Asia-Pacific Journal of Operational Research
Vol. 37, No. 3 (2020) 2050028 (21 pages)

DOI: 10.1142/S0217595920500281

Predictive Modeling for Epidemic Outbreaks: A New Approach and COVID-19 Case Study

Jian Chen*

CreditWise Technology Co., Ltd.
Lujiazui Road No. 166, Mirae Asset Tower Building
16 Block D, Shanghai 200120, China
jian.chen@jhu.edu

Michael C. Fu

Smith Chair of Management Science Department of Decision
Operations and Information Technologies
Robert H. Smith School of Business & Institute for Systems
Research Van Munching Hall University of Maryland
College Park, Maryland 20742, USA
mfu@umd.edu

Wenhong Zhang[a]

Department of Infectious Diseases
Huashan Hospital, Fudan University
12 Wulumuqi Zhong Road, Shanghai 200040, China
zhangwenhong@fudan.edu.cn

Junhua Zheng[b]

Shanghai General Hospital
School of Medicine in Shanghai, Jiao Tong University
86 Wujing Road, Shanghai 200080, China
zhengjh0471@sina.com

Received 7 May 2020
Revised 12 May 2020
Accepted 12 May 2020
Published 22 May 2020

*Corresponding author.

[a]Dr. Zhang is the Head of the Center for Infectious Disease at Huashan Hospital of Fudan University and serves as the leader of Shanghai's Anti-COVID-19 clinical expert team.

[b]Dr. Zheng is the Vice President of First People's Hospital, which is affiliated with Shanghai Jiaotong University. As the leader of the Shanghai medical assistance team to Hubei province after the coronavirus outbreak in Wuhan beginning in January, he and his team worked on the frontlines battling the pandemic for 67 days and nights.

Since the onset of the COVID-19 outbreak in Wuhan, China, numerous forecasting models have been proposed to project the trajectory of coronavirus infection cases. Most of these forecasts are based on epidemiology models that utilize deterministic differential equations and have resulted in widely varying predictions. We propose a new discrete-time Markov chain model that directly incorporates stochastic behavior and for which parameter estimation is straightforward from available data. Using such data from China's Hubei province (for which Wuhan is the provincial capital city and which accounted for approximately 82% of the total reported COVID-19 cases in the entire country), the model is shown to be flexible, robust, and accurate. As a result, it has been adopted by the first Shanghai assistance medical team in Wuhan's Jinyintan Hospital, which was the first designated hospital to take COVID-19 patients in the world. The forecast has been used for preparing medical staff, intensive care unit (ICU) beds, ventilators, and other critical care medical resources and for supporting real-time medical management decisions.

Keywords: COVID-19; coronavirus; epidemics; pandemic; Markov chain; stochastic models; simulation; predictive modeling; analytics; predictive analytics; transition matrix model.

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2), was first identified in China in December 2019 and quickly became a worldwide pandemic, recording close to 5 million confirmed cases and over 300,000 deaths as of May 15th, 2020. Among the tools used in the fight to contain such pandemics, forecast models are critical in helping support not only medical resource management decisions, but also in informing government policies, such as when and where "lock-down" and "stay-at-home" directives should be enacted and lifted.

Traditional epidemiology models use deterministic differential equations to forecast the population dynamics among various states, e.g., susceptible, exposed, infectious, or recovered in the well-known susceptible-exposed-infectious-recovered (SEIR) model recently used to model the COVID-19 outbreak in Wuhan (Wu *et al.*, 2020a). Such models have also been expanded to include additional states for COVID-19, e.g., adding critical states). These models are by design aggregate models that track only the mean populations and do not incorporate stochastic effects directly and are highly sensitive to estimated parameters. For example, a key parameter is the basic reproductive number R_0, or the rate (average number) at which one currently infected person infects new persons; thus, an $R_0 < 1$ indicates that the epidemic is dying out. An inaccurate estimate of R_0 is magnified in poor forecasts, leading to orders of magnitude differences in output.

More recently, epidemiologists have been looking at agent-based models, very familiar to the operations research (OR) community, where individuals can be modeled in detail, e.g., age, gender, health condition, and stochastic characteristics are directly incorporated. Such models can be very useful for studying smaller communities, but since they typically require stochastic (Monte Carlo) simulation (Xu *et al.*, 2015), may face computational challenges in scaling up to large cities or countries when the number of agents becomes large. Also, if the amount of detailed individual data available is limited, then it would be challenging to estimate the model with any degree of confidence.

We propose a discrete-time Markov chain (DTMC) model, where the states of the chain are similar to the states in the compartmental models. Since this DTMC modeling approach is analogous to a financial forecasting model called the transition matrix model (TMM), widely used in credit analysis (Malik and Thomas, 2012; Chen *et al.*, 2018), be it corporate rating migration, or individual consumer behavior, we will also refer to it as the TMM approach. Our proposed approach has the following advantages: it incorporates stochastic features directly while retaining essentially the same states (compartments) as in the compartmental models; its discrete-time nature and degree of modeling detail make it straightforward to estimate model parameters from available data; it is computationally tractable both in terms of parameter estimation and in terms of model output analysis.

From the perspective of operations research (OR) models, an analogy of our proposed approach can be found in queueing systems, where there are three main modeling paradigms, each of which complements the other, depending on what the model is being used for: fluid/diffusion models, Markov chain models and stochastic simulation models (discrete-event or agent-based). Like compartmental models, fluid/diffusion models for queueing systems are deterministic ordinary differential equation models where the state of the system is represented by the population, so it is an aggregate model. At the other extreme are detailed simulation models, but in OR the most commonly taught model at least in coursework is probably the Markov chain model. In queueing systems, the continuous-time model also leads to deterministic ordinary differential equation models, but for the *probability* of taking on each value, i.e., they model the entire probability distribution and not just the aggregated mean population. For the epidemic setting, we believe that the discrete-time Markov chain model lends itself most naturally, for several reasons: the available data makes it easy to estimate model parameters on a daily basis, and the day becomes the obvious discrete-time interval; the existence of absorbing states, e.g., death and cured, the latter state presumably immune for at least the planning horizon of interest; and the computational ease of implementing matrix multiplication for propagating probability state vectors. These factors supported the underlying motivation of finding an approach that fit the anticipated use of the model and which could be easily estimated from the available data, as well as being both flexible and relatively robust to parameter estimation uncertainty.

The remainder of this paper, which builds on Chen *et al.* (2020), is organized as follows. We briefly review compartmental models in Sec. 2, since they are the dominant epidemiological models being used for forecasting, both in the academic (medical) literature and in practice and popular press. In Sec. 3, we present the proposed (modified) discrete-time Markov chain (DTMC) transition matrix model (TMM) used for modeling epidemic outbreaks such as COVID-19. In Sec. 4, we apply the model to the COVID-19 setting using actual data from Hubei province (for which Wuhan is the provincial capital city and which accounted for approximately 82% of the total reported COVID-19 cases in all of China) and discuss model forecasting results, as well as assess model robustness and flexibility. In Sec. 5, we discuss usage of the model and the effect of medical assistance teams dispatched from all over China to Hubei province.

