Results in Physics 28 (2021) 104664

Contents lists available at ScienceDirect

Results in Physics

journal homepage: www.elsevier.com/locate/rinp

Estimation of epidemiological parameters for COVID-19 cases using a stochastic SEIRS epidemic model with vital dynamics

Olusegun M. Otunuga

Department of Mathematics, Augusta University, 1120 15th Street, GE-3018 Augusta, GA 30912, USA

ARTICLE INFO

Keywords:
Compartment disease model
Stochastic disease model
Local lagged adaptive generalized method of moments
Covid-19
Reproduction number
DELPHI model

ABSTRACT

We estimate and analyze the time-dependent parameters: transmission rate, symptomatic recovery rate, immunity rate, infection noise intensities, and the effective reproduction number for the United States COVID-19 cases for the period 01/22/2020-02/25/2021 using an innovative generalized method of moments estimation scheme. We assume the disease-dynamic is described by a stochastic susceptible–exposed–infected–recovered–susceptible (SEIRS) epidemic model, where the infected class is divided into the asymptomatic infected, and symptomatic infectious classes. Stochasticity appears in the model due to fluctuations in the disease's transmission and recovery rates. The disease eradication threshold is derived from the reproduction number. The estimated parameters are used to model the disease outbreak's possible trajectories. Our analysis reveals that current interventions are having positive effects on the transmission and recovery rates. The analysis is demonstrated using the daily United States COVID-19 infection and recovered cases for the period: 01/22/2020-02/25/2021.

Introduction

Several mathematical models [1–10] have been developed to study the transmission of the COVID-19 virus caused by the virus species "severe acute respiratory syndrome coronavirus", named SARS-CoV-2. The virus has caused over thirty million active cases and over 600 million deaths in the United States as of July, 2021. The airborne transmission occurs by inhaling droplets loaded with SARS-CoV-2 particles that are expelled by infectious people. According to the United States Centers for Disease Control and Prevention (CDC),[1] the best way to prevent the disease is to avoid being exposed to the virus. There are two other novel coronaviruses, namely, "severe acute respiratory syndrome coronavirus" (SARS-CoV) and the "Middle East respiratory syndrome coronavirus" (MERS-CoV) that emerged as major global health threats since 2002, according to Wu et al. [3]. Knowledge about the transmission of the infectious disease can help control the spread of the disease. Several vaccinations have been developed to eradicate or reduce the transmission of the virus. On December 11, 2020, the United States Food and Drug Administration (FDA) issued an Emergency Use Authorization (EUA) for the Pfizer-BioNTech Covid-19 vaccine to be used in the United States. The first dose of the vaccine was administered on December 14, 2020. This is the first vaccine administered in the United States. Seven days later, the FDA approved an EUA for the Moderna COVID-19 vaccine in the United States. The third vaccine, The Johnson and Johnson (J&J) vaccine, was first issued in the United States on February 27, 2021. As of July 24, 2021, only 68.7% of adults in the United States have taken at least one of the three vaccination. Data reported by CDC[2] shows that the daily count of total doses of the vaccines administered increases in trend from December 13, 2020 to February, 14, 2021, with a slight decline from February 14 to February 21, and later increases from then to February 25. We study in this work, by estimating and analyzing epidemiological parameters, the impact of these vaccines on the number of cases in the United States.

Parameter identification studies conditions under which observations of a modeled system can be used to identify the value of model parameters [11–13]. Time-dependent model parameters are difficult to identify in non-linear epidemic models. The Kalman filtering [14–16] and the generalized method of moments [17–19] are two well known estimation procedures for stochastic models. The transmission, recovery, loss of immunity rates of the Covid-19 virus are expected to vary with time and depend on many factors such as pneumonia seasonality, mobility, testing rates, mask use per capital weather [2], social behavior, strain-specific factors [11], public health interventions [5], etc. These rates are difficult to estimate, and cannot be

E-mail address: ootunuga@augusta.edu.

[1] https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.

[2] https://covid.cdc.gov/covid-data-tracker/#vaccination-trends. reported July, 24, 2021

[3] The United States Patent *Ladde et al*. with U.S. Patent Number 10719578 was approved on July 21, 2020. A one page summary of the Patent can be found at https://pdfpiw.uspto.gov/.piw?Docid=10719578.

https://doi.org/10.1016/j.rinp.2021.104664
Received 28 October 2020; Received in revised form 4 August 2021; Accepted 5 August 2021
Available online 11 August 2021

estimated from laboratory experiments or surveys of infected populations. In this work, by assuming the transmission of the Covid-19 virus follows a stochastic susceptible–exposed–infected–recovered–susceptible stochastic epidemic model, and further dividing the infected class into asymptomatic and symptomatic class, we use the patented scheme[3] called the lagged adapted generalized method of moments (LLGMM) parameter identification technique [20] to estimate the time-dependent disease reproduction number, time-dependent disease transmission, temporary recovery, temporary immunity rates, and time-dependent infection noise intensities for the COVID-19 virus. We demonstrate the LLGMM procedure for the time-dependent parameters in a stochastic susceptible–exposed–infectious–recovered–susceptible ($SEIRS$) epidemic model using Covid-19 data for the year 2020.

The paper is organized as follows: The deterministic and corresponding stochastic $SEIRS$ models governing the transmission of the Covid-19 disease are described in Section "The SEIRS model". The LLGMM scheme is described in Section "The time varying state and parameter estimation scheme" and it is used to estimate the epidemiological time-varying parameters in the stochastic model. The application of the scheme to $SEIRS$ epidemic model for Covid-19 is discussed in Section "Case Study: COVID-19". The basic reproduction number and effective reproduction number are also estimated and fitted in Section "Case Study: COVID-19". The summary of the work done is discussed in Section "Discussion and further studies".

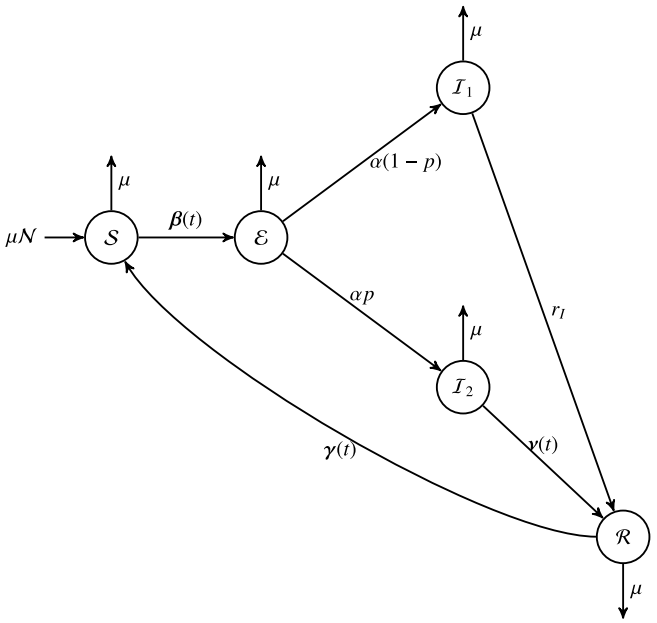

Fig. 1. $SEIRS$ model schematic.

The SEIRS model

We use a deterministic susceptible, infected but not infectious (exposed), asymptomatic infected, symptomatic infectious, and recovered model, with compartments denoted by $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{R}$, respectively, to describe the transmission of the Covid-19 disease in the United States. By exposed, we mean someone who got infected (but not infectious) by being in close contact with someone who has Covid-19. A full list of what counts as close contact is provided by the Centers for Disease Control and Prevention (CDC).[4] The $SEIRS$ model with vital rates governing the transmission of the virus is described by the following deterministic system of nonlinear differential equations:

$$
\begin{aligned}
d\mathcal{S} &= \left(\mu\mathcal{N} - \boldsymbol{\beta}(t)\frac{\mathcal{S}\left(\eta\mathcal{I}_1+\mathcal{I}_2\right)}{\mathcal{N}} - \mu\mathcal{S} + \boldsymbol{\gamma}(t)\mathcal{R}\right)dt, && \mathcal{S}(t_0)=\mathcal{S}_0,\\
d\mathcal{E} &= \left(\boldsymbol{\beta}(t)\frac{\mathcal{S}\left(\eta\mathcal{I}_1+\mathcal{I}_2\right)}{\mathcal{N}} - (\alpha+\mu)\mathcal{E}\right)dt, && \mathcal{E}(t_0)=\mathcal{E}_0,\\
d\mathcal{I}_1 &= \left((1-p)\alpha\mathcal{E} - \left(r_I+\mu\right)\mathcal{I}_1\right)dt, && \mathcal{I}_1(t_0)=\mathcal{I}_{10},\\
d\mathcal{I}_2 &= \left(\alpha p\mathcal{E} - (\boldsymbol{\nu}(t)+\mu)\mathcal{I}_2\right)dt, && \mathcal{I}_2(t_0)=\mathcal{I}_{20},\\
d\mathcal{R} &= \left(\boldsymbol{\nu}(t)\mathcal{I}_2 + r_I\mathcal{I}_1 - (\boldsymbol{\gamma}(t)+\mu)\mathcal{R}\right)dt, && \mathcal{R}(t_0)=\mathcal{R}_0,
\end{aligned}
\tag{1}
$$

where $\mathcal{S}_0 > 0$, $\mathcal{E}_0 > 0$, $\mathcal{I}_{10} > 0$, $\mathcal{I}_{20} > 0$, $\mathcal{R}_0 > 0$, and $\mathcal{S}(t)$, $\mathcal{E}(t)$, $\mathcal{I}_1(t)$, $\mathcal{I}_2(t)$, $\mathcal{R}(t)$ are the population of susceptible, infected but not infectious (exposed), asymptomatic, symptomatic infectious, and temporarily immune (temporarily recovered) individuals, respectively, at time $t \geq t_0$. Here, $\boldsymbol{\beta}(t)$ is the time-dependent transmission or contact rate in response to public health interventions. We assume every individual is in one of these compartments at t_0 with $\mathcal{N} = \mathcal{S}_0+\mathcal{E}_0+\mathcal{I}_{10}+\mathcal{I}_{20}+\mathcal{R}_0$, where $\mathcal{N}$ is the population size of the United States, so that $\mathcal{S}+\mathcal{E}+\mathcal{I}_1+\mathcal{I}_2+\mathcal{R} = \mathcal{N}$ for all time t. The transmission rate, symptomatic recovery rate, and temporary immune rate $\boldsymbol{\beta}(t)$, $\boldsymbol{\nu}(t)$, and $\boldsymbol{\gamma}(t)$, respectively, are assumed to be a positive, bounded, and continuous function of time t. The latency and death rates α and μ, respectively, are each positive constants, with α being the transition rate from the exposed to infected class. The parameter $p \geq 0$ is the fraction of infection cases that are symptomatic, and r_I is the asymptomatic recovery rate. The parameter η accounts for the infectiousness of asymptomatic individuals relative to symptomatic. According to the CDC,[5] a parameter value of $\eta = 50\%$ means an asymptomatic individual is half as infectious as a symptomatic individual, whereas a parameter value of 100% means that an asymptomatic individual is just as likely to transmit infection as a symptomatic individual. See Fig. 1 for a schematic of the SEIRS model described here.

We assume that the transmission, symptomatic recovery, and temporary immune rates $\boldsymbol{\beta}(t)$, $\boldsymbol{\nu}(t)$, and $\boldsymbol{\gamma}(t)$, respectively, of the Covid-19 virus are functions of time and are unknown. We also assume the asymptomatic recovery rate r_I is unknown. By studying the trajectories of the number of Covid-19 infection cases in the United States, we noticed, from the spikes in the trajectory of the number of cases, that it is influenced by some external noises. These noises can be caused by many factors such as the rates at which Covid-19 testing is done, vaccination rates, variability in the number of contacts between infected and susceptible individuals, mask use per capital, social behavior, public health intervention, and so on. Since no much information is known about the temporary immune rates $\boldsymbol{\gamma}(t)$ of the virus, we only assume in this study that the transmission and recovery rates are influenced by Gaussian white noise processes causing the rates to fluctuate around the function $\beta(t)$ and $\nu(t)$, respectively. As mentioned above, external fluctuations may be caused by many factors such as public health interventions and variability in the number of contacts between infected and susceptible individuals, and such random variations can be modeled by a white noise. Thus we assume

$$
\begin{aligned}
\boldsymbol{\beta}(t) &= \beta(t) + \sigma_1(t)\zeta_1(t),\\
\boldsymbol{\nu}(t) &= \nu(t) + \sigma_2(t)\zeta_2(t),
\end{aligned}
\tag{2}
$$

where $\zeta_i(t)$, $i = 1,2$ are independent Gaussian white noise term with mean 0 such that $\zeta_i(t)dt = dW_i(t)$, $\sigma_i(t)$, $i = 1,2$, are the noise intensities for the transmission and symptomatic recovery rates, respectively, $W_i(t)$, $i = 1,2$, are independent standard Weiner process defined on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, $(\mathcal{F}_{t_0})$ is measurable and $(\mathcal{F}_t)_{t\geq 0}$ is right continuous. We also assume that $(\mathcal{S}_0, \mathcal{E}_0, \mathcal{I}_{10}, \mathcal{I}_{20}, \mathcal{R}_0)$ is $\mathcal{F}_{t_0}$-measurable and independent of $W_i(t) - W_i(t_0)$,

[4] https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/quarantine.html, accessed 03.07.2021.

[5] Centers for Disease Control and Prevention, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html, accessed 03.06.2021.

