The Disutility of SEIR Model Forecasts During the COVID-19 Pandemic

Tarini Sudhakar   Ashna Bhansali   John Walkington   David Puelz*

THIS VERSION: SEPTEMBER 26, 2023

Abstract

During the COVID-19 pandemic, several forecasting models were released to predict the spread of the virus along variables vital for public health policymaking. Of these, the Susceptible-Infected-Recovered (SIR) compartmental model was the most common. In this paper, we investigate the forecasting performance of The University of Texas COVID-19 Modeling Consortium SIR model. We consider the following daily outcomes: hospitalizations, ICU patients and deaths. We evaluate the overall forecasting performance, highlight some stark forecast biases, and consider forecast errors conditional on different pandemic regimes. We find that this model tends to *over forecast* over the longer horizons and when there is a surge in viral spread. We bolster these findings by linking them to faults with the SIR framework itself.

Keywords: Compartmental models, SEIR models, COVID-19.

---

*email: David.Puelz@mccombs.utexas.edu. Salem Center for Policy, Department of Finance, & Department of Information, Risk, and Operations Management.

1.   Introduction

Forecasts of the spread of COVID-19 in the United States and across the world have played a significant role in informing policymakers. Employing different mathematical approaches, these forecasting models provide predictions that inform critical decisions such as healthcare administration, allocation of medical supplies, and business and school closures. Some models even attempted to estimate the impact of current and future policies on human behavior and COVID-19 transmission (Cramer et al., 2022).

For the Austin-Round Rock metropolitan statistical area, forecasts given by The University of Texas (UT) COVID-19 Modeling Consortium shaped COVID-19 policies to a significant extent such as the introduction of staged lock downs. These had serious repercussions on the overall economy, spanning the closure of small businesses to setbacks in K-12 students due to hybrid or online education. Statewide, Texas is estimated to have suffered a GDP loss of $106 million and job losses of 1.2 million (Liu et al., 2021). Texas also suffered unprecedented setbacks in student achievement in reading and mathematics due to the mass transition to remote learning. The Texas Education Agency cites a roughly 3-month setback in educational attainment relative to a pre-COVID baseline. From 2019 to 2021, the percentage of Texas students performing at or above their grade level in math dropped from 50% to 35% (Texas Education Agency, 2021). While we have moved on from such strict measures for now, some questions remain unanswered. Did we base these policies on well-sourced data? In the future, can we still rely on these forecasting models to aid our decision-making?

Research from the US COVID-19 Forecast Hub (Cramer et al., 2022) demonstrated that stand-alone models tend to generate large prediction errors, especially when forecasting over a long time horizon. The UT COVID-19 Modeling Consortium model is one such example. Given its impact on COVID-19 policymaking, this paper investigates the UT model's forecasting performance in light of realized data, discovers that forecast errors are biased, predictable, and depend on pandemic regime, and proposes paths forward and alternative approaches. The forecasts considered are for the Austin-Round Rock metropolitan statistical area.

2.   Sorting COVID-19 into boxes: SIR Models

Most researchers adopted the Susceptible-Infected-Recovered (SIR) model and its modifications to forecast COVID-19 metrics due to its simplicity and prevalence among academics and training of epidemiologists (Rahimi et al., 2021). The SIR modeling framework falls under the category of compartmental models which date back to the seminal work of Ross (1916).

In these models, individuals within a closed population are separated into mutually exclusive groups, or compartments, based on their disease status. At any given time, each individual is considered to be in one compartment, but can move to another compartment based on the model parameters. As per the model's assumptions, a susceptible individual will become infected by coming into contact with an infected individual. The individual during the infected period is assumed to be contagious.

After this, the individual advances to a non-contagious state, known as recovery. Recovery may also be death or effective isolation (Tolles and Luong, 2020).

The classical SIR model is based on the following set of differential equations:

$$\frac{\partial S}{\partial t} = -\beta IS, \tag{1}$$

$$\frac{\partial I}{\partial t} = \beta IS - \gamma I, \tag{2}$$

$$\frac{\partial R}{\partial t} = \gamma I, \tag{3}$$

where at time t, S is the fraction of individuals susceptible, I is the fraction of individuals infected, R is the fraction of individuals recovered and β and γ are the rate of transmission and recovery, respectively. Equation 1 models the fraction of people would be susceptible to the virus at a certain point, given rate of transmission and infected individuals. Equation 2 models the instantaneous fraction of infected people, using susceptible individuals and rate of recovery. Equation 3 models the instantaneous fraction of recovered individuals using the rate of recovery.

Modified SIR models, such as Susceptible-Exposed-Infected-Recovered (SEIR), require a more complex set of equations and set of parameters. Due to the presence of little information and lack of reliable data at the beginning of the COVID-19 pandemic, many researchers relied on the classical SIR implementation (Moein et al., 2021). Nonetheless, the set up for this modified model formulation is as follows (Moehlis, 2002). The entire population is divided into the following groups:

- S is the fraction of susceptible individuals who are able to contract the disease,

- E is the fraction of exposed individuals who have been infected but are not yet infectious,

- I is the fraction of infectious individuals who can transmit the disease,

- R is the fraction of recovered individuals who have become immune.

Here, we assume that there are equal birth and death rates μ, α is the mean latency period for the virus, γ is the mean infectious period, the contact rate $\beta(t)$ may optionally be a function of time, and recovered individuals do not contract the disease again. This admits the following set of equations:

$$\frac{\partial S}{\partial t} = \mu - \beta(t)IS - \mu S,$$

$$\frac{\partial E}{\partial t} = \beta(t)IS - (\mu + \alpha)E,$$

$$\frac{\partial I}{\partial t} = \alpha E - (\mu + \gamma)I,$$

Since each variable is defined as a fraction of the entire population, we calculate R from the equation $S + E + I + R = 1$.