2. Brief Review of Epidemiology Compartmental Models

Prevailing epidemiological forecast models relevant to COVID-19 are based on extensions of the susceptible-exposed-infectious-recovered (SEIR) model, which are deterministic continuous-time dynamic models that model the evolution of the aggregate population under consideration, where the population is separated into a fixed number of mutually exclusive "compartments". For example, in the original susceptible-infectious-recovered (SIR) model of Kermack and McKendrick (1927), the compartments are defined as follows:

- Susceptible (S) – not infected yet;
- Infected (I) – assumed infectious with symptoms;
- Removed (R) – recovered (sometimes this is the definition) or deceased.

The three compartments are represented by time-varying functions $S(t)$, $I(t)$, $R(t)$, representing the (average) number in each compartment (state). The following is the simplest set of ordinary differential equations (ODEs) modeling the dynamics:

$$\frac{dS(t)}{dt} = -aR_0\frac{I(t)S(t)}{N},$$

$$\frac{dI(t)}{dt} = a\left(R_0\frac{I(t)S(t)}{N} - I(t)\right),$$

$$\frac{dR(t)}{dt} = aI(t),$$

which has just three parameters ($a > 0$, $R_0 > 0$, and the population N). This system can be solved analytically, but once any realistic features are incorporated into the model, which is the case for real-world applications, numerical simulation is required, e.g., the SEIR model of Wu *et al.* (2020a) used for COVID-19.

A basic model driver in compartmental models (SIR, SEIR, and their extensions) is the parameter R_0, which indicates how contagious an infectious disease is, and is also referred to as the reproduction number, because it represents the average number of people who will contract the disease from one person with that disease. Compartmental models are very sensitive to their model parameters, as well as to initial conditions, so accurate estimates are crucial if these models are to be used for forecasting, which is not necessarily the primary usage for these types of models, especially in the early stages of a new outbreak.

Generally, compartmental models can be very effective epidemiology tools once a disease is well understood and in a mature phase. They are also good theoretical models for reference purposes, e.g., for comparing the infection rate of COVID-19 against other respiratory infectious diseases such as SARS and MERS, by comparing the different values of R_0, once accurate estimates of model parameters can be obtained. However, for forecasting purposes based on relatively sparse data, especially with regards to patient-level outcomes (as opposed to aggregate cases), they may exhibit high sensitivity to the estimated model parameters such as R_0, limiting their robustness in forecasting in the early and middle stages of epidemics. In the SEIR model of Wu *et al.* (2020a), R_0 is estimated "using Markov Chain Monte Carlo methods with Gibbs sampling and non-informative flat prior," so as we will shortly see, it is a more involved process than the parameter estimation process for our proposed approach, which relies directly on available empirical data.

In summary, from a practical perspective, if the goal of the forecasting model is to provide support for decisions such as medical resource planning, including staff (doctors, nurses, cleaning staff, etc.), supplies (ICU beds, ventilators, N95 masks, etc.), and biohazardous waste disposal capacity, alternative modeling approaches may be more aligned with the available data for estimating model parameters. Furthermore, adding new compartments in SIR/SEIR models to allow differentiation between severe cases (requiring hospital beds) and critical cases (requiring ICU beds and ventilators) results in additional model complexity and the introduction of even more parameters that are difficult to estimate from the available data.

3. Discrete-Time Markov Chain Model

We model the patient treatment process as an absorbing Markov chain with the following discrete states (analogous to compartments in the SIR/SEIR traditional differential equation-based models): (under) medical observation, discharged, infected non-severe, infected severe, critical, death, and cured. The potential transitions between the states are shown in Fig. 1, where self-loops are understood but omitted in the diagram for clarity, and the three states outside the treatment boxed labeled discharged, cured and dead are absorbing states. Note that "Infected" (represented as a decision diamond) is not a separate state by itself in the DTMC model, as once a close contact is a confirmed case of infection, it is immediately classified as severe or non-severe. Another state called infected asymptomatic could also be easily added to the model, but since there is sparse data to estimate this state,[1] we have not included it, and those patients would not have entered the medical observation state in the current version of the model.

One of the main differences between the compartment models and the proposed modified DTMC model is that the former relies on parameters that have interpretative meanings, so specific data are required to estimate them, whereas the parameters of the proposed model depend only on transition probabilities between states defined based on medical classifications where available data can be used to

[1]Back in February, asymptomatic cases were not reported in China.

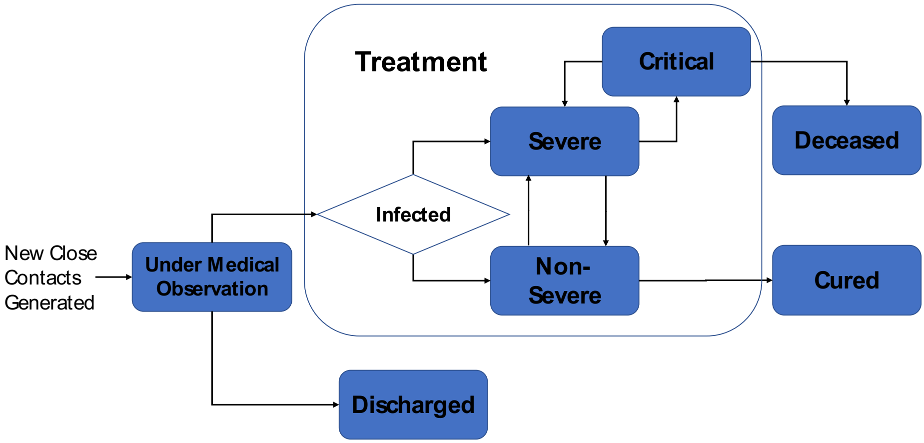

Fig. 1. Patient state transition diagram (self-loop transitions understood but omitted).

estimate them directly. As an example, in the SEIR model of Wu *et al.* (2020a), two parameters are corresponding to the mean latent and infectious periods, which are challenging to estimate accurately from early data.

Note that the states "Cured" and "Discharged" could also have been put together into a single state (e.g., called "Healthy"), but for purposes of parameter estimation and tracking statistics, the cure rate is of separate interest. The three states, Discharged, Deceased and Cured are absorbing states, whereas the others are all transient states, assuming non-zero probabilities for existing transitions in the diagram.

The transition probabilities of the DTMC model can be either determined by regression or simply derived from empirical probabilities. We adopt the second approach when estimating the model in our application for predicting the COVID-19 spread in Hubei province, China. Specifically, we define the states of the model as follows:

- Under Medical Observation (UMO): State of a close contact of a potential infection case, who is traced, identified and put into medical observation, generally in a quarantine facility. From this state, the next day a patient may be confirmed with infection, discharged without infection, or remain UMO (e.g., if the test results have not come back yet).
- Discharged (Dis): Terminal state for a close contact after undergoing medical observation, although one could possibly reenter as a close contact in the generation phase.
- Infected but Non-Severe (INS): State for a patient with mild symptoms (generally treated in makeshift shelter hospitals for Hubei COVID-19 patients). There are

three possible states for the next day: cured, worsened condition to severe, or remaining non-severe.