$i = 1, 2$. By substituting (2) into (1), the dynamics describing the system is interpreted using Stratonovich $SEIRS$ equation with vital rates and stochasticity as

$$\begin{aligned}
dS &= \left(\mu\mathcal{N} - \beta(t)\frac{S(\eta\mathcal{I}_1 + \mathcal{I}_2)}{\mathcal{N}} - \mu S + \gamma(t)\mathcal{R}\right)dt \\
&\quad - \sigma_1(t)\frac{S(\eta\mathcal{I}_1 + \mathcal{I}_2)}{\mathcal{N}} \circ dW_1(t), && S(t_0) = S_0, \\
d\mathcal{E} &= \left(\beta(t)\frac{S(\eta\mathcal{I}_1 + \mathcal{I}_2)}{\mathcal{N}} - (\alpha + \mu)\mathcal{E}\right)dt \\
&\quad + \sigma_1(t)\frac{S(\eta\mathcal{I}_1 + \mathcal{I}_2)}{\mathcal{N}} \circ dW_1(t), && \mathcal{E}(t_0) = \mathcal{E}_0, \\
d\mathcal{I}_1 &= \left(\alpha(1-p)\mathcal{E} - (r_I + \mu)\mathcal{I}_1\right)dt, && \mathcal{I}_1(t_0) = \mathcal{I}_{10}, \\
d\mathcal{I}_2 &= \left(\alpha p\mathcal{E} - (\nu(t) + \mu)\mathcal{I}_2\right)dt - \sigma_2(t)\mathcal{I}_2 \circ dW_2(t), && \mathcal{I}_2(t_0) = \mathcal{I}_{20}, \\
d\mathcal{R} &= \left(r_I\mathcal{I}_1 + \nu(t)\mathcal{I}_2 - (\gamma(t) + \mu)\mathcal{R}\right)dt + \sigma_2(t)\mathcal{I}_2 \circ dW_2(t), && \mathcal{R}(t_0) = \mathcal{R}_0.
\end{aligned} \tag{3}$$

where $\circ$ is the Stratonovich integral symbol. We direct readers interested in a review of stochastic modeling with systems of differential equations, and the interpretation of stochastic dynamic systems as a Stratonovich system to the work of Kloeden [21], Ladde [22], and Méndez et al. [23]. Model (3) performs better than existing SEIRS deterministic model in the sense that it is able to capture fluctuations in the number of susceptible, exposed, infections, and recovered cases. Another advantage the model has over other existing model is the fact that it captures the trajectories of the susceptible, exposed, infected, and recovered population well because the transmission, recovery, and the immunity rates are extracted for each given time rather than as a constant over time. We convert the population sizes $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{R}$ to percentages using the transformation

$$S = \mathcal{S}/\mathcal{N}, \quad E = \mathcal{E}/\mathcal{N}, \quad I_1 = \mathcal{I}_1/\mathcal{N}, \quad I_2 = \mathcal{I}_2/\mathcal{N}, \quad R = \mathcal{R}/\mathcal{N},$$

respectively. Using these transformations, Eq. (3) reduces to

$$\begin{aligned}
dS &= \left(\mu - \beta(t)S(\eta I_1 + I_2) - \mu S + \gamma(t)R\right)dt \\
&\quad - \sigma_1(t)S(\eta I_1 + I_2) \circ dW_1(t), && S(t_0) = S_0, \\
dE &= \left(\beta(t)S(\eta I_1 + I_2) - (\alpha + \mu)E\right)dt \\
&\quad + \sigma_1(t)S(\eta I_1 + I_2) \circ dW_1(t), && E(t_0) = E_0, \\
dI_1 &= \left(\alpha(1-p)E - (r_I + \mu)I_1\right)dt, && I_1(t_0) = I_{10}, \\
dI_2 &= \left(\alpha pE - (\nu(t) + \mu)I_2\right)dt - \sigma_2(t)I_2 \circ dW_2(t), && I_2(t_0) = I_{20}, \\
dR &= \left(r_I I_1 + \nu(t)I_2 - (\gamma(t) + \mu)R\right)dt + \sigma_2(t)I_2 \circ dW_2(t), && R(t_0) = R_0.
\end{aligned} \tag{4}$$

We convert (4) into Itô stochastic differential equation as

$$\begin{aligned}
dS &= \left(\mu - \beta(t)S(\eta I_1 + I_2) - \mu S + \gamma(t)R + \frac{\sigma_1^2(t)}{2}S(\eta I_1 + I_2)^2\right)dt \\
&\quad - \sigma_1(t)S(\eta I_1 + I_2)dW_1(t), && S(t_0) = S_0, \\
dE &= \left(\beta(t)S(\eta I_1 + I_2) - (\alpha + \mu)E - \frac{\sigma_1^2(t)}{2}S(\eta I_1 + I_2)^2\right)dt \\
&\quad + \sigma_1(t)S(\eta I_1 + I_2)dW_1(t), && E(t_0) = E_0, \\
dI_1 &= \left(\alpha(1-p)E - (r_I + \mu)I_1\right)dt, && I_1(t_0) = I_{10}, \\
dI_2 &= \left(\alpha pE - (\nu(t) + \mu)I_2 + \frac{\sigma_2^2(t)}{2}I_2\right)dt \\
&\quad - \sigma_2(t)I_2 dW_2(t), && I_2(t_0) = I_{20}, \\
dR &= \left(r_I I_1 + \nu(t)I_2 - (\gamma(t) + \mu)R - \frac{\sigma_2^2(t)}{2}I_2\right)dt \\
&\quad + \sigma_2(t)I_2 dW_2(t), && R(t_0) = R_0.
\end{aligned} \tag{5}$$

It can be shown, using ideas presented in Mummert and Otunuga [11], that $SEIRS$ stochastic differential equation (5) has a unique (up to equivalence) global solution and the solution will remain within $[0, 1]^5$ whenever it starts from there. Hence, the epidemiologically reasonable region in the SEIRS plane is given by

$$T = \left\{(S, E, I_1, I_2, R) \mid \ 0 \le S + E + I_1 + I_2 + R \le 1\right\}.$$

We assume that all model parameter values are known, except $\beta(t), \nu(t), \gamma(t)$, r_I, and $\sigma_i(t)$, $i = 1, 2$. Also we assume that the infectious $I_1(t)$, $I_2(t)$ and recovered $R(t)$ population data are known and noise-free at discrete time points t_j and can be modeled with the $SEIRS$ model (5). The goal of the work done here is to identify the time-dependent parameters $\beta(t), \nu(t), \gamma(t)$ and $\sigma_i(t)$, $i = 1, 2$, effective and basic reproduction number $\mathcal{R}_e(t)$ and $\mathcal{R}_0$, respectively, specifically at each time t_j for the Covid-19 infection when I_1, I_2, and R are known. To do this, we need to first estimate $E(t)$ and $S(t)$ at each time t using the local lagged adapted generalized method of moments (LLGMM) parameter estimation scheme described in the patent work of Ladde and Otunuga [20]. The estimated parameters are used to model the disease outbreak's possible trajectories.

The time varying state and parameter estimation scheme

In this section, we describe, assuming the infected $I_1(t)$, $I_2(t)$, and recovered $R(t)$ population data are known, how the exposed $E(t)$ and susceptible $S(t)$ population, together with the transmission $\beta(t)$, recovery $\nu(t)$, and temporary immune $\gamma(t)$ rates are estimated using the LLGMM scheme. These estimates are used to calculate the basic and effective reproduction numbers of Covid-19 cases for the United States for the time period: March 1, 2020 to February 25, 2021. The LLGMM method described in the patent is composed of the following components: (1) development of a stochastic mathematical model of the continuous time dynamic process, (2) construction of an equivalent time series model, (3) development of generalized method of moment/observation (GMM) equations, (4) introduction of a conceptual and computational parameter estimation scheme, (5) introduction of a conceptual and computational state estimation scheme, and (6) derivation of ϵ-best sub-optimal state and parameter estimates. The first component is satisfied since we have a continuous stochastic $SEIRS$ epidemic model governing the transmission of the Covid-19 virus. In order to construct equivalent time series model to the stochastic model, we need to be able to specify which state and parameters to estimate. The susceptible and exposed states are unknown, and the models governing the dynamics of the known infected and recovered population depend on the exposed population, recovery rate $\nu(t)$, and the temporary immune rate $\gamma(t)$. We will estimate the states $E(t)$ and $S(t)$ and parameters $\beta(t)$, r_I, $\nu(t)$, and $\gamma(t)$ in two stages. In the first stage, we apply the LLGMM scheme to the stochastic model governing the recovered population $R(t)$ to estimate $\nu(t)$ and $\gamma(t)$. These estimated parameters are now used in the model governing $I_1(t)$ and $I_2(t)$ to estimate the state $E(t)$. The LLGMM scheme is now applied to the model governing $E(t)$ to estimate the state $S(t)$, asymptomatic recovery rate r_I, and the transmission rate $\beta(t)$. In the second stage, we use the estimated states $E(t)$ and $S(t)$, together with the known states $I_1(t)$, $I_2(t)$ and $R(t)$ to update the parameters $\beta(t)$, $\nu(t)$, and $\gamma(t)$, and later use these estimates to simulate the states $S(t)$, $E(t)$, $I_1(t)$, $I_2(t)$, and $R(t)$, and the effective and basic reproduction numbers $\mathcal{R}_e$ and $\mathcal{R}_0$, respectively. The ϵ-best sub-optimal state and parameter estimates are now recorded.

State and parameter estimation process

We explain the state and parameter estimation scheme in two stages in the following subsections. In Section 'State and parameter estimation process: Stage 1', we estimate $\nu(t)$, $\gamma(t)$, r_I, $\sigma_j(t)$, $j = 1, 2$, and $\beta(t)$, together with the state estimates $E(t)$ and $S(t)$ satisfying (5). These estimates are updated in stage 2 of the estimation process in Section 'State and parameter estimation process: Stage 2'. The ϵ-best sub-optimal estimates are derived in Section 'Derivation of ϵ-best sub-optimal state and parameter estimates'.

State and parameter estimation process: Stage 1

Let $U(t) = S(t) + E(t)$. We can extract the data for $U(t)$ using the fact that $S + E = 1 - I_1 - I_2 - R$. The model governing αE is obtained from the model governing S and E in (5) and substituted into the model governing $I_2(t)$ as follows:

$$\begin{aligned}\alpha E &= -\frac{dU}{dt} + \mu - \mu U + \gamma(t)R,\\ dI_2 &= -p\,dU + \left(p(\mu - \mu U + \gamma(t)R) - (\nu(t)+\mu)I_2 + \frac{\sigma_2^2(t)}{2}I_2\right)dt\\ &\quad - \sigma_2(t)I_2 dW_2(t).\end{aligned}$$

In order to estimate the states $E(t)$ and $S(t)$, we first need to estimate the parameters $\nu(t)$ and $\gamma(t)$. To do this, we setup two time series model corresponding to the symptomatic infectious and recovered class as follows:

$$\begin{aligned}dI_2 &= -p\,dU + \left(p(\mu - \mu U + \gamma(t)R) - \left(\nu(t) + \mu - \frac{\sigma_2^2(t)}{2}\right)I_2\right)dt\\ &\quad - \sigma_2(t)I_2 dW_2(t),\\ dR &= \left(r_I I_1 + \left(\nu(t) - \frac{\sigma_2^2(t)}{2}\right)I_2 - (\gamma(t)+\mu)R\right)dt + \sigma_2(t)I_2 dW_2(t).\end{aligned} \tag{6}$$

We first explain the discretization scheme as discussed in Otunuga (2017) [24] and Otunuga (2019) [25]. Let $x = (S, E, I_1, I_2, R)$ be continuous time stochastic process defined on the interval $[t_0 - \tau, T]$, where τ is a continuous delay constant (with corresponding discrete delay constant r). For $t \in [t_0 - \tau, T]$, let $\mathcal{F}_t$ be an increasing sub-sigma algebra of a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ for which $x(t)$ is $\mathcal{F}_t$ measurable. Let $\mathbb{P}$ be a partition of the interval $[t_0 - \tau, T]$ defined by $\mathbb{P} = \{t_k : t_k = t_0 + k\Delta t,\ k \in K(-r, N)\}$, where $K(a,b) = \{i \in \mathbb{Z} : a \le i \le b,\ a, b \in \mathbb{Z}\}$ and $\mathbb{Z}$ is the set of integers. Define m_k as the local admissible sample data observation size at time t_k. For $r \ge 1$, $k \in K(0, N)$, and each $m_k \in K(2, r)$. A m_k- point sub-partition P_k of the closed interval $[t_{k-m_k+1}, t_k]$ is defined as $P_k := \{t_{k-m_k+1}, t_{k-m_k+2}, \ldots, t_k\}$. We estimate the parameters $\nu(t)$ and $\gamma(t)$ at time t_j using known data $(I_2(t))_{t\in P_j}$ and $(R(t))_{t\in P_j}$ by first integrating (6) on the interval $[t_{j-m_j+1}, t_j]$ to get

$$\begin{aligned}\int_{t_{j-m_j+1}}^{t_j} dI_2 &= -p\int_{t_{j-m_j+1}}^{t_j} dU(t)\\ &\quad + \int_{t_{j-m_j+1}}^{t_j}\left(p(\mu - \mu U(t) + \gamma(t)R) - \left(\nu(t) + \mu - \tfrac{\sigma_2^2(t)}{2}\right)I_2(t)\right)dt\\ &\quad - \int_{t_{j-m_j+1}}^{t_j}\sigma_2(t)I_2(t)dW_2(t),\\ \int_{t_{j-m_j+1}}^{t_j} dR(t) &= \int_{t_{j-m_j+1}}^{t_j}\left(r_I I_1(t) + \left(\nu(t) - \tfrac{\sigma_2^2(t)}{2}\right)I_2(t) - (\gamma(t)+\mu)R(t)\right)dt\\ &\quad + \int_{t_{j-m_j+1}}^{t_j}\sigma_2(t)I(t)dW_2(t).\end{aligned} \tag{7}$$

Assuming $\nu(t)$ and $\gamma(t)$ are continuous functions on the closed interval $[t_0 - \tau, T]$ (and hence on $[t_{j-m_j+1}, t_j]$), it follows from the integral mean value theorem that there exist $t_j^*, t_j^{**} \in (t_{j-m_j+1}, t_j)$ such that

$$\int_{t_{j-m_j+1}}^{t_j}\gamma(t)R(t)dt = \gamma(t_j^{**})\int_{t_{j-m_j+1}}^{t_j}R(t)dt,$$

$$\int_{t_{j-m_j+1}}^{t_j}\left(\nu(t) - \frac{\sigma_2^2(t)}{2}\right)I_2(t)dt = \left(\nu(t_j^*) - \frac{\sigma_2^2(t_j^*)}{2}\right)\int_{t_{j-m_j+1}}^{t_j}I_2(t)dt.$$

