

# Unveiling SIR Model Parameters: Empirical Parameter Approach for Explicit Estimation and Confidence Interval Construction

Nanang Susyanto and Jayrold P. Arcede



Volume 5, Issue 1, Pages 54–62, June 2024

Received 26 June 2024, Revised 11 July 2024, Accepted 13 July 2024, Published Online 18 July 2024

To Cite this Article : N. Susyanto and J. P. Arcede, "Unveiling SIR Model Parameters: Empirical Parameter Approach for Explicit Estimation and Confidence Interval Construction", *Jambura J. Biomath*, vol. 5, no. 1, pp. 54–62, 2024, <https://doi.org/10.37905/jjbm.v5i1.26287>

© 2024 by author(s)

## JOURNAL INFO • JAMBURA JOURNAL OF BIOMATHEMATICS



	Homepage	:	<a href="http://ejurnal.ung.ac.id/index.php/JJBM/index">http://ejurnal.ung.ac.id/index.php/JJBM/index</a>
	Journal Abbreviation	:	Jambura J. Biomath.
	Frequency	:	Biannual (June and December)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	<a href="https://doi.org/10.37905/jjbm">https://doi.org/10.37905/jjbm</a>
	Online ISSN	:	2723-0317
	Editor-in-Chief	:	Hasan S. Panigoro
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	<a href="http://ejurnal.ung.ac.id/index.php/jjbm/oai">http://ejurnal.ung.ac.id/index.php/jjbm/oai</a>
	Google Scholar ID	:	XzYgeKQAAAAJ
	Email	:	<a href="mailto:editorial.jjbm@ung.ac.id">editorial.jjbm@ung.ac.id</a>

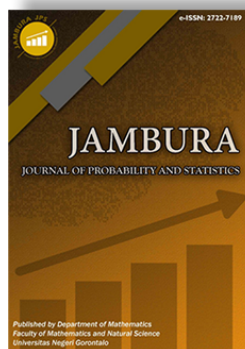
## JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



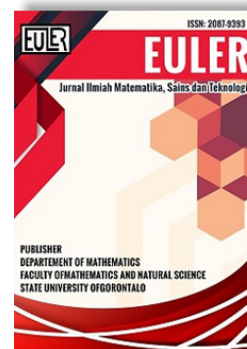
Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

# Unveiling SIR Model Parameters: Empirical Parameter Approach for Explicit Estimation and Confidence Interval Construction

Nanang Susyanto<sup>1,\*</sup>  and Jayrold P. Arcede<sup>2</sup> 

<sup>1</sup>Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia

<sup>2</sup>Department of Mathematics, Caraga State University, Butuan City, Philippines

## ARTICLE HISTORY

Received 26 June 2024

Revised 11 July 2024

Accepted 13 July 2024

Published 18 July 2024

## KEYWORDS

SIR model  
Parameter estimation  
Empirical parameters  
Confidence intervals

**ABSTRACT.** We propose a simple parameter estimation method for the Susceptible-Infectious-Recovered (SIR) model. This method offers explicit estimates of parameters using second-order numerical derivatives to construct empirical parameters. In addition, the method constructs confidence intervals, providing a robust assessment of parameter uncertainty. To validate the accuracy of our method, we applied it to simulated data, in order to demonstrate its effectiveness in accurately estimating the true model parameters. Furthermore, we applied this method to actual COVID-19 case data from the USA, Indonesia, and the Philippines. This application enables the estimation of parameters and reproductive numbers, along with their confidence intervals, thus underscoring the efficacy of our technique. Notably, the parameter estimates obtained through our approach successfully predicted the case numbers in all three countries, confirming its predictive reliability. Our method offers significant advantages in terms of simplicity and accuracy, making it an invaluable tool for epidemiological modeling and public health planning.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of JJBM:* Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

## 1. Introduction

The Susceptible-Infectious-Recovered (SIR) model, a cornerstone in epidemiology, has played a pivotal role in understanding the dynamics of infectious diseases for almost a century. Initially introduced by Kermack and McKendrick in 1927, this model categorizes a population into susceptible (S), infectious (I), and recovered (R) compartments [1]. A visual representation of how individuals transition between these compartments is provided in Figure 1. In this paper, we consider the proportion of individuals in each class, meaning we are working with the population proportions. Moreover, we consider a simple short-term model, i.e., no population turnover, or age structure, inhomogeneities and group behavior. This simplifies the model to a system of ordinary differential equations:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S(t)I(t), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t),\end{aligned}\quad (1)$$

with nonnegative initial conditions  $S(0) = S_0$ ,  $I(0) = I_0$ ,  $R(0) = R_0$ , and domain  $\Omega = \{(x, y, z) \in \mathbb{R}^3 : x, y, z \geq 0 \text{ and } x + y + z = 1\}$  that is positively invariant under System (1). Here,  $\beta$  and  $\gamma$  represent the infection and removal rates, respectively. It is worth noting that in reality, especially in diseases with

a significant mortality rate, individuals in the  $R$  compartment can include both those who have recovered and those who have succumbed to the disease.

Although the SIR model may seem deceptively simple, it has been proven to be an immensely powerful tool for analyzing various spreading phenomena. Its versatility extends beyond infectious diseases to encompass a range of phenomena, including the spread of diseases such as herpes [2], influenza [3], and the recent COVID-19 pandemic [4–6]. Furthermore, in non-disease contexts, the SIR model finds application in investigating the spread of behaviors like smoking [7] and computer viruses [8]. Extensions of the SIR model, such as SIRD, SIRS, SEIR, SEIRS, and SIRDS, have further broadened its applicability. These extensions are employed in analyzing the spread of diseases like dengue, incorporating vector compartments [9], capturing phenomena involving reinfection [10], exploring the relationship between different diseases [11, 12], assessing the impact of policies on disease control [13], and evaluating the effectiveness of control measures or treatments in epidemic models [14–16].

Research on the analysis of the SIR model from a mathematical perspective abounds in the literature. The investigation of equilibrium points and their stability in the SIR model [17] provides insights into the interaction of different species in biological communities. From a global stability standpoint, [18] offers a comprehensive analysis. The stability of models with delays is explored in [19], while [20] delves into the stability of the SIR model considering vaccination and treatment.

In the wake of the COVID-19 pandemic, the SIR model and

\*Corresponding Author.