- Infected and Severe (IAS): State for a patient who develops severe symptoms that require hospitalization and oxygen support (World Health Organization, 2020). There are again three possible states for the next day: worsening to critical, improvement to non-severe, or staying the same.
- Infected and Critical (IAC): State for a patient showing critical symptoms and requiring admission to an ICU (World Health Organization, 2020). There are also three possible states for the next day: death, improvement to severe, staying the same.
- Cured (Cu) and Deceased (De) are two terminal states.

Thus, each day, a patient's state on day n is characterized by state vector, defined as follows:

$$V(n) = [\text{UMO Dis INS IAS IAC Cu De}]',$$

where the value of each element can be viewed as the probability of being in that state, so for each new close contact generated, the elements would sum to 1. For a given individual in a known state, the corresponding element would be equal to 1, and the other elements would be zero. For example, a patient currently in state "infected and non-severe" (INS) would have state vector [0 0 1 0 0 0 0]', and the next day, the state vector could remain the same or transition to either [0 0 0 1 0 0 0]' (severe) or [0 0 0 0 0 1 0]' (cured). Thus, for the entire potentially exposed population, the state vector is defined as the count of people in each state. For example, at an early stage with say 100 patients being tested or treated, out of which 30 are awaiting testing results, 10 are critical, 10 are severe, and 50 are non-severe, the state vector would be [30 0 50 10 10 0 0]'.

Defining the one-step (i.e., daily) transition matrix $P = [p_{i,j}]$, where

$$p_{i,j} = \text{daily transition probability from state } i \text{ to state } j,$$

we have the usual DTMC vector-matrix (one-step and multi-step) dynamic equations:

$$V(n+1) = PV(n),$$

$$V(n+m) = P^m V(n).$$

If the population is limited and the transition matrix is stationary, the above formula will be sufficient in predicting all future outcomes, and of course, since it is an absorbing chain, all the transient states would eventually go to zero, and individuals already in the pool will eventually end up discharged, cured, or deceased (the absorbing states). In epidemic settings such as the COVID-19 situation in early 2020, the population is not fixed, and additional individuals enter into the population via the generation of new close contacts as shown in Fig. 1. Every day, new close contacts are added to the medical observation pool. In our modified DTMC model, these new individuals are generated as a proportion of the current new close contacts, i.e., via

$$\mathrm{NCC}(n+1) = \mathrm{NCC}(n)\mathrm{e}^{NCC\_Change\_Rate},$$

where the NCC change rate parameter is analogous to the basic reproductive number R_0 parameter in SIR-type models, in that when it is positive, the number of infected individuals in the population is increasing, corresponding to an R_0 value greater than 1. Just as when intervention measures cause R_0 to eventually decrease below 1, correspondingly the NCC change rate will become negative. The NCC change rate is a critical parameter in our forecast, but it is readily estimated from available data, because it appears in a single equation, whereas R_0 appears simultaneously in two (or more) equations in SIR and SEIR models (e.g., the very simplest SIR model in Sec. 2 and the SEIR model of Wu *et al.* (2020a)).

This DTMC state transition matrix model can be used for forecasts such as when the infection peak time (maximum number of active infection cases) occurs, as well as patient distributions (critical, severe, non-severe), which can be used for supporting medical resource allocation planning. In the next section, we describe the application of this model to data from the Hubei province (which contains Wuhan) in January and February of 2020.

4. Application of the Model to COVID-19 in Hubei Province, China

Although there are three hospitalization states — non-severe cases (INS), severe cases (IAS), and critical cases (IAC), for the COVID-19 Hubei province data, we had only the patient count in each state but not the actual pairwise transitions among these three states, so we combined the entire hospitalization period into a therapeutic state "infected and being treated" (IAT) — also known as the daily active cases, which minimizes the need for estimation for those unobserved transitions. Instead, we use the proportion of patients of each state (INS, IAS, IAC) within IAT, to forecast the number of non-severe, severe, critical patients. Also, the model will track daily new confirmed cases (DNCC) and cumulative confirmed cases (CCC), since these metrics are tracked in actual data. From all these various data, the parameters of the model can be estimated directly as follows:

- New Close Contacts (NCC) Change Rate : $\alpha_{\mathrm{NCC}} = \ln(\mathrm{NCC}(t)/\mathrm{NCC}(t-1))$;
- UMO Daily Discharge Rate: $p_{\mathrm{UMO,Dis}} = \mathrm{Dis}(t)/\mathrm{UMO}(t-1)$;
- Pr $\{\mathrm{UMO} \to \mathrm{IAT}\}$: $p_{\mathrm{UMO,IAT}} = \mathrm{IAT}(t)/\mathrm{UMO}(t-1)$;
- Pr $\{\mathrm{IAT} \to \mathrm{Deceased}\}$: $p_{\mathrm{IAT,De}} = \mathrm{De}(t)/\mathrm{IAT}(t-1)$;
- Pr $\{\mathrm{IAT} \to \mathrm{Cured}\}$: $p_{\mathrm{IAT,Cu}} = \mathrm{Cu}(t)/\mathrm{IAT}(t-1)$.

Table 1. Empirical data on 2020/2/8 for Hubei Province.

	2020/2/8	10-Day Moving Average	10-Day Max	10-Day Min
NCC Change Rate	−21.4%	3.8%	71.6%	−21.4%
UMO Discharge Rate	16.5%	9.0%	16.5%	4.1%
UMO → Infected/Treated	3.04%	4.54%	5.39%	3.04%
Infected/Treated → Deceased	0.35%	0.63%	0.97%	0.35%
Infected/Treated → Cured	1.40%	0.91%	1.44%	0.60%
Severe Case Ratio	16.5%	14.2%	18.1%	11.5%
Critical Case Ratio	4.65%	4.13%	5.3%	4.1%

And as alluded to above, estimating the count of active severe cases and critical cases requires two additional parameters:

- Proportion of severe patients: $\rho_s = \mathrm{IAS}(t)/\mathrm{IAT}(t)$;
- Proportion of critical patients: $\rho_c = \mathrm{IAC}(t)/\mathrm{IAT}(t)$.

We collected data from Caixin Data (a subsidiary of Caixin Group), which collects the original data from the China National Health Commission. We also collected supplementary data from the Hubei Health Commission, mainly for the severe and critical counts. The data period starts from 2019/12/31 and ends on 2020/2/8, updated daily. Next, we describe how to estimate each of these parameters based on the empirical data that are summarized in Table 1, which provides statistics based on 10-day windows.

As mentioned in the previous section when describing the DTMC model, the NCC change rate (α_{NCC}) is closely linked to R_0 but can be directly estimated from empirical data. This parameter has very strong policy implications, as it measures the effectiveness and efficiency of non-pharmaceutical intervention (NPI) policies and actions. A positive value indicates that the NPI measures are failing, as new close contacts are increasing, whereas a negative value indicates that the NPI measures are effective, as fewer people are contracting the coronavirus daily. The higher the absolute value of the parameter, the more effective the NPI measures are. Since the 8-day empirical average of −6% is volatile, we considered three values for the model: −1%, −5%, and −10%. We could not obtain the one-day probability of the UMO discharge rate in Hubei Province, so we used the national rates. Since the one-day probability was 17% and the 10-day moving weighted average was 13%, we tested the model with both 17% and 13%. In Hubei Province on 2/8, the one-day probability of transition from medical observation to confirmed infection was 2.15%, with a 10-day weighted average of 3.94%, so we used 4% as our model parameter. The latest single-day fatality rate was 0.35% and the 10-day moving weighted average was 0.63%. As this probability was declining, we used the latest value (0.35%) as our model parameter to be cautiously optimistic. The latest one-day cure rate was 1.40%, with a 10-day moving weighted average of 0.91%. As this probability was increasing, we used the latest value (1.40%) as our model parameter. From the

Table 2. Parameter values (in %) used in the model forecast for different scenarios.

