

Theoretical Framework of an Econometric Analysis of Germany's Corona-Data

by Georg Quaas (April-04-2020, slightly revised April-30-2020)

The classic epidemic SIR-model

Based on daily data, the classical epidemiological model (CEM) is given by the following equations (Hethcote 2000: 604):

$$(1) \Delta S = -\beta IS/N$$

$$(2) \Delta I = \beta IS/N - \Delta R$$

$$(3) \Delta R = \gamma I$$

S , I and R are the numbers for the size of three overlap-free classes of the population, S is the class of people susceptible to the virus (at the beginning of the disease and without information about the number of immune people it is identical to the number of the population), I is the class of infectious people (its initial value is nearby zero, for instance 1) and R is the class of cured people with immunity (this number is set to zero for the first 10 days of the explorations).

The logic behind this model becomes clear when defining the proportions of the two population groups with and without active pathogens in the total population: $i = I/N$ and $s = S/N$. These proportions can be interpreted as the probability of accidentally encountering an infectious agent or a susceptible agent in the population. The probability that these persons meet has the probability $i \cdot s$ (Hamer 1906). The model assumes that the frequency of infections increases with the size of the population, so that, except for β , the mean term of equation (2) results. We abbreviate this term as H :

$$(4) \quad isN = \frac{ISN}{N^2} = \frac{IS}{N} = H$$

From a practical point of view, H can be interpreted as the number of possibilities to become infected in a population of the extent N , it is, so to speak, the abstract risk situation. The parameter β reflects the actual infection process with regard to how many people are infected by an infectious agent per time unit (here: per day) on average. This depends on several factors that are not explicitly included in the model, such as population density, the number of daily interactions, and common behaviors (hygiene, shaking hands, etc.). Some of these factors can be influenced pragmatically so that policymakers have access to the ongoing process. However, the effect on the parameter β is recorded with a time delay. In the case of the novel coronavirus, the media initially assumed that each infected person infects three more people during his or her infectious phase. The RKI reports a basic reproduction figure R_0 between 2 and 2.5 in its "coronavirus profile." Initial experience suggests that the infectious phase lasts for about 10 days, so that an average can be set as $\beta = 0.3$, that is, as long as more accurate estimates are unavailable.

There is a fourth variable that does not play an explicit role in the above-formulated classical model, and that is the number of deaths, D . It is often supposed (for instance, an der Heiden, M; Buchholz U 2020: p.1) that the number of deaths is included in the number of recovered. This sounds a bit cynical – but from a statistical perspective, another aspect is important: If the number of deaths is empirically available, the model should be supplemented with it. Both the number of deaths and the

number of recoveries reduce the number of infectious cases; therefore, equation (2) can be more precise as:

$$(2a) \Delta I = \beta IS/N - \Delta R - \Delta D$$

The number of recoveries is determined by equation (3) if not previously estimated by the data producer. For the number of new reported deaths, the above assumption applies:

$$(4) \Delta D = \delta I.$$

Parameter δ measures the current number of newly reported deaths based on the number of infected persons at an earlier time of infection.

There is a similarity of equation (2a) to the corresponding equation of the classic endemic model with vital dynamics that includes births and deaths, but no disease related deaths (Hethcote 2000: 607). According to equation (1), the class of susceptible persons is reduced by the number of new infections of which the disease related deaths are a part. Therefore, one must not change this equation by the help of (4). But the number of deaths reduces in short-term the number of the population. Therefore, in the case of a short-term event of an outbreak of a disease with expectedly high a number of victims, it is useful to consider the population as variable by adding the number of deaths and treating it as time-dependent, like all other variables:

$$(5) S(t) + I(t) + R(t) = N - D(t) = N'(t)$$

This is a second difference to the classic endemic model.

When adapting the model to the data provided by the Robert Koch Institute, it is mandatory to add time delays and time limits to individual variables according to the characteristics of the underlying disease. For example, the number of victims refers to a point in time perhaps 2, 3 or even 4 weeks ago. If and only if it refers to the same point in time, the following applies for logical reasons:

$$\delta = (1 - \gamma)$$

However, since convalescence more often happens and is registered earlier than death, the time-lags and consequently the parameters δ and γ are independent of each other.

The determination of time-lags is mainly an empirical problem of virologists. In the case of COVID-19, not all features with which the model is to be equipped are known. For the unknown parameters, plausible assumptions must be made. Particularly problematic is the duration of infectivity, which is seemingly difficult to determine. It should also be noted that the accumulated number of infected persons reported by the RKI (better: the number of persons reported as infected by the health authorities) does not correspond to the number of infectious persons $I(t)$ which plays a role in the classical model.

The Contact Rate

The core mechanism in this model is the transmission of the virus from one person to another. It is grasp by the contact rate: "If β is the average number of adequate contacts (i.e., contacts sufficient for transmission) of a person per unit time, then $\beta I/N = \beta i$ is the average number of contacts with infectives per unit time of one susceptible, and $(\beta I/N)S = \beta N i s$ is the number of new cases per unit time due to the $S = N s$ susceptibles." (Hethcote 2000: 602)

If we multiply the Contact rate β by the exposure time ($1/\gamma$ = the average infectious period acc. to Hethcote, p. 605), we get the total number of persons R (the replacement number) that are on average infected by an infective that is infectious. "The replacement number R is the actual number of secondary cases from a typical infective, so that after the infection has invaded a population and everyone is no longer susceptible, R is always less than the basic reproduction number R_0 ." (ibid. 604) After a while an infective is (i) recovered or (ii) dead. In both cases he or she is not infectious anymore. It is assumed in the case of COVID-19 that recovered people are immune, at least for a while.

The Fatality Rate

According to China's Emergency Response Epidemiology Team (2020) the case fatality rate is 2.3 percent of the confirmed cases, and the duration of the wave (from beginning of the massive spread till the putative end) is estimated by 15 days on average (p.115).

The Response Team treats the first wave of the epidemic spread as terminated. With this supposition, the fatality rate and the duration of the disease can be defined as follows: "Case fatality rates were calculated as the total number of deaths (numerator) divided by the total number of cases (denominator), expressed as a percent. Observed time was summarized using person-days (PD) and mortality was calculated as the number of deaths (numerator) divided by the total observed time (denominator), expressed per 10 PD." (114-115)

In a dynamic situation the assessment of the fatality rate is more complicated, because the deadly event happens X days after the illness was diagnosed. Dividing the actual number of deaths by the actual number of people infected leads to a misleading underestimation of the death rate at the beginning of an epidemic. It follows that the denominator should be the number of infected persons observed X days before.

The Reproduction number

See study 1 on this website: www.forschungsseminar.de

Literature

an der Heiden, M; Buchholz U (2020): Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland. | DOI 10.25646/6571.2

Hethcote, Herbert W. (2000): The Mathematics of Infectious Diseases, in: SIAM Review, Vol. 42, No. 4. (Dec., 2000), pp. 599-653.

[Daily] RKI-report (01.04.2020).

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020): The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020, China CDC Weekly, Vol. 2, No. 8, pp. 113-122.

Data

Robert Koch-Institut: [Tägliche] Fallzahlen in Deutschland... [from 2020-03-04 to ...]