Additionally, when $\beta = \beta_0 =$ constant, i.e., not a function of time, the reproduction rate is given by

$$R_0 = \frac{\beta_0 \alpha}{(\mu + \alpha)(\mu + \gamma)}.$$

Example: Compartmental model use in China. In China, the initial spread of the virus coincided with a travel rush that comes with the Spring Festival. Yang et al. (2020) accounted for this using a modified SIR model based on domestic migration before and after 23 January and most recent COVID-19 epidemiological data to predict the epidemic progression. They corroborated their forecasts using a Long-Short-Term Model, a type of recurrent neural network trained on the 2003 SARS coronavirus outbreak data used to predict numbers of new infections over time.

To apply the SIR model, they needed to estimate β, σ and γ parameters. β is the product of (k) people exposed each day to infected people and (b) probability of transmission after exposure. σ is the incubation rate or the rate at which exposed individuals develop symptoms (average incubation duration is $1/\sigma$). Since the incubation period of the SARS-CoV-2 had been reported between 2 to 14 days, they chose 7 days to be at the midpoint. γ is the average rate of recovery or death in infected populations. They estimated the SIR model on epidemic data from Hubei to determine (b) probability of transmission and then derived β and γ.

3. COVID-19 Forecasting in the United States: IHME and UT Austin

The Institute for Health Metrics and Evaluation (IHME) COVID-19 Forecasting Team located at the University of Washington created multiple models over the course of the pandemic. The initial model garnered much press and attention and used statistical curve-fitting to estimate hospital bed utilization, ICU admissions, ventilator use, and deaths from 25 March 2020 to 29 April 2020 (IHME COVID-19 Health Service Utilization Forecasting Team and Murray, Christopher JL, 2020).

In this model, the IHME team took a different approach to model death rates compared with the classical compartmental framework. They critiqued SEIR models for their assumption of random mixing between all individuals in the population because under that assumption, millions of COVID-19-related deaths were predicted very early in the United States. Random mixing does not account for behavioral changes and government-mandated social distancing measures. Instead, the IHME team modelled actual COVID-19 death rates since they would indicate virus transmission and fatality rates. According to the authors, deaths were also more accurately reported than cases, especially in limited testing areas as those would allocate tests for severely ill patients first. They also assumed hospitalization and related services to be highly correlated with deaths.

Their next model took on a hybrid approach, estimating revised death rates and then fitting an SEIR model until 26 May 2020 (IHME, 2020b; Friedman et al., 2021).The third model replaced curve-fitting with spline-fitting for the relationship between log cumulative deaths and log cumulative cases, while retaining the SEIR model estimation (IHME, 2020a; Friedman et al., 2021). When tested against other forecasting models for predictive accuracy, this third IHME model possessed the best performance (out of 7 models) (Friedman et al., 2021).

3.1. What is unique about the UT Model?

The UT COVID-19 Modeling Consortium developed a highly-publicized and widely-used model to predict trends in key pandemic variables, including hospitalizations and ICU patients, in the Austin-Round Rock metropolitan statistical area (Spencer Woody et al., 2020).[1]

Motivated by the IHME approach, the UT Model developed an alternative curve-fitting method for forecasting COVID-19 mortality. It incorporated two unique features:

1. *Mobile-phone data to capture social distancing measures:* To capture the effect of changing social distancing measures for each US state on individual-level mobility, the model used local data from mobile-phone GPS traces from SafeGraph.[2]

2. *A correction for the underestimation of uncertainty in the IHME forecasts:* The IHME model estimated cumulative death rates using a least-squares-like procedure on the log scale and calculated confidence intervals based on large-sample statistical theory. For this method to produce valid uncertainty measures, consecutive model errors should be independent of each other. However, this assumption is violated in the IHME fitting procedure as today's cumulative death rate includes yesterday's death rate and an increment. This implies that these two death rates must be correlated. The UT model corrected this by fitting daily death rates using a mixed-effects negative-binomial generalized linear model, accounting for heteroskedasticity and correlation.

But the UT model's forecasts still generated large prediction errors. We analyze the UT forecasting model for hospitalizations, ICU patients, and deaths by following these steps:

1. *Visualize forecasts against realized data*: We plot realized data with forecasts generated by the UT model. While we have daily realized data on hospitalizations, ICU patients, and deaths, we do not have the same for forecasts as the model was not updated daily and had a different schedule for each parameter.

2. *Map forecasting errors from 1 to 20 days ahead*: We compute the forecasting errors out to 20 days ahead, showing the median forecasting error and inter quartile range (IQR) of the error distribution.

3. *Map forecasting errors, conditional upon whether the virus is spreading or waning*: We compute the forecasting errors from 1 to 20 days out, conditional upon whether there is a surge in the viral spread.

   For each parameter of interest, we define rising hospitalizations/ICU patients/deaths as surging and falling hospitalizations/ICU patients/deaths as waning. To smooth out noise in the realized data, we take a 14-day moving average of the realized data and then compute the percent change

---

[1]We can access all forecast commits. since spring of 2020 on their GitHub directory. These forecasts are also visually available to the public on their website.

[2]More information on SafeGraph can be found here.

in the moving averages. Our final definition of a rising (falling) regime is a day that has a 14-day average percent change of greater (less) than zero.