	S1: Optimistic		S2: Cautiously Optimistic		S3: Relatively Pessimistic	
α_{NCC}	−10.0%	−10.0%	−5.0%	−5.0%	−1.0%	−1.0%
$p_{\text{UMO,Dis}}$	17%	10.5%	17%	10.5%	17%	10.5%
$p_{\text{UMO,IAT}}$	4.00%	4.00%	4.00%	4.00%	4.00%	4.00%
$p_{\text{IAT,De}}$	0.35%	0.35%	0.35%	0.35%	0.35%	0.35%
$p_{\text{IAT,Cu}}$	1.40%	1.40%	1.40%	1.40%	1.40%	1.40%
ρ_s	14.50%	14.50%	14.50%	14.50%	14.50%	14.50%
ρ_c	4.50%	4.50%	4.50%	4.50%	4.50%	4.50%

historical data, the proportion of critical cases is relatively stable, whereas the proportion of severe cases fluctuated more. Neither of these showed monotone behavior, so we used the averages (14.50% and 4.50%) of the latest value and 10-day moving weighted average as our model parameters to put more weight on recent observations.

In addition to the parameter values just described, we considered six different scenarios to control for forecast uncertainty, based on three different values of the NCC Change Rate α_{NCC} (−10% optimistic, −5% cautiously optimistic, and −1% relatively pessimistic) and two different values of the UMO discharge rate $p_{\text{UMO, Dis}}$ (17% and 10.5%). In Wu *et al.* (2020b), NCC Change Rate was assumed constant, whereas in follow-up work (Zheng *et al.*, 2020), it was allowed to be time-varying to capture changes due to NPI implementations. Table 2 summarizes the scenarios and parameter values used in our model.

A summary of the daily update equations used in the model are as follows:

$$\text{NCC}(t) = \text{NCC}(t-1)\exp(\alpha_{\text{NCC}});$$

$$\text{DNCC}(t) = \text{UMO}(t-1)p_{\text{UMO,IAT}};$$

$$\text{CCC}(t) = \text{CCC}(t-1) + D\text{NCC}(t);$$

$$\text{UMO}(t) = \text{UMO}(t-1) + \text{NCC}(t) - D\text{NCC}(t) - -\text{Dis}(t);$$

$$\text{Dis}(t) = \text{UMO}(t-1)p_{\text{UMO,Dis}};$$

$$\text{IAT}(t) = \text{IAT}(t-1) + D\text{NCC}(t) - \text{IAT}(t-1)p_{\text{IAT,De}} - \text{IAT}(t-1)p_{\text{IAT,Cu}};$$

$$\text{IAS}(t) = \text{IAT}(t)\rho_s;$$

$$\text{IAC}(t) = \text{IAT}(t)\rho_c;$$

$$\text{INS}(t) = \text{IAT}(t) - \text{IAS}(t) - \text{IAC}(t);$$

$$\text{De}(t) = \text{De}(t-1) + \text{IAT}(t-1)p_{\text{IAT,De}};$$

$$\text{Cu}(t) = \text{Cu}(t-1) + \text{IAT}(t-1)p_{\text{IAT,Cu}}.$$

Table 3. Backtesting of the Hubei province forecast.

Key Performance Metrics	Actual	Forecast					
		S1: Optimistic		S2: Cautiously Optimistic		S3: Relatively Pessimistic	
NCC Change Rate	−9.0%	−10%	−10%	−5%	−5%	−1%	−1%
UMO Discharge Rate	16.0%	17.0%	10.50%	17.0%	10.50%	17.0%	10.50%
Peak NAT	50,633	39,612	47,148	44,082	55,150	62,041	85,502
Peak Date	2020/2/16	2020/2/23	2020/2/28	2020/3/1	2020/3/7	2020/4/6	2020/4/14
Peak NAS	9,289	5,753	6,845	6,400	8,004	9,000	12,402
Peak Date	2020/2/16	2020/2/23	2020/2/28	2020/3/1	2020/3/7	2020/4/6	2020/4/14
Peak NAC	2,492	1,786	2,124	1,986	2,484	2,793	3,849
Peak Case	2020/2/21	2020/2/23	2020/2/28	2020/3/1	2020/3/7	2020/4/6	2020/4/14
CCC on 2/29	66,907	54,189	64,064	60,192	71,596	68,045	81,284

Model validation was carried out using backtesting — analogous to standard procedures used for models used in the financial industry — on the forecast versus the actual numbers for February end. The results for all six scenarios are summarized in Table 3, where actual values and model forecasts are provided for key performance metrics, such as peak value of critical cases, active cases, and month-end total cases. The results indicate that the model gives reasonable forecasts in all six scenarios, an indication of robustness, and does especially well under the "cautiously optimistic" scenarios, which was chosen as the most likely setting, leading to relative errors for peak active cases, peak severe cases, peak critical cases, and February-end total cases of 1.0%, 20.1%, 7.5%, and 1.3%, respectively, relative to the median of the two cautiously optimistic scenarios. Also, a quick sensitivity analysis using the six scenarios quantitatively supports the robustness of the model, as the range of February-end cumulative confirmed cases (last line in Table 3, CCC entry) is from 54 K to 81 K, a maximum-to-minimum ratio of under 1.5, which is substantially less than the ratio found in most compartmental models, which range from 5.3 to over 700.

Next, we take a more detailed look by considering the daily dynamics through the end of February. On February 12th, three days after we published our forecast (Wu *et al.*, 2020b), the Hubei Province health commission changed the diagnosis criteria, and allowed those patients who do not have definitive PCR test results but have clinical symptoms (mainly CT scan results) to be counted as confirmed COVID-19 cases. This changed criterion increased the daily incremental cases by more than 14,000, resulting in the spike showing in Fig. 2.

