Novel Estimators for the Number of Susceptible Individuals in SIR Models of Infectious Epidemics

Michaël Antonie van Wyk*, André Martin McDonald†, David M. Rubin* and Fangfang Zhang‡
* University of the Witwatersrand, Johannesburg, Braamfontein, South Africa
E-mail: {anton.vanwyk, david.rubin}@wits.ac.za Tel: +27-11-717-7267
† Council for Scientific and Industrial Research, Pretoria, South Africa
E-mail: amcdonald@csir.co.za Tel: +27-12-841-4349
‡ Qilu University of Technology, Jinan, China
E-mail: zhff4u@qlu.edu.cn Tel: +86-151-6916-3982

***Abstract*—Accurate estimation of model parameters early in infectious epidemics may improve planning and resource allocation in mitigating the adverse consequences on affected populations. By applying the Peano-Baker series formula and the Cauchy repeated integral formula, we present the development of three novel estimators which facilitate the estimation of the number of susceptible people as a function of time $S(t)$ for an SIR model of an infectious epidemic. Association of these three estimators by combining them produces new estimators. We present the case for the new estimator, $\hat{S}_{1,3}$, derived from the association of two of the original estimators, $\hat{S}_1$ and $\hat{S}_3$, to estimate $S(0)$. This produces an estimate based on the history of net infection rate, $I'(t)$, from time 0 to t. By assuming parameters from the literature for the spread of COVID-19 in Wuhan, we run numerical simulations starting with an infection rate $I'(t)$ and adding filtered Gaussian noise. Discretization produces inaccuracy in bias and variance, however $\hat{S}_1$ and $\hat{S}_3$ yield accurate figures for $S(t)$, despite noise contamination. $\hat{S}_{1,3}$ also yields accurate figures for $S(0)$, with improvements as more observations are accumulated. We plan to investigate other novel estimators and further study their performance with real-world data.**

I. Introduction

The ability to predict the likely course of an infectious epidemic is critical for planning and the allocation of resources to assist in mitigating the adverse effects on the population. This became abundantly evident during the COVID-19 pandemic, and is arguably the most important reason for the need to develop highly predictive models and techniques for the accurate, early estimation of model parameters.

Various deterministic compartmental models have been studied in mechanistic epidemiology. Arguably the simplest of all is the three-compartment Susceptible-Infected-Recovered (SIR) model, an early version of which was described in 1927 [1]. The equations for the SIR model are as follows:

$$S'(t) = -\frac{\beta}{N} S(t) I(t), \tag{1}$$

$$I'(t) = \frac{\beta}{N} S(t) I(t) - \gamma I(t), \tag{2}$$

$$R'(t) = \gamma I(t), \tag{3}$$

where the five parameters are infection rate (β), recovery rate (γ), total population size (N), initial number of susceptible individuals ($S_0 := S(0)$), and initial number of infected individuals ($I_0 := I(0)$).

Adaptations to the SIR model family requiring more than five parameters have been used in modeling epidemics. For example the use of a Susceptible-Exposed-Infected-Recovered (SEIR) model during the early spike in COVID-19 cases in Northern Italy in 2020, highlighted the importance of knowing the model parameters as accurately as possible [2]. Amiri et al. [3] performed parameter estimation on an extended SIR model (SQAIR) which included a quarantined (Q) compartment and a hidden asymptomatic (A) compartment.

While these elaborate SIR variants can provide more detailed modeling, they introduce additional system states that may not be directly observable. Thus, parameter values for these more complex models often need to be estimated through informed guesses, rather than direct measurement.

On one hand, these expanded model parameters allow for more nuanced representations of real-world dynamics. However, the reliance on inferred parameter values introduces greater uncertainty and potential for error compared to the parsimonious five-parameter basic SIR model. Even the simplest SIR model meets von Neumann's upper limit of five parameters and in order to preserve model simplicity, we take it as our point of departure.

In line with this argument, Shayak et al. [4] point out that a number of equally good fits may be achieved with a range of parameter values and argue that the standard deviation, mean time, and total number of cases represent the important characteristics in deciding goodness of fit. They go on to suggest that while greater compartmentalization beyond S, I and R may seem desirable, it potentially gives rise to overparameterization.

