FEDERAL RESERVE BANK OF MINNEAPOLIS | Research Division

WORKING PAPER
No. 768

Measuring the Transmission Rate of COVID-19: What Can We Learn From Cross Country Comparisons?

May 2020

Han Gao
University of Minnesota

Juan Pablo Nicolini
Federal Reserve Bank of Minneapolis and Universidad Di Tella

DOI: https://doi.org/10.21034/wp.768
Keywords: Identification, Epidemiological models
JEL classification: C10, I10

Measuring the Transmission Rate of COVID-19: What Can We Learn From Cross Country Comparisons?

Han Gao
University of Minnesota

Juan Pablo Nicolini
Federal Reserve Bank of Minneapolis and Universidad Di Tella

May 29, 2020

Abstract

We modify a standard SIR epidemiological model to allow for testing and asymptomatic agents. We explore cross country variation's ability to allow for identification of key parameters of the model: the fatality rate and the evolution over time of the normalized transmission rate. We first show that as long as tests are applied only to agents who exhibit symptoms, those parameters cannot be identified. We briefly discuss which additional information may allow for identification. Finally, we also describe conditions under which the normalized transmission rate can be computed with very high accuracy, and how cross country evidence can be used to evaluate the effect of lockdowns on evolution of the effective transmission rate over time.

1 Introduction

Among the uncertainties that make policy decisions difficult during the current COVID-19 pandemic, the true fatality rate of the virus and the rate at which the virus spreads rank at the top. We have daily measures of fatalities and newly detected cases. However, there are reasons to believe that measurement error affects both numbers.

First, there may be fatalities that never make it to the hospital. If this were the case, existing statistics on fatalities a lower bound to the true lethality of the virus. We will mostly abstract from this problem in this note.

Second, there is ample evidence that a sizable fraction of infected persons develop no symptoms, are never tested, and therefore are never detected.For example, as of April 22nd, a total of 82,798 infected persons had been detected in China, a country that completed the first (and hopefully last!) cycle of the disease. The same day, a total of 4,632 fatalities had been recorded. Absent any asymptomatic agents, these numbers imply a fatality rate of 5.6%, which is extremely large. This rate is comparable to the lowest estimates of the Spanish flu, which is considered one of the deadliest pandemics in world history.[1] Most proposals to relax lockdown policies are based on the sensible notion that the number of infected people in China is substantially

[1]About 500 million people are estimated to have been infected, a third of the world population of the time. The death toll is estimated to be as high as 100 million and as low as 17 million. According to the lowest estimate, the fatality rate would be about 3.4%. See Rosenwald, Michael S. (7 April 2020). "History's deadliest pandemics, from ancient Rome to modern America". Washington Post. Archived from the original on 7 April 2020. Retrieved 11 April 2020.

larger than 82,798. The debate is about how much larger than that.

In this note, we study the extent to which cross country evidence can be used to estimate the underlying parameters of a classical SIR model.

A main assumption is maintained throughout – the fatality rate and the speed at which infected agents recover depend on the specificity of the virus, and they are therefore similar across countries, once controlled by observables such as the age structure of the population or the income per capita of the countries.[2] If this were not the case and those two parameters were systematically affected by a virus-country interaction, then the cross country evidence would be useless.

If tests were random and the sample size large enough, very precise estimates of the true number of asymptomatic agents could be obtained. However, because of the urgency created by the crisis, tested agents are mainly people who exhibit symptoms identified with COVID-19. This bias in testing makes statistical inference a difficult task.

We extend the simplest version of the SIR model by assuming that countries set different thresholds for testing, possibly for budgetary reasons. The model then endogenously relates the number of detected infections, the number of fatalities, and the number of tests done by countries (or regions). The endogenous nature of testing is modeled by assuming that tests are performed on agents who exhibit the known virus-related symptoms. Two unknown distributions are key primitives of the model. The first one is the probability

[2]In the analysis below, we will ignore the effect of these observables.

that an infected agent develops certain symptoms. The second one is the probability that a healthy agent develops certain symptoms.

We first show that the assumed pattern of testing implies that the available cross country evidence cannot be used to identify the true fatality rate. To put it differently, the cross country data on fatalities and detected cases are consistent with a large set of values for the fatality rate of the virus including the most extreme estimates, from 0.1% to 2%.

We prove this by showing that there is one more unknown than restrictions imposed by the model. Thus, we can set either one to any arbitrary number, and the model can be used to solve for the other unknowns. The main reason for this result is that the model imposes no structure on the distribution of virus-like symptoms among the non-infected agents, which is a primitive of the model. Absent any restrictions, distributions can be found that rationalize the cross country data. Under this scenario, the data on detected infections are uninformative regarding the parameters of interest: the fatality rate and the evolution of the normalized transmission rate.

We then go on and show how imposing some restrictions on that distribution could potentially allow for identification. We do this by imposing a restriction that is natural within the context of the particular model we use. But the logic can be used with alternative restrictions that could be justified with independent evidence on those distributions.

Finally, we exploit the fact that the model allows for only one degree of freedom to obtain bounds for one of the parameters of interests: the

evolution over time of the effective transmission rate. We illustrate with two examples how this property can be used in making cross country (or cross state) comparisons. In the first, we study the case of the Nordic countries (Denmark, Norway, and Sweden) that adopted very different strategies at the onset of the infections. We set the common fatality rate equal to some of the estimates that had been found in the literature. The model, together with the data for the three countries, can be used to back out the implied values for the other parameters, like the daily rate of transmission. We show that the different behavior over time of the rate of infection across these three countries is quite robust to the different values assumed for the fatality rate. The exercise provides an estimate of the effect of a more severe lock-down (as in Norway) relative to a less severe one (as in Sweden).

In our second example, we apply the model to several states in the U.S.. The exercise allows us to quantify the effect of lockdowns on the transmission rate across states. It shows remarkable impact across states, with fast convergence of the normalized transmission rate to one or below, which implies a decreasing number of active cases. However, a notable exception is Minnesota, where the normalized transmission rate starts growing on April 19th to values substantially higher than one.

We end the paper by discussing the conditions under which the effective transmission rate can be identified in the data, even under severe uncertainty regarding the fatality rate. We show that given the available data, the computation of the effective transmission rate is very insensitive to the true

fatality rate when the fraction of susceptible agents - those who have not yet been exposed to the virus - is large. Thus, while the computations are robust at the onset of an epidemic, they would lose precision in additional waves of infections.