We first focus on daily COVID-19 hospitalizations. Figure 1 displays the raw data. The realized data in black is daily and starts on February 19, 2020 and ends on April 1, 2022. The forecasts in red start on August 1, 2020 and end on April 1, 2022. Although we do have daily realized data, the model was not updated daily. Instead, there are 246 forecast dates, denoted as "commit dates" in the plot and shown in blue. These are dates when the entire model with new forecasts is updated, i.e., when each of the red forecasts begin.
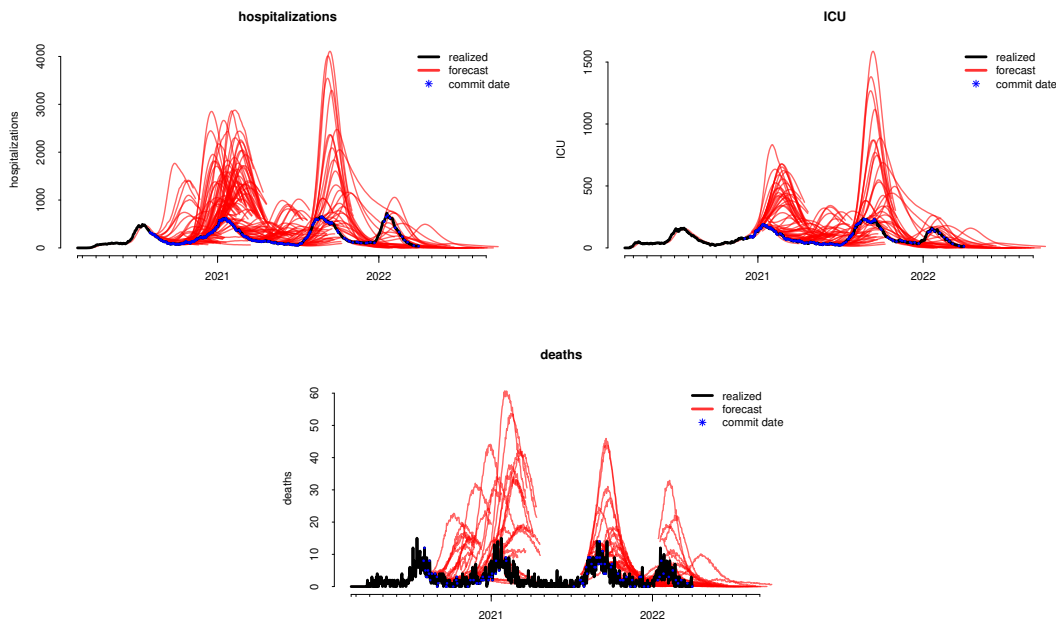

Figure 1. Visualizations of the raw data for hospitalizations, ICU patients, and deaths. The realized data is displayed in black, the forecasts are given in red, and the commit dates (the points at which the forecasts are created) are given in blue along the realized path.

As a first step, we compute the forecasting errors out to 20 days ahead and display them in Figure 2. The median forecasting error is shown in black and the inter quartile range (IQR) of the error distribution is the shaded region. An immediately visible feature is upward bias of the error distribution, even via the median error which is robust to egregiously large forecasts. This implies that the model systematically *over forecasts* hospitalizations.
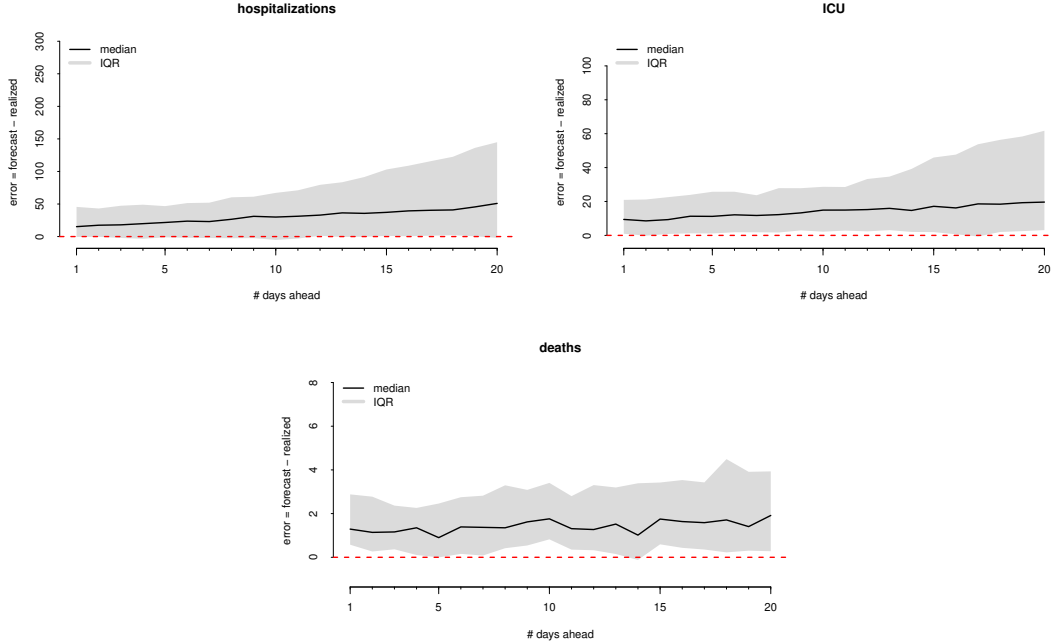
Figure 2. Forecasting errors from 1 to 20 days ahead. The median forecast error for each period ahead is displayed by the black line, and 25th to 75th quantiles of the error distribution are represented by the gray region.