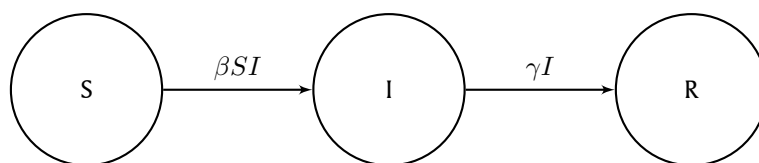


Figure 1. SIR Diagram Transfer

its extensions, coupled with various mathematical tools, have become indispensable for analyzing the spread of the virus. During the early stages of the pandemic, the SIR model was employed to understand the spread of the virus in Indonesia [4, 21], while in France, case analyses were conducted using the model [22]. In Iran, a combination of the SIR model and Gaussian Weighted Regression was utilized to examine the spread of the virus [23]. Investigations into the later stages of the pandemic were carried out in the Philippines, particularly in Davao City [6]. Furthermore, numerous studies have been conducted on COVID-19, employing SIR models in various contexts [13, 24, 25].

In addition to efforts aimed at finding analytic and numerical solutions, as well as investigating the stability of SIR models, parameter estimation is a crucial aspect. While knowing the infection and removal rates enables the determination of solutions, either analytically or numerically, to ascertain the number of individuals in each compartment, these parameters are often unknown in reality and must be estimated from data. Here, we assume that we can observe all the proportions of individuals in each class. The most prevalent method for estimating these parameters is through Least Squares Estimates (LSE) [26–30]. More advanced methods have been proposed using machine learning approaches [31, 32], yet these methods do not provide explicit estimates.

A simpler method outlined in [33] does offer explicit parameter estimates using second-order finite differences. However, it can be sensitive to data, meaning it may fluctuate significantly due to data noise. This issue can be addressed by quantifying the uncertainty of the estimates. We propose a simple method that not only provides explicit parameter estimates but also quantifies uncertainty by generating confidence intervals a first in the field to our knowledge. This method’s potential will be demonstrated through both simulated and actual data on COVID-19’s spread.

The remainder of the paper is structured as follows. Section 2 describes the proposed parameter estimation method. Section 3 contains the application of the method to simulated and real-world data to demonstrate its efficacy. Finally, Section 4 is dedicated to discussing the results, limitations of our methods, and potential improvement.

## 2. Empirical Methods Approach

In this section, we will outline the detailed method through which our approach provides explicit parameter estimates in the SIR model. Our approach relies on the utilization of second-order finite differences to approximate the first derivative of functions. Recalling the definition of the derivative of a differentiable function  $f : (a, b) \rightarrow \mathbb{R}$ , where  $a, b \in \mathbb{R}$  and  $a < b$ , at  $x = c$  can be expressed as

$$f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c-h)}{2h}.$$

If  $f$  is continuously differentiable up to the third order, then the derivative can be approximated as:

$$f'(c) = \frac{f(c+h) - f(c-h)}{2h} + O(h^2).$$

This implies that we can numerically compute (or approximate) the first derivative as:

$$f'(c) \approx \frac{f(c+h) - f(c-h)}{2h}, \tag{2}$$

with the error being of quadratic order, thus termed as the “second-order approximation” method for computing the first derivative.

Let  $a = t_0 < t_1 < \dots < t_n = b$  be the  $n + 1$  time points such that  $t_j - t_{j-1} = h$  for each  $j = 1, 2, \dots, n$  that is called equidistant time points. If  $f_j = f(t_j)$  for each  $j = 0, 1, \dots, n$ , then  $f'(t_j)$  can be approximated using the second-order approximation according to eq. (2) by

$$f'(t_j) \approx \frac{f_{j+1} - f_{j-1}}{2h}.$$

Now, we are ready to apply the second-order approximation to the SIR model. Given equidistant time points  $t_0 < t_1 < \dots < t_n$ , let  $S_j, I_j$ , and  $R_j$  be defined as  $S_j = S(t_j), I_j = I(t_j)$ , and  $R_j = R(t_j)$  for each  $j = 1, 2, \dots, n$ , where all these values are obtained from observations. The primary objective is to estimate  $\beta$  and  $\gamma$  in eq. (1), which best fit these data. Utilizing the approximation in eq. (2) with  $h = 1$ , we can approximate

$$\begin{aligned} S'(t_j) &\approx \frac{S_{j+1} - S_{j-1}}{2}, \\ I'(t_j) &\approx \frac{I_{j+1} - I_{j-1}}{2}, \\ R'(t_j) &\approx \frac{R_{j+1} - R_{j-1}}{2}. \end{aligned}$$

Then, the first and third equations in eq. (1) can be written as:

$$\begin{aligned} \frac{S_{j+1} - S_{j-1}}{2} &\approx \beta S_j I_j, \\ \frac{R_{j+1} - R_{j-1}}{2} &\approx \gamma I_j. \end{aligned}$$

Thus, we can estimate  $\beta$  and  $\gamma$  as follows:

$$\begin{aligned} \beta &\approx \frac{S_{j+1} - S_{j-1}}{2S_j I_j}, \\ \gamma &\approx \frac{R_{j+1} - R_{j-1}}{2I_j}, \end{aligned}$$

for each  $j = 1, 2, \dots, n - 1$ . However, as we assume that  $\beta$  and  $\gamma$  are constants, it is natural to consider

$$\beta_j = \frac{S_{j+1} - S_{j-1}}{2S_j I_j}, \tag{3}$$

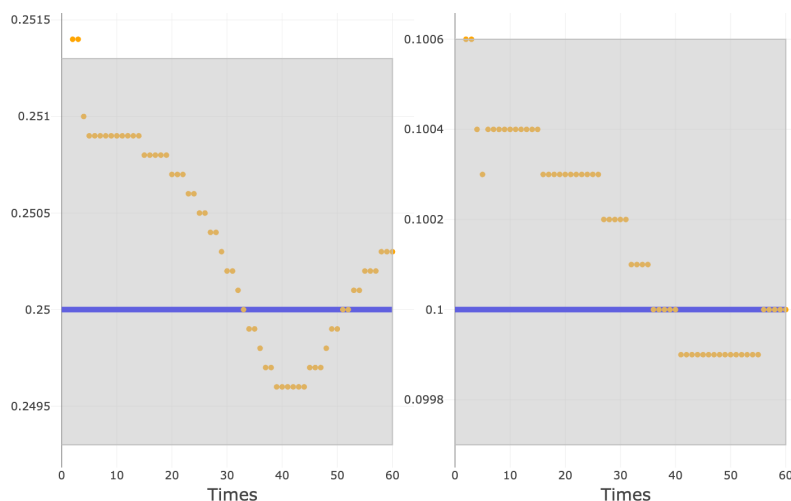


Figure 2. Estimates of  $\beta$  (left) and  $\gamma$  (right) for First set ( $\beta = 0.25, \gamma = 0.1$ ).

and

$$\gamma_j = \frac{R_{j+1} - R_{j-1}}{2I_j}, \tag{4}$$

which we will refer to as the empirical beta and empirical gamma, respectively, serving as "realizations" of  $\beta$  and  $\gamma$ .

