The Evolution of the Identifiable Analysis of the COVID-19 Virus

Vivek Sreejithkumar*

Abstract. It is important to accurately forecast a new infection such as COVID-19 in order to effectively implement control measures. For this purpose, we study whether the epidemiological parameters such as the rate of infection, incubation period, and rate of recovery for the COVID-19 disease can be identified from daily incidences and death data. The data are obtained from the Florida Department of Health, which reports the numbers of daily COVID-19 cases and disease-induced casualties. Two mathematical models that consist of a system of ordinary differential equations are used to simulate the spread of the coronavirus in the Florida population. Structural identifiability analysis is conducted on the models to determine whether the models are well-structured to forecast the outbreak. Analysis revealed that the SEIR model is structurally identifiable, while the social distancing model is not structurally identifiable. If the model is structurally unidentifiable, it may not accurately forecast the pandemic, and in turn, may lead to inaccurate control measures. Then, the practical identifiability of parameter estimates that provide the best fit was investigated using Monte Carlo simulations. The practical identifiability analysis revealed that all of the parameters in the SEIR model are practically identifiable, but the parameters δ, δ_E, and ρ were found to be unidentifiable in the social distancing model. By comparing two models in this project, we were able to determine the effectiveness of social distancing in preventing incidences and saving lives from the disease in Florida. Furthermore, we consider how people's behavior changes over time, and how this may affect the rate of disease spread in the population. To represent this, we develop a recipe to determine the time-dependent transmission rate, $\beta(t)$, from the data and introduce a methodology of how to accomplish this.

Key words. coronavirus, COVID-19, epidemiology, SEIR Model, parameters, mathematical model, MATLAB, infectious disease, Monte Carlo simulations;

AMS subject classifications. 922D30, 922D40

1. Introduction.

The novel coronavirus (COVID-19), caused by the SARS-CoV-2 virus, has been groundbreaking in the realm of infectious diseases, bringing worldwide attention to the importance of reducing disease spread amidst a global pandemic. COVID-19 is a catastrophic infectious disease that is transmitted by close contact from person to person via respiratory droplets [15]. This newly risen coronavirus that has widely spread throughout the world very rapidly was declared as a worldwide pandemic by the World Health Organization (WHO) in March 2020 [2]. Many national agencies, such as the Centers for Disease Control (CDC), have shared advisories to practice social distancing in an effort to limit the amount of person to person contact and slow the spread of the disease. By practicing social distancing and limiting close contacts with other individuals, we work towards reducing the transmission rate of the virus and minimize the amount of people who contract the disease [9]. Epidemiology is the field of study concerned with the spread of such infectious diseases, and mathematics has an immanent connection with epidemiology in representing how quickly infectious diseases spread within a population. Research institutions and national agencies can use methods of mathematical modelling to map the impact of a virus on a population [12]. By

*Florida Atlantic University, 777 Glades Road, Boca Raton, FL, 33431 (vsreejithkum2018@fau.edu).

doing so, it is possible to predict the long-term behavior of the virus in order to take proper precautionary measures. Ideally, conducting mathematical modeling on infectious diseases allows for a quantitative method to evaluate the efficiency of implementing control measures and develop an accurate forecast of the outbreak [11]. There is a large existing literature on mathematical models specific to COVID-19, including a mathematical representation of the impact of non-pharmaceutical interventions (such as social distancing, quarantine, and use of face masks) in curtailing the disease spread [10]. In addition, mathematical assessments can be utilized to simulate the effectiveness of vaccination efforts to curtail the pandemic [3]. In this project, mathematical models are used to represent the transmission of the COVID-19 virus throughout a population, in this case the population of Florida. The mathematical model can then be matched with actual data of COVID-19 cases as reported by the Florida Department of Health to estimate the values of certain epidemiological parameters in the mathematical model, such as the transmission rate, incubation rate, and recovery rate of the disease [13].

After determining the values of the epidemiological parameters in the model, it is important to study whether the parameter estimation is well posed to uniquely reveal the parameters from the observed data, which is what we investigate by conducting identifiability analysis. Primarily, it is crucial to study the identifiability of the model before estimating the parameter values to determine whether the model is structured to reveal its parameters, which is known as structural identifiability [7]. After collecting the estimated values of the parameters in the epidemiological model, the next step is to measure the practical identifiability of the model through Monte Carlo simulations. In Monte Carlo simulations, we introduce error into the dataset and then refit epidemiological model to the data with error 1,000 times to observe how the parameter values react to the introduced error [18]. Results from Monte Carlo simulations may suggest whether the model is practically identifiable or not, and whether the parameter estimations are reliable enough to influence decisions of implementing disease control measures, such as social distancing. In the data from the Florida Department of Health, it is clear that people's behaviors are changing over time, which may have an effect in the rate of disease spread. For example, during the holiday seasons people tend to congregate in large groups more often, and a subsequent rise in COVID-19 incidences is observed. To represent this in the model, a recipe is developed to make the transmission rate change over time, expressed as $\beta(t)$. By computing the time-dependent transmission rate, the model is reflective of how individuals' social behaviors change over time and influence the rate of disease spread [5]. By combining several epidemiological characteristics of COVID-19 in the model, the model is optimized to best represent the spread of the disease throughout the Floridian population.

2. Methods.

The data for this project were collected from the website for the Florida Department of Health [13]. The Florida Department of Health reports weekly cases of individuals infected with the COVID-19 virus and disease-induced deaths on their website – excluding personal information such as names and other factors of identification and is available for use to the public. For this project, data was collected weekly every Monday starting from April 27th, 2020 until July 5th, 2020. After the preliminary mathematical model was constructed, a code was written in the computer software MATLAB to run the epidemiological model in comparison to the COVID-19 incidence data from the Florida Department of Health website. The code and data used for this project has been made available at

https://github.com/viveksreejithkumar/Identifiable-Analysis-COVID-19.

The MATLAB code was instructed to perform curve fitting on the model according to the COVID-19 incidence and death data by manipulating the parameter values until the squared difference between the data points and the model predictions was minimized. After the fitting process was complete, the parameters that provided the best fit of the model to the data were recorded as results. After developing the most accurate mathematical model and estimating the parameter values that provided the best fit to the incidence data, the next step in the project was to measure the identifiability of the parameter estimations. In this research, we would like to analyze the evolution of identifiability analysis as the new infections spread in the population. We achieve this by performing Monte Carlo Simulations for each week data was collected starting from April 20.

To conduct identifiability analysis, Monte Carlo simulations are run on the model by introducing error into both the COVID-19 incidence and death data. A second code on the MATLAB software is written that injects the noise level into the data, and then refits the epidemiological model 1,000 times according to the data points that contain the error. Then, a list of the average relative errors in the parameter estimations is produced, which allows for observation on how the parameter values reacted to the noise levels in the data. If the average relative errors are higher than the measurement errors, then it is concluded that the model is practically identifiable, meaning the parameter estimates are reliable.

Variable	Definition
$N(t)$	Number of total individuals in the population
$S(t)$	Number of susceptible individuals at time t
$S_d(t)$	Number of susceptible individuals who practice social distancing at time t
$E(t)$	Number of exposed individuals at time t
$I(t)$	Number of infectious individuals at time t
R(t)	Number of individuals removed from the population at time t

Table 1
Definition of variables in the outbreak models (3.1) and (3.2)

3. Formulation of Mathematical Models.

3.1. SEIR Model. The Susceptible-Exposed-Infected-Recovered (SEIR) model consists of ordinary differential equations (ODEs) which describe the movement of individuals of the population between four different non-intersecting classes, $S(t), E(t), I(t)$, and $R(t)$. People in the susceptible ($S(t)$) class are the individuals in the population who do not yet have the disease and are prone to contracting it. Usually, the individuals within the exposed ($E(t)$) class are not considered infectious and are not capable of transmitting the disease to other individuals until the incubation period is over, from when they might begin to infect others without realizing they have become infectious. However, the exposed class is considered differently with COVID-19. In this project, we consider the exposed (asymptomatic) class to be able to transmit the disease to other individuals before the incubation period is over. The infected ($I(t)$) class consists of individuals who are currently infected with the disease and

Parameter	Definition
β	Transmission rate of infected individuals
β_E	Transmission rate of exposed individuals
η	Fraction of susceptible individuals who practice social distancing
δ	Factor that reduces the infectious rate of infected individuals who practice social distancing. $(0 < \delta < 1)$
δ_E	Factor that reduces the infectious rate of exposed individuals who practice social distancing. $(0 < \delta_E < 1)$
α	Recovery Rate
ν	Disease-induced death rate
ρ	Rate of social distancing individuals discontinuing social distancing
k	Rate at which the exposed class progress into the infectious class

Table 2
Definition of parameters in the outbreak models (3.1) and (3.2)

transmit the disease to others. The recovered ($R(t)$) class consists of individuals who had the disease in the past and have recovered from the COVID-19 disease.