We next consider visualizing the errors conditional on different stages of the pandemic. Our goal is to answer the question: Does the model systematically *over or under forecast* when the virus is spreading more rapidly or is receding? In other words, is the model more likely to make forecasting errors during variant surges or when the viral spread is waning? To answer this question, we define rising hospitalizations as surging and falling hospitalizations as waning. To smooth out noise in the realized data, we compute the percent change in the realized hospitalizations and smooth with a 14-day rolling window. Our final definition of a rising (falling) regime is a day that has a 14-day average percent change of greater (less) than zero.

The regimes as defined above are displayed in Figure 3. The forecast errors conditional on these regimes (analogous to Figure 2) are displayed in Figures 4 (rising) and 5 (falling). When looking at these figures, an interesting feature emerges. The model makes large positive mistakes during surges (and with larger variance, Figure 4) compared to when the viral spread is waning (Figure 5) and hospitalizations are falling.
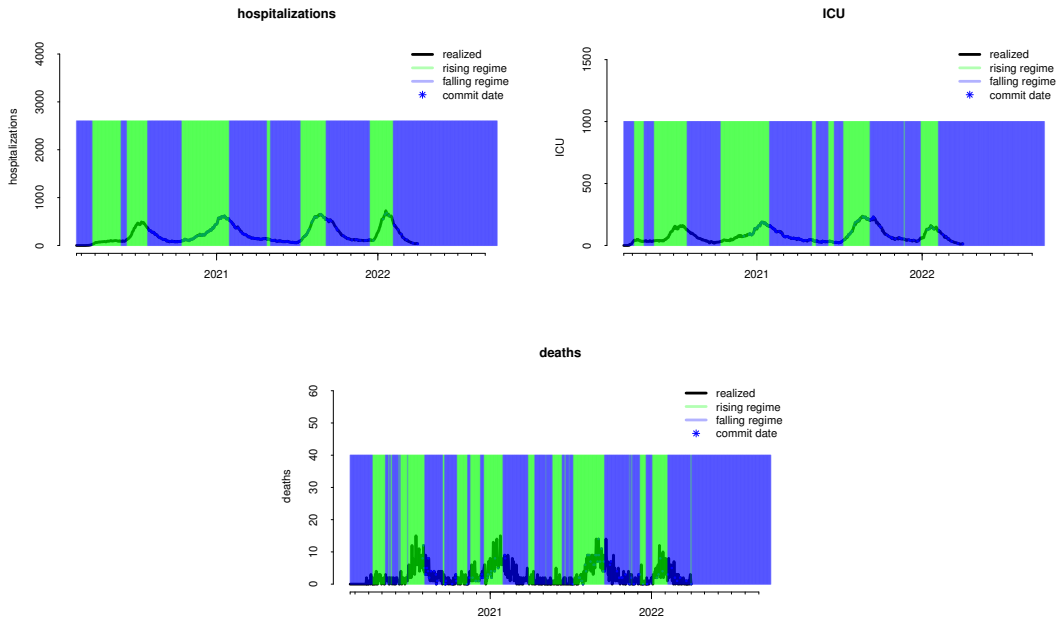
Figure 3. The realized data with "rising" and "falling" regimes overlayed in green and blue, respectively. "Rising" is defined as a day when the 14-day moving average percent change in hospitalizations exceeds zero. "Falling" is defined as a day when the 14-day moving average percent change in hospitalizations is below zero.
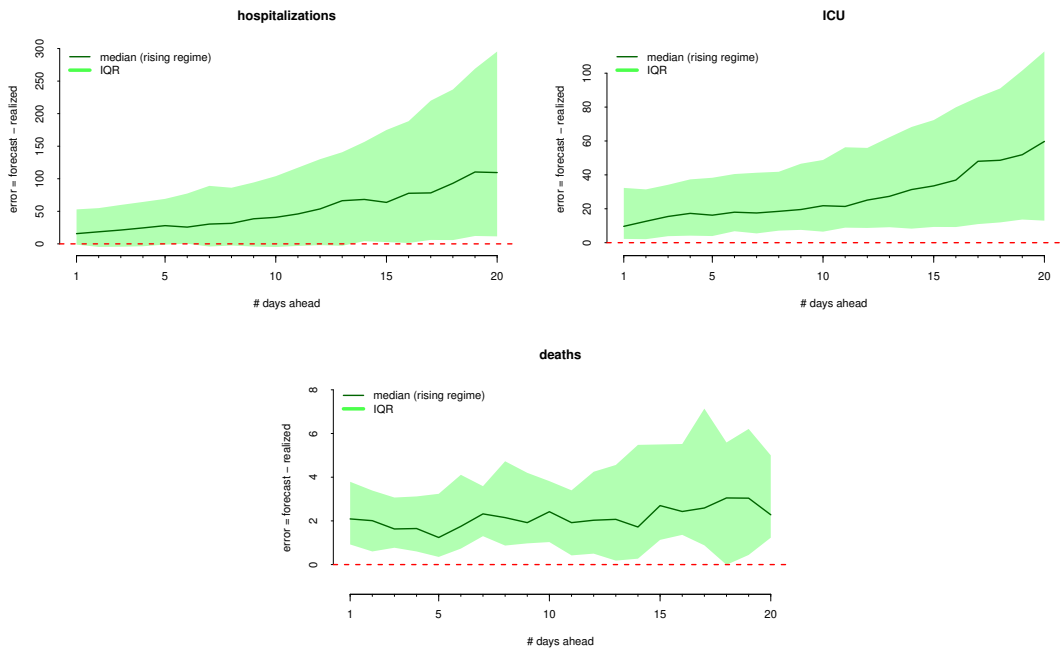

Figure 4. Forecasting errors from 1 to 20 days ahead, conditional upon **forecasting in a rising regime**. The median forecast error for each period ahead is displayed by the bold line, and 25th to 75th quantiles of the error distribution are represented by the shaded region.

