Estimating short term trends in transmission and mortality rates during the Covid 19 Epidemic

John Sibert*
Joint Institute of Marine and Atmospheric Research
University of Hawai'i at Mānoa
Honolulu, HI 96822 U.S.A.

July 19, 2020

Introduction

The sudden advent of the COVID-19 pandemic provoked many political jurisdictions to advise people to "shelter in place" and to practice "social distancing". If this advice has been effective, it should be possible to detect the effects of the advice by comparing changes in numbers of infected people and perhaps changes in transmission rates over time and between areas. The SIR models of epidemic spread divide the affected population into three compartments: Susceptible, Infected and Recovered. SIR models are usually

*sibert@hawaii.edu; johnrsibert@gmail.com

expressed as coupled ordinary differential equations,

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{IS}{N} - \mu S && (1)\\
\frac{dI}{dt} &= \beta\frac{IS}{N} - \mu I - \gamma I && (2)\\
\frac{dR}{dt} &= -\mu R + \gamma I && (3)\\
N &= S + I + R && (4)
\end{aligned}$$

where N is the population size, β is the instantaneous transmission rate ($[t^{-1}]$), μ is the instantaneous mortality rate ($[t^{-1}]$), and γ is the instantaneous recovery rate ($[t^{-1}]$).

Unfortunately, few data sets include data for each of these compartments. The New York Times' "historical" data[1] is an easily accessible source of data. These data comprise daily totals of "cases" and "deaths" for each county in the United States. I assume that the data included as "cases" are a reasonable approximations of the Infected compartment (I) in a SIR model. There are simply no credible data of comparable scope on either the Susceptible or the Recovered compartments.

Model Structure

I make some simplifying assumptions in the face of incomplete data: (1) The entire population is susceptible so that $S/N = 1$. (2) Over the short term, the size of the Susceptible compartment does not change, $\frac{dS}{dt} = 0 = \frac{dN}{dt}$, eliminating the Susceptible compartment. (3) People who recover from a

[1] `https://github.com/nytimes/covid-19-data/`

COVID-19 infection return to the Susceptible compartment, eliminating the Recovered compartment. With these assumptions, and with the addition of a "deaths" compartment, the simplified SIR model is

$$\frac{dI}{dt} = \beta I - \mu I - \gamma I \tag{5}$$

$$\frac{dD}{dt} = \mu I \tag{6}$$

and has state variables that might be matched to available observations.

The data available during the initial stages of the COVID-19 pandemic contain measurement errors of various types. Definitions and methods of detecting and reporting the numbers of infected persons vary between political jurisdictions (or "geographies" in the parlance of the New York Times) and may also change with time. Comparable uncertainties also occur in reporting of deaths caused by COVID-19 infection. There is additional variability in the biosocial processes that mediate disease transmission.

State-space models separate variability in the biosocial processes in the system (transition model) from errors in observing features of interest in the system (observation model). (See Harvey 1990).

The general form of a state-space process or transition model is

$$\alpha_t = T(\alpha_{t-1}) + \eta_t \tag{7}$$

where α_t is the state at time t and the function T embodies the dynamics mediating the development of the state at time t from the state at the previous time with random process error, η_t.

The transition model for the simplified SIR model is constructed from finite difference approximations of equation (5) with associated log-normal random errors.

$$
\begin{aligned}
I_t &= I_{t-\Delta t}\big(1 + \Delta t(\beta_{t-\Delta t} - \mu_{t-\Delta t} - \gamma_{t-\Delta t})\big)e^{\eta_t} && (8)\\
D_t &= \big(D_{t-\Delta t} + \Delta t \mu_{t-\Delta t} I_{t-\Delta t}\big)e^{\eta_t} && (9)
\end{aligned}
$$

where η is a normal random deviate, $\eta \sim N(0, \sigma_\eta)$, representing temporal variability in the biosocial factors that mediate the spread of the pandemic. The recovery rate, $\gamma_{t-\Delta t}$, in equation (8) is computed algebraically as

$$
\gamma_{t-\Delta t} = \beta_{t-\Delta t} - \mu_{t-\Delta t} + (1 - \frac{I_t}{I_{t-\Delta t}}) \tag{10}
$$

I have no particular justification, beyond the parsimony principle, for the assumption that the variance, σ_η, of the processes for I and D, should be the same.