The derivative of the susceptible class ($S(t)$) is the individuals who become infectious per unit of time. Since this is an outbreak model, the rate of change in the susceptible class is negative. The susceptible class gets smaller as more people become infected. Incidences is defined as the number of individuals who become infectious per unit time. As the COVID-19 disease is spread through airborne respiratory droplets when an individual sneezes or coughs, it only takes one contact between an infected person and a susceptible person to transmit the disease. Furthermore, if the virus inhabits an object or a surface and an individual comes into contact with that surface, the disease can be transmitted when the individual subsequently touches their eyes, nose, or mouth. For COVID-19, new incidences occur due to contacts that a susceptible individual has with an infected individual or an infected object or surface. In our model, we are only modelling contacts between individuals (without an environment component). We use β, to represent the rate of infection, which, by definition, is the product of the number of contacts on infected individual makes in the population and the probability that contact results in disease transmission. We use standard incidence to model the new infections per unit time. The exposed class, $E(t)$, accounts for individuals who have been exposed to the virus but have not yet showed any symptoms of the infection (asymptomatic). Even though we don't see individuals in the exposed class demonstrating symptoms of the disease, we consider them as infectious [4]. Thus the new incidences in the model takes the form $\beta SI/N + \beta_E SE/N$. Once an exposed individual completes the incubation period which is given as $\frac{1}{k}$, becomes infectious showing clinical symptoms and moves into the infected class, $I(t)$. An infected individual either recovers from the disease and moves to the recovered class at rate α, or dies from the disease at rate ν. With these assumptions, the model takes the

following form;

$$
\begin{aligned}
S' &= -S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right),\\
E' &= S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right) - kE,\\
I' &= kE - (\alpha+\nu)I,\\
R' &= \alpha I.
\end{aligned}
\tag{3.1}
$$

A flowchart that exhibits the movement of individuals between the 4 classes of the $SEIR$ model (3.1) is presented in Figure 1.

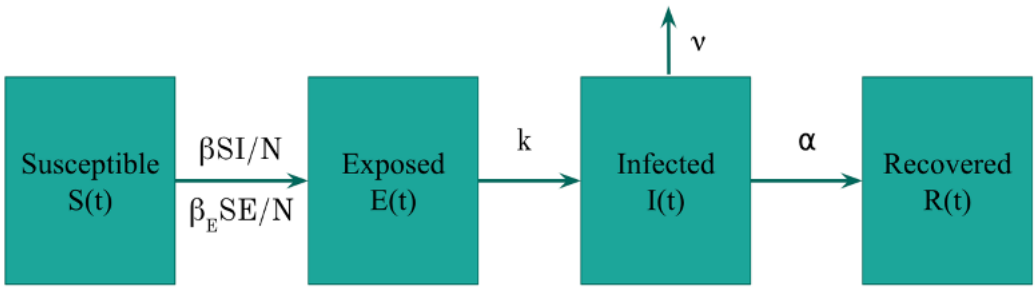

Figure 1. *SEIR Model – Flowchart: A compartmental flowchart illustrating the movement of individuals between the 4 classes of the SEIR model. A susceptible individual can come into contact with the virus and moves into the exposed class. After the incubation period ($\frac{1}{k}$) moves into the infected class. An infected individual either recovers from the disease at rate α and moves into the recovered class or dies from the disease at rate ν.*

3.2. SEIR Model with Social Distancing. To combat the spread of the infectious disease, one preventative measure introduced during the COVID-19 pandemic was social distancing. Specifically, in Florida, a stay-at-home lockdown order was announced by the governor on April 1^{st}, 2020. This order mandated that individuals remain inside their homes during the period of the pandemic, which limited contacts between individuals that could otherwise have possibly increased disease transmission. Over time, as the lockdown period ended and Florida began slowly reopening, a social distancing protocol was mandated to limit the number of disease-spreading contacts. Per social distancing protocols, individuals who leave their homes were required to wear masks and stay 6 feet apart from others to slow the spread of the disease. To account for social distancing individuals, we modify the SEIR model (3.1) to include a social-distancing class ($S_d(t)$) that indicates individuals who follow social-distancing protocols. A susceptible individual either decide to follow social distancing protocols and move to the social distancing class or decide not to practice social distancing and remain in the susceptible class. The number of individuals who are social distancing per unit time is given by the term ηS, and the number of social distancing individuals who decides to quit social distancing per unit time is given by the term ρS_d. Individuals in the susceptible or social distancing class encounter the virus, but we assume that the transmission rate is reduced due to social-distancing. We encounter this by multiplying the transmission rate with $0 < \delta < 1$

when a social distancing individual gets into contact with an infected individual and with $0 < \delta_E < 1$ when a social distancing individual gets into contact with an exposed individual. Taking into account social distancing individuals, new incidences term becomes;

$$S\left(\beta\frac{I}{N}+\beta_E\frac{E}{N}\right)+S_d\left(\beta\delta\frac{I}{N}+\beta_E\delta_E\frac{E}{N}\right).$$

The new incidences first move into the exposed class, and after the incubation period, pass into the infected class. Infected individuals recover from the disease and move into the recovered class or die from the disease. With these assumptions, the social distancing model takes the following form;

$$\begin{aligned}
S' &= -S\left(\beta\frac{I}{N}+\beta_E\frac{E}{N}\right)-\eta S+\rho S_d,\\
S_d' &= -S_d\left(\beta\delta\frac{I}{N}+\beta_E\delta_E\frac{E}{N}\right)-\rho S_d+\eta S,\\
E' &= S\left(\beta\frac{I}{N}+\beta_E\frac{E}{N}\right)+S_d\left(\beta\delta\frac{I}{N}+\beta_E\delta_E\frac{E}{N}\right)-kE,\\
I' &= kE-(\alpha+\nu)I,\\
R' &= \alpha I.
\end{aligned} \tag{3.2}$$

A flowchart that demonstrates the movement of individuals between the 5 classes of the social distancing model (3.2) is given in Figure 2. The flow of individuals throughout the model is very similar to the flowchart of the SEIR Model (3.1), except we add one compartment for individuals who practice social distancing ($S_d(t)$) Although it appears in the social distancing model that social distancing affects only the susceptible populationin reducing their exposure, in reality, social distancing would also reduce the transmissibility of the exposed/infected class. However, for simplicity we only consider the effect of social distancing as a separate susceptible class. For the purpose of identifiability, we apply social distancing to the model as simple as possible.

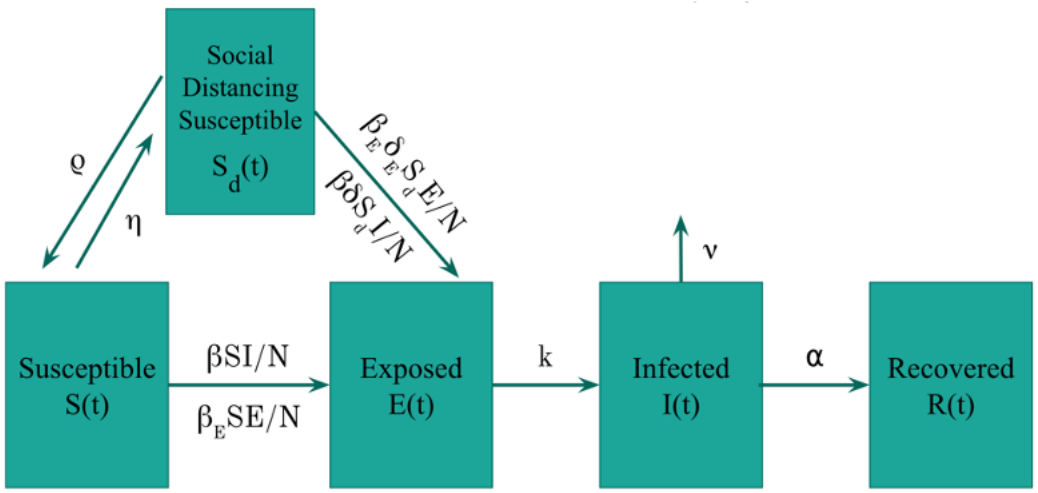

Figure 2. *Social Distancing SEIR Model – Flowchart: A compartmental flowchart illustrating the movement of individuals between the 5 classes of the social distancing model (3.2). Individuals who begin as susceptible either decide to follow social distancing protocols and move to the social distancing class or they do not practice social distancing and remain in the susceptible class. Individuals in the susceptible or social distancing class encounter the virus and then move into the exposed class, then finish their incubation period and pass into the infected class, and then finally recover from the disease and move into the recovered class.*

4. Parameter Estimation.

4.1. Data. The data for this project was collected from the Florida Department of Health website [1]. On this website, there is publicly available data on new daily cases of COVID-19 and on disease-induced deaths in Florida. The website tracks data from the past 30 days and the data are updated daily around 11A.M. EST. For this project, COVID-19 incidence and deaths data from the Florida Department of Health website was collected weekly, on every Monday afternoon. In this project, data collection was conducted weekly (every Monday) starting from April 27th until July 5th, totaling in 12 sets of data. After collecting the new weekly data, the new data was adding to the existing data and re-fitted using the MATLAB computer software each new week. In this way, we observe the evolution of the parameter values in the model each week as new data are available.