These "realizations" of  $\beta$  and  $\gamma$  are explicit once the S, I, and R data are given. We construct the estimates by simply taking the average of those empirical beta and empirical gamma:

$$\hat{\beta} = \frac{1}{n-1} \sum_{j=1}^{n-1} \beta_j, \tag{5}$$

and

$$\hat{\gamma} = \frac{1}{n-1} \sum_{j=1}^{n-1} \gamma_j. \tag{6}$$

To quantify the uncertainty, we first compute the standard deviation for both parameters:

$$\sigma_{\hat{\beta}} = \sqrt{\frac{1}{n-2} \sum_{j=1}^{n-1} (\beta_j - \hat{\beta})^2},$$

$$\sigma_{\hat{\gamma}} = \sqrt{\frac{1}{n-2} \sum_{j=1}^{n-1} (\gamma_j - \hat{\gamma})^2}.$$

Assuming normality, when a significance level  $\alpha \in (0, 1)$  is given, the corresponding  $(1 - \alpha)100\%$  confidence interval for  $\beta$  and  $\gamma$  is simply:

$$\left[ \hat{\beta} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sigma_{\hat{\beta}}, \hat{\beta} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sigma_{\hat{\beta}} \right], \tag{7}$$

and

$$\left[ \hat{\gamma} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sigma_{\hat{\gamma}}, \hat{\gamma} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sigma_{\hat{\gamma}} \right], \tag{8}$$

respectively. Here,  $\Phi^{-1}$  denotes the inverse of the cumulative distribution function of the standard normal distribution. Note that, as mentioned in the Introduction, our estimate and its confidence interval for each parameter are explicit. We provide the R-code for this simple method in the Appendix.

### 3. Results

In this section, we will apply our proposed method to both simulated and real data to see its performance. For all data, all parameters are rounded to the nearest  $10^{-4}$ . This value is sufficiently small because when calculating errors, with the range of  $\beta$  and  $\gamma$  that we have, it will result in around  $10^{-3}$  relative errors.

#### 3.1. Simulated data

We will simulate 3 datasets here. Simulation involves setting the parameter values  $\beta$  and  $\gamma$ , as well as the initial proportions for the S, I, and R individuals. Given time steps  $t_0 < t_1 < \dots < t_n$ , we can obtain the proportion values for S, I, and R using the Runge-Kutta method based on System (1), i.e.,

$$S_0, S_1, \dots, S_n,$$

$$I_0, I_1, \dots, I_n,$$

$$R_0, R_1, \dots, R_n.$$

Subsequently, we will estimate the parameters  $\beta$  and  $\gamma$  explicitly and construct the  $(1 - \alpha)100\%$  confidence intervals based on the S, I, and R values. We will then compare the estimated parameters using our proposed method with the corresponding true values and provide the margins of error.

The parameter values used for simulating the data are taken from [33] for comparison with their method and from [34] to mimic the conditions in that paper. For all datasets, we set  $n = 60$  to represent a two-month period,  $(S_0, I_0, R_0) = (\frac{999}{1000}, \frac{1}{1000}, 0)$ , and a significance level of  $\alpha = 5\%$ .

##### 1. First set

We chose  $\beta = 0.25$  and  $\gamma = 0.1$  as in [33]. Estimates of  $\beta$  and  $\gamma$  for the first dataset are shown in Figure 2, where the orange points represent the empirical parameters and the blue lines represent the true parameters. The left panel displays the estimated values of  $\beta$  with  $\hat{\beta} = 0.2503$ , while the right panel shows the estimates of  $\gamma$  with  $\hat{\gamma} = 0.1002$ . The grey areas represent the 95% confidence intervals for each parameter estimate, which are  $[0.2493, 0.2513]$  for  $\beta$  (with the true value of 0.25) and  $[0.0997, 0.1006]$  for  $\gamma$  (with the true value of 0.1). Thus, the relative margin of error in

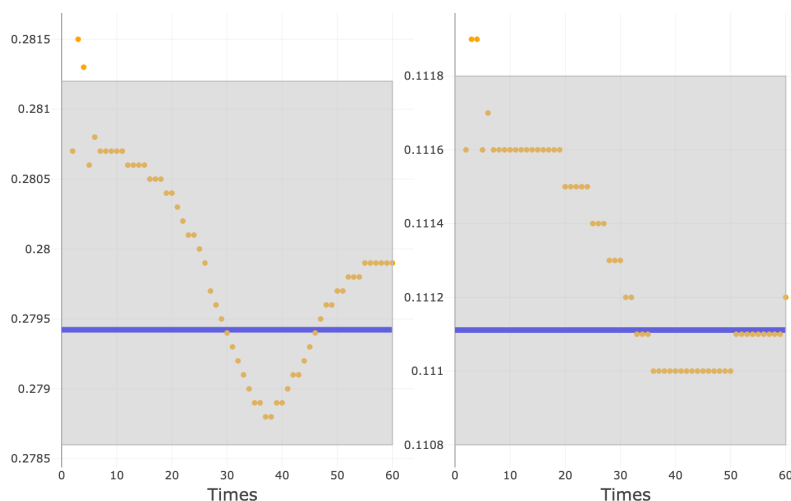


Figure 3. Estimates of  $\beta$  (left) and  $\gamma$  (right) for Third set ( $\gamma = 1/9, R_0 = 2.5148$ )

this data is around  $8 \times 10^{-3}$  for  $\beta$  and  $9 \times 10^{-3}$  for  $\gamma$ , which is sufficiently accurate.

2. Second set

Still adopting parameters from [33], we now chose  $\beta = 0.1$  and  $\gamma = 0.09$ . Interestingly, our proposed method gives estimated values of  $\beta$  and  $\gamma$  that are exactly the same as their true values:  $\hat{\beta} = 0.1$  and  $\hat{\gamma} = 0.09$ . Recall that the empirical parameters are rounded to the nearest  $10^{-4}$ .