One approach to modeling time-dependent rates of transmission and mortality, β and μ, is to treat them as random effects (Skaug and Fournier 2006). Random effects are appropriate if repeating a time series of observations would not yield the same outcome as the initial observations. Random effects are also appropriate when observing the same process in two different areas. I model the β and μ time series as log-normal random walks. I assume that

$$
\begin{aligned}
\log \beta_t &= \log \beta_{t-\Delta t} + \varepsilon; \quad \varepsilon \sim N(0, \sigma_\beta) && (11)\\
\log \mu_t &= \log \mu_{t-\Delta t} + \varrho; \quad \varrho \sim N(0, \sigma_\mu) && (12)
\end{aligned}
$$

The general form of the state-space observation model is

$$
x_t = O(\alpha_t) + \varphi_t \tag{13}
$$

where the function O describes the measurement process with error ε in observing the state α.

I applied separate observation error models for cases and deaths. The observation model for cases is a simple log-normal error

$$\log \varphi_t = \left(\log \frac{1}{\sqrt{2\pi\sigma_I^2}} - \left(\frac{\log I_t - \log \widehat{I}_t}{\sigma_I} \right)^2 \right) \tag{14}$$

where I is the observed number of cases and $\widehat{I}$ is the number of cases predicted by equation 8.

Not all those afflicted by COVID-19 have died; there are far fewer deaths than infections. In addition, the observed time series for both I and D begins at the first recorded case. The first recorded death occurs several days or weeks after the first recorded case. Therefor the deaths time-series inevitably contains a substantial number of recorded zeros. The observation model for deaths accommodates observed zeroes by assuming to be "zero-inflated" log normal likelihood given by

$$\log \varepsilon_t = \begin{cases} D_t > 0: & (1 - p_0) \cdot \left(\log \frac{1}{\sqrt{2\pi\sigma_D^2}} - \left(\frac{\log D_t - \log \widehat{D}_t}{\sigma_D} \right)^2 \right) \\ D_t = 0: & p_0 \cdot \log \frac{1}{\sqrt{2\pi\sigma_D^2}} \end{cases} \tag{15}$$

where D is the observed number of deaths, $\widehat{D}$ is the number of deaths predicted by equation 9, and p_0 is the proportion of observed deaths equal to zero.

Model parameters are estimated by maximizing the joint likelihood of the process errors, observation errors, and random effects.

$$L(\theta, \alpha, x) = \prod_{t=2}^{m} \left[\phi\big(\alpha_t - T(\alpha_{t-1}), \Sigma_\eta\big) \right] \cdot \prod_{t=1}^{m} \left[\phi\big(x_t - O(\alpha_t), \Sigma_\varepsilon\big) \right] \tag{16}$$

Table 1: List of model variables for the simple SIR model. There are two state variables computed from the of estimated parameters and random effects. There are two random effects and five estimated variance parameters.

Variable	Definition
	State variables:
I	Number of infected individuals or "cases"
D	Number of deaths
	Random effects:
β_t	Transmission rate
μ_t	Mortality rate
	Estimated parameters:
σ_I	Infectious compartment estimation standard deviation
σ_D	Deaths compartment estimation standard deviation
σ_η	Standard deviation of transmission and deaths process errors
σ_β	Standard deviation of transmission rate random walk
σ_μ	Standard deviation of mortality rate random walk

where m is the number of days elapsed since the first recorded case, x_t is the vector of daily observations of cases and deaths, α_t is the vector of the daily calculations of the state variables and random effects, and θ is a vector of model parameters (Table 1). The R package TMB (Kristensen et al. 2016 package was used to estimate the parameters of the model. The R and supporting C++ files are available on github.[2]