4.2. Fitting Process. In compact form outbreak models (3.1) and (3.2) can be written as

$$\boldsymbol{x}'(t) = f(\boldsymbol{x}, \boldsymbol{p}) \quad \boldsymbol{x}(0) = \boldsymbol{x}_0,$$

where $\boldsymbol{x}$ represents the state variables, and $\boldsymbol{p}$ represents the model parameters, and $\boldsymbol{x}_0$ represents the initial conditions. For instance for model (3.1) $\boldsymbol{x}$, $f(\boldsymbol{x}, \boldsymbol{p})$ and $\boldsymbol{p}$ take the following form;

$$\begin{aligned} \boldsymbol{x}(t) &= (S(t), E(t), I(t), R(t)), \\ f(\boldsymbol{x}, \boldsymbol{p}) &= \left(-S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right), S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right) - kE, kE - (\nu+\alpha)I, \alpha I\right), \\ \boldsymbol{p} &= (\beta, \beta_E, k, \nu, \alpha). \end{aligned}$$

Similarly, for model (3.2),

$$\begin{aligned}\boldsymbol{x}(t) &= (S(t), S_d(t), E(t), I(t), R(t))\\ f(\boldsymbol{x}, \boldsymbol{p}) &= (-S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right) - \eta S + \rho S_d - S_d\left(\beta\delta\frac{I}{N} + \beta_E\delta_E\frac{E}{N}\right) - \rho S_d + \eta S,\\ &\quad S\left(\beta\frac{I}{N} + \beta_E\frac{E}{N}\right) + S_d\left(\beta\delta\frac{I}{N} + \beta_E\delta_E\frac{E}{N}\right) - kE, kE - (\nu+\alpha)I, \alpha I)\\ \boldsymbol{p} &= (\beta, \beta_E, k, \nu, \alpha, \eta, \delta, \delta_E, \rho).\end{aligned}$$

The observations are the functions of the state variables. Let $y(t) = (y_1(t), y_2(t))$ represent the observations, then

$$y_1(t) = g_1(\boldsymbol{x}(t), \boldsymbol{p}), \quad y_2(t) = g_2(\boldsymbol{x}(t), \boldsymbol{p}),$$

Since we use number of new cases and deaths due to COVID-19, we set the observations as

$$\begin{aligned} y_1(t) &= g_1(x(t), \boldsymbol{p}) = kE \quad \text{new incidences per day,}\\ y_2(t) &= g_2(x(t), \boldsymbol{p}) = \nu I \quad \text{number of deaths due to COVID-19 per day}\end{aligned} \tag{4.1}$$

Next, we formulate the statistical model as

$$Y_1^i = g_1(x(t_i), \boldsymbol{p}) + E_i \; i = 1, \cdots, n \quad Y_2^j = g_2(x(t_j), \boldsymbol{p}) + E_j \; j = 1, \cdots, n \tag{4.2}$$

where E_i and E_j are errors in measurement. Because of the measurement errors the data (random variables) Y_1 and Y_2 do not fall on the smooth path of the observations $g_1(x(t_i), \boldsymbol{p})$ and $g_2(x(t_i), \boldsymbol{p})$. We assume the measurement errors satisfy the following relationship;

$$E_i = g_1(x(t_i), \boldsymbol{p})\epsilon_i \quad E_j = g_2(x(t_j), \boldsymbol{p})\epsilon_j$$

where ϵ_i and ϵ_j are independent and identically distributed with mean zero and constant variance σ_0^2. With this setting, the random variables Y_1 and Y_2 are assumed to have mean as the observations, that is

$$\mathbb{E}(Y_1) = g_1(x(t_i), \boldsymbol{p}) \; i = 1, \cdots, n \quad \text{and} \quad \mathbb{E}(Y_2) = g_2(x(t_j), \boldsymbol{p}) \; j = 1, \cdots, n$$

and the variances of the random variables Y_1 and Y_2 would be

$$Var(Y_1) = g_1(x(t_i), \boldsymbol{p})^2\sigma_0^2 \; i = 1, \cdots, n \quad \text{and} \quad Var(Y_2) = g_2(x(t_j), \boldsymbol{p})^2\sigma_0^2 \; j = 1, \cdots, n$$

To estimate the model parameters, we minimize the sum of the squared differences between the model predictions and the measurements , where ω_1 and ω_2 represent the weights for the least squares.)

$$\hat{p} = \min_p \sum_{i=1}^{n} (\omega_1(Y_1^i - g_1(x(t_i), \boldsymbol{p}))^2 + \omega_2(Y_2^i - g_2(x(t_i), \boldsymbol{p}))^2) \quad \text{subject to} \quad p > 0 \tag{4.3}$$

The MATLAB function `fminsearchbnd` computes this above constraint optimization problem (4.3). The system of ODEs in the epidemiological model is solved using the MATLAB `ode15s` function.

4.3. Identifiability Analysis. Before estimating the model parameters, we first determine whether it is possible to uniquely asses the parameters from the given data. This process is called identifiability analysis and it is an important step in estimating model parameters [8]. The identifiability analysis is usually performed in two steps: first step is the structural identifiability and the second step is the practical identifiability. Structural identifiability is the property of the epidemic model, and it assumes noise-free data. In structurally identifiaility analysis, the objective is to determine whether the model is structured to reveal the parameter estimates. Practical identifiability then is performed with actual data which is contaminated with noise. If an epidemic model is not structurally identifiable, then different sets of parameters would give the same output. In such a case, when the epidemic model is structurally unidentifiable, the estimated parameters would be unreliable and thus one should be cautious in using the model to forecast the future cases or the determine the intervention strategies.

Structural Identifiability Analysis: To study the structual identifiability analysis of the model, we first eliminate the unobserved state variables and obtain a set of monic differential polynomial which only involves the observations. We eliminate the unobserved state variables using DAISY [1] and obtain the input-output equations as presented in the Appendix 6.

The input-output equations are monic differential polynomials whose coefficients consist of model parameters. We see from the first equation in (.1) that

$$\frac{dy_2}{dt} - y_1\nu + y_2(\nu + \alpha) = 0,$$

it is a monic polynomial since the coefficient of the highest term is 1. To obtain a monic polynomial from the second equation in (.1), we divide by the term $\beta^2\nu^2$ all the other coefficients. Let $c(\boldsymbol{p})$ denote the complete set of coefficients in (.1). We say that an epidemic model is structurally identifiable, if the mapping from parameter space into $c(\boldsymbol{p})$ is one-to-one. We define an epidemic model to be structurally identifiable as follows [16],

Definition 4.1. *Let $c(\boldsymbol{p})$ denote the coefficients of the input-output equations (.1), with $\boldsymbol{p} = (\beta, \beta_E, k, \nu, \alpha)$ denoting the parameters of the epidemic model (3.1). Then we say that the epidemid model (3.1) is structured to reveal its parameters from the observations (4.1) if and only if*

$$c(\boldsymbol{p}) = c(\hat{\boldsymbol{p}}) \implies \boldsymbol{p} = \hat{\boldsymbol{p}}$$

Based on this definition, we solve in Mathematica

$$c(\boldsymbol{p}) = c(\hat{\boldsymbol{p}})$$

where $\hat{\boldsymbol{p}} = (\hat{\beta}, \hat{\beta}_E, \hat{k}, \hat{\nu}, \hat{\alpha})$ and obtain

$$\beta = \hat{\beta},\ \beta_E = \hat{\beta}_E,\ k = k,\ \nu = \hat{\nu},\ \alpha = \hat{\alpha}.$$

That is we showed that the epidemic model (3.1) is structured to reveal its parameters uniquely from the observations of daily incidences and deaths. The results from conducting structural identifiability analysis is shared in section **??**.

To perform the identifiability analysis for the second epidemic model (3.2), we first need to eliminate the unobserved state variables and obtain the input-output equations. We used DAISY to obtain the input-output equations, but DAISY was unable to compute input-output equations due to computational limitations. Therefore, for the second epidemic model (3.1) we only perform practical identifiability analysis as outlined in the following section.

Practical Identifiability Analysis: Structural identifiability is a property of the epidemic model itself and it is performed without the actual data. It is based on the observations, which include the data points at the discrete points, and observations are smooth continuous functions. That is Structural identifiability assumes that the data are noise free, and the output is known for every point in time. A model which is structurally identifiable might not be practically identifiable when actual data are used to estimate the parameters. Next we perform practical identifiability analysis using Monte Carlo simulations. We present noise into the observed data at increasing levels to produce 1000 synthetic data sets. Then, the true parameter set $\hat{\boldsymbol{p}}$, obtained through curve-fitting, is used to refit the epidemiological model to the new synthetic data, observing how the parameter values react in response to the noise. In this project, noise was presented to the data collected from the Florida Department of Health each week during the 12-week period and the model was refit 1,000 times at each error level ($0\%, 1\%, 5\%, 10\%,$ and 20% error). Monte Carlo Simulations are performed in this project in the below steps [18].