The spike led us to consider a drastic revision of our model's future forecasts to adjust for the changed criterion. However, after carefully examining the implication of the new criteria, we decided not to change our forecast, as we concluded that the change would not have too much impact on the forecast results for cumulative cases in the long term, because it just confirmed the suspected cases earlier, most

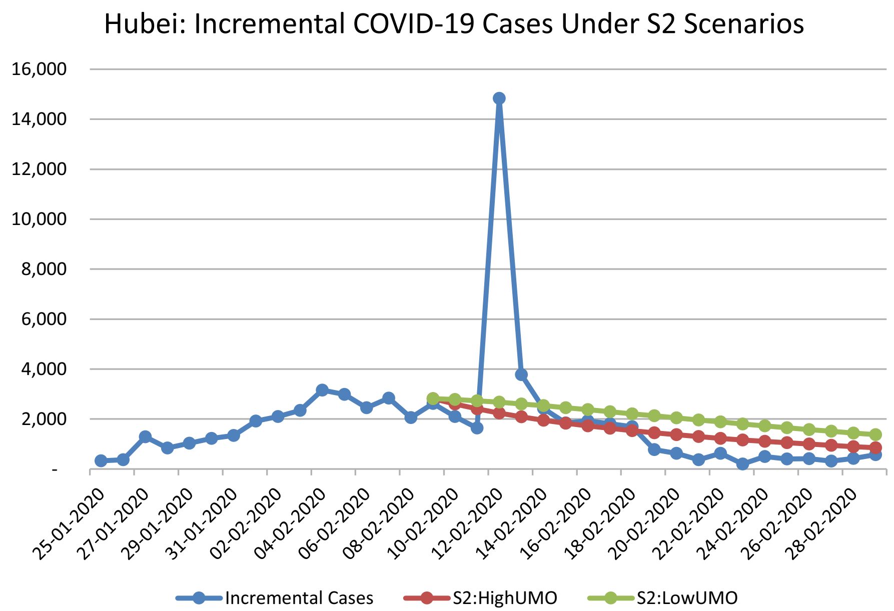

Fig. 2. Cautiously optimistic scenario forecasts of incremental cases.

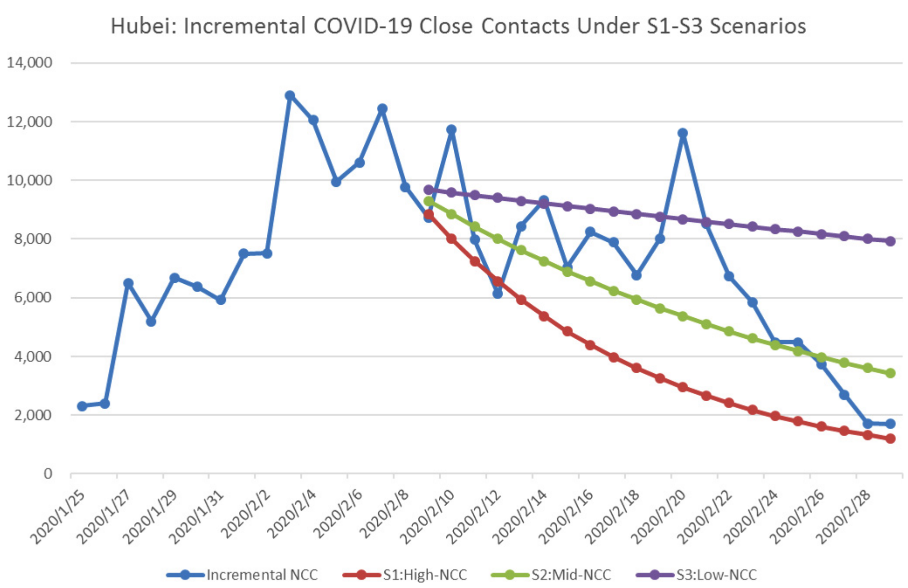

Fig. 3. Scenario forecasts of close contacts.

of which would turn into confirmed cases sooner or later. As long as our forecasts for new close contacts are not heavily affected, the model should be robust to handle this surge in incremental COVID-19 cases. Indeed, Fig. 3 confirms that the

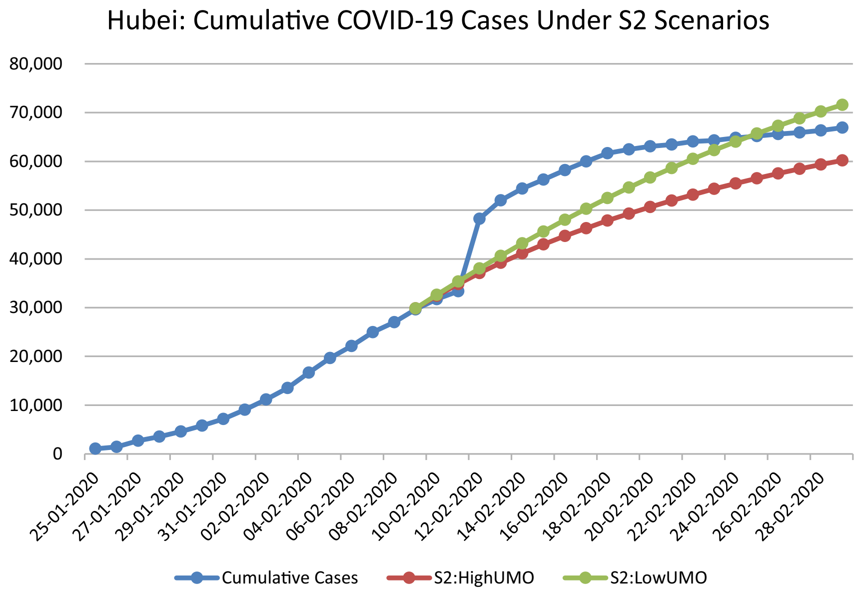

Fig. 4. Cautiously optimistic scenario forecasts of cumulative cases: The effects of the spike in reported cases on February 12 are washed out by the end of the month.

incremental close contacts did not change dramatically as a result of the spike in new confirmed cases. As a result, the cumulative cases are back in the range of the two S2 scenarios by the end of February, as shown in Fig. 4. Also, the critical case numbers are not heavily impacted by the spike, as there is an intermediate state of severe cases between critical cases and new cases; thus, the spike in new cases is not immediately reflected in the critical case numbers. So Fig. 5 shows that the model forecasts up to February 22 are good, after which the results of the arrival of medical assistance teams led to a dramatic decrease. On the other hand, the severe cases forecasts are clearly impacted by the surge, since there is no middle state between the severe state and new cases, and that resulted in moderate underprediction for much of the month, as shown in Fig. 6.

The arrival of armies of medical assistance teams throughout the Hubei province in February clearly had a beneficial effect on reducing mortality rates and conversely increasing cure rates in Hubei province, as shown in Figs. 7 and 8, respectively, where the model forecasts gradually and consistently diverge for both of these cumulative counts (cured and deceased) throughout the month. This is discussed further in the next section.

5. Supporting COVID-19 Medical Management Decisions in Wuhan

In the very early stages of the COVID-19 outbreak, the daily fatality rate of severe (including critical) cases was extremely high, around 8%. On Chinese New Year's

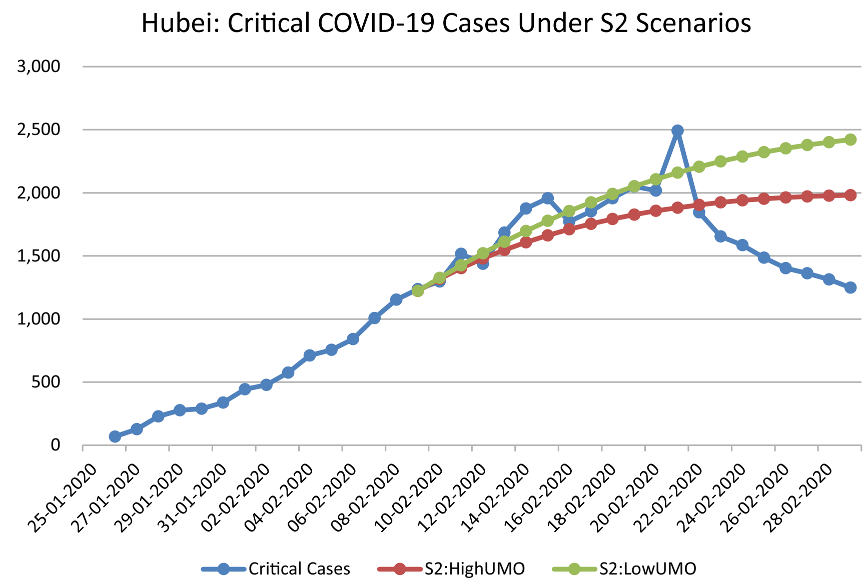

Fig. 5. Cautiously optimistic scenario forecasts of critical cases: Drop starting February 22nd reflecting the arrival of medical assistance teams from the rest of China.

