ there is no visible effect in terms of lost productivity. This can be because once $\lambda_L \geq 0.3$ the whole population gets infected quickly.

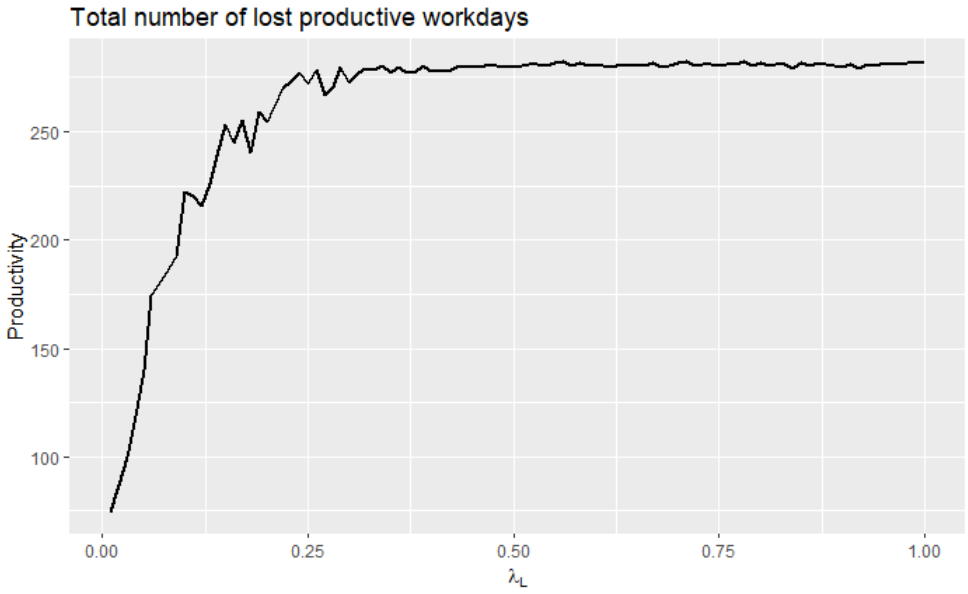

Figure 3.2: 100 simulations have been run for each value of the parameter λ_L. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_G = 0.01$, $D = 1$, $\gamma = 0.75$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$.

Figure 3.3 shows a linear relation between the parameter δ and the loss of productivity. δ is the probability that an individual in state Q return to work. The parameter only affects the number of days spent at home and not how the disease spreads. Because of this, the linear relation is expected.

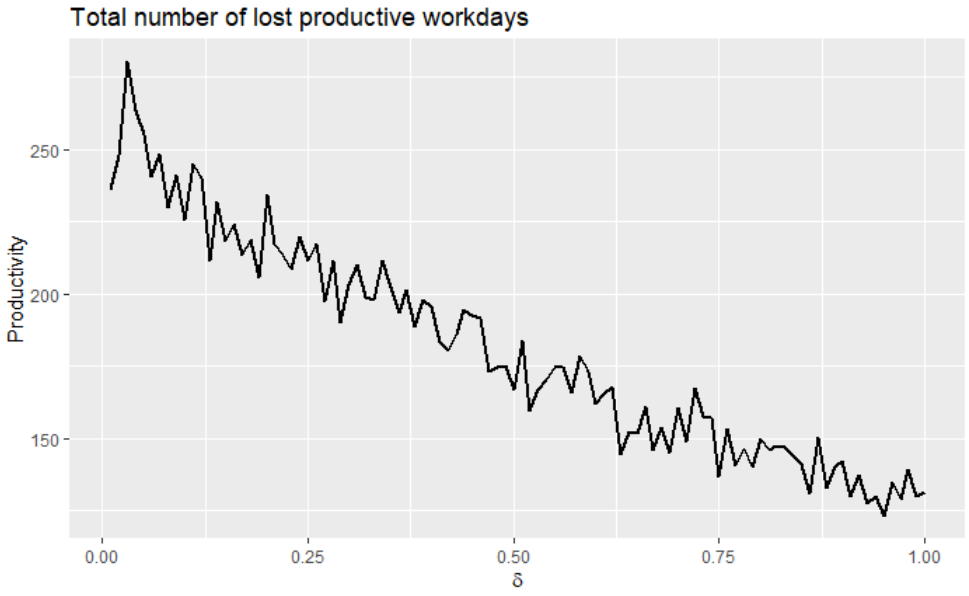

Figure 3.3: 100 simulations have been run for each value of the parameter δ which is the probability that an individual in state Q recovers. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $D = 1$, $\gamma = 0.75$, $\rho = 5$ and $\theta = 0.7$.