1. Obtain solutions of the model using the fitted values as the true parameter vector $\hat{\boldsymbol{p}}$ and acquire the observation vectors $\boldsymbol{g}_1(\boldsymbol{x}(t_i), \boldsymbol{p})$ and $\boldsymbol{g}_2(\boldsymbol{x}(t_i), \boldsymbol{p})$ at the specific time points of the data $\{t_i\}_{i=1}^n$.
2. Generate 1,000 artificial data sets from the statistical model (4.2) with an established measurement error. The synthetic data set is generated from a normal distribution with a mean that is the output vector obtained in step (1.) and a standard deviation that is $\sigma_0\%$ of the mean. Thus,

$$Y_i = g(\boldsymbol{x}(t_i), \hat{\boldsymbol{p}}) + g(\boldsymbol{x}(t_i), \hat{\boldsymbol{p}})\epsilon_i \quad i = 1, 2, \cdots, n,$$

   such that $\mathbb{E}(\epsilon_1) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma_0^2$. Therefore, the random variables y_i have a mean of $\mathbb{E}(y_i) = g(\boldsymbol{x}(t_i), \hat{\boldsymbol{p}})$ and a variance of $\mathrm{Var}(y_i) = g(\boldsymbol{x}(t_i), \hat{\boldsymbol{p}})^2\sigma_0^2$.
3. Fit the model $\boldsymbol{x}'(t) = f(\boldsymbol{x}, \boldsymbol{p}) \quad \boldsymbol{x}(0) = \boldsymbol{x}_0$ to the synthetic data sets generated in Step 2 to approximate the parameters $\boldsymbol{p}_j$ for $j = 1, 2, \ldots 1,000$. The model is fit as follows

$$\boldsymbol{p}_j = \min_{\boldsymbol{p}} \sum_{i=1}^{n} (\omega_1 (Y_1^i - g_1(\boldsymbol{x}(t_i), \boldsymbol{p}))^2 + \omega_2 (Y_2^i - g_2(\boldsymbol{x}(t_i), \boldsymbol{p}))^2)$$

4. Determine the average relative errors (AREs) in estimating the parameter values in the set $\boldsymbol{p}$ by the following

$$ARE(p^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{|\hat{\boldsymbol{p}}^{(\boldsymbol{k})} - \boldsymbol{p_j}^{(\boldsymbol{k})}|}{\hat{\boldsymbol{p}}^{(\boldsymbol{k})}}$$

   where $p^{(k)}$ is the k^{th} parameter of the set $\boldsymbol{p}$, p_j^k is the k^{th} parameter of set p_j and $\hat{\boldsymbol{p}}^{(\boldsymbol{k})}$

is the k^{th} parameter of set $\hat{\boldsymbol{p}}$ of genuine parameters.

5. Repeats steps (1) - (5) for each level of noise $\sigma_0 = (0\%, 1\%, 5\%, 10\%, \text{and } 20\%)$.

By computing the average relative errors (ARE) in parameter estimation in step (4.), we can observe how the parameter values react to the introduced noise. When no noise is introduced into the data $\sigma_0 = (0\%)$, the ARE of the epidemiological parameters in a model that is structurally identifiable should be equal to or almost 0. As the noise in the dataset increases, the ARE of the parameters should respectively increase, and the rate of increased error will determine the practical identifiability of the model. If a parameter in the model is sensitive to measurement error, the parameter will have a significant ARE for a moderate level of noise in the data, and we then consider that parameter as practically unidentifiable. To have practically identifable parameters, we analyze the Monte Carlo Simulations for ARE values that are reasonable for the errors in measurement.

In this project, the steps to perform Monte Carlo simulations are repeated with each increasing level of error in data, and the epidemiological model is re-fit 1,000 times at each error level. The average relative errors (AREs) for each parameter estimation are recorded as results. By analyzing the ARE in parameter estimation in comparison to the level of noise in the data, the practical identifiability of the model and the reliability of each parameter value estimation can be measured. Parameters that demonstrate high levels of average relative error at increasing noise levels are likely to be unreliable estimations and having many unreliable parameter estimations makes a model maintain weak identifiability.

5. Results.

5.1. Structural Identifiability Results. As described in Section 4.3, DAISY is used to study the structural identifiability of the model. We summarize the result from analyzing the structural identifiability of (3.1) in the following Proposition 5.1.

Proposition 5.1. *The epidemic model* (3.1) *is structurally identifiable from daily incidences and deaths.*

5.2. SEIR Model Fitting Results. In the weeks after May 25th, the lockdown period in Florida ended and individuals began to leave their homes and the social distancing rate was significantly lowered compared to the first 6 weeks of data collection. Thus, the social distancing model is no longer appropriate to model the behavior of the disease starting from the week of June 1st. From June 1st to September 13th, data was still collected from the Florida Department of Health but was combined to fit all of the data starting from the week of June 1st with the regular SEIR model (non-social distancing) (1). The curve-fitting process was similar to the process with the first 6 weeks, just with the SEIR model. The parameter values that best curve-fit to the non-social distancing model are the values that minimize the sum of the squared differences between the model predictions and the data as shown by expression (3). For this fit, the data included the number of daily COVID-19 infections and disease-induced deaths reported by the Florida Department of Health beginning from May 2nd until September 13th. Since the Florida Department of Health reports the number of COVID-19 infections and deaths for the past 30 days since the day of data collection, the first set of data that was collected on June 1st begins from May 2nd. The results of this

project include the set of the 5 epidemiological parameter values that provided the best fit for the data collected between June 1st and September 13th. Table 13 below lists the parameter values that best established an accurate fitting of the non-social distancing model to the data starting from June 1st according to the MATLAB computer software.

SEIR Model Parameter Estimations

Date	β	α	ν	β_E	k
05/2/20 - 09/13/20	0.0301	0.2000	0.0020	0.5277	0.5000

Table 3

The parameters that best fit the epidemiological model for incidences and death data reported by Florida Department of Health. The fit is from May 2nd to September 13th.

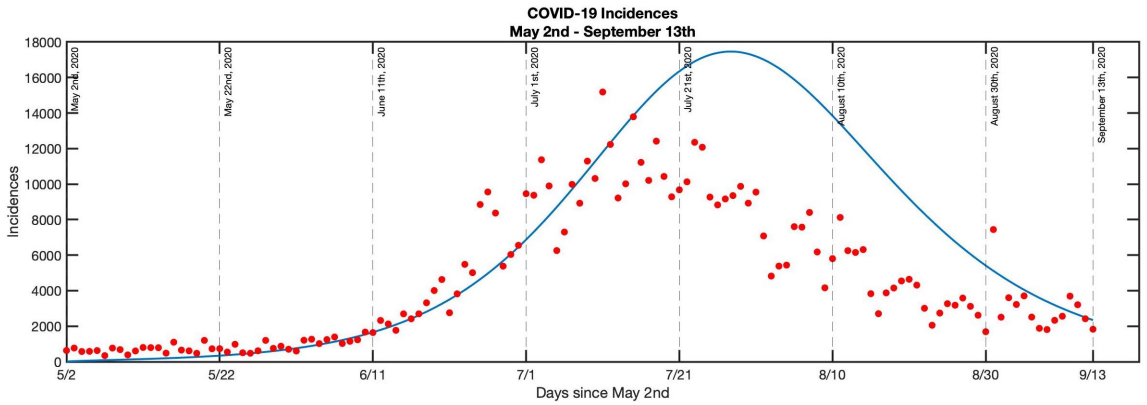

Figure 3. *Plot of the epidemiological model (3.1) (blue line) fitted to the observed incidence data (red dots) from the Florida Department of Health from May 2nd to September 13th, after the lockdown period in Florida.*

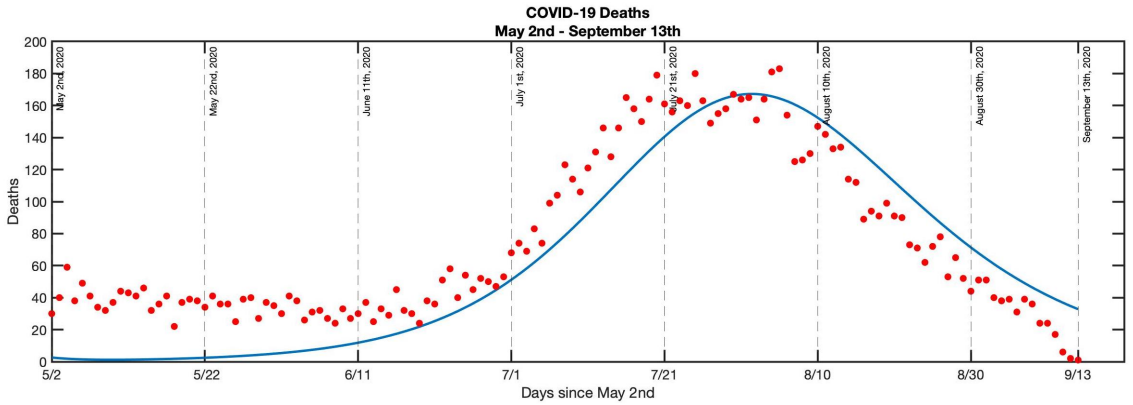

Figure 4. *Plot of the epidemiological model (3.1) (blue line) fitted to the observed death data (red dots) from the Florida Department of Health from May 2nd to September 13th, after the lockdown period in Florida.*

289 **5.3. SEIR Model Monte Carlo Simulation Results.** Similar to the social distancing mod-
290 els, Monte Carlo simulations are conducted on the non-social distancing SEIR model starting
291 from May 2nd to determine the identifiability of the epidemiological model and the reliabil-
292 ity of the parameter estimations. The true parameter values are provided in Table 13. The
293 average relative error for each parameter is presented with an increasing level of error in the
294 data (r). If the average relative error value from the Monte Carlo Simulations is high, the
295 parameter is considered not identifiable. For example, looking at Table 10 we notice that as
296 the error in data increases, the average relative error in all 5 parameters increases very slowly:
297 20% error in the data results in only 13.39% average relative error in β - the parameter with
298 the highest average relative error. This suggests that the parameter estimations from the
299 non-social distancing model are practically identifiable because the average relative errors in
300 all 5 parameters are controlled- many of the average relative errors in parameter estimation
301 are close to 0% even when there is 20% error in data.