[2]simpleSIR4 at `https://github.com/johnrsibert/SIR-Models`

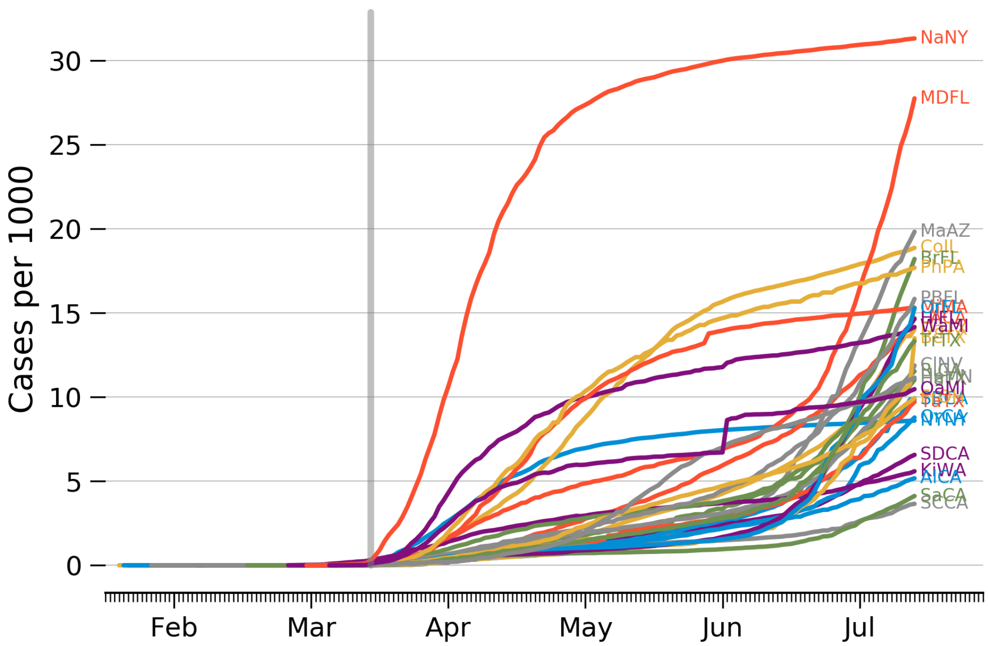


Figure 1: Trends in number of cases per 1000 people in the 30 most populous US counties. The vertical gray bar mark the date of the California shelter in place order.

Results

The trends in number of cases in the thirty largest counties in the United States are shown in Figure (1). These trajectories fall into two more or less distinct groups: those that are concave downward, e.g. New York City (NYNY), and those that are concave upward, e.g. Miami-Dade FL (MDFL).

Prevalence histories for these five counties are shown in Figure 2 where the 11-day moving average of the daily increase in cases demonstrates clearly the efficacy of control measures.