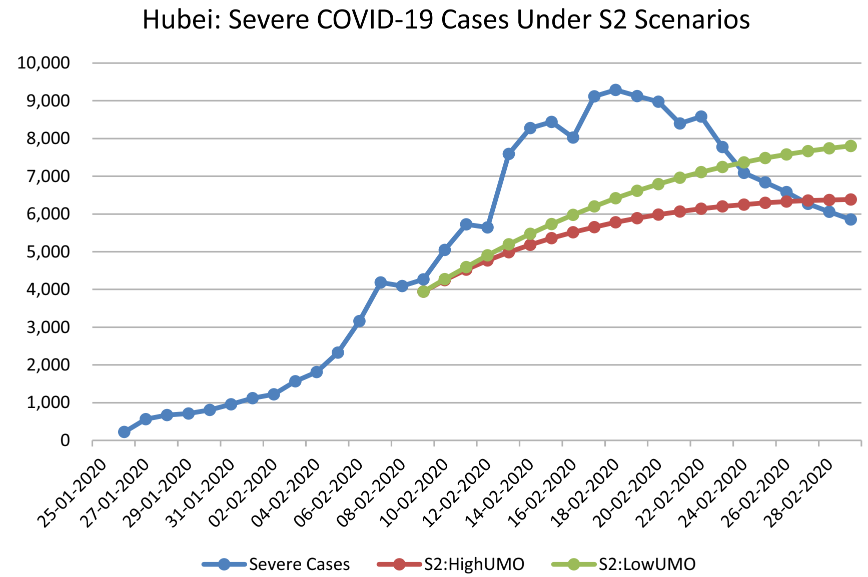

Fig. 6. Cautiously optimistic scenario forecasts of severe cases: Surge in reported cases due to changed criteria (spike in Fig. 2 on February 12) leads to underprediction for most of the month.

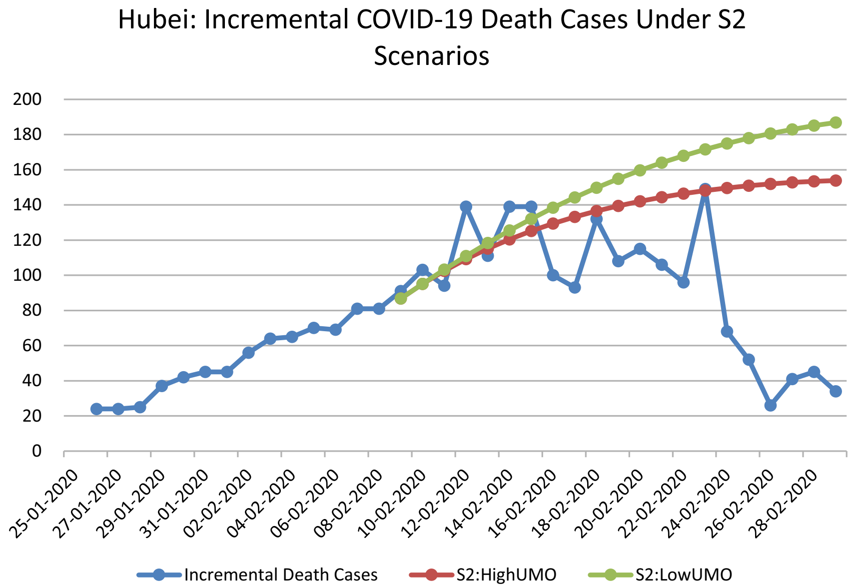

Fig. 7. Cautiously optimistic scenario forecasts of death cases: Effect of medical assistance teams leads to marked decreased mortality rates during February.

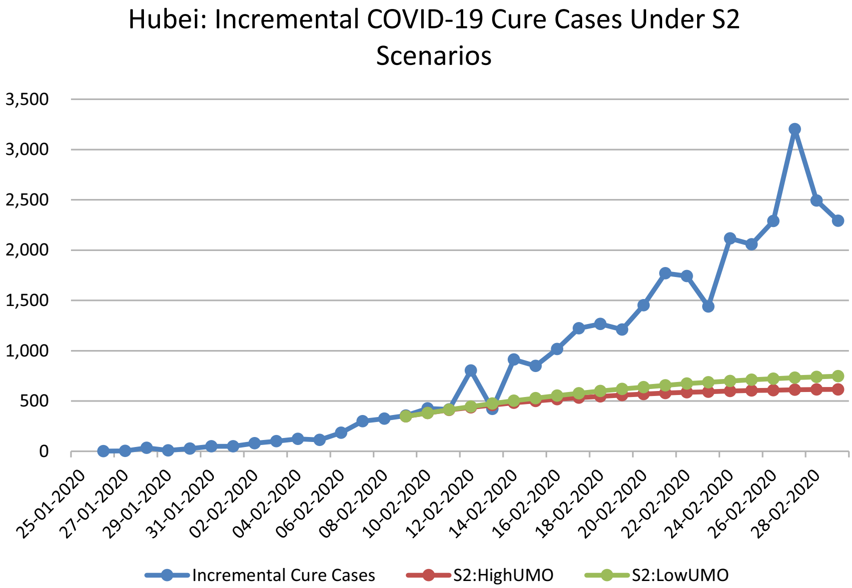

Fig. 8. Cautiously optimistic scenario forecasts of cured cases: Effect of medical assistance teams leads to a marked increase in cure rates during February.

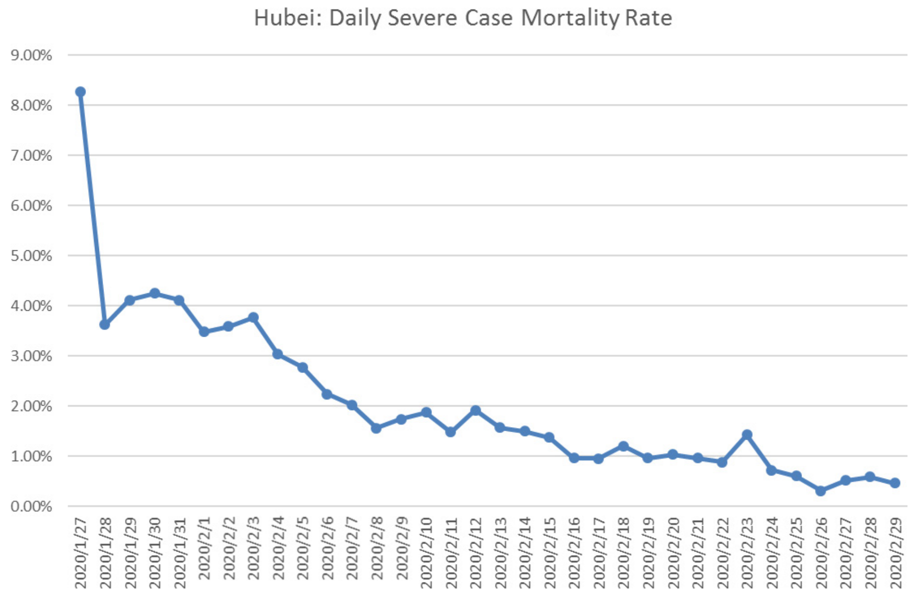

Fig. 9. Mortality rate continual drop (nonstationarity affects model forecasts).