Numerous approaches have been implemented to estimate the simple SIR model parameters in the COVID-19 epidemic, for example [5]–[14]. Rica et al. [8] achieved an order-of-magnitude improvement in goodness of fit compared to a random search algorithm by applying a differential evolution optimisation algorithm to a single differential equation reformulation of the SIR model using COVID-19 data from Chile.

Liu at al. [13] developed a discrete-time Markov chain

approach to an SIR model to estimate the evolution of the COVID-19 pandemic in Wuhan. Despite the relative paucity of data, very early predictions of the course of the epidemic were feasible in terms of guiding public health decisions.

The difficulty in achieving adequate estimates for SIR parameters is not simply a function of data paucity. It is also determined by the type of data available. The I in SIR refers to infected, which is a binary category in that a person is either defined as infected or not.

It is not normally possible to know exactly when the infected state reverts to non-infected for any individual, as this would require a clear definition of what constitutes being "infected" which is often somewhat arbitrary. Also, even if this definition were less arbitrary, such as when it is based on the result of a polymerase chain reaction (PCR) test, it would require regular testing which is often not feasible.

For example, as a COVID-19 infection subsides, a time is reached when the PCR test becomes negative. However a positive test does not necessarily indicate active infection, and it is defined at a specific sensitivity of the cycle threshold of the PCR test [15], which itself may be somewhat arbitrary or be based on the notion of being infectious rather than infected. The additional issue of false positive and false negative tests adds another layer of complexity to the definition.

The incidence rate, namely the rate at which new infections occur per unit time, is a more attainable measure, however this does not provide the overall rate of change of the infected category, $I'(t)$, which, of course, depends on both the incidence rate as well as the recovery rate, which, as we have described, is fraught with ambiguity.

The need for parameter estimation as early as possible in an infectious epidemic is evident. In this paper, we describe the development of a novel family of estimators, and we show their utility in epidemiological parameter estimation in the context of an SIR model. We will consider the situation where the net number of infections per unit time is available. For several reasons, these numbers can never be known exactly and therefore we will view the reported net infection rate $\tilde{I}'(t)$ as a noise contaminated version of the true net number of infections, $I(t)$, expressed mathematically as follows,

$$\tilde{I}'(t) = I'(t) + n(t), \tag{4}$$

where $n(t)$ represent the unknown noise that accounts for the combined errors causing these inaccuracies.

II. Novel Estimators

In this section we present three of a novel family of estimators that we have derived which have application beyond just epidemiology, for example, in radar and sonar where nonlinear differential equations are encountered. However, in epidemiology, the SIR models provide an ideal vehicle for introducing and demonstrating these estimators. For this reason, we will be introducing these estimators using the simplest SIR model. Furthermore, these estimators can only be arrived at by considering continuous-time representations before introducing time-discretization. We begin by presenting the necessary mathematical devices.

A. Preliminaries

The two key analytical tools needed for the work presented here, are the Peano-Baker series and the Cauchy repeated integral formula. The Peano-Baker series provides a method for solving *linear, time-varying* (LTV) ordinary differential equations. For the simplest case, namely for first-order differential equations of the form,

$$x'(t) = a(t)\, x(t), \quad x(0) := x_0 \tag{5}$$

the Peano-Baker series takes the form,

$$x(t) = x_0\, \mathrm{e}^{\int_0^t a(\tau)\, \mathrm{d}\tau}. \tag{6}$$

The second of these analytical devices, is the Cauchy repeated integral formula,

$$\int_a^t \int_a^{\tau_1} \cdots \int_a^{\tau_{n-1}} f(\tau_n)\, \mathrm{d}\tau_n \mathrm{d}\tau_{n-1} \ldots \mathrm{d}\tau_1 = \frac{1}{(n-1)!} \int_a^t (t-\tau)^{n-1} f(\tau)\, \mathrm{d}\tau \tag{7}$$

Finally, since *digital signal processing* (DSP) algorithms are discrete by nature, we will resort to one of the simplest and most versatile methods for converting continuous-time expressions to equivalent discrete-time versions, namely the *forward difference approximation* to the derivative,

$$x'(t) \approx \frac{x(t+\Delta t) - x(t)}{\Delta t} \tag{8}$$

which yields the discrete-time expression

$$x[k] \approx x[k-1] + \Delta t\, x'[k-1], \tag{9}$$

where $x[k] := x(k\,\Delta t)$ and $x'[k] := x'(k\,\Delta t)$ with $k = 0, 1, \ldots$ the discrete-time equivalent of the continuous-time t. Below, we will use the standard notation and write $T := \Delta t$.