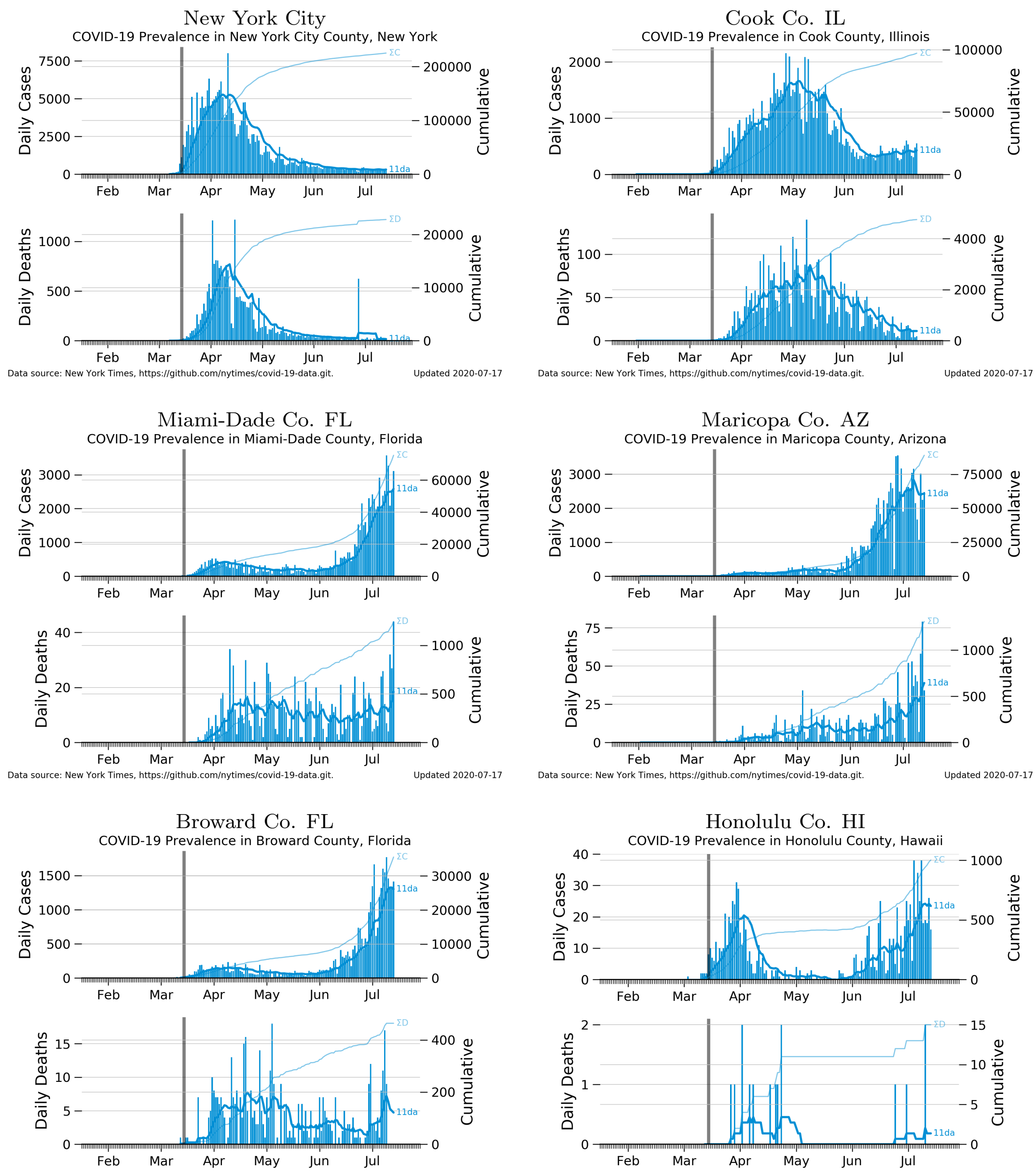