Therefore,

$$\begin{aligned}\int_{t_{j-m_j+1}}^{t_j} dI_2(t) &= -p\int_{t_{j-m_j+1}}^{t_j} dU(t) + \int_{t_{j-m_j+1}}^{t_j} p(\mu - \mu U(t))\,dt\\ &\quad + p\gamma(t_j^{**})\int_{t_{j-m_j+1}}^{t_j}R(t)dt - \left(\nu(t_j^*) - \frac{\sigma_2^2(t_j^*)}{2}\right)\int_{t_{j-m_j+1}}^{t_j}I_2(t)dt\\ &\quad - \mu\int_{t_{j-m_j+1}}^{t_j}I_2(t)dt - \int_{t_{j-m_j+1}}^{t_j}\sigma_2(t)I_2(t)dW_2(t),\end{aligned}$$

$$\begin{aligned}\int_{t_{j-m_j+1}}^{t_j} dR(t) &= r_I\int_{t_{j-m_j+1}}^{t_j}I_1(t)dt + \left(\nu(t_j^*) - \frac{\sigma_2^2(t_j^*)}{2}\right)\\ &\quad \times\int_{t_{j-m_j+1}}^{t_j}I_2(t)dt - \left(\gamma(t_j^{**}) + \mu\right)\int_{t_{j-m_j+1}}^{t_j}R(t)dt\\ &\quad + \int_{t_{j-m_j+1}}^{t_j}\sigma_2(t)I_2(t)dW_2(t),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left[\left(\int_{t_{j-m_j+1}}^{t_j} d\sqrt{I_2(t)} - \mathbb{E}\left(\int_{t_{j-m_j+1}}^{t_j} d\sqrt{I_2(t)}\Big|\mathcal{F}_t\right)\right)^2\Big|\mathcal{F}_t\right]\\ &= \frac{\sigma_2^2(t_j^*)}{4}\mathbb{E}\left[\int_{t_{j-m_j+1}}^{t_j}I_2(t)dt\Big|\mathcal{F}_t\right].\end{aligned}$$

It is clear from the integrals above that the estimates $\nu(t_j^*)$ and $\gamma(t_j^{**})$ are calculated using m_j data set at and to the left of time t_j. Since these estimates are calculated using m_j observation size at and to the left of t_j, we write $\nu(t_j^*) \equiv \hat{\nu}_{t_j,m_j}$, $\gamma(t_j^{**}) \equiv \hat{\gamma}_{t_j,m_j}$, $\sigma_2(t_j^*) \equiv \hat{\sigma}_2(t_j, m_j)$. The third component of the LLGMM scheme is satisfied by discretizing the above integral equations and applying conditional expectation to obtain

$$\begin{aligned}\sum_{k=j-m_j+1}^{j}\mathbb{E}\left[\Delta I_{2_k}\Big|\mathcal{F}_{k-1}\right] &= -p\sum_{k=j-m_j+1}^{j}\mathbb{E}\left[\Delta U_k\Big|\mathcal{F}_{k-1}\right]\\ &\quad + p\sum_{k=j-m_j+1}^{j}\left(\mu - \mu U_{k-1}\right)\Delta t + p\hat{\gamma}_{t_j,m_j}\sum_{k=j-m_j+1}^{j}R_{k-1}\Delta t\\ &\quad - \left(\nu(t_j^*) - \frac{\sigma_2^2(t_j^*)}{2}\right)\sum_{k=j-m_j+1}^{j}I_{2_{k-1}}\Delta t,\\ \sum_{k=j-m_j+1}^{j}\mathbb{E}\left[\Delta R_k\Big|\mathcal{F}_{k-1}\right] &= r_I\sum_{k=j-m_j+1}^{j}I_{1_{k-1}}\Delta t + \left(\hat{\nu}_{t_j,m_j} - \frac{\hat{\sigma}_2^2(t_j,m_j)}{2}\right)\\ &\quad \times\sum_{k=j-m_j+1}^{j}I_{2_{k-1}}\Delta t - \left(\hat{\gamma}_{t_j,m_j} + \mu\right)\sum_{k=j-m_j+1}^{j}R_{k-1}\Delta t,\end{aligned} \tag{8}$$

where $I_{1_{k-1}} = I_1(t_{k-1})$, $I_{2_{k-1}} = I_2(t_{k-1})$, $R_{k-1} = R(t_{k-1})$, $U_{k-1} = U(t_{k-1})$, $\Delta I_{2_k} = I_2(t_k) - I_2(t_{k-1})$, $\Delta R_k = R(t_k) - R(t_{k-1})$, and $\hat{\nu}_{t_j,m_j}, \hat{\gamma}_{t_j,m_j}, \hat{\sigma}_2(t_j, m_j)$ are parameter estimates of $\nu(t), \gamma(t)$, and $\sigma(t)$, respectively, at time t_j. The unique solution $\hat{\nu}_{t_j,m_j}$, $\hat{\gamma}_{t_j,m_j}$, and $\hat{\sigma}_2(t_j, m_j)$ of (8) is obtained as Eq. (9) which is given in Box I.

Let $\Delta W_{j,m_j}^l$ be a Wiener process with mean 0 and variance Δt, where $j \in K(1, N)$ for sample size N, $l \in K(1, L)$ for L simulations, and $m_j \in K(2, r)$ is the observation size. Let $\hat{E}_{t_j,m_j}^l$ be the lth simulated exposed population estimate at time t_j using m_j observation size, $l \in K(1, L)$. Let $I_1(t_j) = I_{1_j}$, $I_2(t_j) = I_{2_j}$, and $R(t_j) = R_j$ be the known asymptomatic infected, symptomatic infected, and recovered population at time t_j, respectively, and set $\hat{I}_{2_{t_j,m_j}}^l = I_{2_j}$ and $\hat{R}_{t_j,m_j}^l = R_j$. We extract $\hat{E}_{t_j,m_j}^l$ from the discretized model governing the infected population in (5) using Monte-Carlo method and Euler scheme as follows:

$$\begin{aligned}\hat{E}_{j-1,m_{j-1}}^l &= \left(\Delta I_2(t_j) + \left(\hat{\nu}_{j-1,m_{j-1}} + \mu - \frac{\hat{\sigma}_2^2(t_{j-1}, m_{j-1})}{2}\right)I_2(t_{j-1})\Delta t\right.\\ &\quad \left. + \hat{\sigma}_2(t_{j-1}, m_{j-1})I_2(t_{j-1})\Delta W_{j,m_j}^l\right)/(\alpha\Delta t),\end{aligned}$$

where $\Delta I_2(t_j) = I_2(t_j) - I_2(t_{j-1})$, $\Delta W_{j,m_j}^l = W_{j,m_j}^l - W_{j-1,m_{j-1}}^l$. We take the average

$$\hat{E}_{j,m_j} = \frac{1}{L}\sum_{l=1}^{L}\hat{E}_{j,m_j}^l$$

as the estimated value of $E(t)$ at time t_j using m_j observation size. From this, the simulated value $\hat{S}_{j,m_j}$ of the susceptible population $S(t)$ at time t_j using m_j observation size is calculated as

$$\hat{\gamma}_{t_j,m_j} = \frac{\left(p\sum_{k=j-m_j+1}^{j}(\mu-\mu U_{k-1}) - \mu\sum_{k=j-m_j+1}^{j} I_{2_{k-1}} + r_I\sum_{k=j-m_j+1}^{j} I_{1_{k-1}}\right)\Delta t - \sum_{k=j-m_j+1}^{j}\mathbb{E}\left[p\Delta U_k + \Delta I_{2_k} + \Delta R_k\middle|\mathcal{F}_{k-1}\right]}{(1-p)\sum_{k=j-m_j+1}^{j} R_{k-1}\Delta t} - \frac{\mu}{1-p},$$

$$\hat{\nu}_{t_j,m_j} = \frac{\left(p\sum_{k=j-m_j+1}^{j}(\mu-\mu U_{k-1}) - \mu p\sum_{k=j-m_j+1}^{j} R_{k-1} + p\, r_I\sum_{k=j-m_j+1}^{j} I_{1_{k-1}}\right)\Delta t - \sum_{k=j-m_j+1}^{j}\mathbb{E}\left[p\Delta U_k + \Delta I_{2_k} + p\Delta R_k\middle|\mathcal{F}_{k-1}\right]}{(1-p)\sum_{k=j-m_j+1}^{j} I_{2_{k-1}}\Delta t} - \frac{\mu}{1-p} + \frac{\hat{\sigma}_2^2(t_j,m_j)}{2},$$

$$\hat{\sigma}_2^2(t_j,m_j) = \frac{4\sum_{k=j-m_j+1}^{j}\left(\Delta\sqrt{I}_{2_k} - \frac{1}{m_j}\sum_{k=j-m_j+1}^{j}\Delta\sqrt{I}_{2_k}\right)^2}{\sum_{k=j-m_j+1}^{j} I_{2_{k-1}}\Delta t}. \tag{9}$$

Box I.

$$\hat{S}_{j,m_j} = 1 - \hat{E}_{j,m_j} - I_{1_j} - I_{2_j} - R_j. \tag{10}$$

Define

$$\Xi^1_{j,m_j} = \left(\hat{E}_{j,m_j} - aI_{1_j} - bI_{2_j}\right)^2,$$

where a and b are constants. The estimates $\hat{m}_j$, $\hat{a}$, $\hat{b}$, and $\hat{r}_I$ of m_j, a, b, and r_I that minimize the square Ξ^1_{j,m_j} are calculated and used to obtain the stage 1 estimate for the exposed class $E(t)$ as $\hat{E}_{j,\hat{m}_j}$ for each $j \in K(1,N)$. The stage 1 estimate $\hat{S}_{j,\hat{m}_j}$ for the susceptible class $S(t)$ is calculated using (10) as

$$\hat{S}_{j,\hat{m}_j} = 1 - \hat{E}_{j,\hat{m}_j} - I_{1_j} - I_{2_j} - R_j.$$

Using the fact that there exists constants $\bar{t}_j, \bar{\bar{t}}_j \in (t_{j-m_j+1}, t_j)$ such that

$$\int_{t_{j-m_j+1}}^{t_j}\beta(t)S(t)\left(\eta I_1(t) + I_2(t)\right)dt = \beta(\bar{t}_j)\int_{t_{j-m_j+1}}^{t_j} S(t)\left(\eta I_1(t) + I_2(t)\right)dt,$$

$$\int_{t_{j-m_j+1}}^{t_j}\frac{\sigma_1^2(t)}{2}S^2(t)\left(\eta I_1(t) + I_2(t)\right)^2 dt = \frac{\sigma_1^2(\bar{\bar{t}}_j)}{2}\int_{t_{j-m_j+1}}^{t_j} S^2(t)\left(\eta I_1(t) + I_2(t)\right)^2 dt,$$

we estimate the remaining parameters $\beta(t)$ and $\sigma_1(t)$ by integrating the equation governing $E(t)$ in (5) on (t_{j-m_j+1}, t_j) as follows:

$$\begin{aligned}\int_{t_{j-m_j+1}}^{t_j} dE(t) &= \beta(\bar{t}_j)\int_{t_{j-m_j+1}}^{t_j} S(t)\left(\eta I_1(t) + I_2(t)\right)dt - (\alpha+\mu)\int_{t_{j-m_j+1}}^{t_j} E(t)dt \\ &\quad - \int_{t_{j-m_j+1}}^{t_j}\frac{\sigma_1^2(t)}{2}S(t)\left(\eta I_1(t) + I_2(t)\right)^2 dt \\ &\quad + \int_{t_{j-m_j+1}}^{t_j}\sigma_1(t)S(t)\left(\eta I_1(t) + I_2(t)\right)dW_1(t).\end{aligned} \tag{11}$$

Following the same argument above, by discretizing and taking the conditional expectation of integrals in (11), the estimates $\beta(\bar{t}_j) \equiv \beta_{j,m_j}$ and $\hat{\sigma}_1(t_j,m_j)$ of $\beta(t)$ and $\sigma_1(t)$, respectively, at time t_j are obtained as Eq. (12) (see Box II).

State and parameter estimation process: Stage 2

In stage 2, we updated the exposed and susceptible state estimates $\hat{E}_{j,\hat{m}_j}$ and $\hat{S}_{j,\hat{m}_j}$, and parameter estimates $\hat{\nu}_{j,\hat{m}_j}$ and $\hat{\gamma}_{j,\hat{m}_j}$ of $\nu(t)$ and $\gamma(t)$, respectively. Similar to the outline above, let ν_{j,m_j} and γ_{j,m_j} be the updated estimates for the parameters $\nu(t)$ and $\gamma(t)$, respectively, at time t_j using m_j data sets at and to the left of t_j. The updated infectious rate ν_{j,m_j} and temporary immune rate γ_{j,m_j} at time t_j are obtained as

$$\nu_{j,m_j} = \frac{\alpha p\sum_{k=j-m_j+1}^{j}\hat{E}_{k-1,\hat{m}_{k-1}}\Delta t + \frac{\hat{\sigma}_2^2(t_j,m_j)}{2}\sum_{k=j-m_j+1}^{j} I_{2_{k-1}}\Delta t - \sum_{k=j-m_j+1}^{j}\mathbb{E}\left[\Delta I_{2_k}|\mathcal{F}_{k-1}\right]}{\sum_{k=j-m_j+1}^{j} I_{2_{k-1}}\Delta t} - \mu,$$

$$\gamma_{j,m_j} = \frac{r_I\sum_{k=j-m_j+1}^{j} I_{1_{k-1}} + \left(\nu_{j,m_j} - \frac{\hat{\sigma}_2^2(t_j,m_j)}{2}\right)\sum_{k=j-m_j+1}^{j} I_{2_{k-1}}\Delta t - \sum_{k=j-m_j+1}^{j}\mathbb{E}[\Delta R_k|\mathcal{F}_{k-1}]}{\sum_{k=j-m_j+1}^{j} R_{k-1}\Delta t} - \mu. \tag{13}$$

In order to update the state values at time t_j, we discretize the Itô differential Eqs. (5) using the implicit Euler scheme [26,27] as described in [11]. For the stochastic model

$$dy = f(t,y,\theta(t))dt + g(t,y,\theta(t))dW(t),$$

where $y = (S,E,I_1,I_2,R)$, f and g represent the drift and diffusion coefficient of system (5) with time-dependent parameters θ, the implicit Euler scheme gives the discretize lth simulated solution $y^l_{j+1} \equiv y^l_{j,m_j}$ satisfying

$$y^l_{j+1} - y^l_j = \varepsilon f(t_{j+1}, y^l_{j+1}, \theta^l_{j+1})\Delta t + (1-\varepsilon)f(t_j, y^l_j, \theta^l_j)\Delta t + g(t_j, y^l_j, \theta^l_j)\Delta W^l_j,$$

for $0 \le \varepsilon \le 1$, where $j = 0,1,2,\ldots,N$ for sample size N, $l = 1,2,\ldots,L$ for L simulations in the Monte-Carlo method, $\theta^l_j \equiv \theta^l_{j,m_j} = \{\beta^*_j, \nu^*_j, \gamma^*_j, \sigma^*_1(t_j), \sigma^*_2(t_j), \eta, \alpha, r_I, \mu\}$ as described in (9), (12), and (13). The simulated state $y^l_j \equiv y^l_{j,m_j}$ for the lth simulation at time t_j using m_j data set at and to the left of m_j can be determined iteratively and the estimate y_{j+1} (regarded as the simulated state estimate) is calculated by taking the average $y_{j+1} = \frac{1}{L}\sum_{l=1}^{L} y^l_{j+1}$.