Parameter Estimations

r (error in data)	β	α	ν	β_E	k
0%	0	0	0	0	0
1%	0.86	0.08	0.12	0.10	0.05
5%	4.03	0.36	0.49	0.45	0.19
10%	7.85	0.78	0.91	0.87	0.31
20%	13.39	1.38	1.42	1.52	0.46

Table 4

Average Relative Errors (AREs) in parameter estimation for errors of 0%, 1%, 5%, 10%, 20% for data from June 1st to September 21st are shown in the table above. The AREs are calculated by formula (4).

302 **5.4. Social Distancing Model Parameter Estimation results.** By curve-fitting the epi-
303 demiological model to the weekly COVID-19 data from the Florida Department of Health,
304 the parameter values that establish the best fit between the model and the data can be deter-
305 mined. The parameter values that best curve-fit to the model are the values that minimize
306 the sum of the squared differences between the model predictions and the data as shown by
307 expression (3). The results of this project include the set of the 9 epidemiological parameter
308 values that provided the best fit for each week, to observe the evolution of certain parameters
309 in the epidemiological model as more data are collected each week. Table 3 below lists the
310 parameter values that best established an accurate fitting to the data for each week according
311 to the MATLAB computer software.

Date	β	η	δ	α	ν	ρ	β_E	δ_E	k
Before the Peak (3/23/20 – 4/5/20)	0.0040	0.4014	0.1495	0.1960	0.0088	0.0530	1.6001	0.1635	0.3957
At the Peak (3/23/20 - 4/20/20)	0.0040	0.3977	0.1386	0.2000	0.0077	0.0531	1.5982	0.1615	0.3935
After the Peak (3/23/20 - 5/25/20)	0.1922	0.2654	0.4564	0.0952	0.0054	0.0059	1.4122	0.0083	0.1976

Table 5

The parameters that best fit the epidemiological model (3.2) for incidences and death data reported by Florida Department of Health. Each fit starts from March 23rd and ends on the date specified in the table.

5.5. Figures. The figures shown below show the evolution of the epidemiological model (blue) in comparison to the observed data (red dots). In each figure, April 27th is represented by the dotted green line.

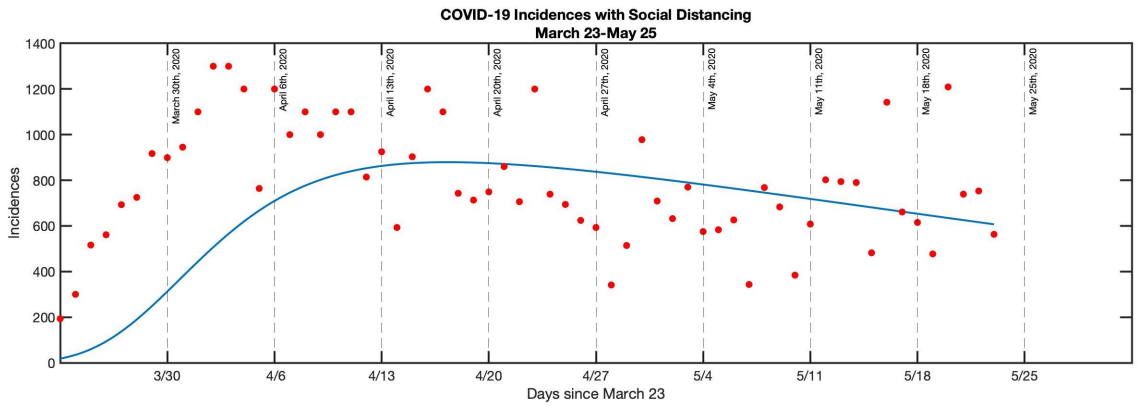

Figure 5. *Plot of the epidemiological model (3.2) (blue line) fitted to the observed incidence data from the Florida Department of Health (red dots).*

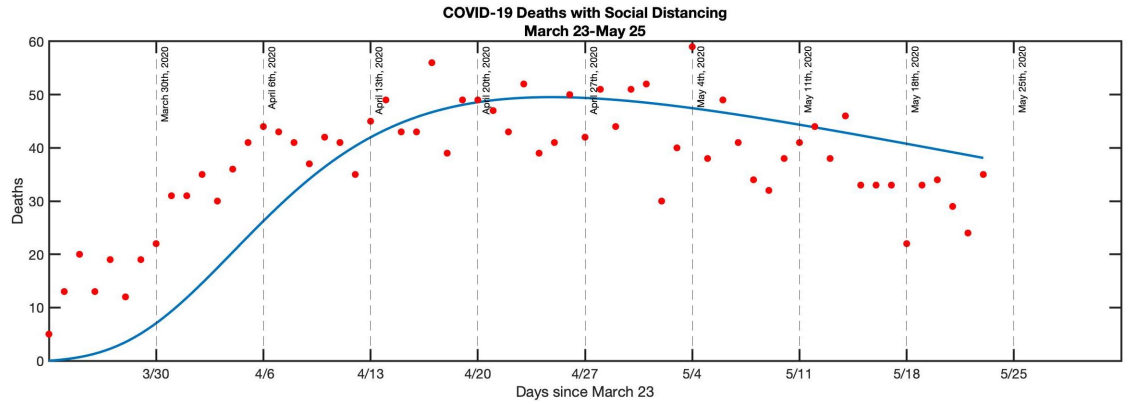

Figure 6. *Plot of the epidemiological model (3.2) (blue line) fitted to the observed death data from the Florida Department of Health (red dots).*

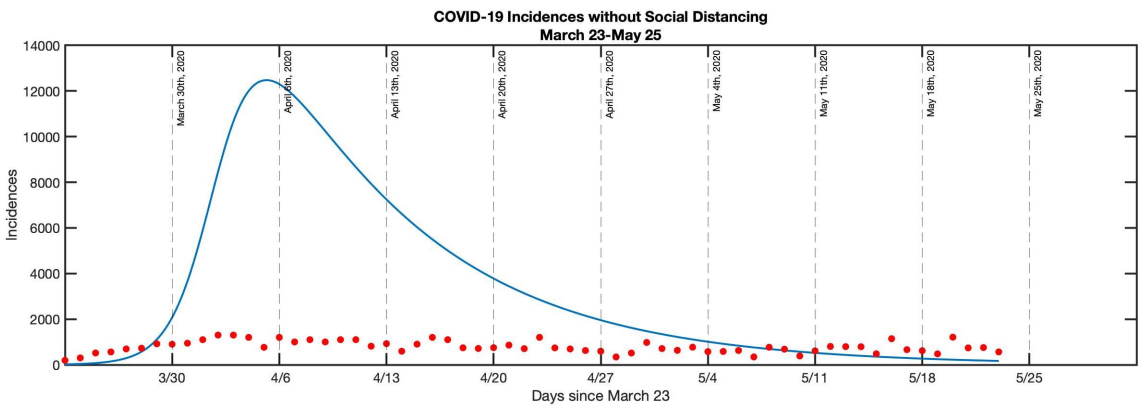

Figure 7. *Plot of the epidemiological model (3.1) (blue line) simulating what COVID-19 incidence rates would have looked like without social distancing compared to the observed death data from the Florida Department of Health (red dots).*

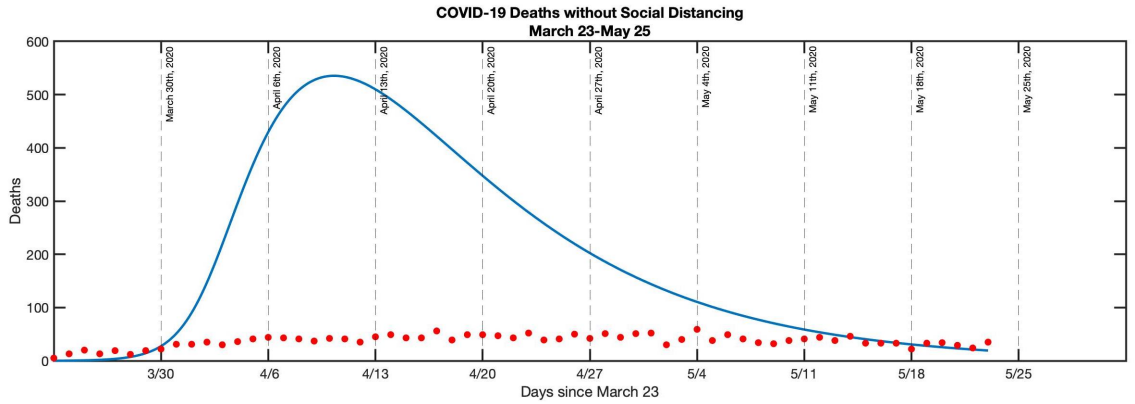

Figure 8. *Plot of the epidemiological model (3.1) (blue line) simulating what COVID-19 death rates would have looked like without social distancing compared to the observed death data from the Florida Department of Health (red dots).*

5.6. Social Distancing Model Monte Carlo Simulation Results. By conducting Monte Carlo simulations, the identifiability of the epidemiological model can be measured, and the reliability of the parameter estimations can be determined [17]. The true parameter values are provided in Table 3. The average relative error for each parameter is presented with an increasing level of error in the data (r). If the average relative error value from the Monte Carlo Simulations is high, the parameter is considered not identifiable. For example, looking at Table 5, we observe that as the error in data increases, the average relative error in δ increases very rapidly; 20% error in the data results in 111% average relative error in δ. This suggests that δ is not practically identifiable. Compared to β, what is observed is that even as the error in data increases to 20%, the average relative error in β remains controlled and is at just 2% error 7. For the week of April 20th, β is considered practically identifiable; the same cannot be concluded for δ.