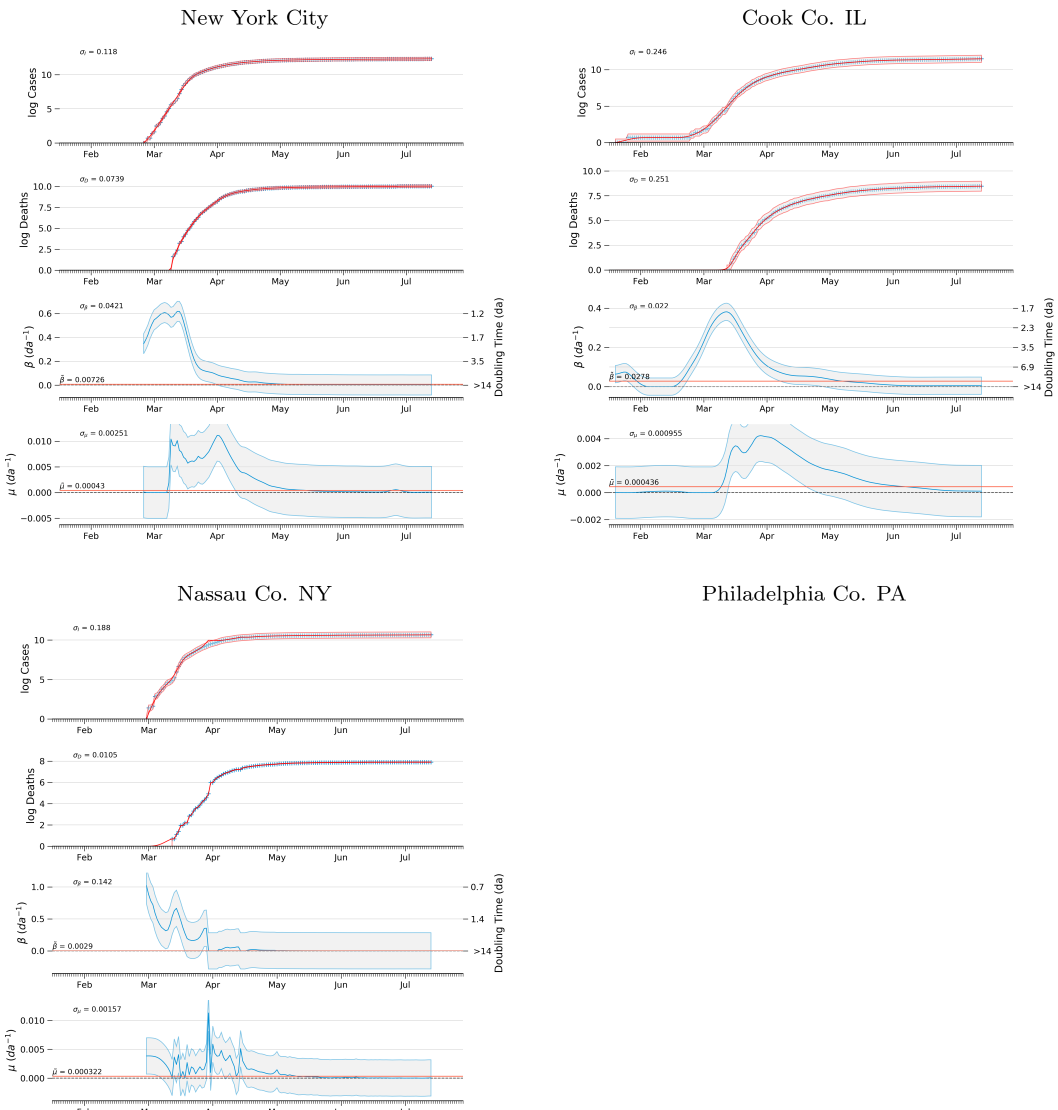