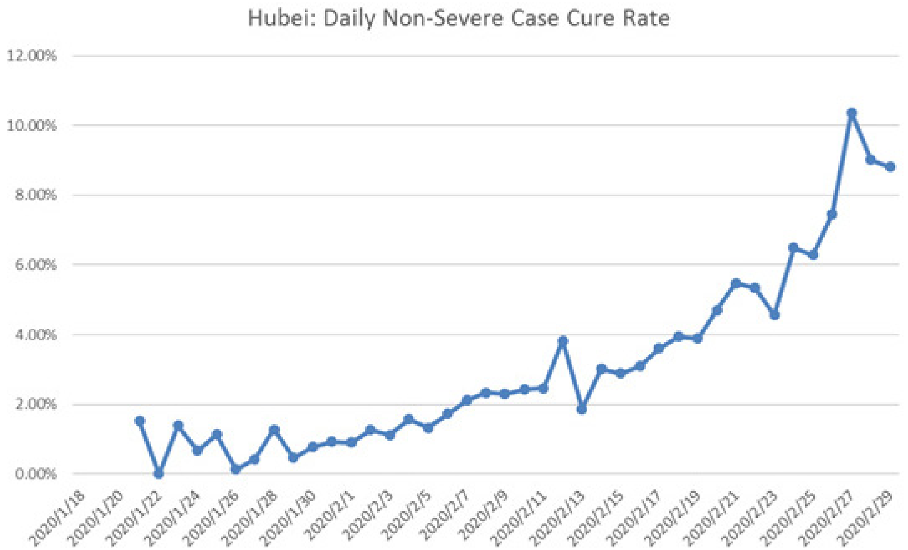

Fig. 10. Cure rate continual improvement (nonstationarity affects model forecasts).

Eve, January 24th, the first medical assistance team was dispatched from Shanghai to Wuhan to help the local medical staff. On February 10th, the TMM forecast was adopted by the first Shanghai medical assistance team (led by Dr. Zheng, one of this paper's authors) in Wuhan's Jinyintan Hospital, the first designated hospital to take

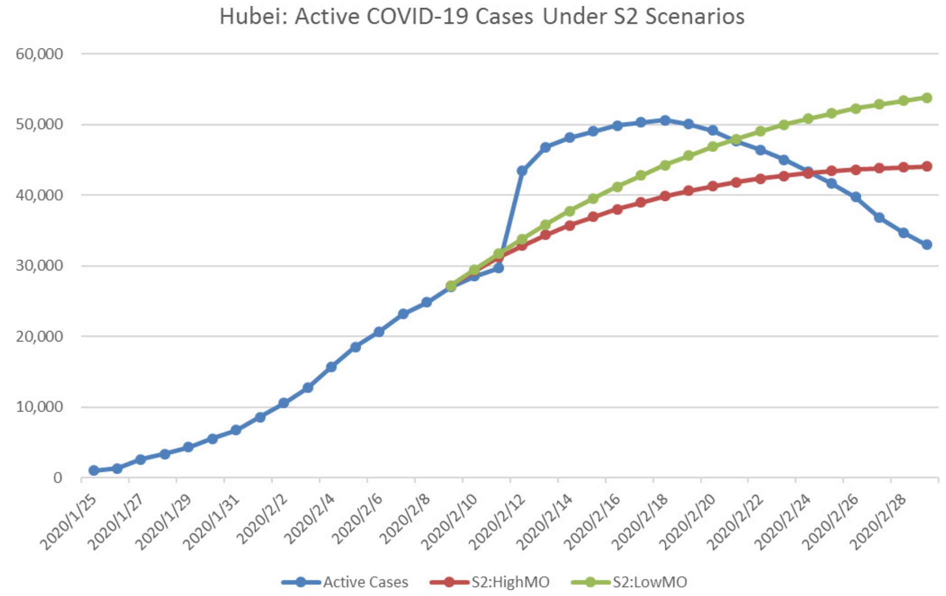

Fig. 11. Actual peak of active cases in mid-February two weeks earlier than model forecasts due to improvement in medical treatment (also reflected in Figs. 9 and 10).

COVID-19 patients in the world. The forecast has been used in preparing medical staff, ICU beds, ventilators, and other critical care medical resources by central and provincial health commissions and local centers of disease control. For example, on February 14th, we published an article,[2] indicating that under the cautiously optimal scenario, medical staff needed for taking care of severe and critical patients could reach 40,000–45,000. Soon after this forecast, more medical assistance teams were dispatched from all over China to Wuhan and other Hubei province cities, reaching more than 42,000 medical staff by the beginning of March. On February 15th, we forecasted the "back-to-normal" date most likely to be mid-April,[3] but due to the extraordinary efforts of these medical assistance teams, the lock-down in Wuhan was lifted on April 8th (and Dr. Zheng was able to return to Shanghai after fighting COVID-19 for 67 days in the epicenter of Wuhan).

During January/February 2020, two hospitals with 1,900 beds were built within two weeks to accept severe and critical patients. As time went on, medical staff became more experienced in treating COVID-19. All these measures helped to reduce the daily severe case fatality rate to a very low level of 0.5%, an almost 94% drop from the very early stages, illustrated in Fig. 9. The model forecasts in Fig. 7 use the higher fatality rate observed on 2/8, resulting in higher death toll forecasts.

[2]http://chenjian.blog.caixin.com/archives/221560.
[3]http://chenjian.blog.caixin.com/archives/221630.

Moreover, starting from 2/5, 16 shelter hospitals were put into use, accepting more than 12,000 non-severe patients in Wuhan. These shelter hospitals were converted from stadiums, shopping malls, convention centers, etc. This measure moved the treatment window earlier and placed patients with mild symptoms in these makeshift hospitals to receive proper medical treatment rather than self-quarantining at home, where there would be significantly heightened risk of family transmission and community transmission. The daily cure rate of non-severe cases increased dramatically from the level of 1% in late January to close to 10% in late February, illustrated in Fig. 10. The model forecasts in Fig. 8 use the much lower cure rate observed on 2/8, resulting in the significantly lower cure forecasts.

Finally, as a result of all these timely measures, the peak of actual active cases was moved earlier to February 16, instead of the date originally estimated as between 3/1–3/7, as illustrated in Fig. 11.

