Estimating a breakpoint in the pattern of spread of COVID-19 in South Korea

Young-Joo Kim[a,1], Myung Hwan Seo[b,1], Hyun-E Yeom[c,*]

[a] Department of Economics, Hongik University, Wausanro 94, Mapo-gu, Seoul 04066, South Korea
[b] Department of Economics, Seoul National University, Gwanak Ro 1, Gwanak Gu, Seoul 08826, South Korea
[c] Department of Nursing, Chungnam National University, Munhwaro 266, Jung-gu, Daejeon 35075, South Korea

A B S T R A C T

Objectives: Amid the global coronavirus disease 2019 (COVID-19) crisis, South Korea has been lauded for successfully preventing the spread of this infectious disease, which may be due to the aggressive implementation of preventive policies. This study was performed to evaluate the pattern of spread of COVID-19 in South Korea considering the potential impact of policy interventions on transmission rates.
Methods: A SIR (susceptible–infected–removed) model with a breakpoint that allows a change in transmission rate at an unknown point was established. Estimated trajectories of COVID-19 from SIR models with and without a breakpoint were compared.
Results: The proposed model with a break fitted the actual series of infection cases much better than the classic model. The estimated breakpoint was March 7, 2020 and the transmission rate dropped by 0.23 after the breakpoint. A counterfactual study based on our estimate indicated that the number of infected could have reached 2 500 000 compared to the peak of 8000 in the observed series.
Conclusions: It is critical to consider a change in the transmission rate to evaluate the trajectory of spread of COVID-19 in South Korea. Our estimation and counterfactual experiments indicate that public health interventions may play a role in determining the pattern of spread of infectious diseases.

Introduction

The highly contagious coronavirus disease 2019 (COVID-19) is causing public health emergencies worldwide and a crisis in global health security. Unlike many other countries that are experiencing ongoing and continuously increasing numbers of infected cases (i.e., high transmission rates), South Korea had an extremely short growth phase and quickly transitioned to a decrease in growth (i.e., low transmission rates).

In order to prevent the spread of infectious diseases, it is critically important for governments to make rapid decisions for the development of coping strategies and to apply them systematically based on detailed monitoring and empirical evidence. The Korea Centers for Disease Control and Prevention

(KCDC) immediately implemented an aggressive tracing procedure for infected individuals and widespread non-selective testing for susceptible populations. Concurrently, the KCDC also initiated a series of public health interventions across the country including the wearing of face masks, social distancing, a voluntary plus mandated quarantine, business shutdowns, and school closures. These policies may have played a key role in successfully preventing a pandemic spread of COVID-19.

The SIR (susceptible–infected–removed) model is widely used in epidemiological research to predict how an infectious disease will spread over time after an outbreak. In the SIR model, the transmission rate is one of the key parameters that determine the evolutionary process of an epidemic; however, it is considered to be constant over time in the SIR model (Kermack and McKendrick, 1927). Although the important role of public health interventions in reducing the transmission rate has been discussed in the literature (Cauchemez et al., 2009; Ferguson et al., 2006), empirical evidence to support this notion is scarce due to the limitation of the classic SIR model that assumes a single transmission rate. In the real world, the transmission rate of an infectious disease could differ depending

* Corresponding author.
  E-mail addresses: y.j.kim@hongik.ac.kr (Y.-J. Kim), myunghseo@snu.ac.kr (M.H. Seo), yeom@cnu.ac.kr (H.-E. Yeom).
[1] Young-Joo Kim and Myung Hwan Seo contributed equally to this work as co-first authors.

on various events such as strong policy interventions or a novel drug discovery. To address this issue, we extended the classic SIR model by incorporating a change in transmission rate at an unknown breakpoint. This study was performed to examine the trajectory of spread of COVID-19 in South Korea and whether it changed at a breakpoint that was possibly related to a policy intervention. A comparison was made of the accuracy of the SIR models, both with and without a breakpoint, in estimating the pattern of spread of COVID-19. Furthermore, through counterfactual experiments, this study highlights the potential contributions of public health interventions to mitigating the spread in South Korea.

Methods

Data source

Data published by the KCDC (Korea Centers for Disease Control and Prevention, 2020) on the daily number of COVID-19 infected and recovered cases from February 2 to April 15, 2020 were analyzed. The first case of COVID-19 in South Korea was reported on January 23, 2020 and the number of infected cases was not released on a daily basis until February 2.