Derivation of ϵ-best sub-optimal state and parameter estimates

We estimated, among the values $\{y^l_j \equiv y^l_{j,m_j}\}_{m_j=2}^{r}$ at time t_j, the value closest to the known real value. Specifically, the correct delay constant m_j must be determined. Let S^l_{j,m_j}, E^l_{j,m_j}, $I^l_{1_{j,m_j}}$, $I^l_{2_{j,m_j}}$ and R^l_{j,m_j} be the simulated susceptible, exposed, asymptomatic, symptomatic infectious and recovered estimate for the lth simulation at time t_j using m_j observations at and to the left of t_j. We take the average

$$S_{j,m_j} = \frac{1}{L}\sum_{l=1}^{L} S^l_{j,m_j}; \quad E_{j,m_j} = \frac{1}{L}\sum_{l=1}^{L} E^l_{j,m_j}; \quad I_{1_{j,m_j}} = \frac{1}{L}\sum_{l=1}^{L} I^l_{1_{j,m_j}};$$

$$I_{2_{j,m_j}} = \frac{1}{L}\sum_{l=1}^{L} I^l_{2_{j,m_j}}; \quad R_{j,m_j} = \frac{1}{L}\sum_{l=1}^{L} R^l_{j,m_j}, \tag{14}$$

$$\hat{\sigma}_1^2(t_j, m_j) = \frac{\sum_{k=j-m_j+1}^{j} \left(\Delta \hat{E}_{k,\hat{m}_k} - \frac{1}{m_j} \sum_{n=j-m_j+1}^{j} \Delta \hat{E}_{n,\hat{m}_n} \right)^2}{\sum_{k=j-m_j+1}^{j} \hat{S}_{k-1,\hat{m}_{k-1}}^2 \left(\eta I_{1_{k-1}} + I_{2_{k-1}} \right)^2 \Delta t},$$

$$\beta_{j,m_j} = \frac{\sum_{k=j-m_j+1}^{j} \mathbb{E}\left[\Delta \hat{E}_{k,\hat{m}_k} | \mathcal{F}_{k-1} \right] + (\alpha+\mu) \sum_{k=j-m_j+1}^{j} \hat{E}_{k-1,\hat{m}_{k-1}} \Delta t + \frac{\hat{\sigma}_1^2(t_j,m_j)}{2} \sum_{k=j-m_j+1}^{j} \hat{S}_{k-1,\hat{m}_{k-1}} \left(\eta I_{1_{k-1}} + I_{2_{k-1}} \right)^2}{\sum_{k=j-m_j+1}^{j} \hat{S}_{k-1,\hat{m}_{k-1}} \left(\eta I_{1_{k-1}} + I_{2_{k-1}} \right) \Delta t}. \quad (12)$$

Box II.

as the simulated value of $S(t)$, $E(t)$, $I_1(t)$, $I_2(t)$, and $R(t)$ at time t_j using m_j observations at and to the left of t_j. Define

$$\Xi_{j,m_j} = \left(S_{t_j} - S_{j,m_j} \right)^2 + \left(E_{t_j} - E_{j,m_j} \right)^2 + \left(I_{1_{t_j}} - I_{1_{j,m_j}} \right)^2 + \left(I_{2_{t_j}} - I_{2_{j,m_j}} \right)^2 + \left(R_{t_j} - R_{j,m_j} \right)^2 \quad (15)$$

as the quadratic mean square error between the known data $\{\hat{S}_{j,\hat{m}_j}, \hat{E}_{j,\hat{m}_j}, I_{1_{t_j}}, I_{2_{t_j}}, R_{t_j}\}$ and the averaged realizations $\{S_{j,m_j}, E_{j,m_j}, I_{1_{j,m_j}}, I_{2_{j,m_j}}, R_{j,m_j}\}_{m_j=2}^{r}$. For any arbitrary small positive number ϵ and for each time t_j, we define the ϵ-sub-optimal admissible set of m_j at t_j as

$$\mathcal{M}_j = \{ m_j \ : \ \Xi_{j,m_j} < \epsilon \}. \quad (16)$$

If $m_j \in \mathcal{M}_j$ gives the minimum value for Ξ_{j,m_j}, then we record m_j as m_j^*. If condition (16) is not met at time t_j, then the value of m_j where the minimum $\min_{m_j} \Xi_{j,m_j}$ is attained is recorded as m_j^*. The ϵ-best sub-optimal parameter and state estimates at time t_j are now recorded as $\theta_{j,m_j^*} = \{\beta_{j,m_j^*}, \nu_{j,m_j^*}, \gamma_{j,m_j^*}, \sigma_1(t_j, m_j^*), \sigma_2(t_j, m_j^*)\}$ and $\{S_{j,m_j^*}, E_{j,m_j^*}, I_{1_{j,m_j^*}}, I_{2_{j,m_j^*}}, R_{j,m_j^*}\}$, respectively.

Case study: COVID-19

Data source

We apply the LLGMM scheme to the $SEIRS$ epidemic model governing the United States COVID-19 cases. The daily COVID-19 cases data was collected from CDC[6] and Center for Systems Science and Engineering (CSSE) at Johns Hopkins University[7] for the COVID-19 periods 04/01/2020 through 03/06/2021. As explained by CDC, the numbers are confirmed COVID-19 cases as reported by U.S. states, U.S. territories, New York City, and the District of Columbia from the previous day. The recovered COVID-19 data cases were collected from the COVID Tracking Project[8] for the COVID-19 periods 04/01/2020 through 02/27/2021 and compared with the CSSE daily data.

The CDC COVID-19 data is given daily, so that $\Delta t = 1$ day. According to CDC, the incubation period for COVID-19 is thought to extend to 14 days, with a median time of 4-5 days of exposure to symptoms onset.[9] McAloon et al. [28] showed in their work that the time from exposure to symptom onset is 6 days. Also, in their analysis based on confirmed cases outside Wuhan, Lauer et al. and Linton et al. [29,30] estimated the median incubation period to be 5.0 days with a range of 2 to 14 days, similar to 5.2 days incubation period with symptom onset between 3 to 6 days reported by Li et al. [31]. For these reasons, measuring time in days, we assume α should be selected in the range $[1/14, 1/2]$ per day and used a default value of $1/6$ day^{-1}. The measure of the infectiousness of asymptomatic cases to symptomatic cases remains highly uncertain. According to the parameter values for the five COVID-19 Pandemic Planning Scenarios produced by CDC,[10] the infectiousness of asymptomatic individuals relative to symptomatic is estimated to be $\eta = 0.75$. From the CIA World Factbook,[11] the current population of the United States as at July 2020 is $332,639,102$. So, we set $\mathcal{N} = 332,639,102$. Also, The annual U.S. birth rates for the year 2020 is 12.4 births per 1000 population, while the death rate is 8.3 death per 1000 population and life expectancy is 80.3 years.[12] Therefore, we set μ to be in the range $\left[\frac{83}{3650000}, \frac{1}{80.3\times365}\right]$. A complete list of values used in the simulation process is given in Table 1.

Parameter and state estimation: Numerical simulation

The COVID-19 infected[6] and recovered[8] data plot are shown in Fig. 2(a) and (b), respectively. Using the $SEIRS$ model, and setting the latency rate $\alpha = 1/6$, the exposed population $\{E(t_j)\}_{j=r+1}^{N}$ is estimated in stage 1 as $\{\hat{E}_{j,\hat{m}}\}_{j=r+1}^{N}$ using the LLGMM scheme with delay constants $r = 90$ days (with $2 \le m_j \le 90$) and $r = 150$ days (with $2 \le m_j \le 150$). The estimates at time t_j are calculated using m_j past data points at and to the left of t_j. These values are estimated in stage 1 of the estimation stage in Section 'State and parameter estimation process: Stage 1' and values plotted in red color in Fig. 4. The Susceptible population $\{S(t_j)\}_{j=r+1}^{N}$ is estimated in stage 1 as $\{\hat{S}_{j,\hat{m}}\}_{j=r+1}^{N}$ using delay constants $r = 90$ days (with $2 \le m_j \le 90$) and $r = 150$ days (with $2 \le m_j \le 150$) with the fact that $S + E + I_1 + I_2 + R = \mathcal{N}$. These values are plotted in red color in Fig. 3. It can be seen from the estimated exposed and susceptible population that the trend of the infected cases increases (decreases) in the months when the number of those who are exposed also increases(decreases) and the number of those susceptible to COVID-19 decreases (increases).

The estimated exposed and susceptible populations $\{\hat{E}_{j,\hat{m}}\}_{j=r+1}^{N}$ and $\{\hat{S}_{j,\hat{m}}\}_{j=r+1}^{N}$, respectively, in stage 1, together with the real infected and recovered population are used to estimate the noise infected intensity $\{\hat{\sigma}_1(t_j, m_j)\}_{j=r+1}^{N}$ and time-dependent transmission rate $\{\beta_{j,m_j}\}_{j=r+1}^{N}$ in stage 1 and also used to update the time-dependent recovery rate $\{\nu_{j,m_j}\}_{j=r+1}^{N}$ and temporary immune rate $\{\gamma_{j,m_j}\}_{j=r+1}^{N}$ in stage 2 of the estimation process (see discussions in Sections 'State and parameter estimation process: Stage 1' and 'State and parameter estimation process: Stage 2'). These time varying parameters are later used in stage 2 in Section 'State and parameter estimation process: Stage 2' to simulate the susceptible, exposed, infected and recovered population. We note here that there are m_j estimates for each values

[6] https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36, accessed 02.27.2021.

[7] https://github.com/CSSEGISandData/COVID-19/tree/master/, accessed 03.07.2021.

[8] https://covidtracking.com/data, accessed 02.27.2021.

[9] https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html#, accessed 10.30.2020.

[10] https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#table-2, accessed 03.06.2021.

[11] https://www.cia.gov/the-world-factbook/countries/united-states/, accessed 03.07.2021

[12] CIA World Factbook, https://www.cia.gov/library/publications/the-world-factbook/, accessed 09.29.2020.

Table 1
SEIRS model parameter values used in application of the LLGMM state and parameter identification scheme.

Parameter	Description	Units	Range	References
α	Latency rate	day^{-1}	$[\frac{1}{14}, \frac{1}{2}]$[b]	[29–31]
μ	Birth / death rate	day^{-1}	$\left[\frac{83}{3650000}, \frac{1}{80.3\times365}\right]$	CIA[c]
η	Asymptomatic infectiousness	day^{-1}	0.75	CDC[a]
r_I	Asymptomatic recovery rate	day^{-1}	1/10.5	Estimated
$\beta(0)$, $\nu(0)$, $\gamma(0)$	Initial conditions		0	
$\sigma_1(0)$, $\sigma_2(0)$	Initial conditions		0	
r	Maximum allowed past data points		90, 150 days	
$\mathcal{N}$	United States population as at July 2020		33, 26, 39, 102	CIA[b]
N	Number of days from 4/1/2020–2/25/2021		205	

[a] https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#table-2, accessed 03.06.2021.
[b] https://www.cia.gov/the-world-factbook/countries/united-states/, accessed 03.07.2021.
[c] CIA World Factbook, https://www.cia.gov/library/publications/the-world-factbook/, accessed 09.29.2020.

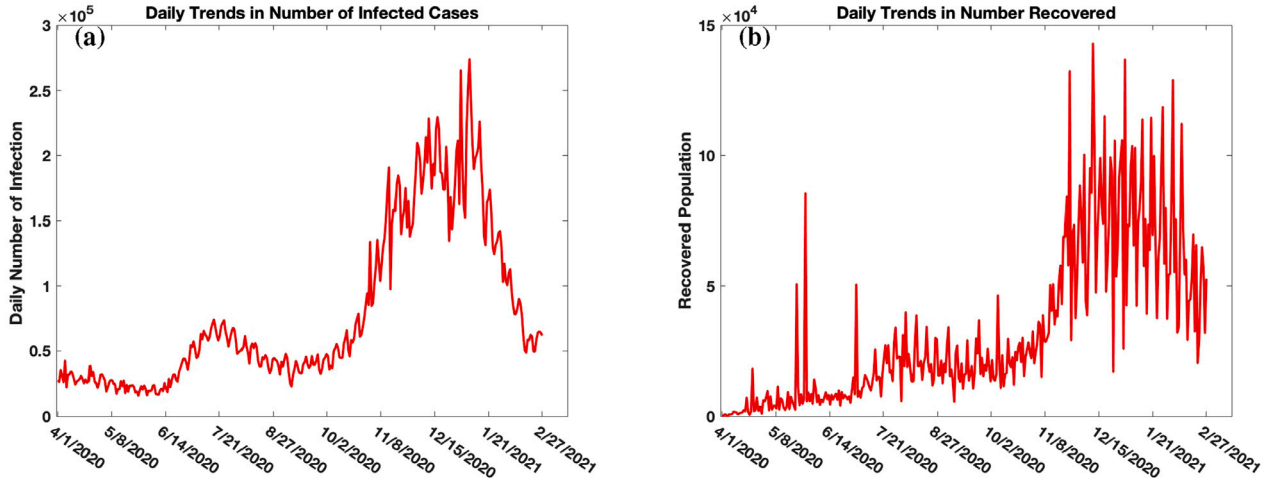

Fig. 2. Real infected[6] and recovered[8] population.

S_{j,m_j}, E_{j,m_j}, $I_{1_{j,m_j}}$, $I_{2_{j,m_j}}$, and R_{j,m_j}. The best estimates, otherwise known as the ϵ-best sub-optimal parameter and state estimates, denoted $\{\beta_{j,m_j^*}\}_{j=r+1}^N$, $\{\nu_{j,m_j^*}\}_{j=r+1}^N$, $\{\gamma_{j,m_j^*}\}_{j=r+1}^N$, $\{\sigma_1(t_j,m_j^*)\}_{j=r+1}^N$, $\{\sigma_2(t_j,m_j^*)\}_{j=r+1}^N$ and $\{S_{j,m_j^*}\}_{j=r+1}^N$, $\{E_{j,m_j^*}\}_{j=r+1}^N$, $\{I_{1_{j,m_j^*}}\}_{j=r+1}^N$, $\{I_{2_{j,m_j^*}}\}_{j=r+1}^N$, $\{R_{j,m_j^*}\}_{j=r+1}^N$, respectively, are calculated in Section 'Derivation of ϵ-best sub-optimal state and parameter estimates' by choosing the observation size m_j^* that satisfies condition (16).