The paper proceeds as follows. In Section 2, we present the model with tests and asymptomatic agents, and we discuss conditions under which the relevant parameters can be identified with available data. In section 3, we show, by means of examples, that it is possible to obtain precise estimates of the evolution over time of the effective transmission rate. Section 4 analyzes a simplified version of the model to explore the theoretical reasons behind the empirical results of Section 3.

2 The Model

We modify a simple SIR model to allow for cases that are undetected because of a lack of symptoms. There is a unit mass of agents who can be in any of the five following possible states.

- S_t : number of agents who have so far not been infected.
- I_t^d : number of agents who have been infected, tested, and detected.
- I_t^n : number of agents who have been infected and were not tested, so they have not been detected.
- R_t^d : number of agents who have been infected, detected, and have either died or recovered.
- R_t^n : number of agents who have been infected, were never detected, and recovered. (As we explain below, our assumptions imply that if the infection is not detected, mortality is zero).

Then, the following identity must hold.

$$S_t + I_t^d + I_t^n + R_t^d + R_t^n = 1$$

We assume that, conditional on being infected, agents recover with probability δ each day. In addition, the fatality rate, again conditional on being infected, is given by q.

2.1 Development of symptoms

In this section, we discuss the testing strategy, which may differ across countries. To simplify, we focus only on tests that detect the virus and do not consider tests for antibodies. In addition, we assume the tests to have no errors.

Every agent in S_t can develop symptoms that are associated with being infected. We let $z \in \mathbb{R}$ be an index that determines the severity of the symptoms. Any agents who transits from S_{t-1} to be infected at time t draws a shock z that determines the symptoms that will be developed, drawn from some distribution with CDF F. This shock is permanent and determines the severity of the infection.

The worse the symptoms of an agent are, the worse the evolution of the infection is. We assume that if symptoms are larger than an unknown threshold z_q , the infection is deadly. The threshold z_q is given by the chemical characteristics of the virus. All infected agents with $z < z_q$ are eventually cured and develop immunity. Thus, the number of fatalities as a fraction of infected agents is given by

$$m = 1 - F(z_q).$$

This threshold, together with the duration of the infection, determines the two exit rates from the infection: the daily mortality rate, q, and the rate at which infected agents get cured, δ. These rates relate to the mortality rate

in the usual way for a Poisson process, given by $m = q/\delta$. The threshold z_q, as well as the rates q and δ, is given by the chemical characteristics of the virus, and we will assume unknown constants.

Non-infected agents in S_t can also develop symptoms, which explains why many agents with symptoms test negative. Specifically, we assume that every period, healthy agents also draw a value for z, but from a different distribution, with CDF $H(z)$. This shock may be correlated over time.

As it is natural with any new virus, neither F nor H are known during the the first months of the pandemic. It is lack of knowledge of these functions that makes identification difficult.

2.2 Endogenous Testing

To allow for endogenous testing, we assume that tests are administered to agents with symptoms. We assume that there is a critical value, $z_c \leq z_q$, such that everybody with $z \geq z_c$ is tested in all countries and all periods. These assumptions imply that all fatal cases are tested and properly identified. As we mentioned in the introduction, there are reasons to believe that this is not the case. To focus on the misreporting of infected individuals, we will nevertheless make this assumption throughout.

In general, countries will test agents more broadly. We model the decision to test as the choice of a country-specific value of $z_t^T < z_c$. The probability of being tested if infected, is modeled by $1 - F_z(z_t^T) = p_t^z$, while the probability of being tested if not infected, is modeled by $1 - H(z_t^T) = p_t$.

This assumption allows for the cross country heterogeneity in testing that we see in the data. A natural interpretation is that richer countries have larger budgets and can afford more testing, but there are other reasons why there are variations in testing across countries. A natural question is whether the cross country heterogeneity in testing allows us to identify the mortality rate of the virus and therefore the fraction of infected agents.

Notice that our assumptions imply that all fatalities have been detected, so no undetected agent dies. We therefore let q_t^d be the fraction of detected agents with $z \geq z_q$.

$$q_t^d = q\frac{I_t^d + I_t^n}{I_t^d},$$

which has the natural interpretations of the mortality rate, conditional on being detected. According to the model, this rate varies across countries, as they set different values for p_{zt}.

The model abstracts from dynamics in the development of symptoms. Once an agent is infected, she draws once and for all a value for z, which determines if she will be tested or not. To simplify, we assume that symptoms last just one day, so agents are tested either immediately or not at all. Therefore, with this form of endogenous testing, there are no transitions from I_t^n toward I_{t+1}^d.

Below, we will consider the possibility that some countries in some periods test agents independently of symptoms. We set a rate τ_t of non-detected agents who are tested. In that case, there would be transitions from I_t^n

toward I^d_{t+1}.

2.3 Rate of Infection

As is standard in SIR models, the rate of infection is assumed to be proportional to the number of encounters between susceptible agents, S_t, and infected agents, $(I^d_t + I^n_t)$. Specifically, the fraction of newly infected agents is given by

$$\beta_t(I^d_t + I^n_t)S_t,$$

where β_t is the rate at which encounters result in infections and is allowed to depend on time, to allow for social distancing or lockdown policies to have an impact on the transmission rate of infections. But not all infected agents will be tested: a fraction p_{zt} are tested and detected, while the rest are never detected. This last group builds, period by period, the set of asymptomatic agents.

2.4 Transitions.

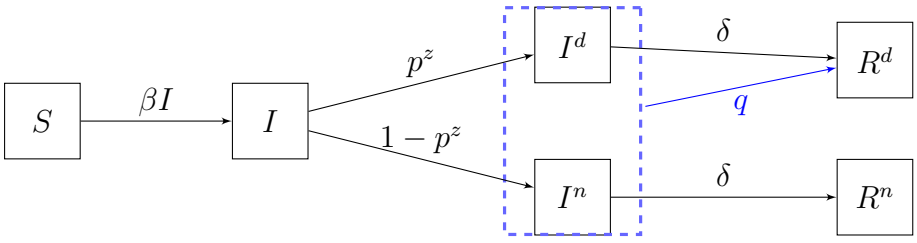


Figure 1

Given the above assumptions, the transitions across different states are as follows.

- An agent in S_t can
  - go to I^d_{t+1} with probability $\beta(I^d_t + I^n_t)p^z$,
  - go to I^n_{t+1} with probability $\beta(I^d_t + I^n_t)(1 - p^z)$,
  - or remain in S_{t+1}.
- An agent in I^d_t can
  - go to R^d_{t+1} with probability $\delta + q\frac{I^d_t + I^n_t}{I^d_t}$,
  - or remain in I^d_{t+1}
- An agent in I^n_t can
  - go to R^n_{t+1} with probability δ,
  - or remain in I^n_{t+1}.