B. Estimator 1

As mentioned previously, we consider the situation where $\tilde{I}'(t)$, a noisy version of the net number of infections per unit time, is available and from which we can obtain the noisy version of its time-integral,

$$\tilde{I}(t) = \tilde{I}(0) + \int_0^t \tilde{I}'(\tau)\, \mathrm{d}\tau. \tag{10}$$

Considering the state equation (1) as a LTV with $I(t)$ contributing toward the coefficient of $S(t)$ there, the Peano-Baker series formula (5) followed by application of the Cauchy repeated integral formula (7), yields the expression,

$$\hat{S}_1(t) = S_0\, \mathrm{e}^{-\frac{\beta}{N}\left(I_0 t + \int_0^t (t-\tau)\tilde{I}'(\tau)\, \mathrm{d}\tau\right)} \tag{11}$$

Now we are in a position to convert this expression to discrete-time, using the forward difference method, to obtain

$$\hat{S}_1[m] = S_0\, \mathrm{e}^{-\frac{\beta T}{N}\left(I_0 m + T\sum_{n=0}^{m-1}(m-n)\tilde{I}'[n]\right)}, \tag{12}$$

which yields an estimator for $S_1[m]$ exclusively in terms of some parameters and of the past and present values of $\tilde{I}'$. In terms of its parameterization, (12) is linear in the primary parameter S_0 but nonlinear in the primary parameters β, N, I_0 and in the secondary parameter T that appears as a result of time-discretization. Observe that the estimator depends nonlinearly on the past and present values of $\tilde{I}'$.

Perhaps the most important observation is that there does not seem to exist any direct discrete-time equivalent method that would produce this estimator. Therefore, it was essential to work with the continuous-time version until a final expression for the estimator was obtained, before converting to discrete-time. Furthermore, even though we assumed uniform sampling here, the expression for the more general case of nonuniform sampling where $T_m := \Delta t_m$ follows easily. Another important aspect which we will discuss here, is that this estimator can be expressed in a computationally efficient recursive form where $\hat{S}_1[m]$ is expressed in terms of $\hat{S}_1[m-1]$.

C. Estimator 2

Once more, we assume we have only $\tilde{I}'(t)$, a noisy net number of infections per unit time and its time integral, and hence we can write (2) as

$$\tilde{I}'(t) \approx \frac{\beta}{N} S(t)\tilde{I}(t) - \gamma \tilde{I}(t), \tag{13}$$

from which $S(t)$ can be made the subject to obtain yet another estimator for $S(t)$,

$$S(t) \approx \frac{N}{\beta}\left(\frac{\tilde{I}'(t)}{\tilde{I}(t)} + \gamma\right) =: \hat{S}_2(t), \tag{14}$$

or, explicitly in terms of $\tilde{I}'$, namely

$$\hat{S}_2(t) = \frac{N}{\beta}\left(\frac{\tilde{I}'(t)}{I_0 + \int_0^t \tilde{I}'(\tau)\,\mathrm{d}\tau} + \gamma\right). \tag{15}$$

Using (9), we obtain the discrete-time estimator,

$$\hat{S}_2[m] = \frac{N}{\beta}\left(\frac{\tilde{I}'[m]}{I_0 + T\sum_{n=0}^{m-1} \tilde{I}'[n]} + \gamma\right). \tag{16}$$

This estimator depends nonlinearly on past and present values of $\tilde{I}'$ and, except for being independent of S_0, it depends nonlinearly on the same parameters as $\hat{S}_1[m]$ does, and also on γ.