Figure 3.4 shows that γ affects the loss of productivity in a rather unusual way. γ is the probability that an infectious individual in state I_2 stays home from work the next day. The loss of productivity is highest for values between 0.25 and 0.75, and low for $\gamma \leq 0.25$ or $\gamma \geq 0.75$. For low values of γ, this would mean that infectious individuals stay at work for the maximum amount of $\rho = 5$ days. After five days they recover without taking any paid sick leave. Very few individuals stay home from work to recover, which leads to the loss of productivity being lower overall. For high values of γ, this would mean that infectious individuals stay home from work early. This leads to fewer

individuals being infected overall and the disease dying out quickly.

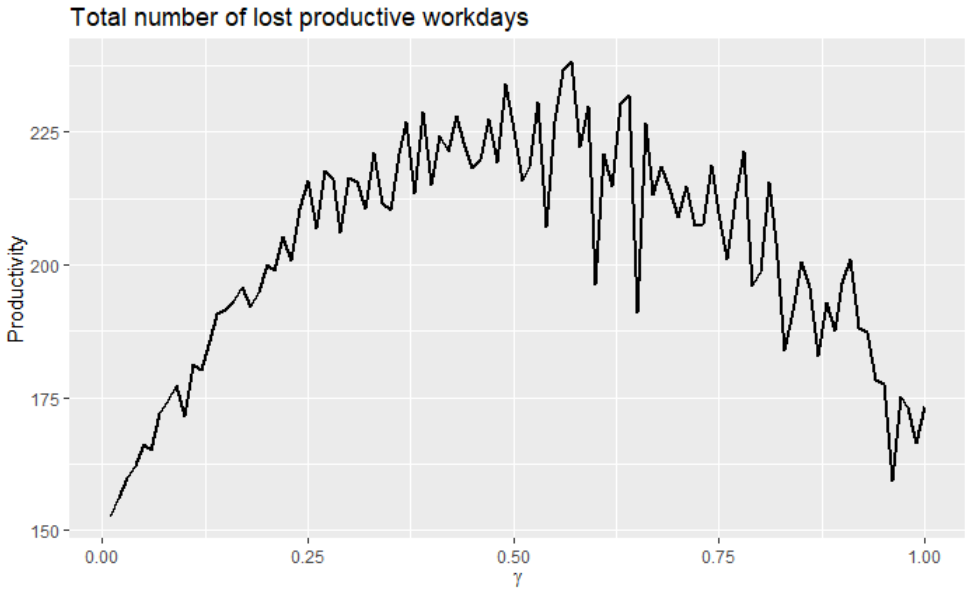

Figure 3.4: 100 simulations have been run for each value of the parameter γ which is the probability that an individual in state I_2 stays home from work. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $D = 1$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$.

Figure 3.5 shows a non-linear relation between the number of days until the disease is observed D and the loss of productivity. For $D = 0$ the loss of productivity is significantly lower. The reason for this is that for $D = 0$ there is a possibility that infectious individuals stay home from work without the possibility of infecting anyone else. The loss of productivity is slightly lower for $D = 4$ when compared to $D \in \{1, 2, 3\}$. This can be explained by the fact that for $D = 4$ infectious individuals will not have the opportunity to stay home from work. This means that the number of days of paid sick leave

is low but the number of days that infectious individuals are still at work is high.