Monte Carlo Simulation Before the Peak: 3/23/20 – 4/5/20

r (error in data)	β	η	δ	α	ν	ρ	β_E	δ_E	k
0%	0.01	0	6.50	0.32	0.19	0.03	0	0	0
1%	0.34	0.64	27.17	0.92	0.58	0.49	0.15	1.62	0.50
5%	4.46	5.25	68.94	3.69	2.92	7.04	2.16	11.02	3.97
10%	16.94	11.83	103.21	7.37	6.05	17.83	5.19	21.28	7.73
20%	36.06	23.95	134.87	13.84	11.61	34.00	10.82	29.75	13.06

Table 6

Average relative errors in parameter estimation for errors of 0%, 1%, 5%, 10%, 20% before the peak (during first 14 days) are shown in the table above. The average relative errors are calculated by formula (4).

Monte Carlo Simulation at the Peak: 3/23/20 – 4/20/20

r (error in data)	β	η	δ	α	ν	ρ	β_E	δ_E	k
0%	0.01	0.07	0.26	0.25	0.14	0.03	0.02	0.16	0.08
1%	0.03	0.26	4.26	0.48	0.46	0.23	0.07	0.51	0.16
5%	0.44	2.79	47.89	1.89	2.34	3.63	1.00	5.90	1.66
10%	1.81	6.18	86.21	4.63	5.28	9.34	2.77	14.01	3.60
20%	2.45	12.68	111.22	10.17	11.02	20.49	6.58	26.36	6.39

Table 7

Average relative errors in parameter estimation for errors of 0%, 1%, 5%, 10%, 20% at the peak (during the week of April 20th) are shown in the table above. The average relative errors are calculated by formula (4).

Monte Carlo Simulation After the Peak: 3/23/20 – 5/25/20

r (error in data)	β	η	δ	α	ν	ρ	β_E	δ_E	k
0%	0	0	0	0	0	0	0	0	0
1%	0.80	1.33	2.16	0.63	0.56	5.82	0.80	24.45	2.15
5%	2.83	4.72	7.19	3.03	2.71	18.92	3.06	69.55	7.56
10%	5.29	8.64	11.32	6.26	5.62	30.65	5.78	101.72	13.23
20%	9.75	13.46	18.64	12.61	11.27	47.64	8.91	177.70	21.04

Table 8

Average relative errors in parameter estimation for errors of 0%, 1%, 5%, 10%, 20% after the peak (during the week of May 25th) are shown in the table above. The average relative errors are calculated by formula (4).

In Table 9, the Monte Carlo Simulation results from Tables 4-6 are summarized in terms of the practical identifiability of the parameters in the model each week. Checkmarks are used to indicate the parameters that were found to be identifiable for each week. From the table, one can conclude that δ and δ_E were definitely not identifiable in this study because the average relative errors were extremely high for the individual Monte Carlo simulations (Tables 4-6). Furthermore, ρ was found to be unidentifiable in this study. The parameters β and η were not identifable before the peak in the data, and k was not identifiable after the peak in the data. The rest of the parameters in the model were considered identifiable.

Summary of Practical Identifiability Analysis

Date	β	η	δ	α	ν	ρ	β_E	δ_E	k
Before the Peak (3/23/20 – 4/5/20)				✓	✓		✓		✓
At the Peak (3/23/20 - 4/20/20)	✓	✓		✓	✓		✓		✓
After the Peak (3/23/20 - 5/25/20)	✓	✓		✓	✓		✓		

Table 9

A summary of tables 4-6 is displayed in the table above. Checkmarks indicate the sections of data for which the parameter was found to be practically identifiable, which can be concluded by observing how rapidly the average relative error increases with respect to the error in data each week (Tables 4-6).

337 **5.7. Effectiveness of Social Distancing Results.** The results of project serve to demon-
338 strate the efficiency of social distancing in preventing the spread of the COVID-19 disease. By
339 taking the total of the weekly incidences and deaths from the Florida Department of Health,
340 a comparison can be conducted between the actual data and the predicted numbers from
341 the non-social distancing model (3.1), which represent the spread of the COVID-19 disease if
342 there was no social distancing enforced, which leads to a greater number of contacts between
343 individuals. By quantifying the number of infected individuals and disease-induced deaths
344 that would have occurred if social distancing did not take place, the number of lives saved and
345 the number of prevented incidences can be measured. As outlined in Table 12 below, in the
346 first 6 weeks of data collection, an average of 184,586 COVID-19 incidences were prevented
347 and 8,588 lives were saved from the efficiency of social distancing guidelines. These results
348 suggest that social distancing plays a crucial role in preventing incidences and disease-induced
349 deaths.

Date	4/20/20	4/27/20	5/4/20	5/11/20	5/18/20	5/25/20
Total of Weekly Incidences (FDOH Data)	25,316	31,141	25,832	22,426	21,088	20,449
Total of Weekly Deaths (FDOH Data)	852	1,212	1,107	1,164	1,112	953
Total of Predicted Weekly Incidences (Model 1)	206,124	209,990	208,447	209,755	209,456	209,976
Total of Predicted Weekly Deaths (Model 1)	7,753	9,107	9,408	10,320	10,076	11,262
Total Incidences prevented by Social Distancing	180,808	178,849	182,615	187,349	188,368	189,527
Total Deaths prevented by Social Distancing	6,901	7,895	8,301	9,156	8,964	10,309

Table 10

Total weekly incidences/deaths in comparison to the non-social distancing model predictions

350 **5.8. Time dependent β.** From the results in Table 4, it can be concluded that the $SEIR$
351 model is structurally identifiable. However, from Figure 3 and 4 it is observed that the model
352 does not match the peak of the incidence data, and the model predicted deaths are much
353 lower than the data at the beginning. The reason why the model does not provide an exact
354 fit may be attributed to the fact that people's behaviors change over time, which influences
355 disease spread, and it is not possible to reflect those behavioral changes with a constant
356 transmission rate. An improved model may consider that the prevalence of social distancing
357 fluctuates as people's behaviors change in time. For example, it is common to see large social
358 gatherings during holiday seasons, which increases the likelihood for disease transmission in
359 the population. The best way to incorporate this in the epidemiological model is to have β,
360 the transmission rate, change in time. The approach to have β change in time was inspired
361 by the approach presented in [19]. This time dependent transmission rate is represented by
362 $\beta(t)$, where $\beta(t)$ is the transmission rate at time t.
363 In this project, the value of $\beta(t)$ is derived from the COVID-19 data collected from the
364 Florida Department of Health. In the model, daily COVID-19 incidences are given by

$$y_1(t) = kE$$

365 Taking the derivative gives

$$
\begin{aligned}
y_1'(t) &= k\left(\beta(t)S\left(\frac{I}{N}+q\frac{E}{N}\right)-kE\right)\\
y_1'(t) &= k\left(\beta(t)S\left(\frac{I}{N}+q\frac{E}{N}\right)-y_1(t)\right)\\
y_1'(t)+ky_1 &= k\beta(t)S\left(\frac{I}{N}+q\frac{E}{N}\right)\\
\beta(t) &= \frac{y_1'(t)+ky_1}{kS\left(\frac{I}{N}+q\frac{E}{N}\right)}.
\end{aligned}
\tag{5.1}
$$

In order to determine y_1', the derivative of the data, the data are smoothed and then interpolated with B-splines. The data are interpolated with B-splines because the data consists of discrete time points (the Florida Department of health reports new daily COVID-19 infections and disease-induced deaths). Without smoothing the data, the derivative is difficult to determine because of the amount of noise in the epidemiological data. A plot of B-spline interpolation of the data without smoothing is given in Figure 9.

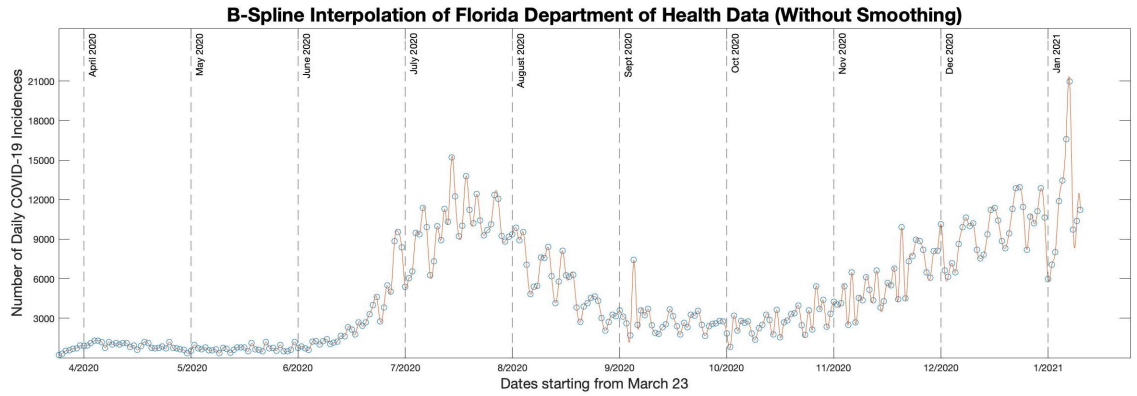

Figure 9. *The data from the Florida Department of Health is interpolated without being smoothed. In this case, it is difficult to determine the derivative y_1' because of the amount of noise in the data.*

As seen in Figure 9, it is not ideal to determine the derivative y_1' from the original interpolated data. To make it easier to compute the derivative, the data are smoothed to account for the noise in the epidemiological data. The data are smoothed using the `smoothdata` function in MATLAB. A plot of the B-spline interpolated data after smoothing is given in Figure 10.