D. Estimator 3

Under the same assumptions as for the previous two estimators, substituting (1) into (2) gives,

$$S'(t) \approx -\tilde{I}'(t) - \gamma \tilde{I}(t). \tag{17}$$

Integrating with respect to time, we get a third estimator for $S(t)$, namely.

$$\hat{S}_3(t) := S_0 + I_0 - \int_0^t \tilde{I}'(\tau)d\tau - \gamma\int_0^t (t-\tau)\tilde{I}'(\tau)\,\mathrm{d}\tau \tag{18}$$

from which we obtain the following discrete-time version of the estimator,

$$\hat{S}_3[m] = S_0 + I_0 - T\sum_{n=0}^{m-1} \tilde{I}'[n] - \gamma T^2 \sum_{n=0}^{m-1} (m-n)\tilde{I}'[n]. \tag{19}$$

As opposed to the previous two estimators, this estimator is linear in the present and past values of $\tilde{I}'$ and for fixed T, it depends linearly on S_0, I_0 and γ. As with estimator $\hat{S}_1$, this estimator can also be expressed recursively.

III. Estimators by Association

Since $S_0 = S(0)$, the initial number of susceptible individuals to an infectious disease in a given population, is extremely important for understanding the disease and for anticipating the impact of a new outbreak, we proceed to consider associations of the three estimators above. In this work, we studied all three possible associations of the above estimators but because of limited space, we only present the association of estimators 1 and 3. Technically the new estimator obtained through this association should be referred to as $\hat{S}_{1,3}(0|t)$ to emphasize that it produces an estimate based on the history of I' from time 0 up to and including t. However, to simplify notation we will simply write this as $\hat{S}_{1,3}(0)$

For the association of estimators $\hat{S}_1(t)$ and $\hat{S}_3(t)$, we assume that both $\hat{S}_1(t) = \hat{S}_3(t)$ and $\hat{S}_1(0) = \hat{S}_3(0)$ hold approximately, and define $\hat{S}_{1,3}(0) = (\hat{S}_1(0) + \hat{S}_3(0))/2$. The continuous-time estimator obtained from this is given by,

$$\hat{S}_{1,3}(0) = \frac{\tilde{I}(t) - \tilde{I}_0 + \gamma(\tilde{I}_0 t + \int_0^t (t-\tau)\tilde{I}'(\tau)\,\mathrm{d}\tau)}{1 - \exp\{-\frac{\beta}{N}(\tilde{I}_0 t + \int_0^t (t-\tau)\tilde{I}'(\tau)\,\mathrm{d}\tau)\}}. \tag{20}$$

Finally, using (9), the discrete-time version follows, namely

$$\hat{S}_{1,3}[0] = \frac{\tilde{I}[m] - \tilde{I}_0 + \gamma T(\tilde{I}_0 m + T\sum_{n=0}^{m-1}(m-n)\tilde{I}'[n])}{1 - \exp\{-\frac{\beta T}{N}(\tilde{I}_0 m + T\sum_{n=0}^{m-1}(m-n)\tilde{I}'[n])\}}, \tag{21}$$

where $\tilde{I}[m] = T\sum_{n=0}^{m-1} \tilde{I}'[n]$. This estimator is nonlinear in the history of I' as well as in its parameters, namely I_0, N, β and T.

We will now study these estimators numerically from an estimation theoretic perspective [16].

IV. Numerical Experiment

This section presents a numerical experiment conducted to characterize the novel estimators derived in section II. The experiment is described in section IV-A. The parameter values selected for the experiment are listed in section IV-B, and the results are presented in section IV-C.

A. Description of experiment

The numerical experiment uses the SIR model introduced in section I, with parameter values that are representative of a COVID-19 outbreak. In the experiment, the differential equations that define the SIR model are solved numerically for the susceptibility (S), infected (I) and recovered (R) counts as functions of time t (where $t \geq 0$ denotes the number of elapsed days since the start of the outbreak). These functions,

TABLE I
PARAMETER VALUES OF THE NUMERICAL EXPERIMENT.

Parameter	Symbol	Value
Infection rate	β	1/15 per day
Recovery rate	γ	4/10 per day
Total population size	N	7501 individuals
Initial number of susceptible individuals	S_0	7491 individuals
Initial number of infected individuals	I_0	10 individuals
Noise gain factor (low intensity)	κ	0.005
Noise gain factor (medium intensity)	κ	0.010
Noise gain factor (high intensity)	κ	0.015
Noise IIR filter pole	ρ	0.75
Sample period	T	1 day

referred to as the true counts, serve as the baseline (i.e., the ideal values) for comparison with the estimated values.