Figure 5. Forecasting errors from 1 to 20 days ahead, **forecasting in a falling regime**. The median forecast error for each period ahead is displayed by the bold line, and 25th to 75th quantiles of the error distribution are represented by the shaded region.
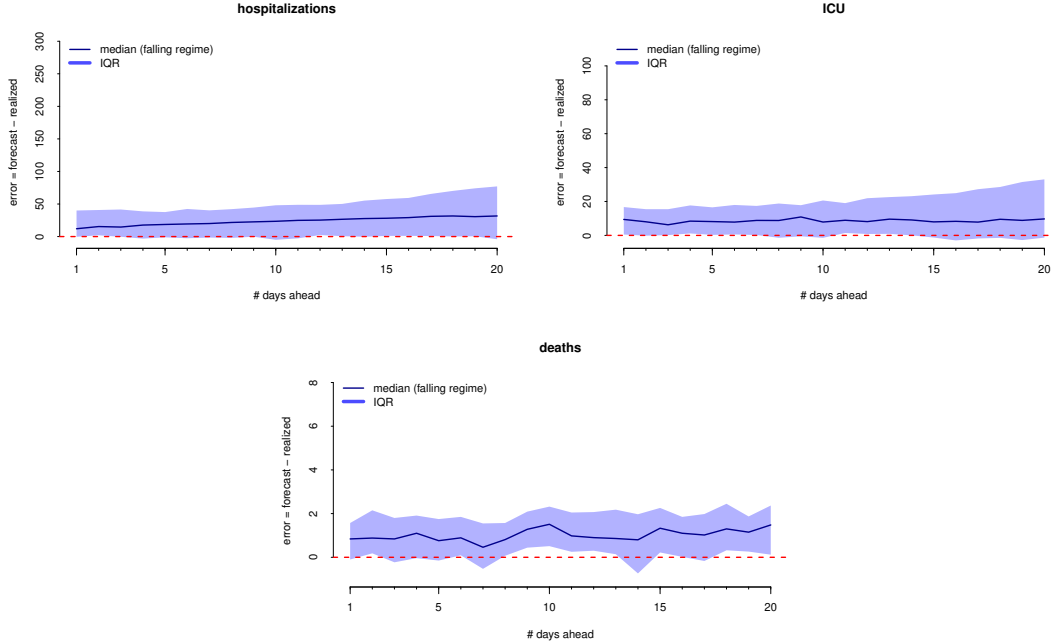
We now investigate the number of ICU patients and its corresponding forecasts. Figure 1 displays the raw data versus the forecasts at each commmit date. Similar to our work on hospitalizations, we calculate the forecasting error up to 20 days ahead as seen in Figure 2. There is a clear upward bias, especially for predictions generated 15 to 20 days ahead. This implies again the model systematically *over forecasts* the number of ICU patients.

We map the errors to the rising and falling regimes defined by the realized ICU admission data. As defined in Section 3.1, rising ICU admissions corresponds to a surge in the viral spread, and falling ICU admissions denote waning viral spread. Figure 3 visualizes these rising and falling regimes given the realized. We see a distinct trend emerge in the rising regime. Errors in the model are larger and more positive when the virus spread is surging compared to when the ICU admissions are falling.

Lastly, we look at the daily COVID-19 deaths. Figure 1 shows the model's forecasts (in red), along with the actual values (in black). This is similar to the plots shown above for hospitalizations and ICU admissions, where the realized values are daily though aperiodic, and the forecasts lines start on their respective commit date.

In Figure 2, we show the forecasting errors of the model computed with forecasts up to 20 days ahead. Similar to the previous error plots, the black line shows the median forecasting error, while the IQR is shaded in gray. Interestingly, deaths exhibit a weakerupward bias than with ICU and hospitalization errors.

Once again, we show the realized data during the "rising" and "falling" regimes in Figure 3. To see if the model has different forecasting errors during a "rising" regime or a "falling" one, we plot the

errors conditional upon the regime when these values were forecast.

Figure 4 shows the model's errors during a "rising" regime, whereas Figure 5 shows the model's errors during a "falling" regime. Unlike with hospitalisation and ICU forecasts, the error is similar and more attentuated in both regimes.

4. Larger issues with the SIR model design

The SIR model fails to properly model certain aspects of disease spread. Melikechi et al. (2022) pointed out that over the last century, many have modified the SIR model to incorporate different compartments for various subpopulations or added new terms that identify unique pathogen transmissions. However, adding too many features may lead to overfitting when making inferences of parameters early on in an epidemic. At the beginning of an epidemic or pandemic, there is a risk of noisy observations. Melikechi et al. (2022) raise the concept of *practical identifiability* which refers to the ability to "discern different parameter values based on noisy observations". Most SIR models employ Monte Carlo simulations, where the model is simulated with pre-determined parameters. Noise is added to the simulated data and then, a fitting procedure is used on the noisy data. With increase in the magnitude of noise, Monte Carlo parameter estimates often display large difference in values, leading to huge uncertainty in the parameters and unreliable inferences.