3. Third set

In these simulations, we choose  $\gamma = 1/9$  to illustrate the COVID-19 removal rate (assuming the infectious period is 9 days). The infection rate  $\beta$  is chosen such that the reproductive number  $R_0 = 2.5148$  as studied in [34]. Hence, we have  $\beta = 0.2794$ . Our proposed method gives  $\hat{\beta} = 0.2799$  with a 95% confidence interval of  $[0.2786, 0.2812]$ , which still contains the true value  $\beta = 0.2794$ . For the removal rate, our proposed method results in  $\hat{\gamma} = 0.1113$  with a 95% confidence interval of  $[0.1108, 0.1118]$ , also containing the true value  $\gamma = 1/9$ .

From the relative margin of error point of view, in this third set, we get around  $9 \times 10^{-3}$  for both parameters, which is again sufficiently accurate. The complete graphs of estimates of  $\beta$  and  $\gamma$  are shown in Figure 3, where the orange points represent the empirical parameters and the blue lines represent the true parameters. The left panel displays the estimated values of  $\beta$ , while the right panel shows the estimates of  $\gamma$ .

3.2. Real data

We will apply our proposed method to real data on the COVID-19 pandemic from three different countries, representing various periods of the pandemic. All data are sourced from Johns Hopkins University and Medicine (<https://coronavirus.jhu.edu/>). The first dataset is COVID-19 data from the USA representing the early pandemic phase, which will be compared to the method presented in [33]. The second and third datasets are COVID-19 data from the authors' countries, i.e., Indonesia and the Philippines, representing mid- and late-pandemic periods, respectively. This diverse temporal sampling is intended to capture the impact

of COVID-19 variants.

Since our proposed method assumes  $\beta$  and  $\gamma$  to be constant, while in reality they can be time-varying, at each time point we estimate the parameters using the data from the last 14 days, as done in [33]. In general, at each time step, we use the last 14 days as training data and the next 7 days as testing data. Therefore, as the time step changes, our parameter estimates will also vary over time because they are based on different training data. Additionally, we estimate the reproductive number ( $R_0$ ) and its confidence interval using a similar method as we did for  $\beta$  and  $\gamma$ . Recall that we have empirical beta eq. (3) and empirical gamma eq. (4). Thus, we can define the empirical basic reproduction number,  $R_{0,j}$  as

$$R_{0,j} = \frac{\beta_j}{\gamma_j}.$$

This empirical  $R_{0,j}$  can be used to determine the  $R_0$  estimate and its confidence interval. We remark that  $\beta$  and  $\gamma$  are unobserved, and thus need to be estimated from the data. Therefore, we cannot compute the error of the parameter estimates. Rather, we compute the mean absolute prediction error (MAPE) [35], so that at each time step, after we use the data from the last 14 days to estimate the parameters, we can predict the number of cases for the next seven days and consequently, compare it to the actual number of cases. At each time step, the MAPE is computed as:

$$MAPE(t_j) = \frac{1}{7} \sum_{i=1}^7 \left| \frac{\hat{I}_{t_j+i} - I_{t_j+i}}{I_{t_j+i}} \right|$$

where  $\hat{I}_{t_j+i}$  is the predicted number of cases at time period and  $I_{t_j+i}$  is the true number of cases at  $t_j + i$ . Finally, we take the mean of all these individual predictions and call it the final MAPE. Here, the number of cases can easily be determined by multiplying the predicted proportion given by the model by the total population.

1. USA data

As used in [33], we utilized COVID-19 data from the USA in the early period of the pandemic, from June 23, 2020, until September 21, 2020. While their paper only provides the