The experiment involves the computation of the true infected rate I' by substituting the true susceptibility count S and true infected count I into (2). An ensemble of B discrete-time noise-contaminated versions $\tilde{I}'_b[m] := I'(mT) + n_b[m]$, $b = 1, 2, \ldots, B$, of the true infected rate I' are computed. Each noise function n_b is a sample function of a zero-mean Gaussian random process filtered using an all-pole low-pass infinite impulse response (IIR) filter with a single real-valued pole $\rho > 0$. To simulate realistic noise-contaminated curves, where the intensity of the noise fluctuation is proportional to the count of infected individuals, the variance $\sigma^2[m]$ of the random variate $n_b[m]$ is selected to be proportional to $I(mT)$ according to the expression

$$\sigma^2[m] = \kappa I(mT), \tag{22}$$

where κ is called the *noise gain factor*. The power of the above noise process is proportional to the noise gain factor selected.

In the numerical experiment, the discrete-time estimators $\hat{S}_1$ and $\hat{S}_3$, and the discrete-time estimator $\hat{S}_{1,3}$ of S_0, are computed by substituting each noise-contaminated infected rate function $\tilde{I}'_b$ in the ensemble into (12), (19) and (21), respectively. The experiment characterizes each estimator according to its mean and variance, which are computed by averaging over the ensemble.

B. Experiment parameters

The parameters for the experiment are selected as outlined in Table I. The SIR model parameters selected were reported for the first COVID-19 outbreak in Wuhan, China [14]. In the experiment, all population counts are normalized such that a count of unity represents 10^3 individuals.

C. Results

Time plots of the susceptible (S), infected (I) and recovered (R) counts of the numerical example are plotted in Figure 1. The figure shows that count of infected individuals reaches its peak value on day 25, after which the infected count decays to zero. After 92 days, the number of infected individuals, as a fraction of the total population, falls below 1%.

The true infected rate I' and true infected count I of the numerical example are plotted as functions of time in Figure 2.

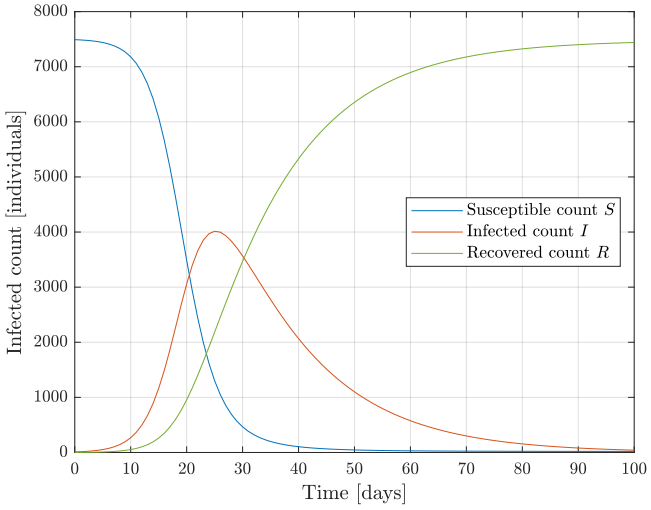

Fig. 1. Time plots of the susceptible (S), infected (I) and recovered (R) counts of the experiment before noise contamination.

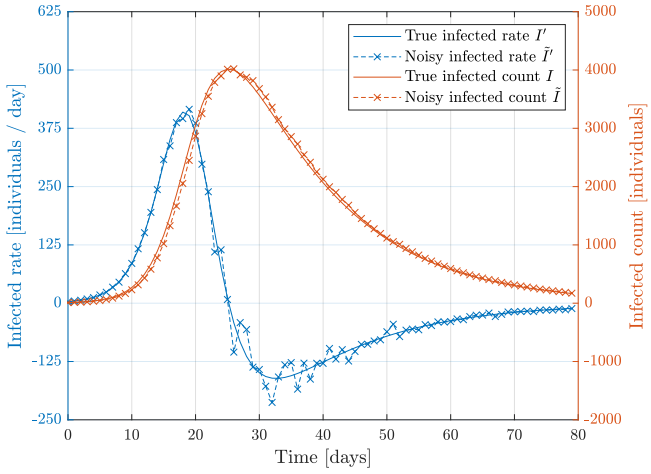

Fig. 2. Time plots of the true infected rate (I') and true infected (I) counts of the experiment, and noise-contaminated examples $\tilde{I}'$ and $\tilde{I}$.