Lemoine (2021) dissected missteps of a popular SEIR model built by Flaxman et al. (2020), in which they analyzed the effect of non-pharmaceutical interventions on deaths due to the virus. Flaxman et al. (2020) used partial pooling of information between countries, with both individual and shared effects on the time-varying reproduction number. According to them, pooling allowed for more information to be used and helped overcome country-specific idiosyncrasies in the data to enable more-timely estimates. The authors argued that government lockdowns were the interventions that made a bulk of the impact on controlling the viral spread. Lemoine, however, pointed out that the overall effect of government interventions on deaths in Sweden was similar to that in other countries. Why is this important? Sweden was the one country in the analysis that did not have a full lockdown. When he reproduced their analysis, Lemoine found that the country-specific effect for Sweden that the model ignored was almost as large as the effect of a full lockdown, a feature that the authors failed to present in their findings. Based on this, Lemoine argued that such SIR models have been unable to deliver useful inputs for policymakers. But more importantly, epidemiologists have failed to acknowledge this fact, by not ascribing the failure of their models to the right causes.

There are three major critiques for the SIR models: assumption of homogeneous mixing, assumption of closed population, and latency period of infection.

4.1. Homogeneous mixing

SIR models assume random mixing between all individuals in a given population. Based on this assumption, each individual has the same amount of contact as everyone else. Such a model would not be able to account for a higher contact rate at hospitals or lower contact rate for quarantined

individuals (Meyers and Dimitrov, 2010). There is a lack of adequate inclusion of individual behavioral and social influence in SIR models. Infectious disease epidemics have a substantial social aspect and public health implication. Homogeneous mixing assumes an equal probability of transmission between two people regardless of their age or location (Del Valle et al., 2013). This can fail to take into account age-dependent or location-dependent risk. We need to include varying degrees of interventions such as social distancing, stay-at-home, and shelter-in-place orders at different times and across different regions. The assumption of homogeneous mixing of S with I state individuals in the SIR model is therefore invalid during COVID-19.

This flaw of SIR models may explain why the errors tend to be so high with the UT Austin model. While homogeneous mixing can be helpful for projecting the number of cases, it can lead to large errors both in the early stages of the epidemic and in calculating the final epidemic size (Del Valle et al., 2013). Homogeneous mixing often overestimates the epidemic's size, and can lead to more interventions than needed. While the UT model accounts for different age groups and risk factors, it still models disease transmission through an SIR framework where the base assumption remains homogeneous mixing. That is, all individuals within each group would have the same susceptibility to infection, and all individuals within each infection status compartment would have the same infectiousness.

Even with more spatially explicit meta-population models, homogeneous mixing at smaller scales such as within a state, county, or city level is still questionable, as some people can stay home and some are essential. In addition, the regional variability of individual sentiment and behavior, for example, whether to obey or enforce these orders, is essential to determine in order to predict the trajectory of the COVID-19 pandemic, but is generally not included in the SIR models (Chen et al., 2020).

Assumption of homogeneous mixing can lead to overestimating health service needs by not accounting for behavioral changes and government-mandated actions. In Wuhan, strict social distancing was instituted on January 23, 2020, and by the time new infections reached 1 or fewer a day (March 15, 2020), less than 0.5% of the population was infected. At the time, SIR models generally suggested that 25% to 70% of the population to be infected. (IHME COVID-19 Health Service Utilization Forecasting Team and Murray, Christopher JL, 2020).

Since most SIR models consider a single R_0 value, they miss unexpected social behavior changes and are unable to follow the alterations. For instance, social gatherings have a great impact on disease spread. A religious event in Malaysia, held from February 27 to March 3, was supposed to be the source of viral spread in India and Pakistan (Moein et al., 2021; Dimitrov and Meyers, 2010).

Chen et al. (2020) argue that since SIR models are formulated at the population level, we face an important discrepancy between patient-level data and population-level modeling. Exposed (E) and infectious (I) compartments characterize the disease spread at the population level, ignoring individual clinical variations in patients. Due to its broadness, the E compartment assumes everyone exposed is unable to infect others, and the I compartment does not account for varying levels of severity among patients such as asymptomatic, mild, and severe stages. Given these assumptions, SIR models cannot pin down and quantify the impact of super-spreaders who can lead to a disproportionately

large number of new cases. Super-spreading can be due to individual clinical characteristics such as super-shedding of virus or behavioral aspects like super-contacting. Neither are addressed well by SIR models built at a population level. This flaw in the SIR models may also explain some of the inaccuracy in the UT model, as not accounting for patient-level differences may cause them to overlook individual behaviors and *over forecast*.

4.2. Closed population

The focus of SIR models is often placed on the estimation of the basic reproduction number R_0 (Katul et al., 2020). But what should be addressed is the assumption of a closed population in SIR models. Most regions do not follow complete isolation, making them vulnerable to changes in the neighboring communities. SIR models also consider recovered individuals to be immunized. This assumption contrasts with the possibility of the reactivation of the virus or reinfection of previously infected individuals (Moein et al., 2021). Similarly, it does not account for asymptomatic individuals. With the closed population assumption, Ding et al. (2021) argue that standard SIR models miss out on the fact that pre-symptomatic and asymptomatic cases can spread the disease between populations through travel. Researchers have addressed this at multiple levels: within state, country, and even globally.

Studies such as Kucharski et al. (2020) and Wu et al. (2020) estimated cases in Wuhan, China by accounting for people moving in and out of the city. Kucharski et al. (2020) did so by assuming that once exposed, a part of the population would travel internationally. To account for international travellers, they used the number of outbound travellers (assuming 3,300 per day before travel restrictions were imposed on 23 January 2023, and zero after), relative connectivity of different countries, relative probability of reporting a case outside Wuhan and reporting within Wuhan and internationally.