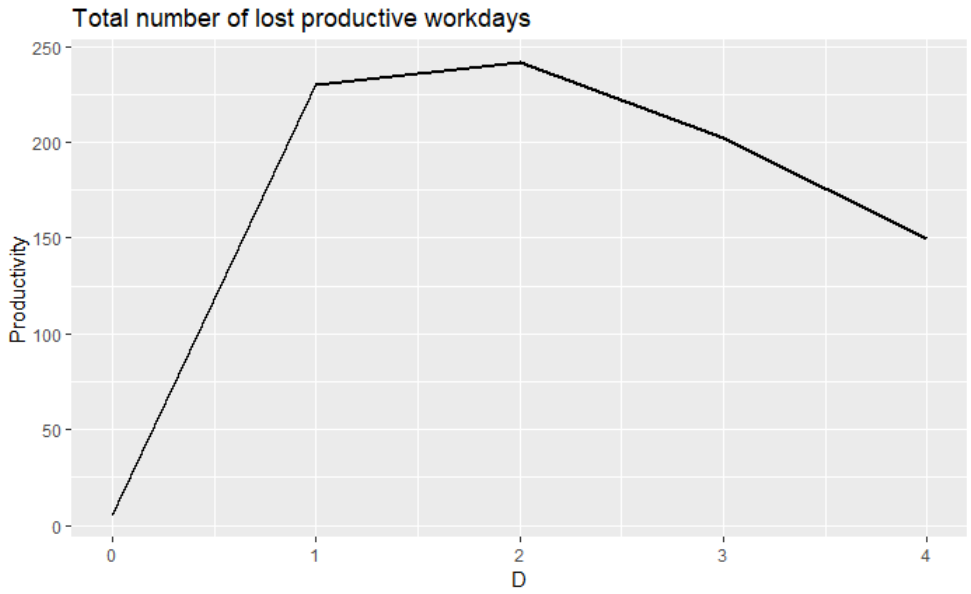

Figure 3.5: 100 simulations have been run for each value of the parameter D which is the amount of time until the disease is observed. The values range from 0 and 4 in one day increments. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $\gamma = 0.75$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$.

From Figure 3.6 we see that the maximum amount of possible sick days ρ, results in more productivity lost the larger the value of ρ. If ρ is larger than 10 it does not result in further loss of productivity. This is because after 10 sick days almost all infectious individuals have recovered. Increasing the roof on the maximum number of sick days beyond 10 does not impact the productivity negatively.

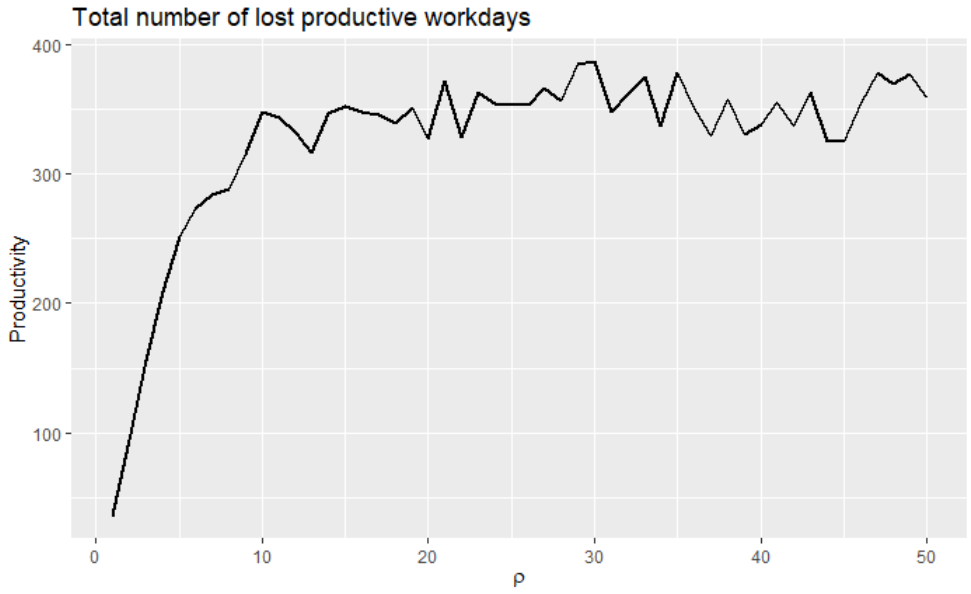

Figure 3.6: 100 simulations have been run for each value of the parameter ρ which is the maximum number of days an individual can be sick. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $D = 1$, $\gamma = 0.75$, $\delta = 0.25$ and $\theta = 0.7$.