Finally, both R^d_t and R^n_t are absorbing states.

As discussed above, newly infected agents are given by $S_t\beta(I^n_t + I^d_t)$, of which a fraction p^z is tested and detected. A fraction $(1 - p_z)$ is not tested and remains undetected. The non-infected agents in S_t who get tested get a negative testing result and remain in S_t. Note that given the way we model testing, the infected agents who have $z < z^T$ have a fatality rate of zero.

3 Solution of the model

Given

1. initial conditions $\{S_0, I_0^d, I_0^n, R_0^d, R_0^n\}$, where all are ≥ 0 and such that

$$S_0 + I_0^d + I_0^n + R_0^d + R_0^n = 1,$$

2. and parameters δ, q, β_t, p_t^z,

the following equations solve for the sequences $\{S_t, I_t^d, I_t^n, R_t^d, R_t^n\}_{t=1}^{\infty}$:

$$S_t - S_{t+1} = \beta_t (I_t^d + I_t^n) S_t, \tag{1}$$

$$I_{t+1}^d - I_t^d = \beta_t (I_t^d + I_t^n) p_t^z S_t - \delta I_t^d - q \left(I_t^d + I_t^n \right), \tag{2}$$

$$I_{t+1}^n = I_t^n + \beta_t (I_t^d + I_t^n)(1 - p_t^z) S_t - \delta I_t^n, \tag{3}$$

$$R_{t+1}^d = R_t^d + \delta I_t^d + q \left(I_t^d + I_t^n \right), \tag{4}$$

$$R_{t+1}^n = R_t^n + \delta I_t^n. \tag{5}$$

Equation (1) equates the change in the mass of susceptible agents to change in the mass of newly infected ones. Equation (2) equates the net inflow into the pull of detected agents to that of the newly infected who get tested, minus the previously detected agents who either recovered or died. Similarly, equation (3) equates the net inflow into the asymptomatic agents to that of the newly infected agents who do not get tested, minus the previously

infected who recovered. Recall that undetected agents never develop serious symptoms and do not die. Finally, equations (4) and (5) describe the law of motion for the two absorbing states. Again, recall that only detected agents are exposed to mortality risk and that

$$q\left(I_t^d + I_t^n\right) = q_t^d I_t^d,$$

where the conditional risk, q_t^d, is endogenous to policy, since it depends on the intensity of the tests.

Exogenous testing If we were to allow for exogenous testing, where a fraction τ_t of non-detected agents get tested, then the law of motion for I_t^d and I_t^n would change as follows

$$I_{t+1}^d - I_t^d = \beta_t(I_t^d + I_t^n)p_t^z S_t - \delta I_t^d - q\left(I_t^d + I_t^n\right) + \tau_t I_t^n, \tag{6}$$

$$I_{t+1}^n = I_t^n + \beta_t(I_t^d + I_t^n)(1 - p_t^z)S_t - \delta I_t^n - \tau_t I_t^n. \tag{7}$$

As we show below, the data we have are not enough to identify the model, even assuming $\tau_t = 0$ for all t. We proceed to consider that case first. We allow for exogenous testing at the end of next section.

3.1 Identification

The problem we address is the inverse of the one we described above.[3] In that problem, given initial conditions and parameter values, we explained how to solve for the sequences $\{S_t, I_t^d, I_t^n, R_t^d, R_t^n\}_{t=1}^{\infty}$. During an epidemic like the one that is ongoing, we do observe the sequences $\{I_t^d, R_t^d\}_{t=1}^{T}$. In addition, given our assumption that all fatalities are well measured, we also observe $q\left(I_t^d + I_t^n\right)$. However, we do not independently observe q and $\left(I_t^d + I_t^n\right)$. It is therefore impossible to evaluate the current number of truly infected agents.

In addition, we do have a relatively precise range of estimates for δ, since data on the time it takes for infected people to recover are available.

The first main question we address is whether the model and the cross country evidence be used to identify q and therefore I_t^n. As we mentioned in the introduction, the answer is no.

We prove this by showing that given any initial conditions $\{S_0, I_0^d, I_0^n, R_0^d, R_0^n\}$, any possible value for q can be made consistent with the data through proper selection of the values of the other unknown parameters, β_t, p_t^z.

To see this, notice first that equation (3) holds by construction: the three terms on the right-hand side are observable. Indeed, it is this equation that is used to compute the observable $\{I_t^d\}_{t=0}^{T}$. Then, once q is fixed, and given that we observe $q\left(I_t^d + I_t^n\right)$, we can compute the total number of infected people $\left(I_t^d + I_t^n\right)$.

[3] We follow the strategy of "counting equations" to show that the model cannot be identified. It is similar to the strategy of showing there are redundant fiscal instruments in order to implement a given Ramsey allocation.

Next, using the value for δ, equations (4) and (5) and some initial condition can be used to compute $\{R_t^d, R_t^n\}_{t=1}^T$, while the identity

$$S_t + I_t^d + I_t^n + R_t^d + R_t^n = 1$$

can be used to solve for $\{S_t\}_{t=1}^T$. Finally, equation (1) can be used to compute β_t, while equation (3) can be used to solve for p_t^z. Finally, using data on I_t^d, we can compute q_t^d and I_t^n.

It is reasonable to assume that both q and δ are common to all countries, since they are specific to the chemical structure of the virus. But the logic above holds for every country, even if we impose that those two parameters are common across countries.

Notice that it is possible to compute the values for β_t, p_t^z and S_t without using information on I_t^d, which is used only at the end to residually compute I_t^n and q_t^d.

As the data are different across countries, the exercise will deliver country-specific series for $\{S_t, \beta_t, p_t^z\}_{t=1}^\infty$ and for $\{I_t^n, R_t^d, R_t^n\}_{t=1}^\infty$. Through the lens of the model, these differences respond to differences in mitigation policies - which affect the value for β_t - and in testing policies – which affect the value for p_t^z.

It is this identification problem that explains the high degree of uncertainty regarding the best course of action for policy. It is also the main justification for the need to perform random tests at a reasonable scale, as

has been forcefully argued by many scientists. These tests would allow for the independent estimation of the amount of undetected cases, I_t^n, and therefore the fatality rate, q, and the rest of the parameters of the model.