The figure includes an example of a discrete-time noise-contaminated infected rate curve $\tilde{I}'$. This curve was generated using the process outlined in section IV-A, with a medium intensity noise gain factor $\kappa = 0.01$. The figure shows that the noise fluctuation is largest over the time period where the infection count reaches its peak value.

The noise-contaminated infected count curve $\tilde{I}$, obtained via discrete-time numerical integration of $\tilde{I}'$, is plotted in Figure 2. This curve exhibits a smaller noise fluctuation due to the low-pass filtering effect of integration. Furthermore, the noise-contaminated infected count curve exhibits a time delay relative to the ideal infected count curve. This time delay affects the accuracy of the estimators, as demonstrated below.

Figure 3 presents time plots of the bias and variance of the discrete-time estimator $\hat{S}_1$, as computed numerically for the low, medium and high intensity noise gain factors κ. The

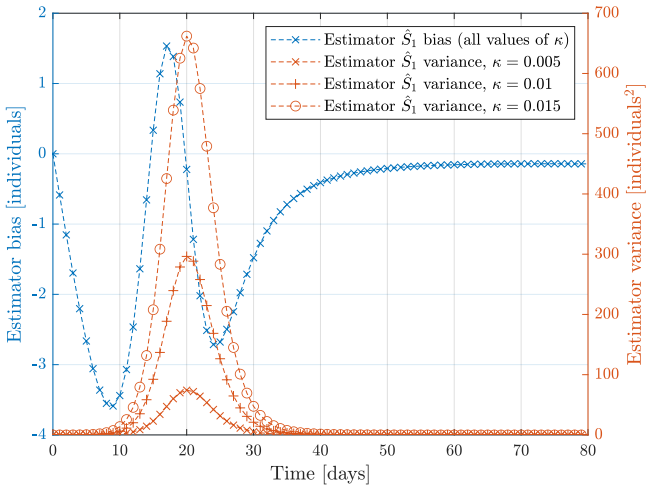

Fig. 3. Bias and variance of estimator $\hat{S}_1$, plotted as a function of time and for distinct noise gain factors κ.

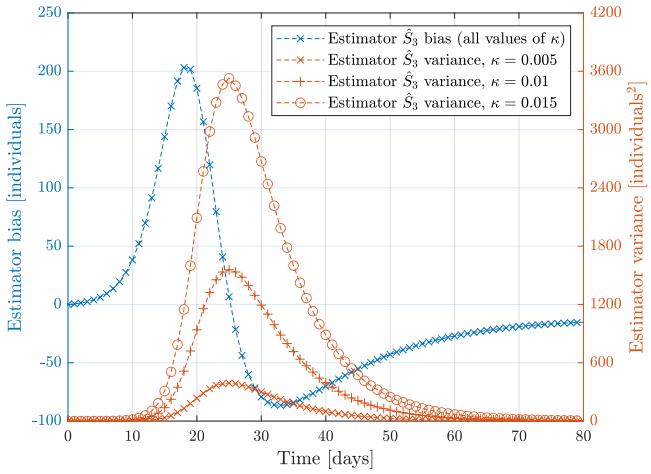

Fig. 4. Bias and variance of estimator $\hat{S}_3$, plotted as a function of time and for distinct noise gain factors κ.

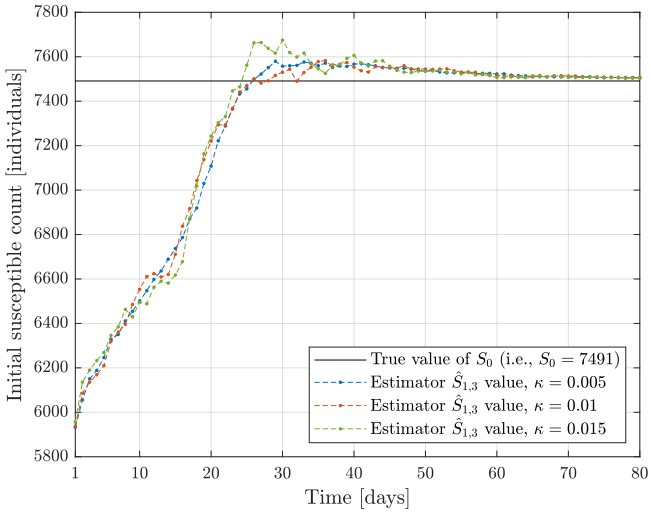

Fig. 5. Examples of time plots of the estimator $\hat{S}_{1,3} \approx S_0$, for distinct noise gain factors κ.