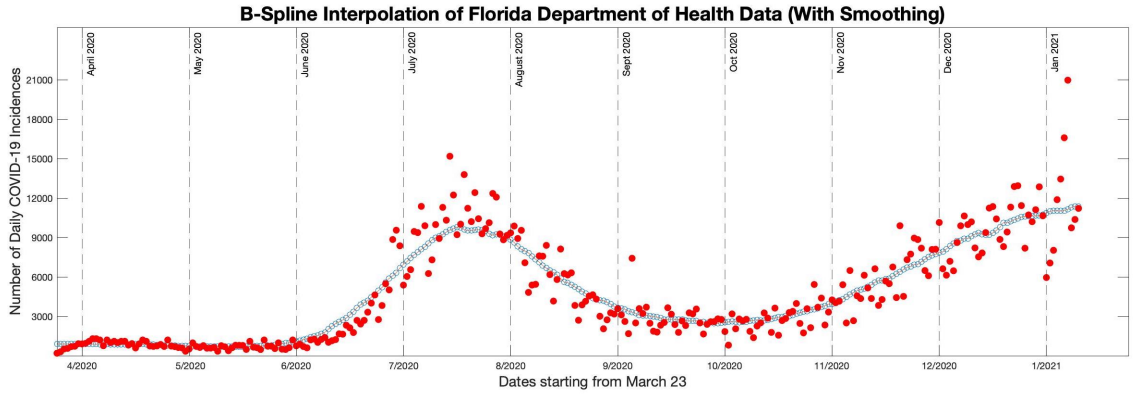

Figure 10. *Interpolation of the data from the Florida Department of Health after being smoothed. By smoothing the data, the amount of noise in the data are reduced, making it easier to determine the derivative* y_1'.

In this project, y_1' and kE come from the data collected from the Florida Department of Health. To get the derivative y_1' the epidemiological data must be first smoothed and then interpolated with B-splines to take the derivative. After determining the derivative y_1', we can express $\beta(t)$ as the transmission rate of the disease at time t. This is demonstrated in the rewritten SEIR model with a time dependent transmission rate as below.

$$\begin{aligned} S' &= -\beta(t)S\left(\frac{I}{N} + q\frac{E}{N}\right), \\ E' &= \beta(t)S\left(\frac{I}{N} + q\frac{E}{N}\right) - kE, \\ I' &= kE - (\alpha + \nu)I, \\ R' &= \alpha I. \end{aligned} \tag{5.2}$$

Using model (5.2), one can predict COVID-19 cases in Florida from March 23rd over a prolonged period of time while accounting for behavioral changes in the population that may influence the rate of disease transmission. The model predicts very well, as shown in figure 11 below. In this plot, the gray bars represent epidemiological data from the Florida Department of Health, and the blue curve is the model prediction from model (5.2). In this model, the value of $\beta(t)$ is determined from the data, which a recipe is provided for (5.1), and the values of the other parameters, including $k, \alpha,$ and ν, are determined from the curve-fitting process. Since typical SEIR models have a constant value for β, those models can only predict 1 wave of the disease. However, because model (5.2) uses a time dependent β value, the second wave of the disease in Florida can also be predicted, which is shown in figure 11.

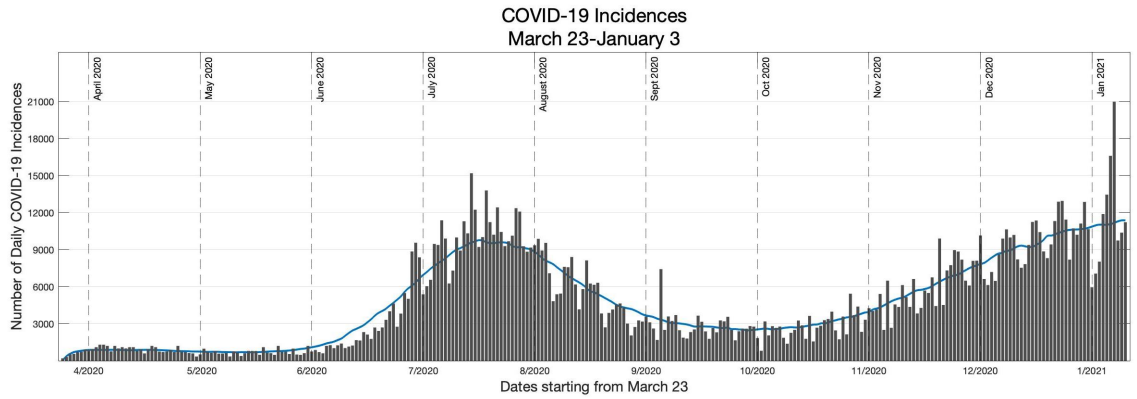

Figure 11. *Plot of the epidemiological model with a time dependent transmission rate (5.2) (blue line) fitted to the data from the Florida Department of Health representing the number of daily COVID-19 incidences from March 23rd to January 3rd.*

6. Discussion. The novel coronavirus (COVID-19) has made a broad impact as a global pandemic, spreading to over 210 countries and leading to over 131 million confirmed cases and 2.85 million deaths (as of April 2021) [14]. In order to effectively implement control measures to prevent the spread of the disease, it is crucial to forecast the behavior of the disease spread and determine the epidemiological characteristics of the disease. We utilize an SEIR model to determine epidemiological characteristics such as the rate of infection, incubation period, and rate of recovery for the COVID-19 disease in the state of Florida according to incidence and disease-induced death data from the Florida Department of Health.

During the COVID-19 pandemic, one major public health guideline advertised by the CDC was "social distancing", which consists of non-pharmaceutical control measures to prevent the spread of the disease by maintaining a physical distance (at least 6 feet) between individuals and limiting the number of person-to-person contacts. By developing a social distancing model to represent the disease spread, we can quantify the efficiency of social distancing in mitigating transmission of the disease and simulate the effect on the disease spread if there was no social distancing in the population. However, because the Monte Carlo Simulations of the social distancing model established high average relative errors in parameter estimation, the social distancing model was found to be not practically identifiable, suggesting that the model is not good at forecasting the outbreak and should not be used to determine control measures to slow the spread of the disease. To quantify the effectiveness of social distancing, we calculate the difference between the predicted weekly incidences and deaths and the weekly sum of the Florida Department of Health data to determine how many incidences were prevented and how many lives were saved by social distancing. We found that an average of about 185,000 weekly COVID-19 incidences were prevented and an average of about 8,500 lives were saved because of the effectiveness of social distancing, between the weeks of April 20th and May 25th. This suggests the benefit of implementing non-pharmaceutical control measures such as social distancing in an effort to mitigate disease transmission.

Because the lockdown period ended in Florida in the weeks starting from June 1st, we introduce a different model to simulate the disease spread without social distancing, the

typical SEIR model. With low average relative errors in the Monte Carlo Simulation, the SEIR model was found to be both structurally and practically identifiable, increasing the confidence in model predictions and the confidence in using the SEIR model to determine possible control measures in an effort to slow the disease spread. Since the SEIR model was structurally identifiable, we provided a recipe on how to determine $\beta(t)$, the transmission rate dependent on time, using incidence data from the Florida Department of Health. However, we have not investigated the identifiability of the time-dependent $\beta(t)$.

Mathematical modeling of infectious diseases is crucial in understanding the epidemiological characteristics of diseases, advising the implementation of disease control measures, and forecasting the spread of the disease. To incorporate the time-dependent transmission rate, which is derived from the data, in disease forecasting, the disease can only be forecasted for a short period of time without data in the future. The value that is found as $\beta(t)$ for the end of the data collection period can be assumed to be constant for the next following week. It is not advised to predict ahead of that because people's behaviors might change later in time. However, shortly in the future (such as within the next week), people's behaviors may linger. Another application of the model is to analyze vaccination efforts, in a manner as presented by Glenn Webb [6]. The effectiveness of vaccination efforts can be measured by modifying the SEIR model to include a vaccinated class, and then a time-dependent vaccination rate can be determined following a similar methodology as presented here to find the time-dependent transmission rate. This can be used to measure the effectiveness of vaccination efforts in a population for a short period of time. Future directions include using the model with the time-dependent transmission rate to analyze the vaccination efforts in Florida.

Appendix.