Why do cross country data not allow for identification? At first sight, it may seem surprising that the ample cross country variation in testing do not allow for identification. This impression may be reinforced by the notion that in the previous analysis, the only use of the test, was the identification of I_t^d. Thus, one may argue that data on the fraction of S_t agents that tested negative have not been used.

This is in fact true. We do have information on the number of tests done daily per million inhabitants in each country, which we denote ω_t. The following equation must then hold in equilibrium:

$$p_t^z \beta_t S_t (I_t^d + I_t^n) + p_t S_t (1 - (I_t^d + I_t^n)\beta_t) = \omega_t.$$

The only value of that information is to identify the fraction of still susceptible agents who developed symptoms above z_t^T, the threshold value for the symptoms such that agents get tested. In terms of the notation of the model, those data allow for identification of $p_t = 1 - H(z_t^T)$. That information can be used to estimate the unknown CDF $H(z_t^T)$, but that is of no use for learning the values of the key parameters q and β_t.

Notice that the logic above hinges on the ability to find a distribution $H(z_t^T)$ that rationalizes the data. To the extent that no restriction is imposed

on $H(z_t^T)$, the logic follows.

Restrictions on $H(z_t^T)$. We now discuss how independent knowledge of the distribution $H(z_t^T)$ may be used to identify a fraction of exogenous tests that could be used to estimate the true fatality rate. In the context of the model, a reduction in z_t^T implies that the threshold for testing is reduced, which means that more infected agents get tested, so p_t^z increases.

Consider the case in which the ordering of the one-dimensional index, z, is the same for both infected and non-infected agents. In this case, as the threshold z_t^T is reduced, it follows that $p_t = 1 - H(z_t^T)$ goes up, the same way p_t^z does.

Under this assumption, the model implies that p_t^z and p_t must move in the same direction. But nothing in the identification algorithm for the parameters implies that this has to be the case. Thus, in order to fit the data, we must allow for a new variable. One candidate is to allow for a fraction of tests that do not depend on the symptoms of the tested agents and that allow for a flow of agents from I_t^n toward I_{t+1}^d, so that p_t^z and p_t do move in the same direction. This possibility is in line with anecdotal evidence that in some cases, countries did test agents in order to detect at least some asymptomatic agents in order to isolate them and thereby – mitigate the infection rate.

We interpret these tests as the exogenous component of the tests done, which is blurred in the data. Under this interpretation, countries do some

exogenous testing, but they report the sum of the endogenous tests plus the exogenous ones.

What we propose now is a mechanism that would use the model to identify data points in which tests had some exogenous component. If there is enough of this exogenous component in the data on tests, it may then be used to try to estimate the true fatality rate. We now illustrate how this process may work out.

The restriction that p_t^z and p_t must move in the same direction can be written

$$(p_t^z - p_t)\,(p_t^z - p_t) \geq 0. \tag{8}$$

We let τ_t be the rate at which non-detected agents are tested independently of their symptoms. Then, the following equation must hold:

$$p_t^z \beta_t S_t (I_t^d + I_t^n) + p_t S_t (1 - (I_t^d + I_t^n)\beta_t) + \tau_t (S_t + I_t^n) = \omega_t, \tag{9}$$

where ω_t is the fraction of agents tested at time t.

The law of motion for the infected and detected agents evolves as in (6) repeated here for convenience:

$$\beta_t (I_t^d + I_t^n) p_t^z S_t + \tau_t I_t^n = I_{t+1}^d - I_t^d + \delta I_t^d + q\left(I_t^d + I_t^n\right). \tag{10}$$

The problem then becomes finding values for $\{p_t^z, p_t, \tau_t\}_{t=0}^T$ that satisfy restrictions $(8)\,, (9)\,,$ and $(10)\,.$ Clearly, there may be many solutions, given

that restriction (8) is an inequality.

The best way for the model to accommodate the largest possible share to the endogenous component of the tests is to choose $\{\tau_t\}_{t=0}^{T}$ in order to minimize

$$\sum_t \tau_t,$$

subject to (8) , (9) and (10) , plus the restriction that $\tau_t \geq 0$ for all t.

There will be instances when the solution implies $\tau_t = 0$ for some periods or country observations. In fact, if in computing the - unique - solution for $\{\beta_t, p_t^z, p_t\}_{t=1}^{\infty}$ obtained when $\tau_t = 0$ for all t and all countries, it turns out that p_t^z and p_t commove always, so restriction (8) is satisfied, then the solution to that minimum problem would imply $\tau_t = 0$ for all t and all countries. The hope is that the solution implies that $\tau_t > 0$ for enough period or country observations that variation can be used to estimate the true fatality rate.

Two caveats are in order. First, restriction (8) follows directly from the assumptions of the model. And it makes total sense, given that in the model, symptoms live in a single-dimensional space. If symptoms live in multidimensional sets, this monotonicity property may not necessarily follow from the fact that $H(z_t^T)$ is a CDF. Second, given that the minimization problem gives the largest possible chance to the endogenous component of the tests, what we obtain is just a lower bound for the true value for τ_t.

4 Computing the effective transmission rate

Despite the difficulties discussed above, the model can still be used to discipline calibration exercises. The reason why is that, as discussed above, the modeler has one, but only one, degree of freedom in choosing parameters.

Our parameters of interest are q, β_t and, to the extent that one is willing to impose restrictions on the CDF $H(z_t^T)$, also p_t^z and p_t. As argued above, once you choose one of them - q, for example - the model can be used to compute the others - $\{\beta_t,\, p_t^z, p_t\}_{t=0}^T$, in this example. Thus, the structure of the model does impose partial discipline in quantitative analysis.

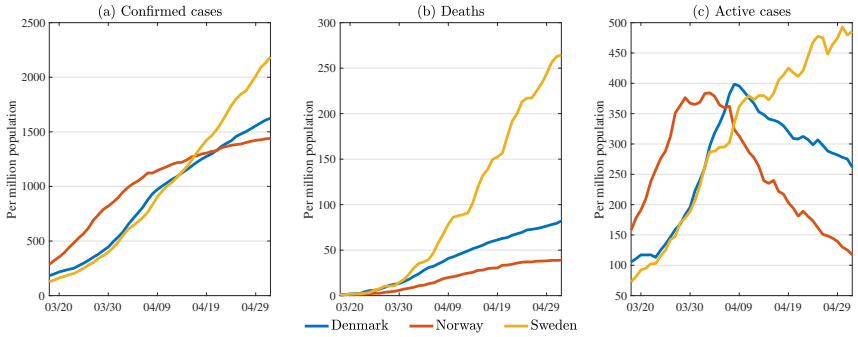
In what follows, using cross country evidence, we provide an example of how the model can be used to learn about the effect of different mitigation policies.