Figure 3.7 shows a linear relation between θ and the loss of productivity. θ is a multiplier on the productivity of infectious individuals. It does not affect the epidemic in any way. Since it only acts as a multiplier on a part of the workers the linear relation is expected.

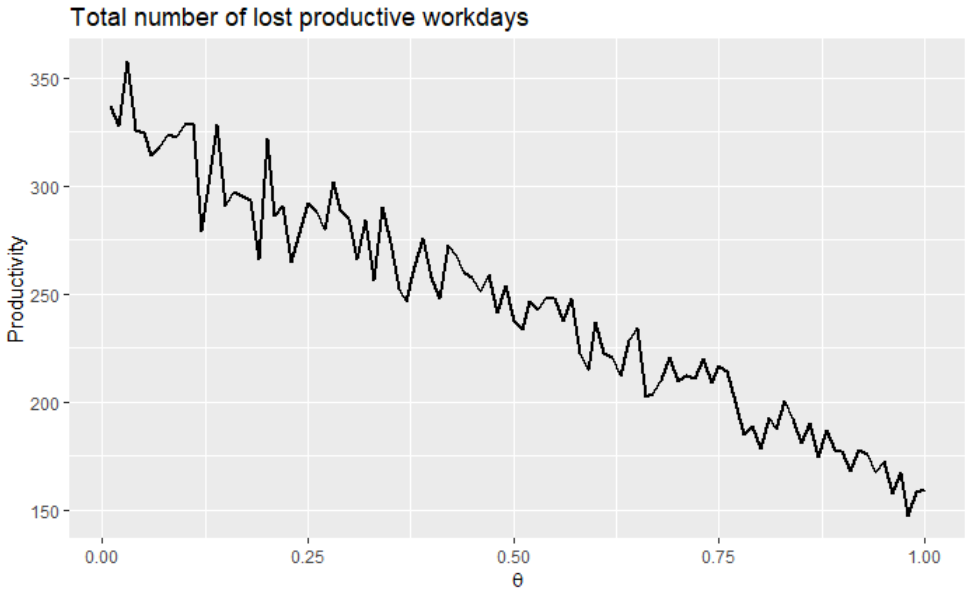

Figure 3.7: 100 simulations have been run for each value of the parameter θ which is the productivity of an infectious individual. The values range from 0.01 to 1 in increments of 0.01. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $D = 1$, $\gamma = 0.75$, $\delta = 0.25$ and $\rho = 5$.

Simulations have been done for two different sets of parameters. The difference between the two sets of parameters is the value for parameter D. In one of the sets we have $D = 1$ and in the other $D = 2$. All other parameters are the same in both sets, $\lambda_L = 0.1$, $\lambda_G = 0.01$, $\gamma = 0.75$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$. Figure 3.8 shows that for $D = 2$ the epidemic is much larger. The epidemic for $D = 1$ although smaller does last longer. This might seem counter-intuitive. The reason is that if most of the population gets infected early they will also recover early, hence the epidemic will die out faster. This is looked at more thoroughly in a study by Lashari, Serafimović and Trapman[6] on the duration of an SIR epidemic on a configurated model. In the study, they show that a decrease in R_0 might increase the duration of the epidemic.