Preliminary estimation using the classic SIR model

The SIR model has been proposed as a continuous time model (Kermack and McKendrick, 1927), which we discretize to fit the daily observations as follows:

$$\Delta R_t = \gamma I_{t-1}$$

$$\Delta I_t = \beta \frac{I_{t-1} S_{t-1}}{N} - \gamma I_{t-1},$$

where S_t, R_t, I_t are the number of susceptible, recovered, and infected cases, respectively, at time t, and $N = S_t + R_t + I_t$ is the total population. By construction, $\Delta S_t + \Delta R_t + \Delta I_t = 0$. The framework of the basic SIR model is built upon simplifying assumptions, including a closed population and no reinfection, such that individuals who recover from infection are no longer susceptible (Weiss, 2013). The parameters β and γ are the case transmission rate and recovery rate, respectively. The rate of transition from the susceptible state to the infected state is the transmission rate, and the rate of transition from the infected state to the recovered state is the recovery rate. The model implies that the two parameters determine the epidemic dynamics of an infectious disease and that policy interventions operate through manipulation of the values of the parameters. We first applied the standard SIR model to assess how well it explained the actual pattern of spread of infected cases.

The true value of the recovery rate is not identifiable with only the infected series and is difficult to estimate reliably, without a random experiment, when the number of infected is small. An approximate might be a ratio between the numbers of recovered and infected, only after the latter has grown significantly. Trimming the first 49 observations of our time-series, we computed the ratio as 0.51 and set as the recovery rate and estimated β throughout the paper. Specifically, the parameter estimate was obtained by minimizing the sum of squared error (SSE) $\sum_{t=1}^{n} (I_t - I_t(\beta, \gamma))^2$ with respect to β, where I_t is the actual number of infected at time t, and $I_t(\beta, \gamma)$ is the value predicted by the SIR model. We also considered other recovery rates in the range of 0.12–0.58.

Advanced estimation by the SIR model incorporating a breakpoint

The transmission rate of a contagious disease may not be constant over time once various policy interventions have been introduced to mitigate an epidemic. Evidence for a reduced rate of transmission resulting from public health policies over the course of an epidemic has been discussed previously; for example, see Cauchemez et al. (2009), Chladná et al. (2020), and Anderson et al. (2020). When evaluating mitigation strategies, it is difficult to set (i.e., fix) the date of the change in the transmission rate a priori, due to lags in policy effects and/or a mix of different policies. Considering these features, we introduced a structural break in the transmission rate as done by Kass-Hout et al. (2012) to explain daily emergency room visits due to an influenza-like illness. However, the recovery rate may not change quickly, as new treatment methods typically do not appear during the early period of an epidemic.

In this regard, we fixed γ at 0.51 as in the preceding section and introduced a change in the transmission rate at an unknown time to the SIR model by replacing β with $\beta_t(t_0) = \beta_1 1\{t \le t_0\} + \beta_2 1\{t > t_0\}$ and estimated both the breakpoint date, t_0, and before-and-after transmission rates, β_1 and β_2, respectively, from the time-series of the number of infected. Specifically, we minimized SSE $\sum_{t=1}^{n} (I_t - I_t(\beta_t(t_0), \gamma))^2$. Although previous studies, such as that by Liu and Stechlinski (2011), have introduced a time-varying β into the SIR model to accommodate a known time-varying feature like seasonality in infection, none have estimated an unknown break date.

After estimating the parameters by the minimum distance principle, two counterfactual analyses were conducted. In the first, we estimated the trajectory of the number of infected when there was no policy implemented to reduce the transmission rate and thus there was no break in transmission rate. In the second, we estimated the trajectory of the number of infected when policies were delayed by 1 week, i.e., when the break occurred 1 week later than the estimated breakpoint date. These counterfactual experiments highlighted the effect of active policies.

Results

Pattern of spread based on the classic SIR model

We first applied the formula proposed by the classic SIR model to the actual data observed between February 2 and April 15. Using the time-series of the number of cases and the SIR model, we estimated the transmission rate β to be 0.6, and from this
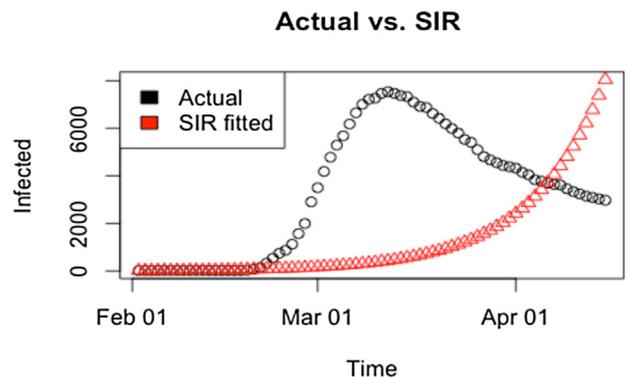

Figure 1. Comparison of actual COVID-19 cases (in black) and the cases predicted (in red) by the classic SIR model. Time is represented along the x-axis and the number of infected cases is represented along the y-axis.

estimated rate we predicted the number of cases. In Figure 1, the black line shows the number of cases observed and the red line shows the number of cases predicted by the SIR model with an estimated transmission rate of 0.6 and $\gamma = 0.51$.