4.1 Numerical examples

Following an outbreak of COVID-19 cases in early March, three Nordic countries took different routes in terms of mitigation policies. While Norway and Denmark opted for relatively severe lockdowns of their economies, Sweden chose a more flexible one, based mostly on promoting social distancing, but left most of its economy open.

The countries' outcomes, as shown in Figure 2, have also been different. The figure shows daily data starting on March 18th, the day when – by chance, one may guess - two of the three countries reached one fatality per

Figure 2: Cumulative reported cases



Note: Active cases projected by assuming $\delta = 0.1$

million inhabitants.[4] Figure 2.a shows the number of detected cases per million inhabitants for the three countries. Figure 2.b shows the number of fatal cases, also per million inhabitants for the same three countries. Finally, Figure 2.c shows the number of active detected cases, assuming a recovery rate $\delta = 0.1$, which implies an average period of infection of 10 days. We use this value as a benchmark.

The data have been obtained from the Johns Hopkins University[5]. The data source also contains data on recovered agents. However, it is incomplete, and there are many reasons to believe that the data collection process is not homogeneous across countries. Thus, we do not use data on recoveries.

Recall that as we will not impose any restriction on $H(z_t^T)$, we will not be using data on tests, Thus, data on detected infections allow us only to estimate the value for q_t^d, which is not a parameter of interest. Thus, the

[4]Denmark reached that number only on March 19th. We still chose to start the three countries on March 18th.

[5]Source: `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`.

following exercise uses data only on fatalities. We therefore let

$$\begin{aligned} I_t &= I_t^d + I_t^d, \\ R_t &= R_t^d + R_t^d. \end{aligned}$$

4.2 The data

The criterion we chose to start the analysis for each country was the day it reached one fatality per million. As it turned out, that date was March 18th for both Norway and Sweden and March 19th for Denmark. Given the proximity of the dates, we then decided to start the three countries on March 18th.

Simple inspection of the data, as in Figure 1, shows very large volatility in new daily fatalities. Feeding the raw data to the model also delivers volatile series that are harder to interpret. Thus, we smooth the data. In Appendix 1, we explain in detail our procedure.

We normalize the data on daily fatalities by the size of the population, and express them per million inhabitants.

4.3 Calibration of the recovery rate δ

This parameter depends on the chemical characteristics of the virus and their interaction with the human body. Medical evidence suggests that the average duration of the infection period ranges between four and 14 days. As our

benchmark, we use an average of 10 days, but we will also show results for values of five and 20 days. As the recovery rate is the inverse of the average duration, we set $\delta = 0.1$ for our benchmark (and will check robustness for $\delta = 0.2$ and $\delta = 0.05$).

4.4 The numerical exercise

We now show that it is still possible to learn from the different experiences in the three Nordic countries. We do so by fixing a value for the mortality rate, m, assumed to be the same across countries. This implies taking a stand on the single free parameter the model allows.

Once that parameter is chosen, we use the model to solve for the other parameters for each of the three countries, which allows for a comparison of the values for the evolution over time of the transmission rate, β_t. Clearly, the result does depend on the assumed value for the unknown mortality rate, m. But the exercise can be repeated for a range of reasonable values of the mortality rate, and thus a range of values for the mitigation rate will result.

The results are remarkably robust to the value of m and indicate a much stronger response of the mitigation efforts in Norway than those in Sweden; the results for Denmark suggest a mixed outcome, with an initial increase in the mitigation rate and a subsequent drastic reduction.

4.5 The algorithm

To make clear the nature of the exercise, we now summarize all the steps described so far that we apply to the data of the three countries:

1. We first smooth the data to obtain series of $\{I_t q\}_{t=0}^{T}$.

2. We choose $\delta = 0.1$, which is equivalent to an average duration of the infection for patients who recover in 10 days. (We also solve the cases of $\delta = 0.05$ and $\delta = 0.2$.)

3. We assume two possible values for m , 1% to 0.1%, a range that has been considered in the literature.[6]

4. Given m and δ, we compute $\{I_t\}_{t=0}^{T}$.

5. Given an initial value for R_0, the series $\{I_t, q\}_{t=0}^{T}$, and the assumed value for δ, we compute $\{R_t\}_{t=0}^{T}$.

6. Given $\{I_t, R_t\}_{t=0}^{T}$, we compute $\{S_t\}_{t=1}^{T}$, using the identity

$$S_t + I_t + R_t = 1.$$

7. We use the initial condition S_0 and $\{I_t, S_t\}_{t=1}^{T}$ on equation

$$S_t - S_{t+1} = \beta_t I_t S_t$$

[6] See Atkeson [2020]

to compute $\{\beta_t\}_{t=0}^{T}$.

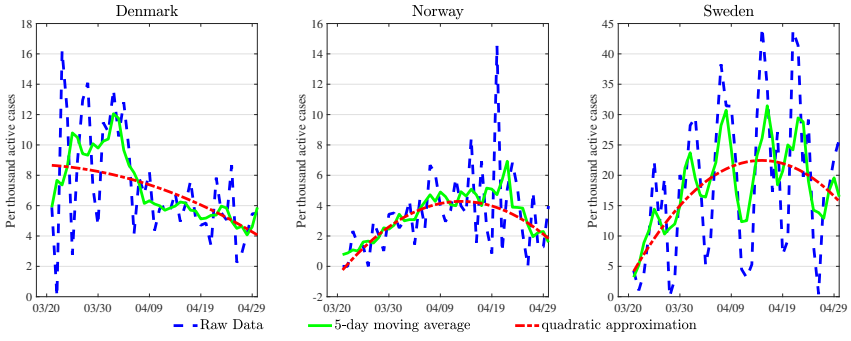
4.6 Results

We first discuss the case of the three Nordic countries, then go on to discuss the comparison across states.

4.6.1 Case 1: Three Nordic countries

In Figure 3, we show the evolution of the value for q_t^d for the three Nordic countries the observations were smoothed through a three-day equal-weight moving average process. This is the estimated value obtained when using $\delta = 0.1$. The data for the three countries start on March 18th and end on April 30th.

Figure 3: Fatality rates by country



As can be seen, the estimate is high for Sweden, intermediate for Denmark, and low for Norway. Given that the assumed true fatality rate is the same across countries, these results suggest that the fraction of undetected

agents is larger in Sweden than in Denmark, while that fraction is the smallest in Norway. Incidentally, this is consistent with the fact that by April 26th, the number of tests per million people was 30.310 in Norway, 26.900 in Denmark, and 9.357 in Sweden.