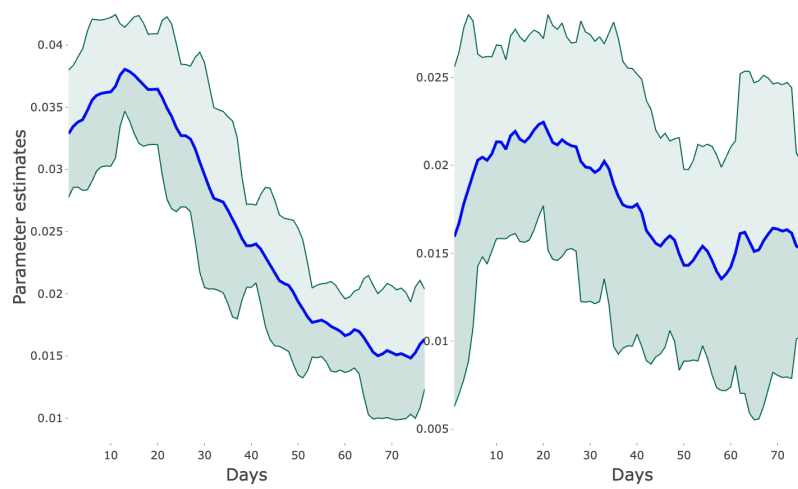


Figure 4. Estimates of  $\beta$  (left) and  $\gamma$  (right) for USA data

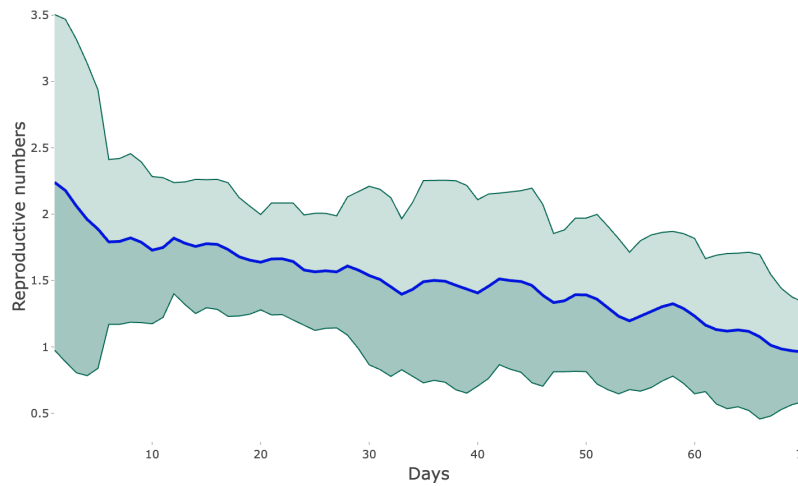


Figure 5.  $R_0$  for USA data from 23 June 2020 to 21 September 2020

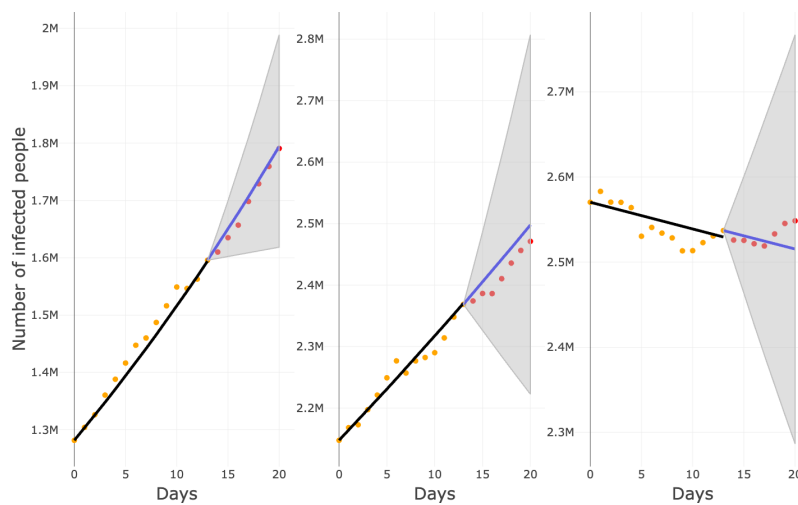


Figure 6. Cases predictions for USA data

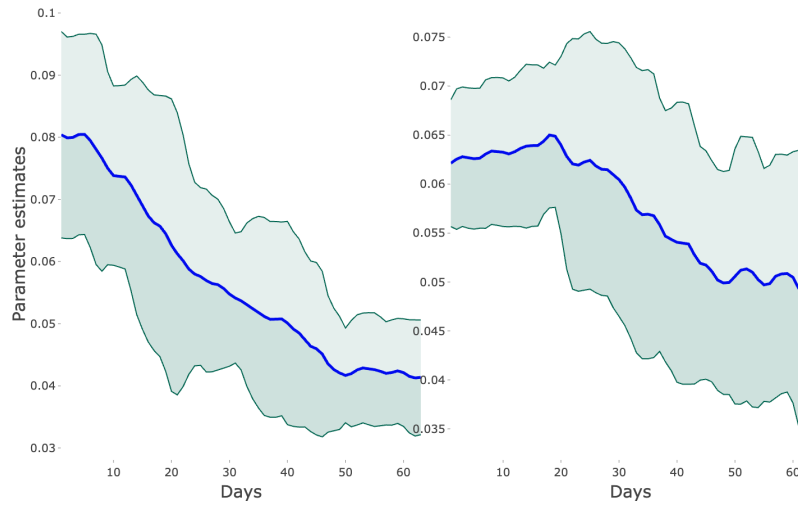


Figure 7. Estimates of  $\beta$  (left) and  $\gamma$  (right) for Indonesia data

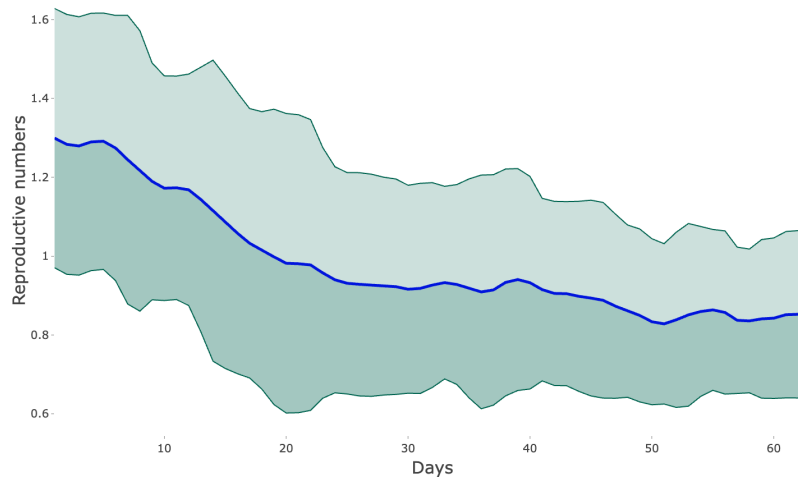


Figure 8.  $R_0$  for Indonesia data from 7 January to 7 April 2021

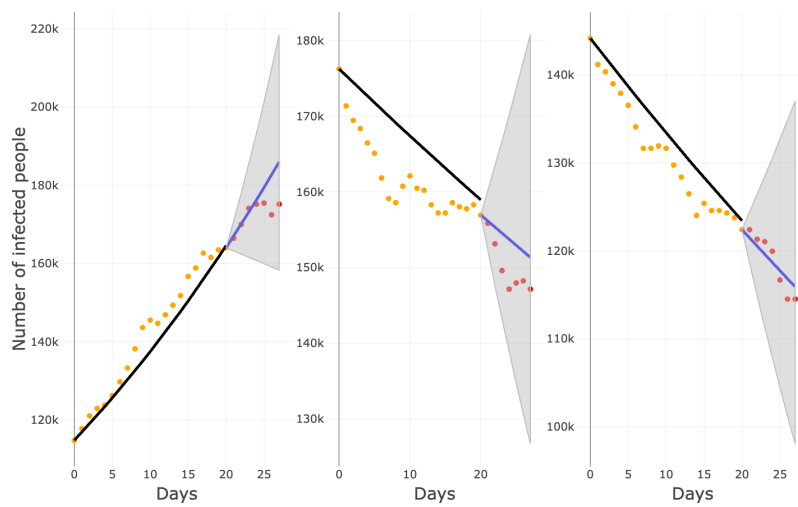


Figure 9. Cases predictions for Indonesia data

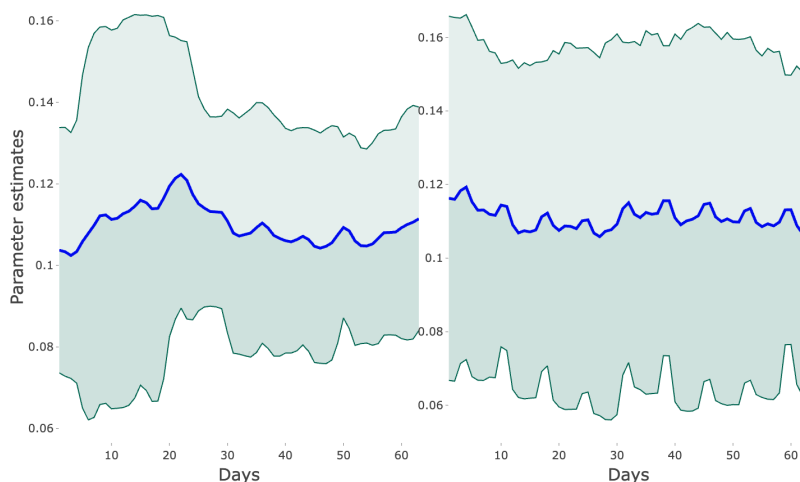


Figure 10. Estimates of  $\beta$  (left) and  $\gamma$  (right) for Philippines data