The simulated susceptible $\{S_{j,m_j^*}\}_{j=r+1}^N$ and exposed $\{E_{j,m_j^*}\}_{j=r+1}^N$ population are shown (in blue) in Figs. 3 and 4, respectively. The results in Fig. 3(a) and (b) are derived using maximum delay constants $r = 90$ and $r = 150$ days, respectively, corresponding to 3 and 5 months past data. We noticed a decline in the number of susceptible individuals in the time periods 10/01/2020-12/24/2020. These periods fall around the festive Thanksgiving and Christmas in the United States when people travel around the country. We also notice an increasing trend in the number of individuals susceptible to the disease from December 2020 to February 2021. These period coincides with the period when the daily count of total Covid-19 vaccine doses administered is also increasing. Further research is being done to analyze the correlation between these two occurrences and to discuss factor(s) contributing to the incline and decline in the susceptible population.

The blue plot in Fig. 4 is the trajectory of the simulated exposed population $\{E_{j,m_j^*}\}_{j=r+1}^N$. The side by side comparison of the trajectories of the estimated $\{\hat{E}_{j,\hat{m}_j}\}_{j=r+1}^N$ and simulated $\{E_{j,m_j^*}\}_{j=r+1}^N$ exposed population in Fig. 4 shows that the LLGMM algorithm fits well and the algorithm seems to prefer using many past data points as seen in Fig. 16, with better result in Fig. 4(b) where the maximum delay constant is $r = 150$ days. We see that the higher the maximum delay constant r, the better the simulation results. The results in Fig. 4(a) and (b) are derived using maximum delay constants $r = 90$ and $r = 150$, respectively.

Fig. 5(a) shows the comparison of the real[6] (red) and simulated $\{I_{j,m_j^*}\}_{j=r+1}^N = \{I_{1_{j,m_j^*}}\}_{j=r+1}^N + \{I_{2_{j,m_j^*}}\}_{j=r+1}^N$ (blue) infected infectious population using delay constant $r = 90$. Fig. 5(b) shows the comparison of the real[6] (red) and simulated $\{I_{j,m_j^*}\}_{j=r+1}^N$ (blue) infected infectious population using delay constant $r = 150$. The number of infected and exposed cases rise in the month of October, November and December. As explained above, these periods fall around the festive Thanksgiving and Christmas in the United States when people travel around the country. We noticed a decline in the number of exposed and infected individuals starting around late December 2020 to late February 2021 when the daily count of total Covid-19 vaccine doses administered is increasing. Analysis is currently being done to investigate if there is any correlation between the daily counts of exposed, infected, recovered individuals and the daily count of total Covid-19 vaccine doses administered.

Fig. 6(a) shows the comparison of the recovered[8] (red) and simulated $\{R_{j,m_j^*}\}_{j=r+1}^N$ (blue) recovered population using delay constant $r = 90$. Fig. 6(b) shows the comparison of the real (red) and simulated $\{R_{j,m_j^*}\}_{j=r+1}^N$ (blue) recovered population using delay constant $r = 150$.

Fig. 7(a)–(b) shows the trajectory of the transmission rate $\{\beta_{j,m_j^*}\}_{j=r+1}^N$ using delay constants $r = 90$ days and $r = 150$ days, respectively. We see that on the average, there are spikes in the transmission rate during the month of November and early December, 2020 when the infection cases rise. This also confirms that the transmission of Covid-19 cases rises during the Thanksgiving and Christmas period in

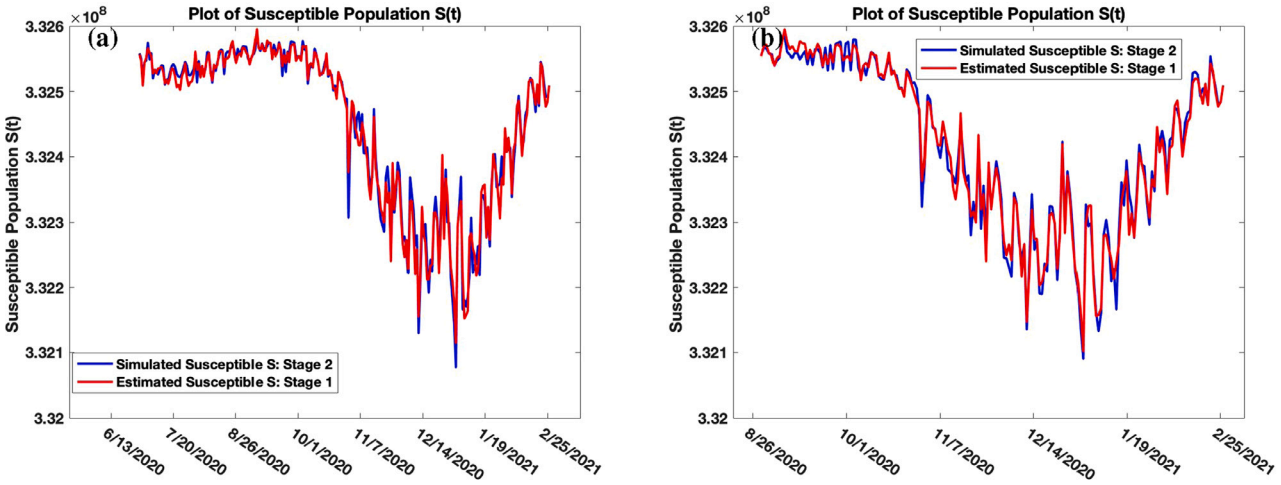

Fig. 3. Simulated (blue) and estimated (red) susceptible values with maximum time delays $r = 90$ and $r = 150$ days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

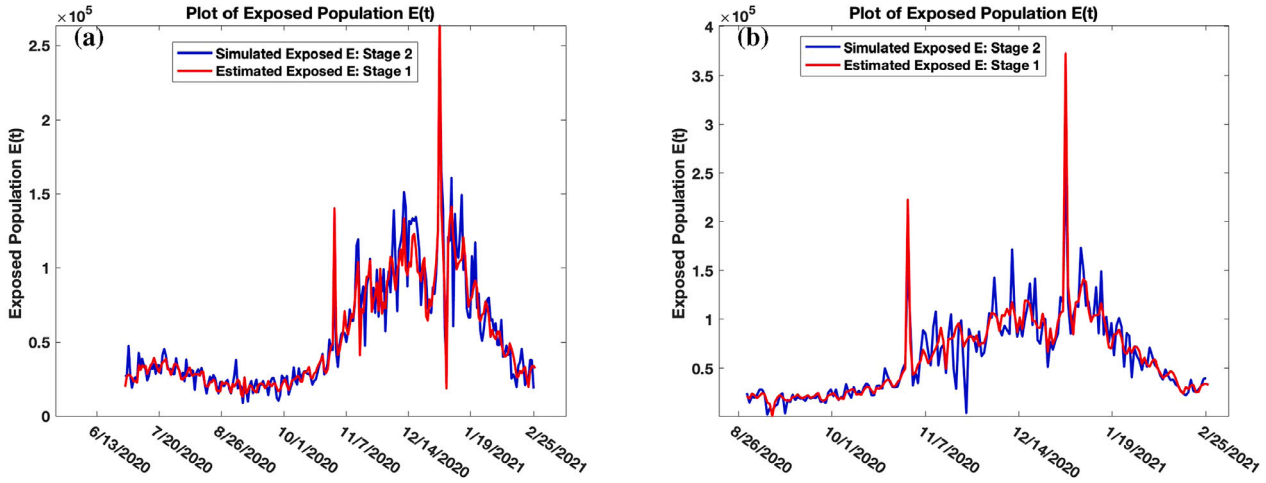

Fig. 4. Simulated (blue) and estimated exposed values (red) with maximum time delays $r = 90$ and $r = 150$ days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the United States. As discussed in the introduction section, the FDA[13] issued an Emergency Use Authorization (EUA) for the Pfizer-BioNTech, the Moderna, and the Johnson & Johnson COVID-19 vaccines on December 11, 2020, December 18, 2020, and February 27, 2021, respectively, in the United States. We see a decline in the transmission rate of the disease (as evidenced by the decline in the estimated and fitted trajectory of $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$ and $\bar{\beta}(t)$ in Figs. 7 and 9, respectively,) starting from late December, 2020. This decrease occured after vaccines were administered. The average of the transmission rate using $r = 90$ days delay constant is 0.1073 per day. In their work, Lingzhi et al. [32] used the DELPHI v2p8 (Differential Equations Leads to Predictions of Hospitalizations and Infections) model to predict $\beta(t)$ and claim it is of the form $\frac{2}{\pi}\arctan\left(-\frac{(t-\bar{a})}{\bar{b}}\right)+1$, for some constants $\bar{a}$ and $\bar{b}$. We verify the validity of their claim by fitting the equation

$$\bar{\beta}(t) = a + b\arctan\left(-(t-c)\right). \tag{17}$$

[13] https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-frequently-asked-questions, accessed 03.09.2021.

Table 2
Parameter for the transmission rate curve fitting using the DELPHI model.

Parameter	Estimate	95% confidence interval
a	0.7485	(0.5169, 0.9802)
b	0.4151	(0.2664, 0.5638)
c	−1.831	(−3.271, −0.3916)
RMSE	0.0697	

to the data $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$. The parameters a, b, and c are estimated in Table 2 and the trend of $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$ compared with the DELPHI model in Fig. 8(a).

Also in their work, Eikenberry et al. [33] and Tang et al. [34] claim that $\beta(t)$ follows a decreasing exponential model of the form

$$\bar{\bar{\beta}}(t) = \beta_1 + (\beta_0 - \beta_1)e^{-r(t-t_0)}, \tag{18}$$

where β_0 is the initial contact rate at the initial time t_0 (July 1, 2020), β_1 is the minimum contact rate, and r is the rate at which contact decreases. We check to see if our result agrees with theirs by fitting equation (18) to the data $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$. The parameters r, β_i, $i = 0, 1$ are

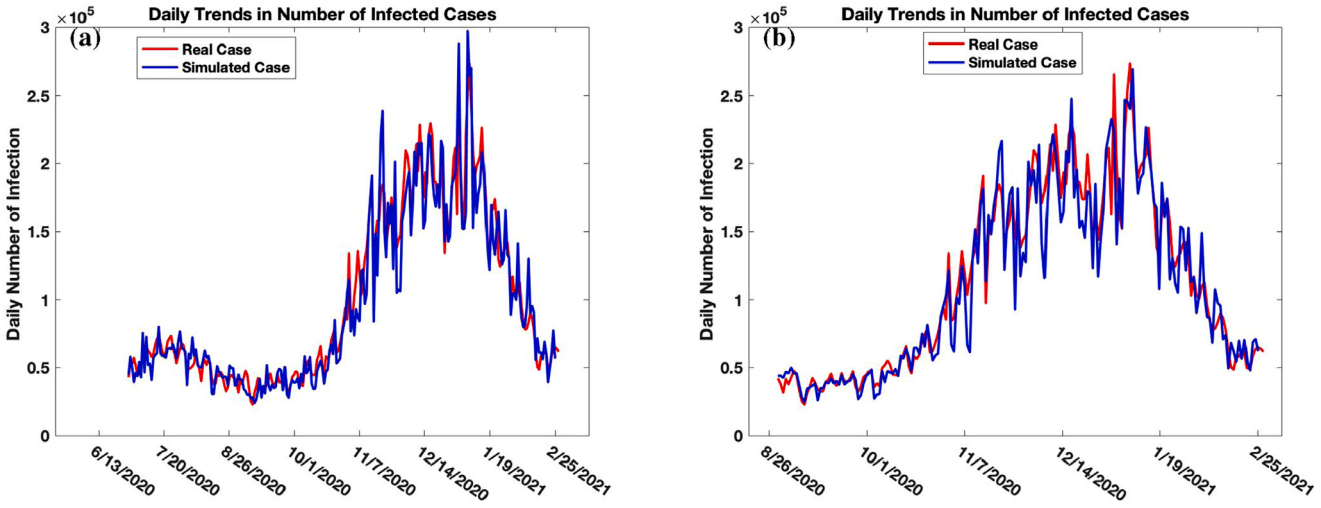

Fig. 5. Simulated (blue) and real (red) infectious values with maximum time delays $r = 90$ and $r = 150$ days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

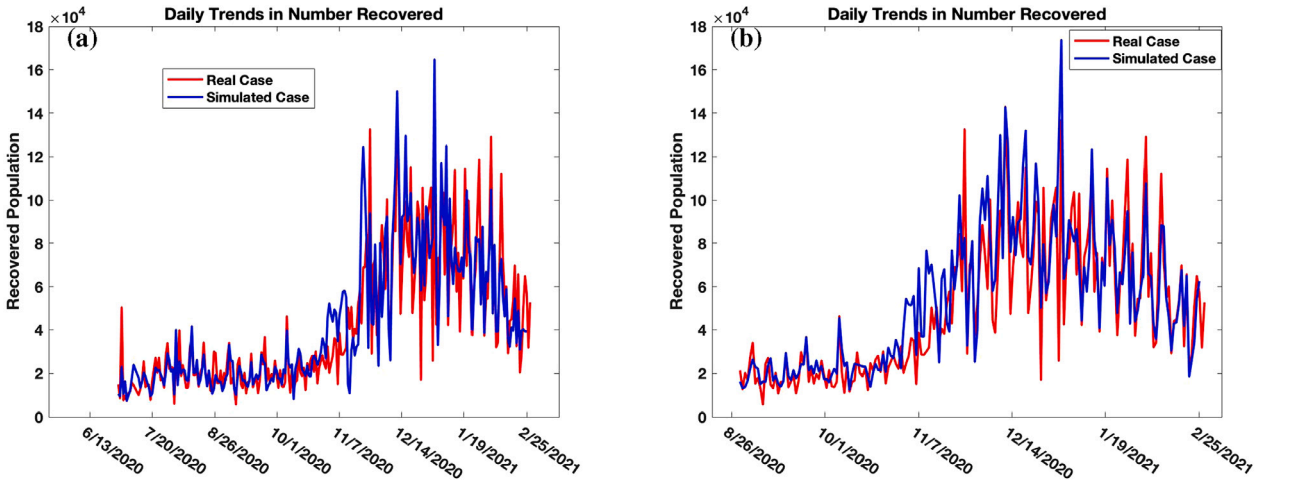

Fig. 6. Simulated (blue) and real (red) recovered values with maximum time delays $r = 90$ and $r = 150$ days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Parameter for the transmission rate curve fitting using the exponential model (18).