Figure 2: Prevalence trajectories for six US counties. Blue bars indicate daily increases increases in cases and deaths; blue lines indicate 11 day moving averages of daily increases; pale blue lines indicate cumulative numbers (labeled ΣC and ΣD); vertical gray bar marks the California shelter in place order. *remove annotations.*

Model results are shown in figures 3 and 4. The model reproduces the observed numbers of cases and deaths almost exactly. The '+' symbols in the Cases and Deaths graphs represent the observed cases (I) and deaths (D) from the data. The red lines overlaying the symbols are model predictions ($\widehat{I}$) and ($\widehat{D}$) of in cases and deaths. σ_I and σ_D are the estimated standard deviations for cases and deaths likelihood contributions, equations (14) and (15). The shaded areas bounded by red outlines are ± 2 estimated standard deviations around the estimated trends. These standard deviations are, in some cases, equivalent to an error of approximately one case or death.

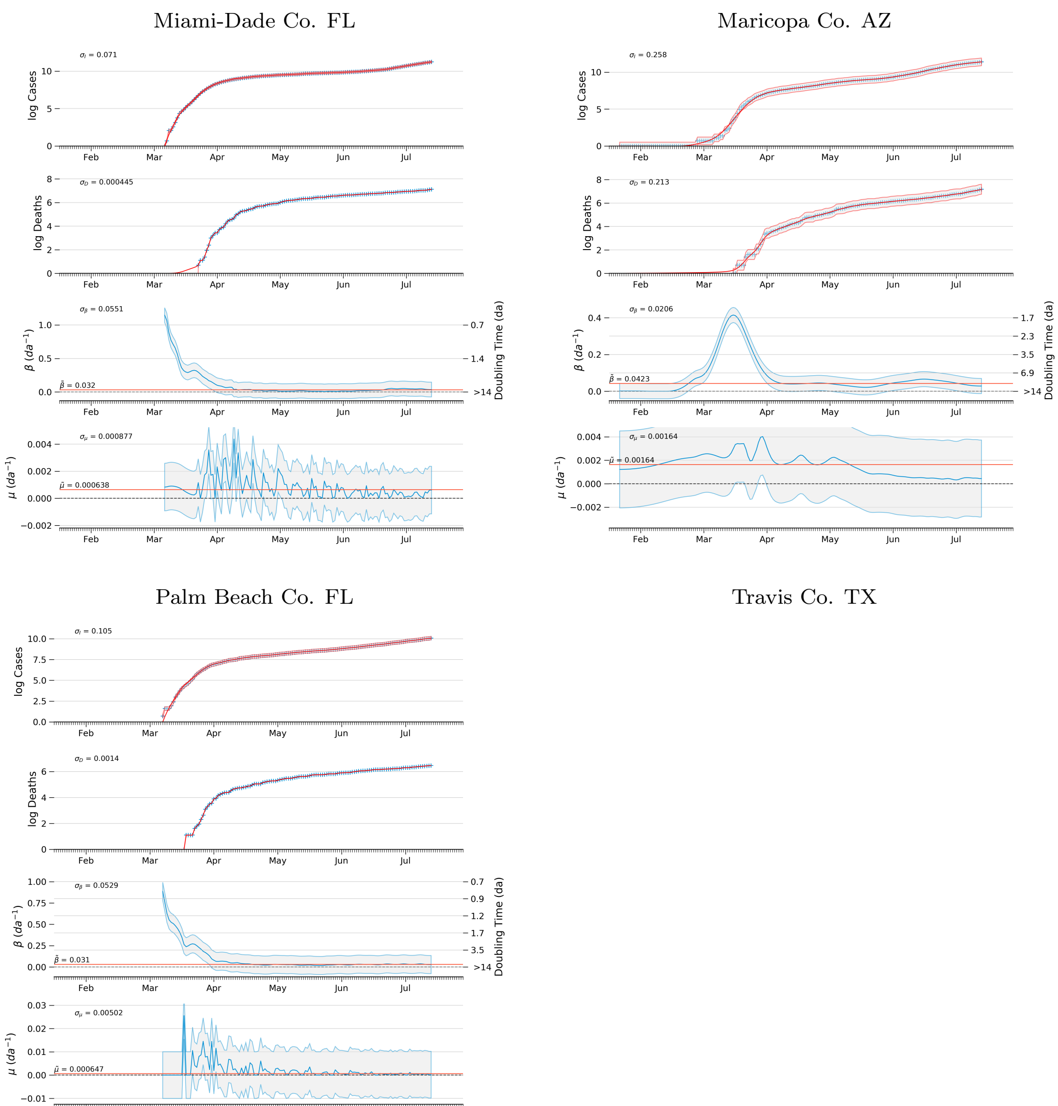
The solid blue lines in the β and μ plots are the estimated transmission and death rate random effects. The shaded areas bounded by blue outlines are estimated random effects ± 2 standard deviations of the generating random walks. The red lines labeled $\tilde{\beta}$ and $\tilde{\mu}$ are the medians of the two random effects.

The estimated transmission rates are very high at the beginning of each time series, exceeding 1da^{-1}in Miami-Dade County, equivalent to a doubling time of less than one day.