From Figure 1, we see that the trend in actual infected cases is clearly different from the trend predicted by the classic SIR model. Specifically, the actual number of infected cases began to increase sharply in late February; this then quickly reversed and began slowing down after reaching a peak in early March. However, the number of infected cases predicted by the classic SIR model gradually increased throughout March, with its predicted trajectory well below the trajectory of observed cases. Starting in early April, the predicted trajectory began to increase swiftly, and by mid-April, the SIR model trajectory predicted more cases than were actually observed. Thus from mid-April, the two trajectories headed in opposite directions, consequently leading to large prediction errors. These findings indicate that the classic SIR model was not able to explain the trend in COVID-19 over time in South Korea. As noted previously, this is likely because the classic SIR model does not take into account the impact of events that may affect the transmission rate of infection, such as public health interventions.

Advanced estimation using the SIR model with an unknown break

As an alternative to the classic SIR model with a single constant transmission rate, we extended the SIR model to allow multiple transmission rates with a breakpoint. Using this approach, it was found that the transmission rate changed from 0.71 to 0.48 at the estimated breakpoint date of March 7, 2020. Figure 2 shows this predicted trajectory (in blue) with the observed data (in black).

Compared to the classic SIR model, the fit of the model improved dramatically when including a breakpoint and allowing for multiple transmission rates (as seen in Figure 2). The observed and predicted trajectories move tightly together over the entire time period, with both peaking in the first week of March. In the first phase, with a high transmission rate of 0.71, the growth of contagion accelerated quickly until the estimated breakpoint of March 7. After this point, the contagion process shifted into a new phase with a lower transmission rate of 0.48 and growth slowed substantially. In both phases, the suggested model was successful in mirroring the epidemic trajectory of COVID-19 in South Korea.

As noted earlier, the breakpoint is estimated by minimizing the SSE, the error referring to the prediction error derived from the difference between the observed number of cases and the
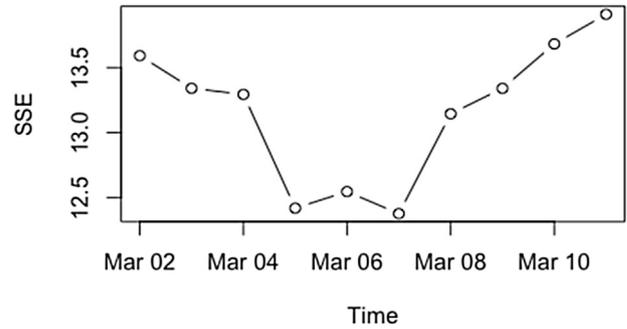

Profiled SSE along break dates

Figure 3. Profiled sum of squared error (SSE) along candidate break dates by the SIR model with a break. Time is represented along the *x*-axis and SSE along the *y*-axis.

predicted number of cases based on the model. For a graphical illustration of the estimated breakpoint, we plotted the profile of the SSE around the timeframe of the estimated break. As shown in Figure 3, the SSE drops between March 5 and March 7 and reaches the minimum on March 7, confirming that the change in transmission rate, in terms of breakpoint, occurred on March 7. This was day 44 following the first case of COVID-19 in South Korea.

As a further assessment of the robustness of the SIR model with a break, we examined whether our findings were sensitive to recovery rates other than 0.51. Considering recovery rate values in the range of 0.12 to 0.58, we consistently found March 7 to be the estimated breakpoint where the transmission rate changes and the same magnitude of change (-0.23) in the transmission rate.

Counterfactual experiments on the dynamics of spread of COVID-19

Situation in the absence of a breakpoint

Using the estimated model parameters β_1 (the transmission rate) and t_0 (the breakpoint), we predicted the dynamics of the epidemic under different policies. First, in the scenario in which mitigation policies were not introduced or the imposed mitigation measures were ineffective, the initial transmission rate remained constant until the end of the period. In Figure 4, the red line represents the predicted number of cases based on the initial transmission rate of 0.71. For ease of comparison, the black line is plotted to indicate the observed number of cases. Under the initial
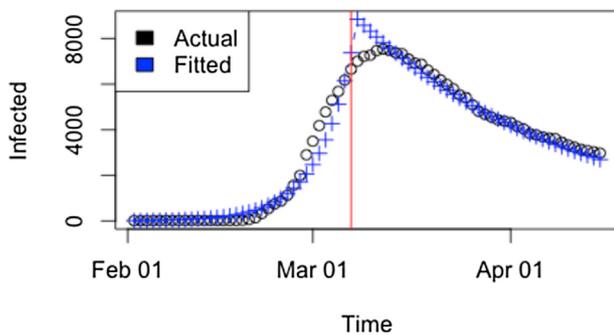