Parameter	Estimate	95% confidence interval
β_0	0.2487	(0.2209, 0.2766)
β_1	0.1005	(0.09828, 0.1026)
r	0.1438	(0.1073, 0.1802)
RMSE	0.01635	

estimated in Table 3 and the trend of $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$ compared with the exponential model in Fig. 8(b).

Our result shows the exponential model fits the transmission rate $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$ better than the DELPHI v2p8 model, with a root mean square of 0.01635. However, these two models fail to capture the trajectory of the transmission rate for recent months. It also fails to capture spikes in the transmission rate. For these reasons, by studying the trajectory of the transmission rate from the month of July 2020 to February 2021, we propose a mixed Gaussian model of the form

$$\tilde{\beta}(t) = \sum_{j=1}^{M} a_j \exp\left(-\frac{(x-b_j)^2}{c_j^2}\right), \tag{19}$$

where a_j, b_j, c_j, M, $j = 1, 2, \ldots, M$, are constants estimated in Table 4 below.

The trend of the transmission rate $\{\beta_{j,m_j^*}\}_{j=r+1}^{N}$ is compared with that of the proposed model (19) and the graph plotted in Fig. 9. The proposed model (19) gives a better root mean square error of 0.0135 when compared with that of the DELPHI (17) and exponential (18) models.

Our analysis shows that the transmission rate of the Covid-19 virus started decreasing around the month of December 28, 2020. Several factors might have contributed to the decline in the transmission rate of the virus at this time. Studies show that the first dose of the vaccination was administered on December 14, 2020. We are currently studying the cause of the decline by comparing the trajectory of the transmission

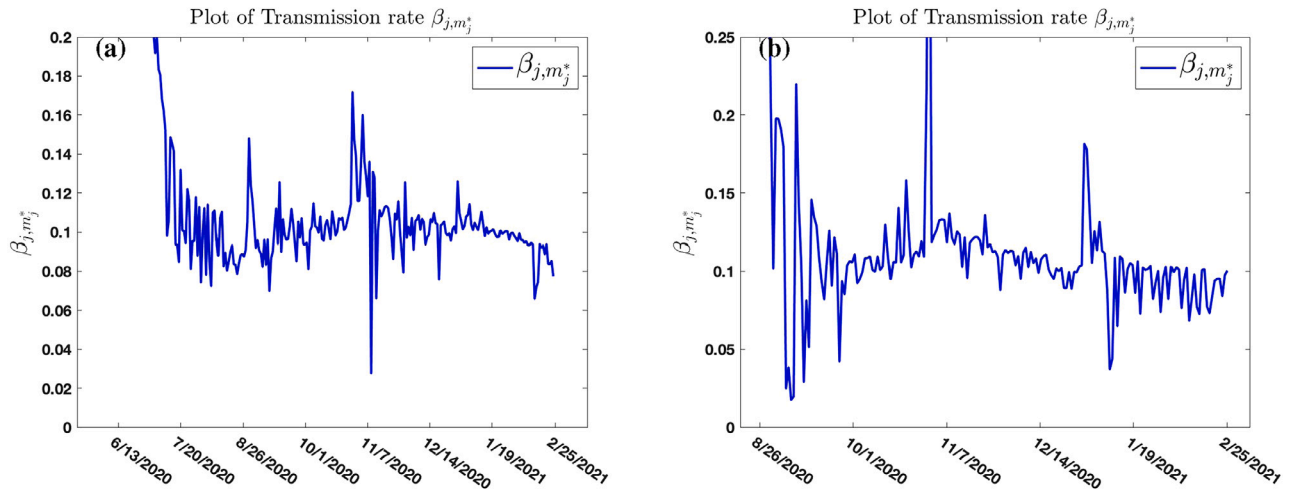

Fig. 7. Transmission rate $\beta(t)$ with maximum time delays $r = 90$ and $r = 150$ days.

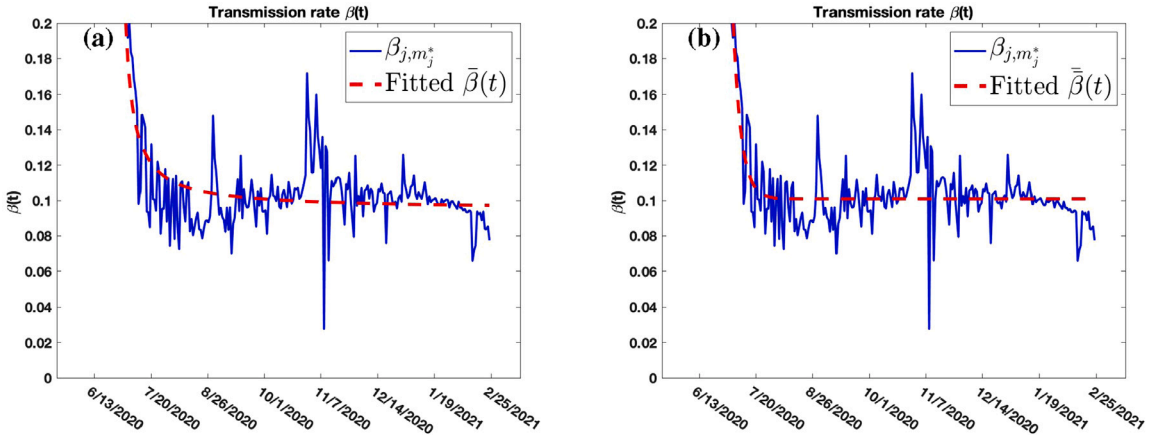

Fig. 8. Transmission rate β_{j,m_j^*} with fitted curves (17) and (18) using delay constant $r = 90$ days.

Table 4
Parameter for the transmission rate curve fitting using the Gaussian model (19).

Parameter	Estimate	95% confidence interval
M	4	
a_1	4.198	(−72.64, 81.04)
b_1	−70.2	(−423.1, 282.7)
c_1	38.62	(−49.78, 127)
a_2	0.03928	(0.02667, 0.05189)
b_2	122.3	(120.8, 123.7)
c_2	5.661	(3.452, 7.87)
a_3	0.1022	(0.09784, 0.1066)
b_3	123.1	(95.45, 150.7)
c_3	252	(163.9, 340.1)
a_4	0.009858	(−0.000252, 0.01997)
b_4	196.1	(178, 214.2)
c_4	25.75	(−9.674, 61.18)
RMSE	0.0135	

rate $\beta(t)$ with that of the number of vaccination records provided by CDC.

Fig. 10(a)–(b) shows the trajectory of the symptomatic recovery rate $\{\nu_{j,m_j^*}\}_{j=r+1}^{N}$ using delay constants $r = 90$ days and $r = 150$ days, respectively. We calculate the average of the symptomatic recovery rate to be 0.0907 using $r = 90$ days. We notice here that this average is close to the estimated asymptomatic recovery rate $r_I = 1/10.5$ estimated in Table 1. We fit the curve

$$\bar{\nu}(t) = \sum_{j=1}^{3} \bar{a}_j \exp\left(-\frac{(x-\bar{b}_j)^2}{\bar{c}_j^2}\right), \tag{20}$$

where $\bar{a}_j$, $\bar{b}_j$, $\bar{c}_j$, $j = 1, 2, \ldots, 3$ are constants estimated in Table 5. Our fit suggests that after a long decline, the symptomatic recovery rate for Covid-19 started increasing around the month of November 23, 2020 to the end of our analysis (February 25, 2021). The increase speeds up (graph concaves up) until January 18, 2021 and starts slowing down (graph concaves down) after that date. Studies on the cause of the increase is ongoing.

Fig. 11 shows the graph of the symptomatic recovery rate $\{\nu_{j,m_j^*}\}_{j=r+2}^{N}$ together with the fitted curve $\bar{\nu}(t)$ in (20).

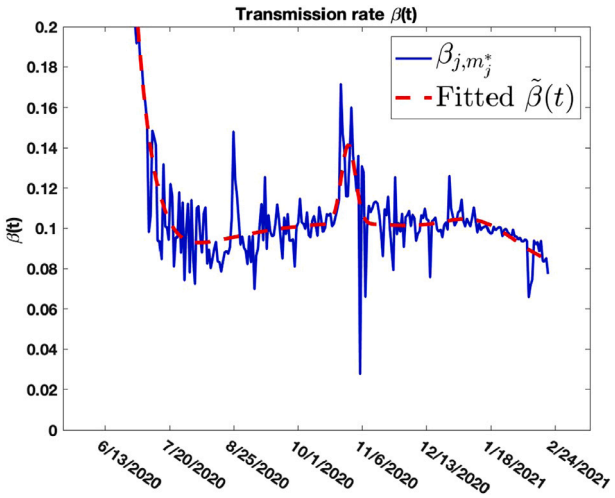

Fig. 9. Transmission rate β_{j,m_j^*} with fitted curve (19) using delay constant $r = 90$ days.

Table 5
Parameter for the symptomatic recovery rate curve fitting using the Gaussian model (20).

Parameter	Estimate	95% confidence interval
$\bar{a}_1$	0.1211	(0.08941, 0.1527)
$\bar{b}_1$	251.9	(130.7, 373.2)
$\bar{c}_1$	132.3	(−77.91, 342.5)
$\bar{a}_2$	0.06153	(−0.03422, 0.1573)
$\bar{b}_2$	89.18	(45.54, 132.8)
$\bar{c}_2$	50.47	(−12.58, 113.5)
$\bar{a}_3$	0.07061	(−0.01364, 0.1549)
$\bar{b}_3$	35.94	(17.94, 53.94)
$\bar{c}_3$	34.91	(21.37, 48.44)
RMSE	0.01505	

Fig. 12(a) and (b) shows the trajectory of the temporary immune rate $\{\gamma_{j,m_j^*}\}_{j=r+1}^{N}$ using delay constants $r = 90$ days and $r = 150$ days, respectively. The trend shows the trajectory of the temporary immune rate behaves like that of the function $t^n e^{-bt}$. We calculate the average temporary immune rate to be 0.2139 using $r = 90$ days delay constant.

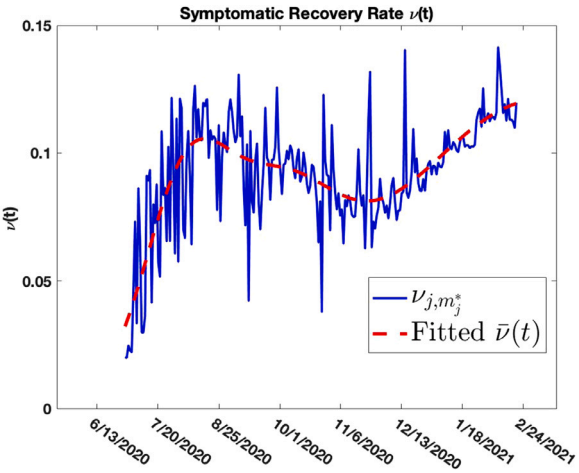

Fig. 11. Fitted recovery rates with maximum time delay $r = 90$ days.

Table 6
Parameter for temporary immune curve fitting.

Parameter	Estimate using $\bar{\gamma}_1$	95% confidence interval
a_1	0.0009599	(−0.002554, 0.004474)
b_1	0.2142	(0.09627, 0.3322)
c_1	0.2036	(0.1885, 0.2187)
n	3.292	(1.287, 5.296)
RMSE	0.105	

We fit the curve

$$\bar{\gamma}_1(t) = a_1 e^{-b_1 t} t^n + c_1 \tag{21}$$

to the data $\{\gamma_{j,m_j^*}\}_{j=r+2}^{N}$ in Fig. 13. The parameters a_1, b_1, c_1, and n are estimated in Table 6.

Fig. 13 shows the graph of the temporary immune rate $\{\gamma_{j,m_j^*}\}_{j=r+2}^{N}$ together with the fitted curves $\bar{\gamma}_i(t)$, $i = 1, 2$, in (21).

Fig. 14(a)–(b) shows the trajectory of the transmission noise intensity $\{\sigma_1(t_j, m_j^*)\}_{j=r+1}^{N}$ using delay constants $r = 90$ days and $r = 150$ days, respectively. We see spikes in the transmission noise intensity during the months of June–July, November–December for the case where $r = 90$ during Thanksgiving and Christmas festive periods. This

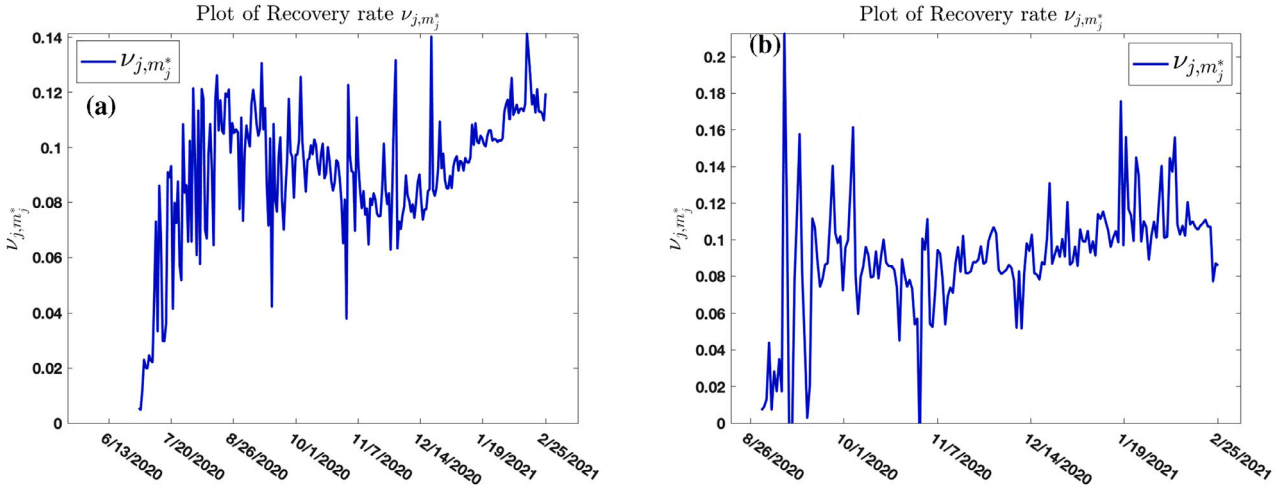

Fig. 10. Recovery rates $\nu(t)$ with maximum time delays $r = 90$ and $r = 150$ days.