Input output equation of SEIR model (3.1)

(.1)

$$0 = \frac{dy_2}{dt} - y_1\nu + y_2(\nu+\alpha)$$

$$\begin{aligned}
0 = {} & \frac{d^3y_1}{dt^3}\frac{dy_1}{dt}y_1^2\beta^2\nu^2 + \frac{d^3y_1}{dt^3}\frac{dy_1}{dt}y_1y_2\beta\nu k(2\beta_E-\nu) + \frac{d^3y_1}{dt^3}\frac{dy_1}{dt}y_2^2\beta_E k^2(\beta_E-\nu) + \frac{d^3y_1}{dt^3}y_1^3\beta^2\nu^2 k \\
& + \frac{d^3y_1}{dt^3}y_1^2y_2\beta\nu k^2(2\beta_E-\nu) + \frac{d^3y_1}{dt^3}y_1y_2^2\beta_E k^3(\beta_E-\nu) - \frac{d^2y_1}{dt^2}^2 y_1^2\beta^2\nu^2 + 2\frac{d^2y_1}{dt^2}^2 y_1y_2\beta\nu k(-\beta_E+\nu) \\
& + \frac{d^2y_1}{dt^2}^2 y_2^2\beta_E k^2(-\beta_E+2\nu) - 2\frac{d^2y_1}{dt^2}\frac{dy_1}{dt}^2 y_1\beta^2\nu^2 + \frac{d^2y_1}{dt^2}\frac{dy_1}{dt}^2 y_2\beta\nu k(-2\beta_E-\nu) + \frac{d^2y_1}{dt^2}\frac{dy_1}{dt}y_1^2\beta\nu^2 k(-4\beta-\beta_E+\nu) \\
& + \frac{d^2y_1}{dt^2}\frac{dy_1}{dt}y_1y_2\nu k(\beta\beta_E\nu+\beta\beta_E\alpha-5\beta\beta_E k-\beta\nu^2-\beta\nu\alpha+3\beta\nu k-\beta_E^2 k-\beta_E\nu k) \\
& + \frac{d^2y_1}{dt^2}\frac{dy_1}{dt}y_2^2\beta_E k^2(\beta_E\nu+\beta_E\alpha-\beta_E k+\nu^2+\nu\alpha+3\nu k) + \frac{d^2y_1}{dt^2}y_1^3\beta\nu^2k^2(-\beta_E+\nu) \\
& + \frac{d^2y_1}{dt^2}y_1^2y_2\nu k^2(\beta\beta_E\nu+\beta\beta_E\alpha+\beta\beta_E k-\beta\nu^2-\beta\nu\alpha-\beta_E^2 k-\beta_E\nu k) \\
& + \frac{d^2y_1}{dt^2}y_1y_2^2\beta_E k^3(\beta_E\nu+\beta_E\alpha+\beta_E k+\nu^2+\nu\alpha-\nu k) + 2\frac{dy_1}{dt}^4\beta^2\nu^2 + \frac{dy_1}{dt}^3 y_1\beta\nu^2k(3\beta+3\beta_E-\nu) \\
& + \frac{dy_1}{dt}^3 y_2\nu k(-4\beta\beta_E\nu-4\beta\beta_E\alpha-\beta\beta_E k+\beta\nu^2+\beta\nu\alpha-2\beta\nu k-\beta_E^2 k+\beta_E\nu k) \\
& + \frac{dy_1}{dt}^2 y_1^2d^2k(\beta\beta_E\nu+\beta\beta_E\alpha+5\beta\beta_E k-\beta\nu k+2\beta_E^2 k-\beta_E\nu k) \\
& + \frac{dy_1}{dt}^2 y_1y_2\nu k(-\beta\beta_E\nu^2-2\beta\beta_E\nu\alpha-7\beta\beta_E\nu k-\beta\beta_E\alpha^2-7\beta\beta_E\alpha k-3\beta\beta_E k^2+\beta\nu^2k+\beta\nu\alpha k-3\beta_E^2\nu k-3\beta_E^2\alpha k- \\
& + \frac{dy_1}{dt}^2 y_2^2\beta_E k^2(\beta_E\nu^2+2\beta_E\nu\alpha+\beta_E\nu k+\beta_E\alpha^2+\beta_E\alpha k-\beta_E k^2+\nu^2k+\nu\alpha k+2\nu k^2) \\
& + 2\frac{dy_1}{dt}y_1^3\beta_E\nu^2k^2(\beta\nu+\beta\alpha+\beta k+2\beta_E k-\nu k) \\
& + \frac{dy_1}{dt}y_1^2y_2\beta_E\nu k^2(-2\beta\nu^2-4\beta\nu\alpha-3\beta\nu k-2\beta\alpha^2-3\beta\alpha k-6\beta_E\nu k-6\beta_E\alpha k-2\beta_E k^2+2\nu^2k+2\nu\alpha k) \\
& + \frac{dy_1}{dt}y_1y_2^2\beta_E k^3(2\beta_E\nu^2+4\beta_E\nu\alpha+\beta_E\nu k+2\beta_E\alpha^2+\beta_E\alpha k+\nu^2k+\nu\alpha k) \\
& + y_1^4\beta_E\nu^2k^3(\beta\nu+\beta\alpha+2\beta_E k-\nu k) + y_1^3y_2\beta_E\nu k^3(-\beta\nu^2-2\beta\nu\alpha-\beta\alpha^2-3\beta_E\nu k-3\beta_E\alpha k+\nu^2k+\nu\alpha k) \\
& + y_1^2y_2^2\beta_E^2k^4(\nu^2+2\nu\alpha+\alpha^2).
\end{aligned}$$

REFERENCES

[1] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angió, *Daisy: a new software tool to test global identifiability of biological and physiological systems*, Comput Methods Programs Biomed, 88 (2007), pp. 52– 61, https://doi.org/10.1016/j.cmpb.2007.07.00.

[2] D. Cucinotta and M. Vanelli, *Who declares covid-19 a pandemic*, Acta Biomed, 91 (2020), pp. 157–160, https://doi.org/10.23750/abm.v91i1.9397.

[3] E. Iboib, C. N. Ngonghala, and A. B. Gumel, *Will an imperfect vaccine curtail the covid-19 pandemic in the U.S.?*, Infectious Disease Modelling, 5 (2020), https://doi.org/10.1016/j.idm.2020.07.006.

[4] B. Ivorra, M. R. Ferrández, M. Vela-Pérez, and A. M. Ramos, *Mathematical modeling of the spread of the coronavirus disease 2019 (covid-19) taking into account the undetected infections. the case of china*, Communications in Nonlinear Science and Numerical Simulation, 88 (2020), https://doi.org/10.1016/j.cnsns.2020.105303.

[5] Z. Liu, P. Magal, O. Seydi, and G. Webb, *A covid-19 epidemic model with latency period*, Infectious Disease Modelling, 5 (2020), pp. 323– 337, https://doi.org/10.1016/j.idm.2020.03.003.

[6] Z. Liu, P. Magal, and G. Webb, *Predicting the number of reported and unreported cases for the covid-19 epidemics in china, south korea, italy, france, germany and united kingdom*, Journal of Theoretical Biology, 509 (2021), https://doi.org/10.1016/j.jtbi.2020.110501.

[7] G. Massonis, J. R. Banga, and A. F. Villaverde, *Structural identifiability and observability of compartmental models of the covid-19 pandemic*, Annual Reviews in Control, (2020), https://doi.org/10.1016/j.arcontrol.2020.12.001.

[8] H. Miao, X. Xia, A. S. Perelson, and H. Wu, *On identifiability of nonlinear ode models and applications in viral dynamics*, SIAM Review, 53 (2011), pp. 3– 39, https://doi.org/10.1137/090757009.

[9] NewsCAP, *Social distancing is critical in reducing the spread of covid-19, according to recent analyses*, AJN, American Journal of Nursing, 120 (2020), p. 15, https://doi.org/10.1097/01.NAJ.0000694524.92148.71.

[10] C. N. Ngonghala, E. Iboib, S. Eikenberry, M. Scotch, C. R. MacIntyre, M. H.Bonds, and A. B. Gumel, *Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus*, Mathematical Biosciences, 325 (2020), https://doi.org/10.1016/j.mbs.2020.108364.

[11] K. Roosa and G. Chowell, *Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models.*, Theoretical Biology and Medical Modelling, 16 (2019), pp. 1– 15, https://doi.org/10.1186/s12976-018-0097-6.

[12] V. Sreejithkumar and N. Tuncer, *Identifiability analysis of the human h1n1 influenza virus*, FAU Undergraduate Research Journal, 10 (2021), pp. 50–50, https://journals.flvc.org/faurj/article/view/128941.

[13] Florida Department of Health, *Florida's covid-19 data and surveillance dashboard*, 2020, https://floridahealthcovid19.gov (accessed 2020-17-06).

[14] U.S. Centers for Disease Control and Prevention, *Covid data tracker*, n.d., https://covid.cdc.gov/covid-data-tracker/ (accessed 2021-25-08).

[15] U.S. Centers for Disease Control and Prevention, *Clinical questions about covid-19: Questions and answers*, February 2020, https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html#COVID-19-Risk (accessed 2020-17-06).

[16] N. Tuncer, H. Gulbudak, V. L. Cannataro, and M. Martcheva, *Structural and practical identifiability issues of immuno-epidemiological vector-host models with application to rift valley fever*, Bulletin of Mathematical Biology, 78 (2016), pp. 1796–18271, https://doi.org/10.1007/s11538-016-0200-2.

[17] N. Tuncer and T. T. Le, *Structural and practical identifiability analysis of outbreak models*, Mathematical Biosciences, (2018), pp. 1—- 18, https://doi.org/10.1016/j.mbs.2018.02.004.

[18] N. Tuncer, M. Marctheva, B. LaBarre, and S. Payoute, *Structural and practical identifiability analysis of zika epidemiological models*, Bulletin of Mathematical Biology, 8 (2018), pp. 2209– 2241, https://doi.org/10.1007/s11538-018-0453-z.

[19] G. Webb, *A covid-19 epidemic model predicting the effectiveness of vaccination*, Mathematics in Applied Sciences and Engineering, 2 (2021), https://doi.org/10.5206/mase/13889.