Besides those qualitative characteristics, in order to make quantitative progress, we feed these data into the model. As mentioned above, to avoid the large day-to-day fluctuations, for each country we will feed the model a value for $I_t^d q_t^d$ that is derived from the quadratic approximation we also present in Figure 3.

As mentioned above, our benchmark case is given by $\delta = 0.1$. We tried two alternative values, $q = 0.001$ and $q = 0.0001$, that correspond to two values for the fatality rate considered in the literature, of 0.1% and 1%.

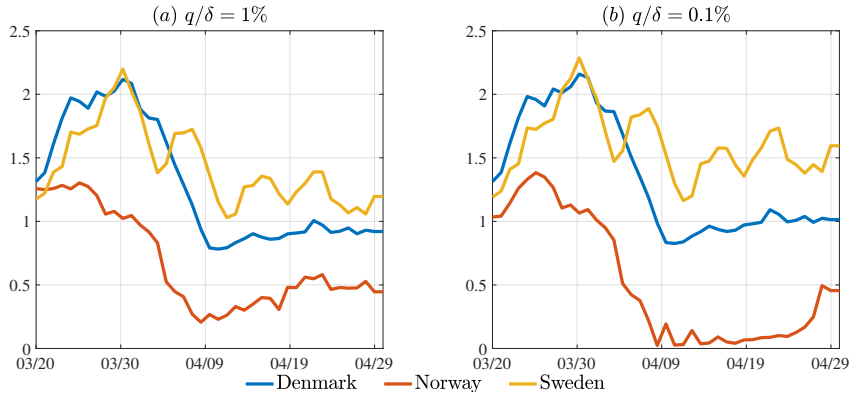
Our variable of interest is the transmission rate, β_t. That parameter determines the gross flows to the pool of infected agents. On the other hand, the parameter δ is the major determinant of the gross flows out.[7] Thus, what really matters is the size of β_t relative to the value of δ. In the following figures, we therefore plot the ratio β_t/δ, known as the normalized transmission rate.

The main results for the high mortality rate ($q = 0.001$) and low mortality rate ($q = 0.0001$) cases are reported in Figures 4a and 4b respectively. Several conclusions arise from the figures.

1. A very different pattern emerges between Norway and Sweden. For the low mortality rate case, by March 25th, Norway starts very close to a

[7]The fatality rate also matters, but it is very small relative to δ.

Figure 4: Normalized transmission rates for the case $\delta = 0.1$



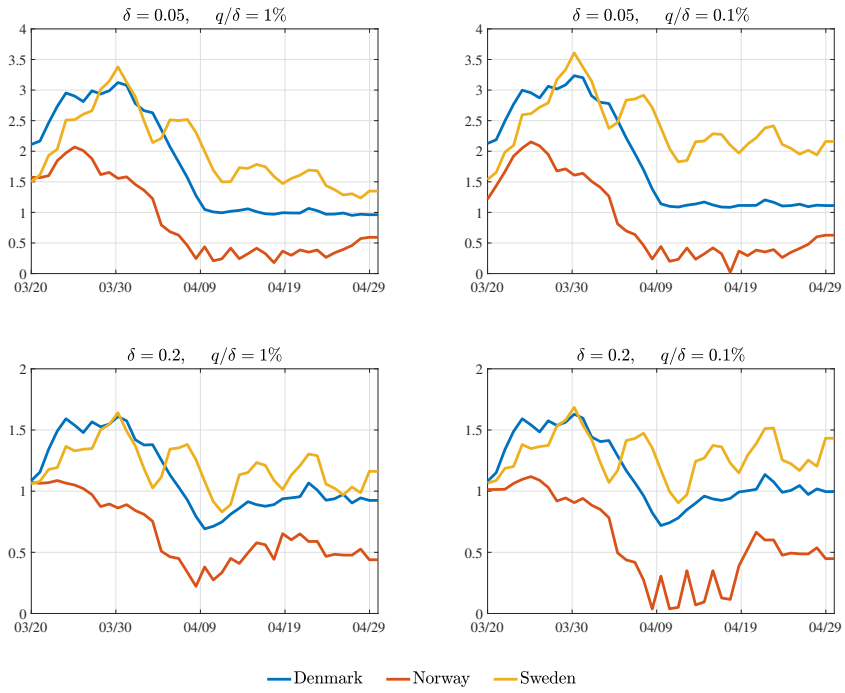
normalized transmission rate of 1 and climbs to a peak of around 1.3. It then starts a steady decline that crosses the value of 1 by April 1st, remaining substantially below that value from that point forward. By contrast, Sweden starts at around the same value as Norway, and by March 31st, climbs to peak over 2. It then starts a volatile decline, remaining substantially above Norway by over 0.5 and always above the threshold 1.

2. The case of Denmark is in between. It behaves like Sweden till around April 5th, but it then goes below the threshold 1 on April 9th, remaining slightly below it from that point forward.

3. The results are robust to a 10-fold difference in the mortality rate. All rates are slightly lower, except for Sweden, where the drop is more important. The difference between Norway and Sweden is even larger

in this case.[8]

Figure 5: Normalized transmission rates for cases $\delta = 0.05$ and $\delta = 0.2$



In Figure 5, we reproduce the results using alternative values for the recovery rate. Specifically, in figures on the first row, we solve the case of $\delta = 0.05$, while in figures on the second row, we solve the case of $\delta = 0.2$. As it can be seen in Figure 5, the results are quite robust to these alternative assumptions.