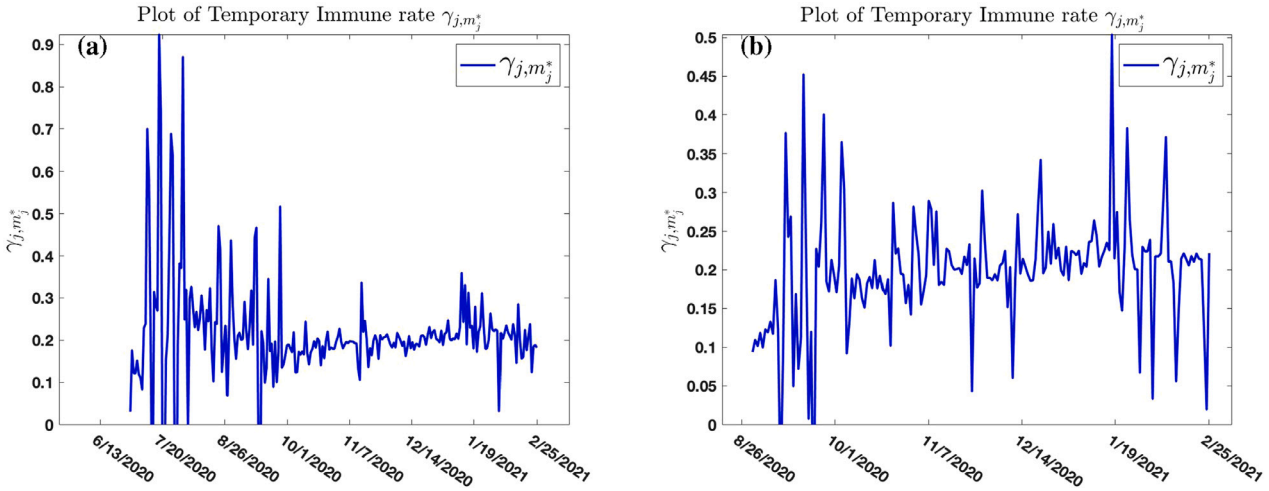

Fig. 12. Temporary immune rate $\gamma(t)$ with maximum time delays $r = 90$ and $r = 150$ days.

is the time period when the infection cases rise due to increase in the transmission rate of COVID-19.

Fig. 15(a)–(b) shows the trajectory of the recovery noise intensity $\{\sigma_2(t_j, m_j^*)\}_{j=r+1}^{N}$ using delay constants $r = 90$ days and $r = 150$ days, respectively.

Fig. 16 shows the sample observation size m_j^* used at each time t_j in the LLGMM procedure, as explained in Section 'Derivation of ϵ-best sub-optimal state and parameter estimates' and determined by the computation itself. Fig. 16(a) and (b) shows the distribution of the time delay using maximum time delay $r = 90$ and $r = 150$, respectively. Notice that the algorithm seems to prefer using many past data points as evidenced in Figs. 3, 4, 5, and 6.

The basic reproduction number $\mathcal{R}_0$ and the effective reproduction number $\mathcal{R}_e(t)$

In order to understand the outbreak dynamics of COVID-19, taking into consideration vital dynamics, we study a powerful quantitative concept that can be used to characterize the contagiousness and transmissibility of the infectious disease [5]. The basic reproduction number (denoted $\mathcal{R}_0$), is the expected number of secondary cases produced by a typical infectious individual in a completely susceptible population. Unfortunately, the number lacks the ability to explain the effects of public health interventions [35]. Realistically, population will rarely be totally susceptible to an infection in the real world. Some contacts will be immune, for example, due to prior infection which has conferred life-long immunity, or as a result of previous immunization. Therefore, not all contacts will become infected and the average number of secondary cases per infectious case will be lower than the basic reproduction number $\mathcal{R}_0$. We quantify these effects, together with the public health interventions for every point in time, using what we call the effective reproduction number (denoted $\mathcal{R}_e(t)$). The effective reproduction number $\mathcal{R}_e(t)$ is the expected number of secondary cases produced by a typical infectious individual in a population where not everyone is susceptible [2]. This number changes dynamically in response to public health interventions. The estimated contact/transmission rate $\beta(t)$ and temporary recovery rate $\nu(t)$ plotted in Figs. 7 and 10, respectively, are used to estimate the reproduction numbers for the COVID-19 as follows:

Using the next generation matrix approach [36], we first obtain the basic reproduction number $\mathcal{R}_0$ at the onset of the pandemic (that is, at t_0). This result is now generalized to calculate the effective reproduction number $\mathcal{R}_e(t)$. Denote $\beta(t_0) = \beta_0$ and $\nu(t_0) = \nu_0$ as the contact rates and symptomatic recovery rates at the onset of the pandemic, respectively. Using the notations in van den Driessche et al. [36], the non-negative

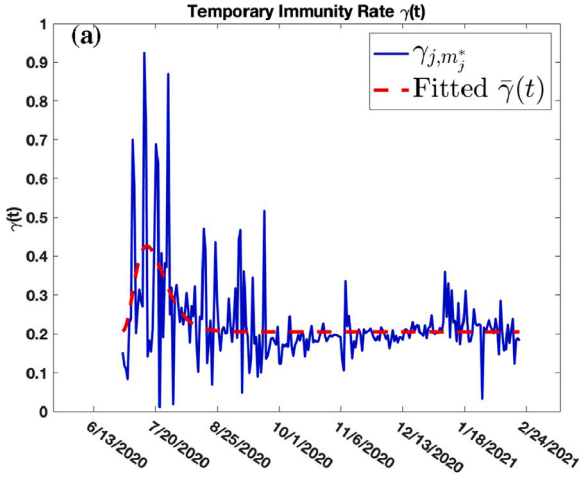

Fig. 13. Fitted Temporary Immune rates $\gamma(t)$ with maximum time delay $r = 90$ days.

matrix F representing matrix of new infection terms and the non-singular matrix V representing transfer of individuals in and out of a particular compartments, are obtained as

$$F = \begin{bmatrix} 0 & \beta_0\eta & \beta_0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} \alpha+\mu & 0 & 0 \\ -(1-p)\alpha & r_I+\mu & 0 \\ -\alpha p & 0 & \nu_0+\mu \end{bmatrix},$$

respectively. The basic reproduction number $\mathcal{R}_0$ is obtained by calculating the spectral radius of the next generation matrix FV^{-1} and obtained as

$$\mathcal{R}_0 = \beta_0 \frac{\alpha}{\alpha+\mu}\left(\frac{\eta(1-p)}{r_I+\mu} + \frac{p}{\nu_0+\mu}\right). \tag{22}$$

Here, $\frac{\alpha}{\alpha+\mu}$ is the fraction of individuals progressing from the exposed class to the infectious class (probability of surviving the exposed period and progressing to infectious class), and $\frac{1}{r_I+\mu}$ and $\frac{1}{\nu(t_0)+\mu}$ are the mean time spent being asymptomatic and symptomatic infectious, respectively, at the start of the epidemic. From this, in the presence of public health interventions (such as mask wearing, 6-feet distancing, vaccinations, et.c.), we obtain the effective reproduction number $\mathcal{R}_e(t)$

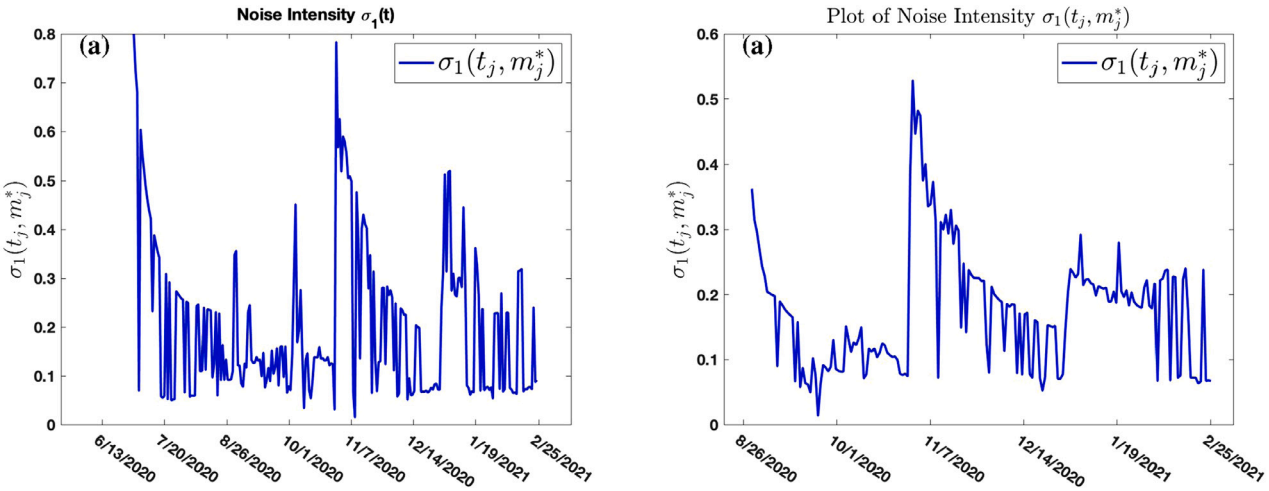

Fig. 14. Transmission rate noise intensity $\sigma_1(t)$ with maximum time delays $r = 90$ and $r = 150$ days.

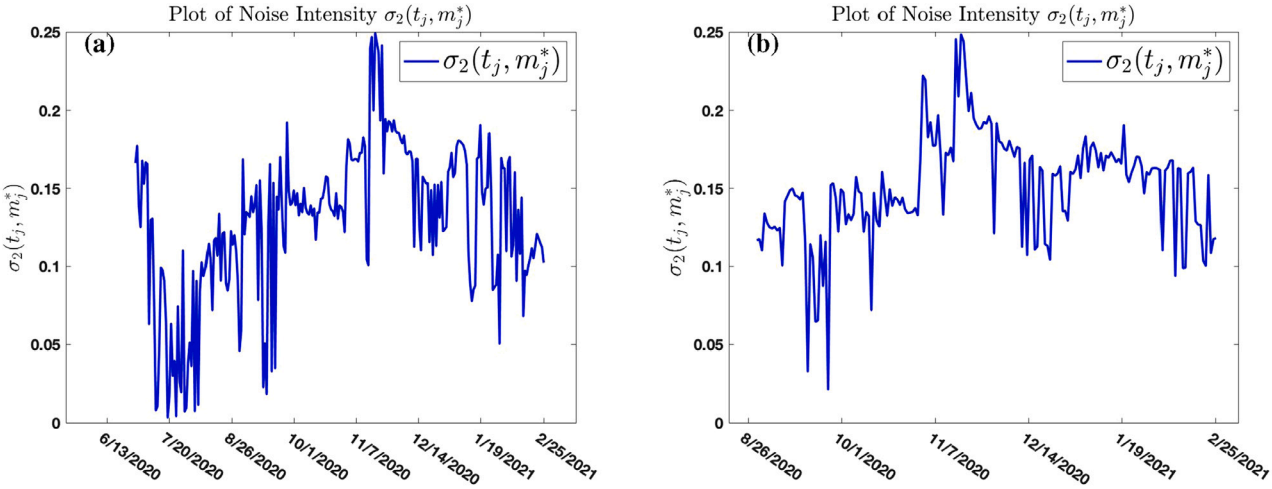

Fig. 15. Recovery noise intensity $\sigma_2(t)$) with maximum time delays $r = 90$ and $r = 150$ days.

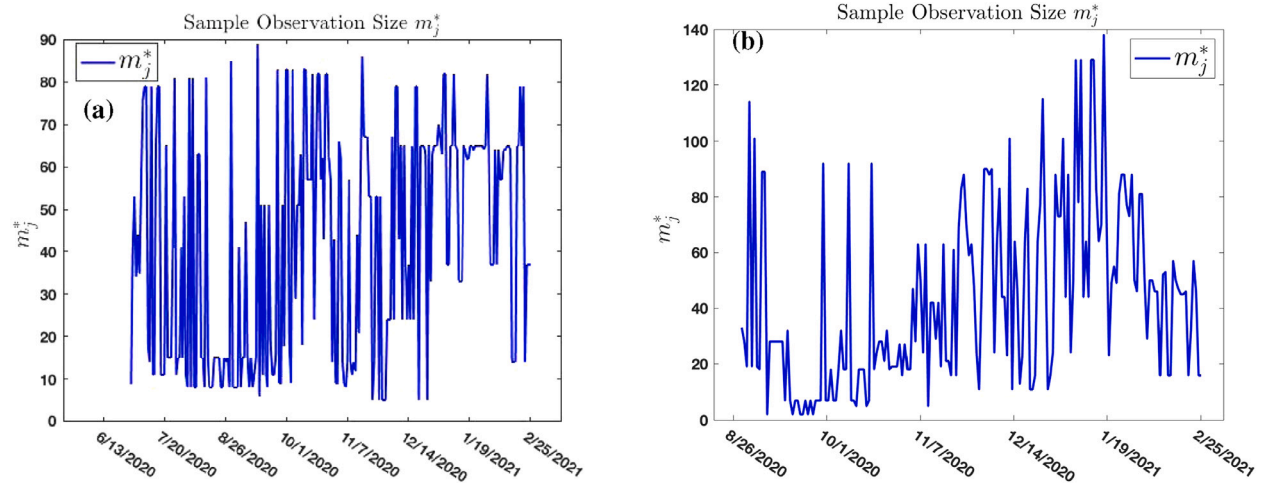

Fig. 16. Sample observation size m_j^* with maximum time delays $r = 90$ and $r = 150$ days.

as

$$\mathcal{R}_e(t) = \beta(t)\frac{\alpha}{\alpha+\mu}\left(\frac{\eta(1-p)}{r_I+\mu} + \frac{p}{\nu(t)+\mu}\right), \tag{23}$$

where $\beta(t) \equiv \beta_{j,m_j^*}$, $\nu(t) \equiv \nu_{j,m_j^*}$ are estimated and fitted in Tables 5, 4, and Figs. 7, 10.

Table 7
Parameter for the effective reproduction number curve fitting.

Month	Average	95% confidence interval
August 2020	0.9683	(0.8417, 1.0948)
September 2020	0.9590	(0.8460, 1.0720)
October 2020	0.9536	(0.9120, 0.9952)
November 2020	1.2468	(1.0982, 1.3954)
December 2020	1.0395	(0.9709, 1.1080)
January 2021	0.9596	(0.9298, 0.9906)
February 2021	0.7567	(0.7197, 0.7936)

Table 8
Parameter for the reproduction number curve fitting.