Figure 2. Comparison of actual COVID-19 cases (in black) and the cases predicted (in blue) by the SIR model with a break. Time is represented along the *x*-axis and the number of infected cases is represented along the *y*-axis. The vertical red line indicates the breakpoint (March 7, 2020) where there was a change in the rate of transmission.
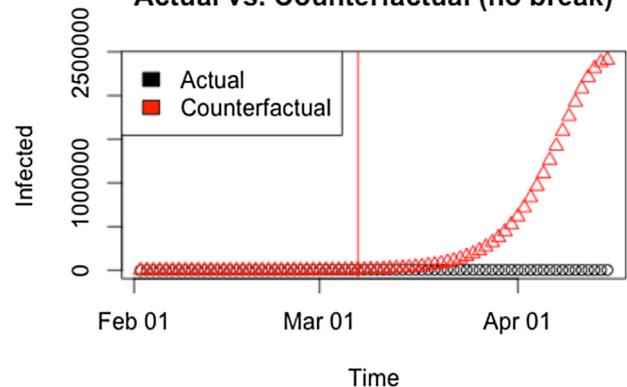

Figure 4. Comparison of COVID-19 cases (in black) and cases predicted (in red) by the SIR model if the initial transmission rate β_1 was maintained without the break. Time is represented along the *x*-axis and the number of infected cases is represented along the *y*-axis. The vertical red line indicates the breakpoint (March 7, 2020).
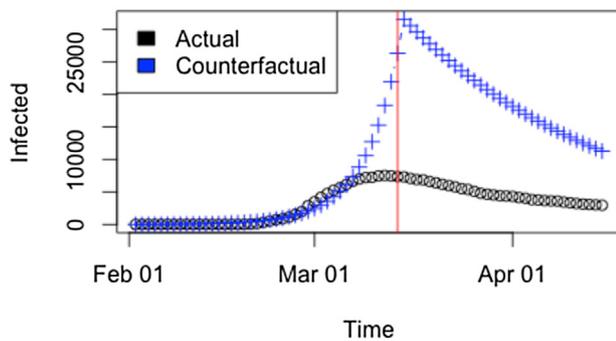
Actual vs. Counterfactual (late break)

Figure 5. Comparison of COVID-19 cases (in black) and cases predicted (in blue) by the SIR model with a break, with a 1-week delay in the break. Time is represented along the x-axis and the number of infected cases is represented along the y-axis. The vertical red line indicates the breakpoint (March 14, 2020) where there was a change in the rate of transmission.

transmission rate without a break, the number of cases was predicted to increase up to 2 500 000 at its peak.

Situation for a delayed breakpoint

We further explored policy impacts on the dynamics of contagion by considering a 1-week delay in policy interventions. Under this scenario, the transmission rate was reduced from 0.71 to 0.48 at a later date than the estimated breakpoint of March 7; specifically, the breakpoint was moved to 1 week later. Figure 5 shows that with a 1-week delay in the change of the transmission rate, the number of infected cases continues to increase until March 14 and reaches a maximum of 30 000, which is more than three times the actual number observed at the peak.

Discussion

With the outbreak of COVID-19 that continues to spread around the world, South Korea has drawn international attention for two unique features. One, the pattern of spread of COVID-19 in South Korea is notably different from what has been observed in most other countries. Two, the government implemented systematic policy interventions and mitigation measures including the provision of nationwide testing through the tracing of infected individuals from very early on.

It was found that the classic SIR model was not able to explain the trend in COVID-19 over time in South Korea. As noted previously, this is likely due to the fact that the classic SIR model does not take into account the impact of events that could affect the transmission rate of infection, such as policy interventions. Thus, we estimated the trajectory of COVID-19 cases using a SIR model that allows for different transmission rates at an unknown breakpoint based on the assumption that the unique pattern of spread in Korea was related to policy interventions. It was found that the pattern of spread predicted by the proposed model better matches the actual trend of the epidemic than the one predicted by the classic SIR model. In addition, the findings based on various prefixed recovery rates supported that the estimated breakpoint was robust, occurring around the first week of March. The findings demonstrate that this SIR model equipped with multiple transmission rates and a breakpoint is a simple yet powerful tool for predicting the trajectory of a seemingly unpredictable novel infectious disease.