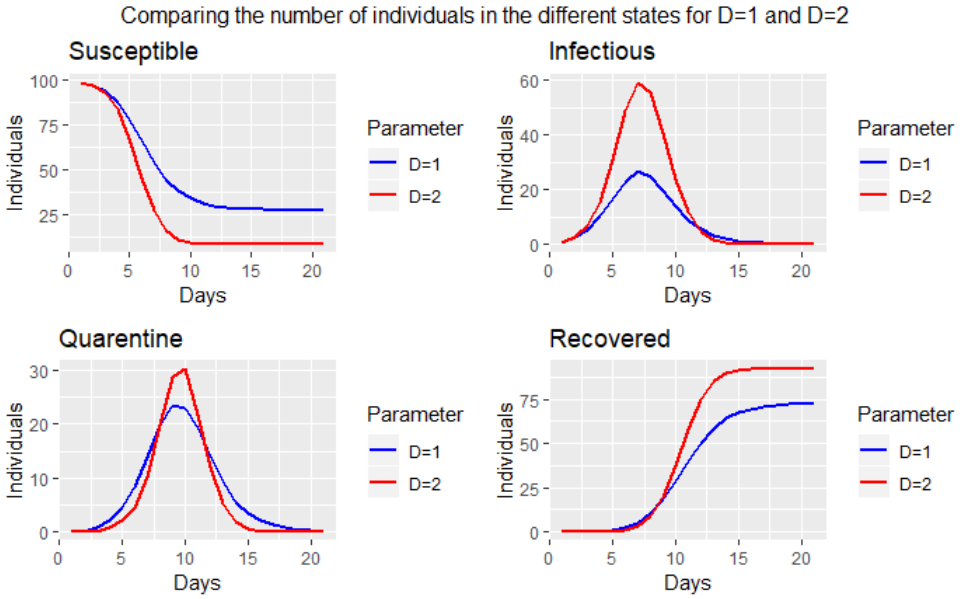

Figure 3.8: 100 simulations have been run for $D = 1$ and $D = 2$. The average number of individuals in each state is displayed for every state. All other parameters remain constant: $\lambda_L = 0.1$, $\lambda_G = 0.01$, $\gamma = 0.75$, $\delta = 0.25$, $\rho = 5$ and $\theta = 0.7$.

Chapter 4

Discussion and improvements

4.1 Discussion

The simulations of how the different parameters affect the productivity gave some different results. The parameters λ_L, λ_G and ρ all affected the productivity in a similar way. The parameters greatly increased the loss of productivity as their value increases, but only until a certain point. After that point there seems to be no further effect from the parameters on the loss of productivity. The parameters γ and D both affect the productivity in a non-linear way. The loss of productivity is lower for small and large values. The loss of productivity is high for midrange values of the parameters. This is because when the value is small the disease dies out quickly with few individuals infected. If the value is high the epidemic spreads to the whole population quickly and then dies out quickly. If however, the value is midrange it results in the epidemic being dragged out. The disease would then stay around longer, which would result in more productivity lost. Fig-

ure 3.8 showed that the set of parameters with $D = 2$ resulted in a larger but shorter epidemic. The reason is that if most of the population gets infected early they will also recover early, hence the epidemic will die out faster. This is further looked at in a study by Lashari, Serafimović and Trapman [6] where they show that a decrease in R_0 might increase the duration of the epidemic.

4.2 Improvements

In this thesis, we have simulated a simple epidemic model and looked at how the spread of disease impacts the workplace in terms of productivity. The model is as mentioned very simple. It has a closed population and only handles discrete time increments that represent a whole work day. If one would like to look at this type of thesis more thoroughly some suggestions for improvement would be to have a population that is not closed so that new employees can get hired and old ones quit. Another improvement would be to look at continuous time and include something about the probability of being sent home or leave during the day. Productivity and how sickness affects it can also be looked into in more detail.

4.3 References

1. Andersson, H. and Britton, T., "Stochastic epidemic models and their statistical analysis", Springer, 2000.

2. Britton, T., "Stochastic epidemic models: a survey", Mathematical Biosciences, 2010.

3. Alm, S. E. and Britton, T., "Stokastik Sannolikhetsteori och statistikteori med tillämpningar", Liber, 2008.

4. Ball, F., Mollison, D. and Scalia-Tomba, G., "Epidemics with two levels of mixing", Ann. Appl. Prob. **7**, 46-89, 1997.

5. Pellis, L., Ball, F. and Trapman, P., "Reproduction numbers for epidemic models with households and other social structures I: Definition and calculation of R_0". Mathematical Biosciences, 2012.

6. Lashari, A. A., Serafimović, A. and Trapman, P., "The duration of an SIR epidemic on a configuration model", in Lashari, A. A. "Stochastic epidemics on random networks", Stockholm University. (2019).