Parameter	Estimate	95% confidence interval
a_1	0.7617	(0.2798, 1.244)
b_1	118.8	(117.5, 120.1)
c_1	2.512	(0.6267, 4.398)
a_2	5.003	(3.475, 6.532)
b_2	0.2489	(−2.381, 2.878)
c_2	6.056	(3.582, 8.53)
a_3	13.78	(−404.2, 431.7)
b_3	−726.4	(−10640, 9185)
c_3	557.1	(−2378, 3492)
a_4	−1.103	(−4.083, 1.877)
b_4	40.46	(−49.76, 130.7)
c_4	66.16	(−21.8, 154.1)
RMSE	0.3613	

Information about the magnitude of the reproduction number at each time t helps to understand the outbreak dynamics of an infectious disease [5]. This result is used to analyze the outbreak dynamics of Covid-19 cases in the United States.

Fig. 17 shows the number of new infections caused by one Covid-19 infectious individual in a population that is not completely susceptible due to public health interventions. This was derived from (23), where $\beta(t) \equiv \beta_{j,m_j^*}$, $\nu(t) \equiv \nu_{j,m_j^*}$ are the ϵ-best estimates for the transmission rate β_{j,m_j^*} and temporary recovery rates ν_{j,m_j^*}. These rates are plotted and fitted in Figs. 7, 10 and Tables 5, 4, respectively. The estimated average effective reproduction number in the month of August 2020 to February 2021 are given in Table 7.

We see from the graph estimates that the effective basic reproduction number falls below 1 from January 10, 2021 to the end of our simulation period (February 25, 2021). Studies concerning the possible decline in $\mathcal{R}_e(t)$ in these months are ongoing, but we suspect the introduction of vaccines into the system might be one of the major contributor to the decline. We see from the plot of the effective reproduction number that the regions where there are spikes correspond to regions in the plot of the transmission and recovery rates β_{j,m_j^*} and ν_{j,m_j^*} where there are spikes. As it is well known [37,38], if $\mathcal{R}_0 < 1$, then the number of infection cases converges to zero on the long run (that is, the virus will stop spreading on the long run). Likewise, $\mathcal{R}_0 > 1$ means epidemic is growing on the long run. The result above shows that on the average, epidemic was growing in the month of November and December 2020, and started to decline in the month of January and February, 2021. This is consistent with the analysis of the transmission and symptomatic recovery rate of the disease. As seen in Fig. 17(b), we fit the curve

$$\bar{\mathcal{R}}_e(t) = \sum_{j=1}^{4} a_j \exp\left(-\left(\frac{t-b_j}{c_j}\right)^2\right) \tag{24}$$

to the data $\{\mathcal{R}_e(t_j)\}_{j=r+2}^{N}$ in Fig. 17(a). The parameters a_j, b_j, c_j, d_j, $j = 1, 2, 3, 4$, are estimated in Table 8.

Discussion and further studies

In this study, the outbreak dynamics of Covid-19 is discussed and analyzed using a Stochastic SEIRS dynamic epidemic model with case study of COVID-19. In the case study, the local lagged adapted generalized method of moments (LLGMM) parameter identification scheme was used to extract the exposed and susceptible population and to identify the time-dependent transmission, recovery, temporary immunity rate, infection noise intensities, and the reproduction numbers (basic and effective reproduction numbers) from the COVID-19 data from March 2020 to February 2021 using time delays $r = 90$ and $r = 150$ days (equivalent to three and five months past data, respectively). Valuable information on the susceptible and exposed population, together with the epidemiological parameters and the reproduction numbers of COVID-19 are obtained with the help of the LLGMM scheme.

From the graph of the infection cases in Fig. 2, we see that the trajectory of the infected population is similar to that of the estimated exposed population. Also, the trend of the estimated susceptible population shows that as the number of infection cases increase (decrease), the estimated number of susceptible population decreases (increases). Using the estimated exposed and susceptible population, we estimated the time varying parameters in the $SEIRS$ epidemic model. We see from the parameter estimates that there are spikes in the transmission rate $\beta(t)$ of Covid-19 in the month of November and December 2020. The number of infection and exposure to Covid-19 also increase in the same months. These months correspond to festive periods in the United States when people travel around the country. Also, these number reduces, starting from the month of December 24, to the end of our analysis period. These period corresponds to the period when the daily count of total Covid-19 vaccine doses administered increases. A model was derived to fit the time-dependent transmission rate (with a root mean square error (RMSE) of 0.0135). The model is compared with the work of Lingzhi Li et al. [32], Eikenberry [33] and Tang et al. [34] and shown to fit the transmission rate better.

The recovery rate $\nu(t)$ of Covid-19 was estimated for the cases where $\alpha = 1/6$, with $\mu = \frac{1}{80.3\times365}$ and time delays $r = 90$ and $r = 150$. A model was derived to fit the time-dependent recovery rate (with a root mean square error of 0.01505). Our studies show that the recovery rate increases around early December 2020 to the end of our analysis period (February 25, 2021). Studies of the cause of increase is ongoing.

The temporary immune rate $\gamma(t)$ was also estimated using time delays $r = 90$ and $r = 150$. A model was derived to fit these estimates.

Finally, we estimated the basic and effective reproduction numbers for the Covid-19 virus. These numbers are calculated using the Next Generation Matrix approach. Our studies show that the effective reproduction number fluctuates around the number 1, and rises above 1 on the average in the month of November and December, 2020. That is, epidemic was growing in the months of November and December, 2020 during the Thanksgiving and Christmas period. The number of infection reduces on the average in the month of January 2021 and mostly in February, 2021 probably due to the introduction of the Pfizer and Moderna Covid-19 vaccines in the United States, which was first administered on December 14, 2020. We discovered that on the average, using latency rate $\alpha = 1/6$, epidemic was more pronounced in the months of November, December 2020, and January 2021 than in other months. The growth reduces in the month of January and February. We see a decline in the transmission rate of the disease starting late December, 2020 to February, 2021. A possible cause of the decline might be attributed to the Pfizer-BioNTech and Moderna Covid-19 vaccine. We are currently studying the cause of the decline by checking if there is any correlation between the estimated transmission rate and the number of vaccination in the United States. A Gaussian model was used to fit the reproduction number with a root mean square of 0.3613.

Further studies on the Covid-19 outbreak is ongoing and results of our findings will be published as it becomes available.

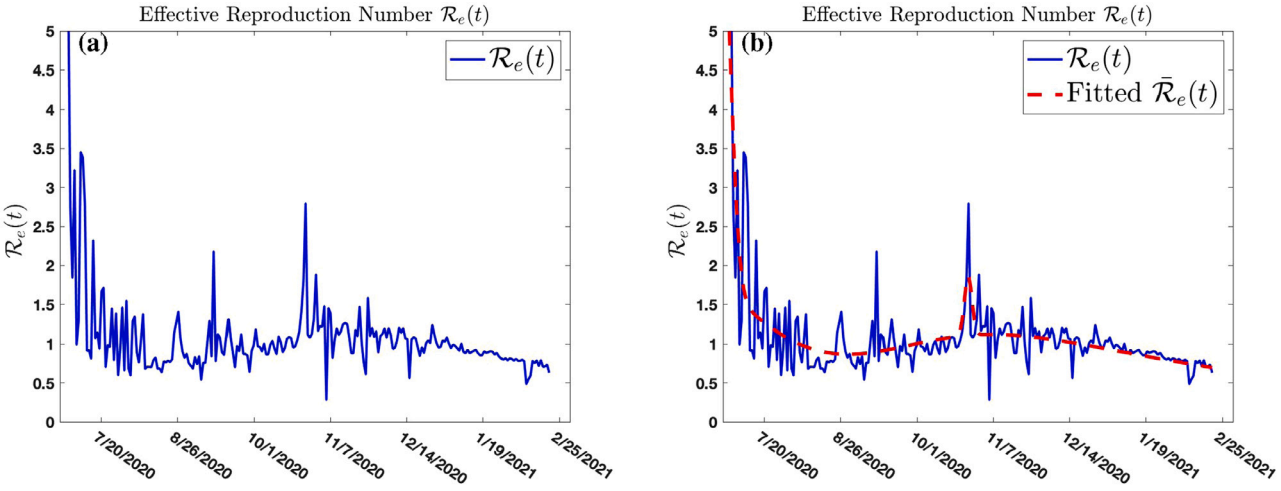

Fig. 17. Effective Reproduction number $\mathcal{R}_e(t)$ using latency rate $\alpha = 1/6$.

CRediT authorship contribution statement

Olusegun M. Otunuga: Carried out the research work, Drafted the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials

The data used in the analysis can be found on the CDC[6] website.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] IHME COVID-19 Forecasting Team. Global projections of potential lives saved from COVID-19 through universal mask use. Nat Med 27, 94–105. Preprint at medRxiv.
[2] IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United States. 2020, http://dx.doi.org/10.1101/2020.10.08.20209510.
[3] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;35(10225):689–97.
[4] Stutt ROJH, Retkute R, Bradley M, Gilligan CA, Colvin J. A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic. Proc R Soc Lond Ser A Math Phys Eng Sci 2020;476:20200376.
[5] Kevin Linka, Mathias Peirlinck, Ellen Kuhl, 66. The reproduction number of COVID-19 and its correlation with public health interventions. Comput Mech 2020;1035–50.
[6] Okuonghae D, Omame A. Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria. Chaos Solitons Fractals 2020;139:110032.
[7] Ndairou F, Area I, Nieto JJ, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solitons Fractals 2020;135:109846.
[8] Musa SS, Qureshi S, Zhao S, Yusuf A, Mustapha UT, He D. Mathematical modeling of COVID-19 epidemic with effect of awareness programs. Infect Dis Model 2021;6:448–60.
[9] Memon Z, Qureshi S, Memon BR. Assessing the role of quarantine and isolation as control strategies for COVID-19 outbreak: a case study. Chaos Solitons Fractals 2021;144:110655.
[10] Peter OJ, Qureshi S, Yusuf A, Al-Shomrani M, Idowu AA. A new mathematical model of COVID-19 using real data from Pakistan. Results Phys 2021;24:104098.
[11] Mummert A, Otunuga OM. Parameter identification for a stochastic SEIRS epidemic model: case study influenza. J Math Biol 2019;79(2). http://dx.doi.org/10.1007/s00285-019-01374-z.
[12] Pollicott M, Wang H, Weiss H. Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. J Biol Dyn 2012;6:509–23.
[13] Otunuga OM. Time varying parameter estimation scheme for a linear stochastic differential equation. Int J Stat Probab 2017;6(5):84–100.
[14] Cazelles B, Chau NP. Adaptive dynamic modeling of HIV/AIDS epidemic using extended Kalman filter. J Biol Systems 1995;3(3):759–68.
[15] Julier SJ, Uhlmann JK. Unscented filtering and nonlinear estimation. Proc IEEE 2004;92(3):401–22.
[16] Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng 1960;82(1):35–45.
[17] Hansen LP. Large sample properties of generalized method of moments estimators. Econometrica 50(4): 1029–54.
[18] Jeisman J. Estimation of the parameters of stochastic differential equations [Ph.D. dissertation], Queensland University of Technology; 2005.
[19] Hurn S, Jeisman J, Lindsay K. Seeing the wood for the trees: A critical evaluation method to estimate the parameters of stochastic differential equations. J Financ Econom 2007;5(3):390–455.
[20] Ladde GS, Otunuga OM, Ladde NS. Local lagged adapted generalized method of moments Dynamic Process, U.S. Patent Number: 10719578.
[21] Kloeden PE, Platen E. Numerical solution of stochastic differential equations. New York: Springer-Verlag; 1995.
[22] Ladde AG, Ladde GS. An introduction to differential equations: stochastic modeling, methods and analysis, vol. 2. World Scientific Publishing; 2013.
[23] Méndez V, Campos D, Horsthemke W. Stochastic fluctuations of the transmission rate in the susceptible-infected-susceptible epidemic model. Phys Rev E 2012;86:011919.
[24] Otunuga OM, Ladde GS, Ladde NS. Local lagged adapted generalized method of moments and applications. Stoch Anal Appl 2017;35(1):100–43.
[25] Otunuga OM, Ladde GS, Ladde NS. Local lagged adapted generalized method of moments: an innovative estimation and forecasting approach and its applications. Time Ser Econom 2019;2019.
[26] Schurz H. Numerical regularization for SDEs: Construction of nonnegative solutions. Dynam Systems Appl 1997;5.
[27] Cyganowski S, Grune L, Kloeden PE. Maple for stochastic differential equations. In: Blowey James F, Coleman John P, Craig Alan W, editors. Theory and numerics of differential equations. Springer-Verlag Berlin Heidelberg; 2001.
[28] McAloon CG, Collins A, Hunt K, Barber A, Byrne A, Butler F, et al. The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. 2020, medRxiv.
[29] Lauer SA, Grantz KH, Bi Q. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Ann Intern Med 2020.
[30] Linton NM, Kobayashi T, Yang Y. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. J Clin Med 2020;9. http://dx.doi.org/10.3390/jcm9020538, [PMID: 32079150].

[31] Li Q, Guan X, Wu P. Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020. http://dx.doi.org/10.1056/NEJMoa2001316, [PMID: 31995857].
[32] Lingzhi Li M, Bouardi HT, Lami OS, Trichakis NT, Trikalinos T, Bertsimas D. Overview of DELPHI Model V2.8-Covid Analytics.
[33] Eikenberry Steffen E, Mancuso Marina, Iboi Enahoro, Phan Tin, Eikenberry Keenan, Kuang Yang, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model 2020;5:293–308.
[34] Biao Tang, Nicola Luigi Bragazzi, Qian Li, Sanyi Tang, Yanni Xiao, Jianhong Wu. An updated estimation of the risk of transmission of the novelcoronavirus (2019-nCov). Infect Dis Model 2020;5:248–55.
[35] Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R_0). Emerg Infect Disease 2019;25:1–4.
[36] van den Driessche P, Watmough James. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 2002;180:29–48.
[37] Otunuga OM. Global stability for a 2n + 1 dimensional HIV/AIDS epidemic model with treatments. Math Biosci 2018;5:138–52.
[38] Otunuga OM. Qualitative analysis of a stochastic SEITR epidemic model with multiple stages of infection and treatment. Infect Dis Model 2020;5:61–90.