Figure 5 hides another difference that follows trivially but runs in favor

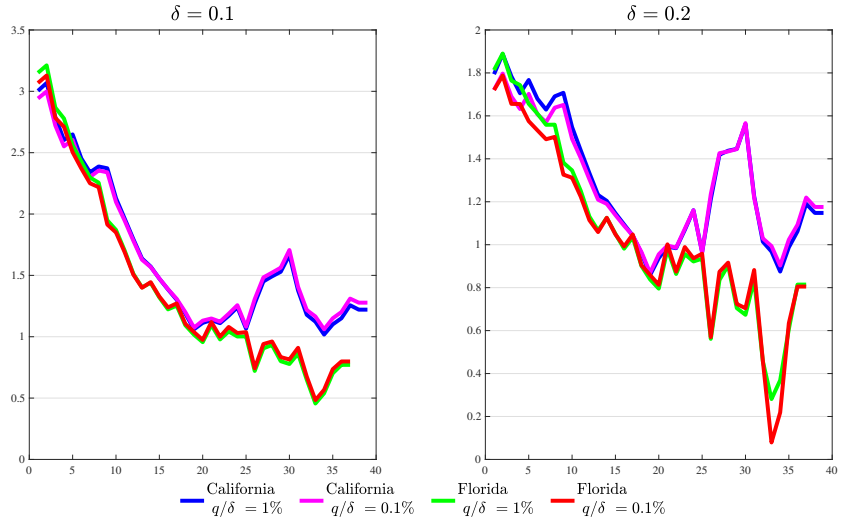
[8]There is a spike at the end of the period in Norway. This is because the q_t^d estimated is very close to q, which according to the model, implies a large reduction in I_t^n. This probably reflects the fact that the assumed m may be too large, or that the q_t^d measured on those days may be artifically low, owing to sample size.

of Sweden. The share of susceptible agents, S_t, is now lower in Sweden, so that country is closer to herd immunity.

4.6.2 Case 2: Several US states

We now illustrate the methodology for some US states. The data source is the New York Times[9]. For each state, we chose the initial day to be the one when it reaches 10 fatalities per million inhabitants. In all cases, we use two values for the recovery rate, $\delta = 0.2$ and $\delta = 0.1$. In addition, we also use the same values for q that we used for the Nordic countries, corresponding to values for the fatality rate $m = 1\%$ and $m = 0.1\%$.

Figure 6: California and Florida



In Figure 6, we show the results for California and Florida. The left panel shows the case of $\delta = 0.1$, while the right panel shows the results for $\delta = 0.2$.

[9]Source: `https://github.com/nytimes/covid-19-data`.

The x-axis shows the days since the initial date of observation. Both states initially behave similarly, starting at high values, but with converging to the value of 1 in about 20 days. They then hover around the value, with some spikes that appear temporary. The results are very robust to the value of delta.

Figure 7: Georgia and Washington

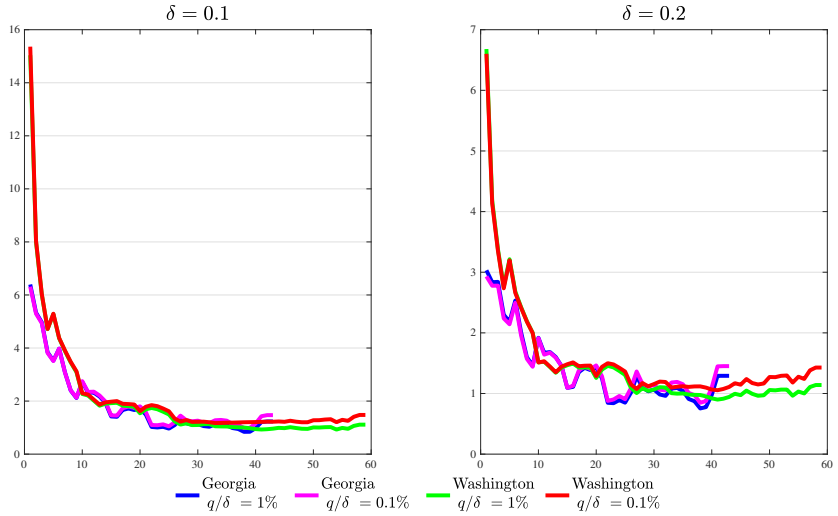
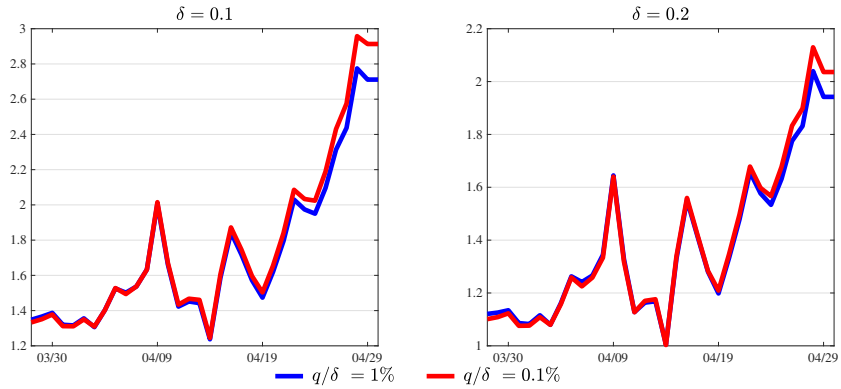


Figure 7 shows the results for Georgia and Washington. They start with values for the transmission rate that are much higher than the previous cases, but they also rapidly converge to values very close to 1. In both cases, by the last days of April (the end of our sample), there seems to be a small increase in transmission. Again, the results are quite robust to the values assumed for δ or m.

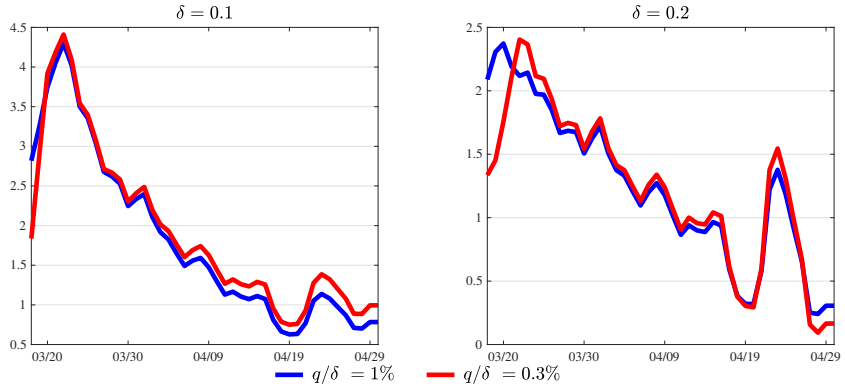
Figure 8 shows the case of Minnesota, which is quite different from the cases that we solved. It starts with normalized transmission rates that are

Figure 8: Minnesota



very close to 1, but after the third week, it exhibits an increasing trend that appears worrisome. This behavior is very robust to alternative values for δ or m.

Figure 9: New York



Finally, Figure 9 shows the case of New York. This is also a special case. From the point of view of our analysis, the main reason is that the number of fatalities relative to the population is above 0.12% by May the 3rd. Therefore, imposing a fatality rate relative to the infected population

of 0.1% makes no sense in this case. Thus, for New York, we used $m = 1\%$ - as before - but we chose as an alternative value $m = 0.3\%$. The general behavior is similar to the other cases: it starts high, but it then converges to 1. However, compared with other states - notice, for instance, the difference with Florida and California in Figure 6 - it starts at higher values and it takes longer to converge (25 or 30 days, depending on the case).[10] This is consistent with the events that we saw unfold in New York.