figure reveals that the bias of the estimator is small over the entire time period considered. Here, the peak bias value is equal to only 0.03% of the susceptible count (on day 17). This demonstrates the high accuracy of estimator $\hat{S}_1$.

The bias curve of estimator $\hat{S}_1$ in Figure 3 exhibits a large fluctuation over the time period that coincides with the peak of the infection rate curve $\tilde{I}'$ (days 10 to 30), but appears to converge slowly towards zero after day 30. It was found that this slow convergence is due to the time delay caused by discrete-time numerical integration of the noise-contaminated infected rate $\tilde{I}'$. Furthermore, the figure shows that the variance of estimator $\hat{S}_1$ reaches its peak on day 20. Larger estimator variance is observed as the noise gain factor κ is increased.

Figure 4 presents plots of the bias and variance of the discrete-time estimator $\hat{S}_3$, computed numerically for the low, medium and high intensity noise gain factors κ. A comparison of Figures 3 and 4 shows that estimator 3 exhibits a significantly larger bias and variance than estimator 1. However, the bias is not excessive; the peak bias value of estimator 3 is equal to 4.4% of the susceptible count (on day 18).

The bias curve of estimator $\hat{S}_3$ exhibits a large fluctuation over the same time period as estimator $\hat{S}_1$ exhibits maximum fluctuation. The figure shows that the bias of estimator $\hat{S}_3$ also exhibits slow convergence towards zero over time, due to the time delay caused by discrete-time numerical integration.

Figure 5 presents time plots of the discrete-time estimator $\hat{S}_{1,3}$ (here, the value of $\hat{S}_{1,3}$ on day m is an estimate of the initial susceptible count S_0 on day zero, as obtained by accumulating observations up to day m). Three curves, each corresponding to low, medium and high intensity noise gain factors, are plotted. The numerical results show that the estimator $\hat{S}_{1,3}$ converges to the true value of S_0 (plotted as the solid black line in the figure) over time, for each noise gain factor. Furthermore, an increase in the noise gain factor yields larger fluctuation in the estimator value over time.

Figure 6 presents time plots of the mean and variance of estimator $\hat{S}_{1,3}$, as computed numerically for the low, medium and high intensity noise gain factors κ. The figure shows that the bias of the estimator converges towards zero after day 33, and that a higher noise gain factor increases the variance of estimator $\hat{S}_{1,3}$. The variance of the estimator reaches its peak value on day 24. Thereafter, the estimator variance converges towards zero. The time convergence of the estimator bias and variance towards zero proves that the estimation accuracy improves as more observations are accumulated.

V. Conclusions

The numerical results for estimator bias and variance show that the novel estimators $\hat{S}_1$ and $\hat{S}_3$ yield accurate figures for the number of susceptible individuals over time, despite noise contamination of the observations. The novel estimator $\hat{S}_{1,3}$ also yields accurate figures for the initial number of susceptible individuals S_0. In particular, the results show that the accuracy

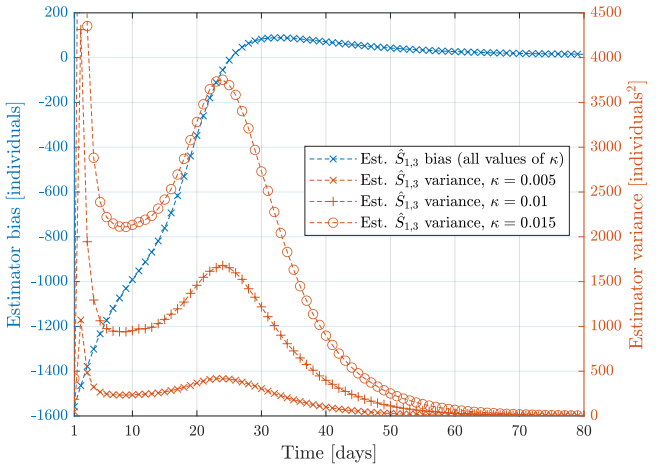

Fig. 6. Bias and variance of estimator $\hat{S}_{1,3} \approx S_0$, plotted as a function of time and for distinct noise gain factors κ.

of the latter estimator improves over time as more observations are accumulated.