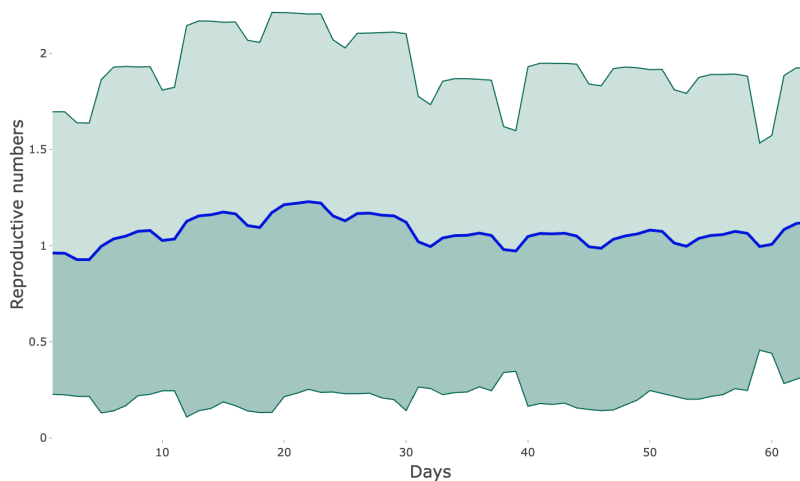


Figure 11.  $R_0$  for Philippines data from 4 May to 2 August 2021

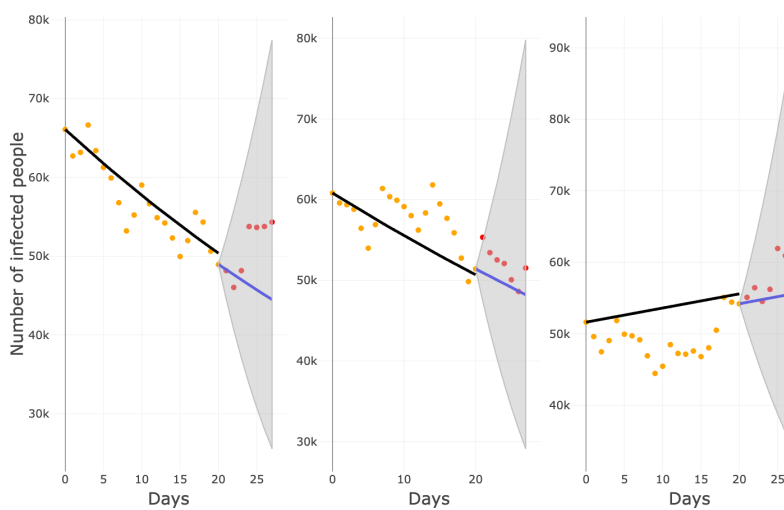


Figure 12. Cases predictions for Philippines data



reproductive numbers, in this paper, we provide explicit estimates of the parameters, as seen in Figure 4, which are used to estimate the reproductive numbers shown in Figure 5. The 95% confidence intervals for these estimates are provided in the shaded areas.

As seen in Figure 5, the trend of reproductive numbers is similar to [33]. A notable addition is that we provide the 95% confidence interval for each time step, enhancing the robustness of our analysis.

In Figure 6, the left panel displays the number of cases and the corresponding predictions using data from June 23 to July 6 as the training data (orange points) and their model (black lines), along with predictions using data from July 7 to July 13 as the testing data (red points) and their corresponding prediction values (blue lines). We assume that the number of population is 331 millions. The 95% confidence intervals are represented by the grey areas. Note that the time step for this case is July 13. The middle and right panels exhibit the same information, but for time steps August 6 and September 6, respectively. These time steps are chosen to represent the beginning, middle, and end of the period from June 23 to September 21, 2020. The Mean Absolute Prediction Error (MAPE) for this data is 1.39%, indicating high accuracy. Moreover, all points in the testing sets fall within our 95% confidence intervals.

## 2. Indonesia data

For Indonesia's data, we considered COVID-19 cases from January 7 to April 7, 2021, to match the length of the USA data. We assume that the population is 272 million. Employing the same methodology as for the USA case, we present the parameter estimates in Figure 7, the reproductive numbers in Figure 8, and the number of cases in Figure 9. In all figures, the 95% confidence intervals are displayed in shaded areas. The Mean Absolute Prediction Error (MAPE) for this data is 2.75% indicating high accuracy. As expected, the reproductive numbers dropped below 1, leading to a decreasing trend in the number of cases in the later part of this period. Again, all testing data are within the 95% prediction intervals.

## 3. Philippines data

The last dataset we used is the COVID-19 cases in the Philippines from May 4 to August 2, 2021, with a 90-day time window. Assuming a population of 113 million, we applied the same settings as before. Our proposed method gives the parameter estimates in Figure 10 and the corresponding reproductive numbers in Figure 11. We observe that the reproductive numbers are around 1, indicating a relatively stagnant trend. However, the confidence intervals are quite wide, which may be due to data collection or reporting issues. The Mean Absolute Prediction Error (MAPE) is around 6.94%, which is larger than that of the USA and Indonesia data but still relatively accurate. The corresponding figures for the number of cases are given in Figure 12. Interestingly, all testing data still fall within our 95% confidence intervals.

## 4. Discussions

We have proposed a simple yet powerful method for estimating parameters in the SIR model, providing explicit estimates

and associated confidence intervals. This method not only allows for the accurate calculation of model parameters but also facilitates the estimation of the reproductive number with corresponding confidence intervals. Our approach has proven successful in accurately retrieving true parameters from simulated datasets and has demonstrated high predictive accuracy in forecasting actual COVID-19 case numbers using real-world data.

While traditional methods for parameter estimation in epidemiological models often involve complex computational techniques, our method stands out for its simplicity and directness. Furthermore, without requiring with extensive computation, our technique provides a quick and straightforward means of obtaining parameter estimates. This can be particularly advantageous in public health scenarios where timely decision-making is crucial.