5 Explaining the previous results

In this section, we show that the computation of the transmission rate using data on fatalities is quite robust to the assumed value for the true fatality rate. In the limit in which the size of the susceptible agents is arbitrarily close to 100%, the estimate of the normalized transmission rate is independent of the value for m. In order to do so, we use the model in which the detected and the undetected agents are grouped together. Thus, this version of the model implies that given

1. initial conditions $\{S_0, I_0, R_0\}$, where all are ≥ 0 and such that

$$S_0 + I_0 + R_0 = 1,$$

[10]The comparison with Washington and Georgia may seem puzzling, given the very high estimated cases in the early days. We have not made a detailed comparison, but the estimates for the early days are less reliable, given that samples are smaller for those days, so we did not pursue this further.

2. and parameters δ, q, β_t,

the following equations solve for the sequences $\{S_t, I_t, R_t\}_{t=1}^{\infty}$:

$$S_t - S_{t+1} = \beta_t I_t S_t$$

$$I_{t+1} - I_t = \beta_t I_t S_t - \delta I_t - q I_t \tag{11}$$

$$R_{t+1} = R_t + \delta I_t + q I_t.$$

We have data on $F_t = qI_t$ for $t = 0, 1, ..., T$, and we know δ. Assume $q = \widetilde{q}$. Then, our estimate for $\widetilde{I}_t$ is

$$\widetilde{I}_t = \frac{F_t}{\widetilde{q}} \text{ for } t = 0, 1, ..., T.$$

We assume that $R_0 = 0$, so

$$\widetilde{R}_t = (\delta + \widetilde{q}) \sum_{j=0}^{t} \widetilde{I}_j.$$

In addition,

$$\widetilde{S}_t = 1 - \widetilde{I}_t - \widetilde{R}_t,$$

and

$$\widetilde{S}_t - \widetilde{S}_{t+1} = \widetilde{\beta}_t \widetilde{I}_t \widetilde{S}_t.$$

Now, let $q = k\widetilde{q}$ for $k > 1$. Then, the evolution of infections is given by

$$\widehat{I}_t = \frac{F_t}{k\widetilde{q}} = \frac{\widetilde{I}_t}{k} \quad \text{for } t = 0, 1, ..., T.$$

A known property of this model - that the growth rate of infections is independent of the fatality rate - follows:

$$\frac{\widehat{I}_{t+1} - \widehat{I}_t}{\widehat{I}_t} = \frac{\frac{\widetilde{I}_{t+1}}{k} - \frac{\widetilde{I}_t}{k}}{\frac{\widetilde{I}_t}{k}} = \frac{\widetilde{I}_{t+1} - \widetilde{I}_t}{\widetilde{I}_t}.$$

In addition, equation (11) implies

$$\frac{\widehat{I}_{t+1} - \widehat{I}_t}{\widehat{I}_t} = \widehat{\beta}_t \widehat{S}_t - \delta - \widehat{q},$$

and

$$\frac{\widetilde{I}_{t+1} - \widetilde{I}_t}{\widetilde{I}_t} = \widetilde{\beta}_t \widetilde{S}_t - \delta - \widetilde{q}.$$

so,

$$\widehat{\beta}_t \frac{\widehat{S}_t}{\widetilde{S}_t} + \frac{\widetilde{q}(1-k)}{\widetilde{S}_t} = \widetilde{\beta}_t. \tag{12}$$

The solution for $\widehat{R}_t$ is given by

$$\widehat{R}_t = (\delta + \widetilde{q}k) \sum_{j=0}^{t} \widehat{I}_j = (\delta + \widetilde{q}k) \sum_{j=0}^{t} \frac{\widetilde{I}_t}{k},$$

which implies

$$\widehat{R}_t = \left(\frac{\delta + k\widetilde{q}}{k\delta + k\widetilde{q}}\right)\widetilde{R}_t.$$

The solution for the set of susceptible agents is then given by

$$\widehat{S}_t = 1 - \widehat{I}_t - \widehat{R}_t = 1 - \frac{\widetilde{I}_t}{k} - \left(\frac{\delta + k\widetilde{q}}{\delta + \widetilde{q}}\right)\frac{\widetilde{R}_t}{k}.$$

Thus,

$$\frac{\widehat{S}_t}{\widetilde{S}_t} = \frac{1}{k}\frac{k - 1 + \widetilde{S}_t + \widetilde{R}_t\frac{\widetilde{q}(1-k)}{\delta+\widetilde{q}}}{\widetilde{S}_t}.$$

Replacing in equation (12) above

$$\widehat{\beta}_t\frac{\widehat{S}_t}{\widetilde{S}_t} + \frac{\widetilde{q}(1-k)}{\widetilde{S}_t} = \widetilde{\beta}_t$$

we obtain

$$\widehat{\beta}_t\frac{k - 1 + \widetilde{S}_t}{k\widetilde{S}_t} + \frac{(1-k)}{k}\frac{\widetilde{R}_t}{\widetilde{S}_t}\frac{\widetilde{q}}{\delta + \widetilde{q}} - \frac{\widetilde{q}(k-1)}{\widetilde{S}_t} = \widetilde{\beta}_t$$

When $\widetilde{S}_t$ is close to 1 so $\widetilde{R}_t$ is close to zero, we obtain

$$\widehat{\beta}_t - \widetilde{q}(k-1) \simeq \widetilde{\beta}_t.$$

So, the difference between the two values for the parameters we used is just 0.0009. As the function is continuous, the estimate is quite insensitive to k when $\widetilde{S}_t$ is close to 1. Even for values of $\widetilde{S}_t = 0.9$, the difference between

the two estimates is less than10%.

Thus, in the early stages of the epidemic, one can obtain very precise estimates of the transmission rates. However, as the epidemic progresses, or in second waves of the epidemic when R is large, identifying the evolution of β_t requires more precise information regarding the fatality rate. This explains the behavior of the estimates in figures 4 to 9, since the data analyzed correspond to the outburst of the epidemic.

References

Andrew Atkeson. What will be the economic impact of COVID-19 in the U.S. ? – Rough estimates of disease scenarios. *Federal Reserve Bank of Minneapolis Staff Report*, 595, 2020.