While the time delay caused by discrete-time integration required to compute the estimates was identified as a source of estimator bias, further analysis showed that the impact of this error may be reduced by shortening the time between subsequent observations of the infected rate.

It is anticipated that the application of these and other estimators in the accurate determination of parameters at an early stage in infectious epidemics, will enhance the capacity to plan and allocate resources to mitigate the ravages on populations.

Planned future work on these estimators include detailed analysis and evaluation on real-world data.

Acknowledgment

We acknowledge the support of the Carl and Emily Fuchs Foundation.

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.

[2] J. M. Carcione, J. E. Santos, C. Bagaini, and J. Ba, "A simulation of a COVID-19 epidemic based on a deterministic SEIR model," *Frontiers in Public Health*, vol. 8, p. 230, 2020.

[3] A. H. Amiri Mehra, M. Shafieirad, Z. Abbasi, and I. Zamani, "Parameter estimation and prediction of COVID-19 epidemic turning point and ending time of a case study on SIR/SQAIR epidemic models," *Computational and Mathematical Methods in Medicine*, vol. 2020, no. 1, p. 1 465 923, 2020.

[4] B. Shayak, S. Jahedi, and J. A. Yorke, *Ambiguity in the use of sir models to fit epidemic incidence data*, 2024. arXiv: 2404.04181 [q-bio.PE]. [Online]. Available: https://arxiv.org/abs/2404.04181.

[5] M. Lounis and D. K. Bagal, "Estimation of SIR model's parameters of COVID-19 in Algeria," *Bulletin of the National Research Centre*, vol. 44, no. 1, p. 180, 2020.

[6] A. Alsayed, H. Sadir, R. Kamil, and H. Sari, "Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020," *International journal of environmental research and public health*, vol. 17, no. 11, p. 4076, 2020.

[7] D. Prodanov, "Analytical parameter estimation of the SIR epidemic model. applications to the COVID-19 pandemic," *Entropy*, vol. 23, no. 1, p. 59, 2020.

[8] S. Rica and G. A. Ruz, "Estimating SIR model parameters from data using differential evolution: An application with COVID-19 data," in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, IEEE, 2020, pp. 1–6.

[9] N. A. Kudryashov, M. A. Chmykhov, and M. Vigdorowitsch, "Analytical features of the SIR model and their applications to COVID-19," *Applied Mathematical Modelling*, vol. 90, pp. 466–473, 2021, ISSN: 0307-904X. DOI: https://doi.org/10.1016/j.apm.2020.08.057. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0307904X20304984.

[10] C. S. Carlson, D. M. Rubin, V. Heikkilä, and M. Postema, "Extracting transmission and recovery parameters for an adaptive global system dynamics model of the COVID-19 pandemic," in *2021 IEEE AFRICON*, IEEE, 2021, pp. 455–458.

[11] D. Prodanov, "Analytical solutions and parameter estimation of the SIR epidemic model," *Mathematical Analysis of Infectious Diseases*, pp. 163–189, 2022.

[12] F. G. Schmitt, "An algorithm for the direct estimation of the parameters of the SIR epidemic model from the i(t) dynamics," *The European Physical Journal Plus*, vol. 137, no. 1, p. 57, 2022.

[13] T. Liu, J. Huang, Z. He, *et al.*, "A real-world data validation of the value of early-stage SIR modelling to public health," *Scientific Reports*, vol. 13, no. 1, p. 9164, 2023.

[14] 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. 11 103–11 115, 2023.

[15] T. P. Velavan and C. G. Meyer, "COVID-19: A PCR-defined pandemic," *International Journal of Infectious Diseases*, vol. 103, pp. 278–279, 2021.

[16] H. L. van Trees, *Detection, estimation and modulation theory*, 2^{nd} ed., New York, NY: John Wiley & Sons, 2001.