In reality, the parameters in the current model are not constants and may vary over time. Although our proposed method can estimate these time-varying parameters by re-estimating at every time step, a more refined approach that accommodate parameter variation also warrants. These methods should provide simple and explicit estimates to be computed quickly. Additionally, the method's performance is contingent as always on the quality and granularity of available data, and its efficacy in different epidemiological contexts could vary.

Further studies could explore the integration of our estimation method with other epidemiological models, such as SEIR or SIRS, and the like to widen its applicability to diseases with different transmission dynamics.

**Author Contributions.** Susyanto, N.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, visualization. Arcede, J. P.: Validation, investigation, resources, writing—review and editing.

**Acknowledgement.** The authors are thankful to the editors and reviewers who have supported us in improving this manuscript.

**Funding.** This research did not receive any external funding.

**Conflict of interest.** The authors declare no conflict of interest.

**Data availability.** The data used in this research is open data provided by Johns Hopkins University and Medicine (<https://coronavirus.jhu.edu/>)

## References

- [1] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927. DOI:10.1098/rspa.1927.0118
- [2] D. Tudor, "A deterministic model for herpes infections in human and animal populations," *SIAM Review*, vol. 32, no. 1, pp. 136–139, 1990. DOI:10.1137/1032003
- [3] B. J. Coburn, B. G. Wagner, and S. Blower, "Modeling influenza epidemics and pandemics: insights into the future of swine flu (h1n1)," *BMC Medicine*, vol. 7, no. 1, pp. 30, 2009. DOI:10.1186/1741-7015-7-30
- [4] N. Nuraini, K. Khairudin, and M. Apri, "Modeling simulation of covid-19 in indonesia based on early endemic data," *Communication in Biomathematical Sciences*, vol. 3, no. 1, pp. 1–8, 2020. DOI:10.5614/cbms.2020.3.1.1
- [5] R. Saxena, M. Jadeja, and V. Bhateja, "Propagation analysis of covid-19: An sir model-based investigation of the pandemic," *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 11103–11115, 2023. DOI:10.1007/s13369-021-05904-0

- [6] L. E. Anonuevo *et al.*, “Transmission dynamics and baseline epidemiological parameter estimates of coronavirus disease 2019 pre-vaccination: Davao city, philippines,” *PLOS ONE*, vol. 18, no. 4, pp. 1–20, 2023. DOI:10.1371/journal.pone.0283068
- [7] G. Zaman *et al.*, “Optimal strategy of vaccination & treatment in an sir epidemic model,” *Mathematics and Computers in Simulation*, vol. 136, no. 3, pp. 63–77, 2017. DOI:10.1016/j.matcom.2016.11.010
- [8] L. Yang and X. Yang, “A new epidemiological model for computer viruses,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 6, pp. 1935–1944, 2014. DOI:10.1016/j.cnsns.2013.09.038
- [9] S. Side *et al.*, “Analysis and simulation of siri model for dengue fever transmission,” *International Journal Of Scientific & Technology Researchers*, vol. 13, no. 3, pp. 340–351, 2020. DOI:10.17485/ijst/2020/v13i03/147852
- [10] M. Zheng *et al.*, “Non-periodic outbreaks of recurrent epidemics and its network modelling,” *Scientific Reports*, vol. 5, no. 1, pp. 16010, 2015. DOI:10.1038/srep16010
- [11] H. Fahlana *et al.*, “Analysis of a coendemic model of covid-19 and dengue disease,” *Communication in Biomathematical Sciences*, vol. 4, no. 2, pp. 138–151, 2021. DOI:10.5614/cbms.2021.4.2.5
- [12] N. Nuraini *et al.*, “The impact of covid-19 quarantine on tuberculosis and diabetes mellitus cases: A modelling study,” *Tropical Medicine and Infectious Disease*, vol. 7, no. 12, 2022. DOI:10.3390/tropicalmed7120407
- [13] S. D. Silver, P. van den Driessche, and S. Khajanchi, “A dynamic multistate and control model of the covid-19 pandemic,” *Journal of Public Health*, vol. 136, no. 3, pp. 63–77, 2023. DOI:10.1007/s10389-023-02014-z
- [14] D. Lestari *et al.*, “Qualitative behaviour of a stochastic hepatitis c epidemic model in cellular level,” *Mathematical Biosciences and Engineering*, vol. 19, no. 2, pp. 1515–1535, 2021. DOI:10.3934/mbe.2022070
- [15] D. Lestari *et al.*, “A minimum principle for stochastic control of hepatitis c epidemic model,” *Boundary Value Problems*, vol. 2023, no. 1, pp. 52, 2023. DOI:10.1186/s13661-023-01740-3
- [16] L. Hanum, D. Ertiningsih, and N. Susyanto, “Sensitivity analysis unveils the interplay of drug-sensitive and drug-resistant glioma cells: Implications of chemotherapy and anti-angiogenic therapy,” *Electronic Research Archive*, vol. 32, no. 1, pp. 72–89, 2023. DOI:10.3934/era.2024004
- [17] G. Zaman and I. Jung, “Stability techniques in sir epidemic models,” *PAMM*, vol. 7, no. 1, pp. 2030063–2030064 2007. DOI:10.1002/pamm.200701147
- [18] F. S. Alshammari and F. T. Akylidiz, “Global Stability for Novel Complicated SIR Epidemic Models with the Nonlinear Recovery Rate and Transfer from Being Infectious to Being Susceptible to Analyze the Transmission of COVID-19,” *Journal of Function Spaces*, vol. 2021, no. 1, pp. 5207152, 2021. DOI:10.1155/2021/5207152
- [19] C. Huang, J. Cao, F. Wen, and X. Yang, “Stability analysis of sir model with distributed delay on complex networks,” *PLoS ONE*, vol. 11, no. 8, 2016. DOI:10.1371/journal.pone.0158813
- [20] A. Elazzouzi, A. L. Alaoui, M. Tilioua, and A. Tridane, “Global stability analysis for a generalized delayed sir model with vaccination and treatment,” *Advances in Difference Equations*, vol. 2019, no. 1, pp. 532, 2019. DOI:10.1186/s13662-019-2447-z
- [21] A. Hasan *et al.*, “Superspreading in early transmissions of covid-19 in indonesia,” *Scientific Reports*, vol. 10, no. 1, pp. 22386, 2020. DOI:10.1038/s41598-020-79352-5
- [22] H. Salje *et al.*, “Estimating the burden of SARS-CoV-2 in France,” *Science*, vol. 369, no. 6500, pp. 208–211, 2020. DOI:10.1126/science.abc3517
- [23] R. RamÁrez-Aldana, J. C. Gomez-Verjan, and O. Y. Bello-Chavolla, “Spatial analysis of covid-19 spread in iran: Insights into geographical and structural transmission determinants at a province level,” *PLOS Neglected Tropical Diseases*, vol. 14, no. 11, pp. 1–15, 2020. DOI:10.1371/journal.pntd.0008875
- [24] M. Shahzad *et al.*, “Dynamics models for identifying the key transmission parameters of the covid-19 disease,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 757–765, 2021. DOI:10.1016/j.aej.2020.10.006
- [25] L. Bougoffa, S. Bougouffa, and A. Khanfer, “Approximate and parametric solutions to sir epidemic model,” *Axioms*, vol. 13, no. 3, pp. 201, 2024. DOI:10.3390/axioms13030201
- [26] I. Dattner and C. A. J. Klaassen, “Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters,” *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 201, 2015. DOI:10.1214/15-EJS1053
- [27] J. P. Arcede *et al.*, “Accounting for symptomatic and asymptomatic in a seir-type model of covid-19,” *Math. Model. Nat. Phenom.*, vol. 15, no. 1, pp. 34, 2020. DOI:10.1051/mmnp/2020021
- [28] T. E. Simos *et al.*, “Real-time estimation of r0 for covid-19 spread,” *Mathematics*, vol. 9, no. 6, pp. 664, 2021. DOI:10.3390/math9060664
- [29] V. Srivastava *et al.*, “A systematic approach for covid-19 predictions and parameter estimation,” *Personal and Ubiquitous Computing*, vol. 27, no. 3, pp. 675–687, 2023. DOI:10.1007/s00779-020-01462-8
- [30] A. Osi and N. Ghaffarzadegan, “Parameter estimation in behavioral epidemic models with endogenous societal risk-response,” *PLoS Computational Biology*, vol. 20, no. 3, pp. e1011992, 2024. DOI:10.1371/journal.pcbi.1011992
- [31] S. Shaier, M. Raissi, and P. Seshaiyer, “Data-driven approaches for predicting spread of infectious diseases through dinns: Disease informed neural networks,” *Letters in Biomathematics*, vol. 8, no. 1, pp. 71–105, 2022. DOI:10.48550/arXiv.2110.05445
- [32] J. Gadewadikar and J. Marshall, “A methodology for parameter estimation in system dynamics models using artificial intelligence,” *Systems Engineering*, vol. 27, no. 2, pp. 253–266, 2024. DOI:10.1002/sys.21718
- [33] M. Medvedeva *et al.*, “Direct estimation of sir model parameters through second-order finite differences,” *Mathematical Methods in the Applied Sciences*, vol. 44, no. 5, pp. 3819–3826, 2021. DOI:10.1002/mma.6985
- [34] E. Soewono, “On the analysis of covid-19 transmission in wuhan, diamond princess and jakarta-cluster,” *Communication in Biomathematical Sciences*, vol. 3, no. 1, pp. 9–18, 2020. DOI:10.5614/cbms.2020.3.1.2
- [35] N. Reich, J. Lessler, K. Sakrejda, S. Lauer, S. Iamsirithaworn, and D. Cummings, “Case study in evaluating time series prediction models using the relative mean absolute error,” *The American Statistician*, vol. 70, no. 3, pp. 285–292, 2016. DOI:10.1080/00031305.2016.1148631