Wu et al. (2020) first inferred the R_0, basic reproduction rate of the virus, of COVID-19 and outbreak size in Wuhan from 1 December 2019 to 25 January 2020, on the basis of confirmed cases exported from Wuhan to cities outside of mainland China, whose symptom onset date had been reported to fall from 25 December 2019, to 19 January 2020. They also forecasted the spread of COVID-19 within and outside Mainland China, accounting for public health interventions and outbound travellers by air, train, and road during the Spring Festival. While they assumed that travel behaviour was not affected by disease and so, international case exportation occurred according to a non-homogeneous process, their work still addresses mobility across cities and countries for modelling COVID-19.

Chinazzi et al. (2020) used a global epidemic and mobility model (GLEAM) to model the international spread of COVID-19, for varying transmissibility and air traffic reductions. The model uses a metapopulation network approach with real-world data, where the world is divided into sub-populations centered around major transportation hubs such as airports. The sub-populations are connected to each other by individuals traveling daily through them. COVID-19 transmission within each sub-population is modelled through a susceptible-latent-infectious-recovered compartmental framework.

Ding et al. (2021) differentiate their study by incorporating granular changes in air traffic and

simulate varying travel restrictions. They focus on data from Canada that showed, a large number of flights going to and fro the country even with travel restrictions and decrease in air traffic. They propose a modified SIR model that accounts for a dynamic flight network, by estimating imported cases using air traffic volume and positive testing rate. Their model has an "open population setting" where people are free to travel in and out of the population.

Not accounting for lack of closed population leads to poor estimation of forecasts. Depending on the assumptions and circumstances of the model, this assumption can lead to both underestimation and overestimation. For instance, Chowell and Nishiura (2014) show that in the case of the Ebola virus, varying R_0 numbers in an SEIR model were due to different assumptions of international or domestic spread of the virus and lack of good quality data.

In our analysis of the UT Austin model, we captured the model's systematic tendency to *over forecast*. UT's initial R_0 values in 2020 were "best guesses" as seen in Wang et al. (2020). They assumed R_0 to be 2.2 but did not provide a clear source for how they came to this number.[3] Tec et al. (2020) estimate R_0 using the basic infectiousness of the disease, number of people that are susceptible to infection, and impact of social distancing, mask wearing and other measures to slow transmission. Potentially, one of reasons why the UT model *over forecasts* is due to their social distancing data. They used mobility trends from SafeGraph data and regressed the transmission rate of the virus on the first two principal components from a principal component analysis (PCA) on eight independent mobility variables, such as home dwell time and visits to universities, bars, grocery stores, museums and parks, medical facilities, schools, and restaurants. If the PCA components did not appropriately capture the variation in the mobility, such as inter-Austin movement, then the transmission rate would lead to a poor estimation of the viral spread.

4.3. Latency period of infection

The SIR model also does not incorporate the latent period between when an individual is exposed to a disease and when that individual becomes infected and contagious. This is because the only categories in the model are susceptible, infectious, and recovered. The S(E)IR model tries to accounts for this parameter by creating an category for people who are exposed but not yet contagious. However, even S(E)IR models are oversimplified, and they will need to model other time-dependent factors, such as the introduction of community mitigation strategies (Tolles and Luong, 2020).

Change in model parameters E, I or R at time t is dependent on a fraction of E and I at time t. As per Liu (2021), this means that after being exposed to the virus on a particular day, an individual may become contagious or recover on that same day. However in reality, an exposed individual will become infectious only after a latent period and recover after an infectious period (Liu, 2021). This timing issue with the compartmental models can cause the forecasts to be inaccurate, especially for further days out.

Even SEIR models, which have the exposed category to model this latency period, may not be doing so correctly for this reason. The latency period has to be set to the correct amount of time, or

---

[3]The authors linked a source, but it is not open access.

the model may *over/under forecast*. A study using COVID-19 data in Tennessee found that "optimal" latency period is 2.40 days, which is close to the mean latency period of 2.52 days which was estimated from the data of seven countries (Patterson and Wang, 2022). If UT Austin's SEIR model did not correctly model this latency period, this could explain some of the model's error.

5. Conclusion

Given the wide use of compartmental models to describe the transmission dynamics of COVID-19 and other diseases, we must carefully consider their limitations when using them to inform public health interventions. In particular, homogeneity assumptions underlying these models do not accurately reflect heterogeneity of the population, and estimates of key parameters such as R_0 are often noisy and unreliable. Additionally, these models do not account for the impact of non-pharmaceutical interventions on disease transmission or capture the complex interactions between the virus, people, and the environment.

Historically, compartmental models were used as descriptive tools instead of real-time models for prediction and decision-making. This change in "use-case" partially explains their significant failure. Chen et al. (2020) point out that in the initial papers describing the SIR approach, the model was applied after the epidemic had ended. But SIR models have little room for new evidence without modifying the model structure and estimation of R_0. Additionally, it was difficult to model the initial part of COVID-19 spread using SIR models, compared to other models, since we had limited or no information on aspects such as asymptomatic transmission, super-spreaders and unreported cases. Moreover, constructing a long-term forecasting model is a major challenge because of the lack of data. The SIR model has a tendency to underestimate peak infection rates and substantially overestimate the persistence of the epidemic after the peak has passed (Pereira et al., 2022).

Intuitively, prediction accuracy can be increased by synthesizing the forecasts of many unique models. This represents one important area of future research to improve the usefulness of compartmental models. Cramer et al. (2022) find that an ensemble model provides more accurate short-term forecasts of hospital and ICU admissions compared to the individual models alone. The ensemble model also has a lower prediction error and better calibration than the individual models, suggesting it could be more effective in real-time decision-making for healthcare systems.