One of the important contributions of this study is the estimation of an unknown breakpoint at which the transmission rate changes. This will allow us to identify effective strategies

among the measures implemented over the course of an epidemic. In early February, before the estimated breakpoint of March 7, various policy interventions were initiated through coordinating governmental agencies (Ministry of Health and Welfare, 2020). For example, the daily testing capacity was expanded from 762 to 20 075 between February 7 and March 2. The distribution of publicly funded face masks started at designated shops and centers in Daegu and Gyeongbuk on February 27, expanding to all regions of the country as of February 28. The provision of various public services was suspended as a means of social distancing, including local public libraries from February 7 and national libraries and museums from February 24. As of February 27, public kindergartens were closed, and the start of the school year was delayed from March 2 through April for students in all grade levels. Taken together, the extent of social distancing and mitigation measures was expanded in the last week of February, approximately 2 weeks ahead of the estimated breakpoint of March 7.

Governments worldwide play an active role in attempting to control contagion and minimize health costs. The findings of this study shed light on the significance of prompt decision-making and systematic implementation of appropriate policy interventions in regulating transmission rates and mitigating spread during an epidemic, particularly of a highly contagious infectious disease such as COVID-19. The experimental simulations in this study indicate that in the absence of immediate preventive public strategies, the transmission rate in South Korea would have been much higher, comparable to the levels seen in other countries. Considering that the early stage of the epidemic began in the metropolitan city of Daegu and the surrounding area Gyeongbuk, which has a population of over 5 million people, reaching that predicted number in the absence of any interventions is entirely conceivable. Given that the realized incidence of cases at the peak was 8000, the estimation indicated that a decrease in the transmission rate by 0.23 had a tremendous impact on the trajectory of this infectious disease.

The study findings substantiate the important role of policies and public health measures in limiting the spread of an infectious disease as discussed in the recent literature on COVID-19. For example, Ferguson et al. (2020) and Marais and Sorrell (2020) suggested combined interventions such as rapid case-finding with widespread testing in conjunction with strict quarantine and population-wide social distancing as the most effective strategies. Given the estimated breakpoint, further estimation that exploits more detailed information on specific policy intervention strategies is warranted to verify factors contributing to the successful mitigation of the epidemic spread of COVID-19 in South Korea. We also suggest an estimation of the pattern of spread of COVID-19 in other countries based on our SIR model with a breakpoint, in order to build up robust evidence for the estimation accuracy.

Funding source

Ethical approval

Approval was not required.

Conflict of interest

The authors declare no conflict of interest.

References

Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. Lancet 2020;395:931–4, doi:http://dx.doi.org/10.1016/S0140-6736(20)30567-5.

Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, et al. Closure of schools during an influenza pandemic. Lancet Infect Dis 2009;9:473–81, doi:http://dx.doi.org/10.1016/S1473-3099(09)70176-8.

Chladná Z, Kopfová J, Rachinskii D, Rouf SC. Global dynamics of SIR model with switched transmission rate. J Math Biol 2020;80:1209–33, doi:http://dx.doi.org/10.1007/s00285-019-01460-2.

Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature 2006;442(July (7101)):448–52, doi:http://dx.doi.org/10.1038/nature04795.

Ferguson N, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. , doi:http://dx.doi.org/10.25561/77482 16 March 2020.

Kass-Hout TA, Xu Z, McMurray P, Park S, Buckeridge DL, Brownstein JS, et al. Application of change point analysis to daily influenza-like illness emergency department visits. J Am Med Inform Assoc 2012;19:1075–81, doi:http://dx.doi.org/10.1136/amiajnl-2011-000793.

Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A Math Phys Sci 1927;115:700–21, doi:http://dx.doi.org/10.1098/rspa.1927.0118.

Korea Centers for Disease Control and Prevention. 2020. . [Accessed 15 April 2020] http://www.cdc.go.kr/cdc_eng.

Liu X, Stechlinski P. Pulse and constant control schemes for epidemic models with seasonality. Nonlinear Anal Real World Appl 2011;12:931–46, doi:http://dx.doi.org/10.1016/j.nonrwa.2010.08.017.

Marais BJ, Sorrell TC. Pathways to COVID-19 'community protection'. Int J Infect Dis 2020;96:496–9, doi:http://dx.doi.org/10.1016/j.ijid.2020.05.058 In this issue.

Ministry of Health and Welfare. Coronavirus disease-19, Republic of Korea. 2020. http://ncov.mohw.go.kr/en.

Weiss HH. The SIR model and the foundations of public health. Mater Matemàtics 2013;3:1–17.