## Appendix

The R function `pars` implements explicit parameter estimation using the empirical parameter approach for  $\beta$  and  $\gamma$ , as described in eqs. (3) and (4), respectively. Confidence intervals are constructed according to eqs. (7) and (8) for a given significance level  $\alpha \in [0, 1]$ , with a default value of 5%. To avoid numerical issues, the empirical values of  $\beta$  and  $\gamma$  are rounded to the nearest  $10^{-\text{dig}}$ , with a default value of  $\text{dig} = 4$ .

The function requires input data for  $S$ ,  $I$ , and  $R$ , and allows for setting the significance level and rounding parameter `dig`. The output is a list containing:

1. `empbeta`: values of empirical  $\beta$  as given in eq. (3).
2. `empgamma`: values of empirical  $\gamma$  as given in eq. (4).
3. `estbeta`: explicit estimate of  $\beta$  as given in eq. (5).
4. `estgamma`: explicit estimate of  $\gamma$  as given in eq. (6).
5. `[lowerbeta, upperbeta]`: confidence interval for  $\beta$  as given in eq. (7).
6. `[lowergamma, uppergamma]`: confidence interval for  $\gamma$  as given in eq. (8).

Here is the R code:

```
pars <- function(S,I,R,alpha=0.05,dig=4){
n <- length(S); n1 <- n-1; dt <- 1
t <- 0:n1
Sj <- S[-n]; Sj1 <- S[-1]; Sjp1 <- Sj1[-1]; Sjm1 <- Sj[-n1]
Ij <- I[-n]; Ij1 <- I[-1]; Ijp1 <- Ij1[-1]; Ijm1 <- Ij[-n1]
Rj <- R[-n]; Rj1 <- R[-1]; Rjp1 <- Rj1[-1]; Rjm1 <- Rj[-n1]
Sj <- Sj1[-n1]; Ij <- Ij1[-n1]; Rj <- Rj1[-n1]
#Calculate empirical beta and gamma
betahat <- (Sjm1-Sjp1)/(2*Ij*Sj*dt); betahat <-
round(betahat,dig)
gammahat <- -(Rjm1-Rjp1)/(2*Ij*dt); gammahat <-
round(gammahat,dig)
#Calculate estimates
mb <- mean(betahat); mg <- mean(gammahat)
#Calculate standard deviation for beta and gamma
sb <- sd(betahat); sg <- sd(gammahat)
za <- qnorm(1-alpha/2)
#Calculate confidence intervals
lb <- mb-za*sb; ub <- mb+za*sb
lg <- mg-za*sg; ug <- mg+za*sg
#Return results
temp <- list(empbeta=betahat,empgamma=gammahat,estbeta=mb,
estgamma=mg, lowerbeta=lb,upperbeta=ub,
lowergamma=lg,uppergamma=ug)
return(temp)
}
```