6. Conclusions and Ongoing Research

We introduced a new discrete-time Markov chain (DTMC) transition matrix model (TMM) for modeling epidemic outbreaks that directly incorporates stochastic behavior. Parameter estimation for the model is straightforward, so we applied the model using COVID-19 data from Hubei province, for which it provided reasonably accurate forecasts, with sensitivity analysis illustrating its robustness properties in terms of far less sensitivity to parameter misspecification than traditional epidemiological compartmental models. As a result, the model has been adopted by the first Shanghai assistance medical team in Wuhan's Jinyintan Hospital, the first designated hospital to take COVID-19 patients in the world, and the forecasts have been used for preparing and allocating medical staff, ICU beds, ventilators, and other critical care medical resources and for supporting medical management decisions.

The proposed approach can forecast all the states in the COVID-19 pandemic, including the intermediate states (medical observation, mild cases, severe cases, critical cases), and the terminal states (discharged from medical observation, cured, deceased). As mentioned earlier in the model formulation, the model applied here to COVID-19 does not include asymptomatic cases, which can be readily handled by adding one more intermediate state, but was purposely excluded since the data for the asymptomatic population is not available. Also, the major purpose of the model is to predict active cases, i.e., patients who are receiving treatment in medical facilities, and the patient distribution of severe/critical cases to better manage medical resources.

Similar to the finance academic community where the literature on asset pricing is dominated by stochastic partial differential equation models for which stylized models are used to generated closed-form solutions and big-picture insights, the prevalent modeling paradigm in the academic epidemiological research literature also values theoretical models, in this case, based on systems of deterministic ordinary differential equations. The first author has spent over two decades in the finance industry and implemented many models for investment decisions, and feels that the model proposed here follows the same vein of industry relying on more practically implementable models. The proposed approach is a preliminary attempt to advocate models that are tailored to the available data and anticipated usage of the model in decision making, whether it be strategic policy, supply chain planning, or hospital operations. Thus, the focus is on flexibility, ease of implementation, and robustness rather than theoretical elegance.

In terms of ongoing work, Zheng *et al.* (2020) draws upon the experience from Hubei and other provinces in China during the early stages of the COVID-19 outbreak to improve the TMM model by incorporating different levels of preventive policy efficiency. This led to forecasts for Italy, South Korea, and Iran, which were posted online[4] on March 9. The forecasts for Italy were channeled to an Italian cabinet member on the same day, indicating a very dire situation with a forecast of more than 190,000 likely cases by April end with weak intervention efforts. The Italian government implemented a national “lock-down” policy on the next day.

References

Chen, J, J Xiang and T Yang, (2018). Re-default risk of modified mortgages, *International Real Estate Review*, 21(1), 1–40.

Chen, J, MC Fu, W, Zhang and J Zheng (2020). Modeling the COVID-19 Spread: An OR/MS View. *OR/MS Today*, June 2020.

Kermack, W and A McKendrick (1927). A contribution to mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, 115, 700–721.

Malik, M and LC Thomas (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28(1), 261–272.

World Health Organization (2020). Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected, Interim Guidance, 13 March 2020.

Wu, JT, K Leung and GM Leung (2020a). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet*, 395, 689–697.

Wu, K, J Zheng and J Chen (2020b). Utilize state transition matrix model to predict the novel corona virus infection peak and patient distribution, medRxiv 2020.02.16.20023614; doi: https://doi.org/10.1101/2020.02.16.20023614.

Xu, J, E Huang, C-H Chen and L-H Lee (2015). Simulation optimization: A review and exploration in the new era of cloud computing and big data. *Asia-Pacific Journal of Operational Research*, 32(3), 2015.

Zheng, Z, K Wu, Z Yao, J Zheng and J Chen (2020). The prediction for development of COVID-19 in global major epidemic areas through empirical trends in China by utilizing state transition matrix model, medRxiv 2020.03.10.20033670; doi: https://doi.org/10.1101/2020.03.10.20033670.

[4] http://chenjian.blog.caixin.com/archives/223401.

Biography

Jian Chen is the founder and CEO of CreditWise Technologies, Co. Ltd. He also serves as the senior adviser of Caixin Insight Group. His specialization includes applications of Discrete Event Dynamic Systems (DEDS) modeling in financial engineering, credit risk management, and infectious disease. He previously held the positions of Managing Partner of RQuest Financial Services Group, Managing Director in IFE Group, Risk Modeling Director in Freddie Mac, Director of Credit Risk Management in Fannie Mae. Dr. Chen holds the academic position of adjunct professor at the Shanghai Institute of Advanced Finance (SAIF). He previously served as an adjunct professor at Johns Hopkins Carey Business School. Dr. Chen holds a BS in Electrical Engineering from Xi'an Jiaotong University, an MS in Electrical Engineering from Shanghai Jiaotong University, and a PhD in Management Science with a concentration on Computational Finance from the Robert H. Smith School of Business, University of Maryland at College Park.

Michael C. Fu holds the Smith Chair of Management Science in the Decision, Operations and Information Technologies department of the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and an affiliate appointment in the Department of Electrical & Computer Engineering, all at the University of Maryland, College Park. He received degrees in math and EECS from MIT in 1985 and a PhD in Applied Math from Harvard University in 1989. He is the co-author/editor/co-editor of six books, including the *Handbook on Simulation Optimization*. He served as Department Editor for Stochastic Models and Simulation at *Management Science*, as Simulation Area Editor for *Operations Research*, and on the Editorial Boards of *Mathematics of Operations Research*, *INFORMS Journal on Computing*, *IIE Transactions*, *and Production and Operations Management*. He served as the Operations Research Program Director at the U.S. National Science Foundation from 2010–2012 and in 2015. He is a Fellow of IEEE and INFORMS.

Wenhong Zhang is Professor and Director of the Department of Infectious Disease at Huashan Hospital at Fudan University, where he also serves as head of the clinical center for infectious disease and liver diseases. Professor Zhang serves as the Chair of the Society of Shanghai Infectious Diseases Physicians and the Secretary-General of the Chinese Association of Infectious Diseases. He has worked as a visiting scholar and postdoctoral fellow at the Department of Microbiology at the University of Hong Kong, Harvard Medical School, and Chicago State University. He has in-depth experience diagnosing and treating various emerging infectious diseases. Since January, Dr. Zhang has served as the head of the Experts Team of Medical Treatment of COVID-19 patients in Shanghai, and he has emerged as a leading expert on the prevention and treatment of COVID-19 in the country. Dr. Zhang graduated from Shanghai Medical University.

Junhua Zheng is the Vice President of First People's Hospital, which is affiliated with Shanghai Jiaotong University. Dr. Zheng is also a professor and a doctoral supervisor at Shanghai Jiaotong University. He is a member and the secretary-general of the Chinese Urological Association, and an editorial advisory board member of medical academic journals including the Chinese Medical Journal and the Chinese Journal of Urology. With an MD degree, Zheng has been dedicated to the research of minimally invasive treatment and infiltration and metastasis of renal carcinoma for the last ten years, and he is the chief editor of the first book regarding organ reservation in China. Dr. Zheng led the Shanghai Medical Assistance Team to Wuhan after the coronavirus outbreak in Wuhan beginning in January. He and his team have worked on the frontlines battling the pandemic since January 24.