In this paper, we investigated predictive failures of the SIR framework and its derivative models. While our empirical analysis utilized forecasts from the model constructed at UT Austin, our conclusions apply to all compartmental models generated from sets of differential equations. The epidemiological community should scrutinize this framework and promote new and innovative techniques that go beyond differential equations and adopt solutions from economics, statistics, machine learning, and beyond. A pandemic forecasting model can always be marketed as unique, but if it relies on differential equations and compartmental modeling, this structural bias will outweigh most novel modifications. We hope this manuscript contributes to a rebirth of research in epidemiological modeling so the best techniques can be elevated and utilized in future pandemics.

References

Chen, S., P. Robinson, D. Janies, and M. Dulin (2020). Four challenges associated with current mathematical modeling paradigm of infectious diseases and call for a shift. In *Open forum infectious diseases*, Volume 7, pp. ofaa333. Oxford University Press US.

Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science 368*(6489), 395–400.

Chowell, G. and H. Nishiura (2014). Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine 12*, 1–17.

Cramer, E. Y., E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. Castro Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang, et al. (2022). Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the united states. *Proceedings of the National Academy of Sciences 119*(15), e2113561119.

Del Valle, S. Y., J. M. Hyman, and N. Chitnis (2013). Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. *Mathematical biosciences and engineering: MBE 10*, 1475.

Dimitrov, N. B. and L. A. Meyers (2010). Mathematical approaches to infectious disease prediction and control. In *Risk and optimization in an uncertain world*, pp. 1–25. INFORMS.

Ding, X., S. Huang, A. Leung, and R. Rabbany (2021). Incorporating dynamic flight network in seir to model mobility between populations. *Applied Network Science 6*(1), 1–24.

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, et al. (2020). Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature 584*(7820), 257–261.

Friedman, J., P. Liu, C. E. Troeger, A. Carter, R. C. Reiner Jr, R. M. Barber, J. Collins, S. S. Lim, D. M. Pigott, T. Vos, et al. (2021). Predictive performance of international covid-19 mortality forecasting models. *Nature communications 12*(1), 2609.

IHME (2020a). Covid-19: What's new for may 29, 2020: Main updates on ihme covid-19 predictions since may 26, 2020.

IHME (2020b). Covid-19: What's new for may 4, 2020: Main updates on ihme covid-19 predictions since april 29, 2020.

IHME COVID-19 Health Service Utilization Forecasting Team and Murray, Christopher JL (2020). Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator-days and deaths by us state in the next 4 months. *MedRxiv*, 2020–03.

Katul, G. G., A. Mrad, S. Bonetti, G. Manoli, and A. J. Parolari (2020). Global convergence of covid-19 basic reproduction number and estimation from early-time sir dynamics. *PLoS One 15*(9), e0239800.

Kucharski, A. J., T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, et al. (2020). Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases 20*(5), 553–558.

Lemoine, P. (2021). The case against lockdowns. *Center for the Study of Partisanship and Ideology*.

Liu, L., S. S. Shwiff, S. A. Shwiff, and M. Miller (2021). Impact of covid-19 on the us and texas economy: A general equilibrium approach.

Liu, X. (2021). Analytical solution of a new seir model based on latent period-infectious period chronological order. *medRxiv*.

Melikechi, O., A. L. Young, T. Tang, T. Bowman, D. Dunson, and J. Johndrow (2022). Limits of epidemic prediction using sir models. *Journal of Mathematical Biology 85*(4), 36.

Meyers, L. A. and N. Dimitrov (2010). Mathematical approaches to infectious disease prediction and control. *INFORMS, Tutorials in Operations Research*.

Moehlis, J. (2002). An sis model-apc/eeb/mol 514 tutorial 4: Seasonal epidemic models.

Moein, S., N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, and Y. Gheisari (2021). Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan. *Scientific reports 11*(1), 4725.

Patterson, B. and J. Wang (2022). How does the latency period impact the modeling of covid-19 transmission dynamics? *Mathematics in Applied Sciences and Engineering 3*(1), 60–85.

Pereira, I. G., A. G. d. S. Junior, D. P. Aragão, E. V. de Oliveira, A. A. Bezerra, F. d. A. Pereira, J. G. F. Costa, J. S. Cuno, D. H. dos Santos, J. M. Guerin, et al. (2022). Epidemiology forecasting of covid-19 using ai—a survey. *Computational Intelligence for COVID-19 and Future Pandemics: Emerging Applications and Strategies*, 89–120.

Rahimi, I., F. Chen, and A. H. Gandomi (2021). A review on covid-19 forecasting models. *Neural Computing and Applications*, 1–11.

Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character 92*(638), 204–230.

Spencer Woody, M. T., M. Dahan, K. Gaither, M. Lachmann, S. J. Fox, L. A. Meyers, and J. Scott (2020). Projections for first-wave covid-19 deaths across the us using social-distancing measures derived from mobile phones.

Tec, M., M. Lachmann, S. J. Fox, R. Pasco, S. Woody, J. Starling, M. Dahan, K. Gaither, J. Scott, and L. A. Meyers (2020). Austin covid-19 transmission estimates and healthcare projections.

Texas Education Agency (2021). Impacts of covid-19 and accountability updates for 2022 and beyond. *2021 Texas Assessment Conference*.

Tolles, J. and T. Luong (2020, 06). Modeling Epidemics With Compartmental Models. *JAMA 323*(24), 2515–2516.

Wang, X., R. Pasco, K. Pierce, Z. Du, S. Fox, and L. A. Meyers (2020). Covid-19 healthcare demand projections: Austin, texas. *UT COVID-19 Model. Consortium*.

Wu, J. T., K. Leung, and G. M. Leung (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet 395*(10225), 689–697.

Yang, Z., Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, et al. (2020). Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of thoracic disease 12*(3), 165.