New York City
Cook Co. IL
Nassau Co. NY
Philadelphia Co. PA
log Cases
log Deaths
β (da⁻¹)
μ (da⁻¹)
Doubling Time (da)
Feb
Mar
Apr
May
Jun
Jul
σI = 0.118
σD = 0.0739
σβ = 0.0421
β̂ = 0.00726
σμ = 0.00251
μ̂ = 0.00043
σI = 0.246
σD = 0.251
σβ = 0.022
β̂ = 0.0278
σμ = 0.000955
μ̂ = 0.000436
σI = 0.188
σD = 0.0105
σβ = 0.142
β̂ = 0.0029
σμ = 0.00157
μ̂ = 0.000322

Figure 3



Miami-Dade Co. FL
Maricopa Co. AZ
Palm Beach Co. FL
Travis Co. TX
log Cases
log Deaths
Doubling Time (da)
Feb
Mar
Apr
May
Jun
Jul

Figure 4

Table 2: Model results. Estimating β and μ trends as random effects with computed γ. Data updateed from https://github.com/nytimes/covid-19-data.git 2020-07-17.

County	n	p_0	f	C	σ_η	σ_β	σ_μ	σ_I	σ_D	$\tilde{\gamma}$	$\tilde{\beta}$	$\tilde{\mu}$
Nassau, NY	133	0.0896	-3590	1	0.183	0.142	0.00157	0.188	0.0105	-2.18e-08	0.00290	0.000322
New York City, NY	137	0.0942	-2180	10	0.298	0.0421	0.00251	0.118	0.0739	-3.39e-08	0.00726	0.00043
Honolulu, HI	132	0.188	-3680	10	0.272	0.0236	0.00108	0.265	0.246	-6.07e-08	0.0179	8.69e-05
Cook, IL	174	0.303	-2770	10	0.38	0.022	0.000955	0.246	0.251	-2.29e-07	0.0278	0.000436
Harris, TX	133	0.104	-1900	10	0.241	0.0299	0.000958	0.282	0.253	-3.76e-08	0.0304	0.000292
Palm Beach, FL	126	0.0787	-3940	1	0.155	0.0529	0.00502	0.105	0.0014	-2.42e-08	0.0310	0.000647
Miami-Dade, FL	127	0.125	-1750	1	0.157	0.0551	0.000877	0.071	0.000445	-2e-08	0.0320	0.000638
Tarrant, TX	128	0.0775	-1850	10	0.273	0.0298	0.00105	0.263	0.21	-4.01e-08	0.0369	0.000409
Broward, FL	132	0.0827	-1610	10	0.208	0.0199	0.00194	0.254	0.24	-3.17e-08	0.0371	0.000473
Travis, TX	125	0.111	-2390	10	0.397	0.0236	0.0024	0.241	0.0643	-2.91e-08	0.0375	0.000363
Dallas, TX	128	0.0698	-1520	10	0.222	0.0222	0.00186	0.277	0.177	-2.73e-08	0.0382	0.000622
Hillsborough, FL	137	0.181	-1940	10	0.113	0.219	0.00477	0.0562	0.0309	-8.13e-08	0.0401	0.000599
Maricopa, AZ	172	0.312	-2040	10	0.213	0.0206	0.00164	0.258	0.213	-4.26e-07	0.0423	0.00164
Bexar, TX	155	0.25	-2370	10	0.199	0.0391	0.00103	0.237	0.246	-8.51e-08	0.0448	0.000363
Median	132.5	0.1075	-2110	10	0.2175	0.02985	0.001605	0.2435	0.1935	-3.575e-08	0.03445	0.000433

Discussion

Whether the available data are sufficiently informative to enable estimation of the model parameters is a critical aspect of the evaluation of any statistical model. The speed at which the COVID-19 pandemic spread during the first quarter of 2020 means that the length of the time series doubled during the development of this model. The capability of the model improve conveniently during the model development period, but whether the improvement is attributable to changes in model structure or to the increase in the length of the time series is unclear. This ambiguity influenced the development of the model.

Sibert 2017; Nielsen and Berg 2014; Chen et al. 2020

References

Chen, Yi-Cheng, Ping-En Lu, Cheng-Shang Chang, and Tzu-Hsuan Liu (2020). "A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons". In: pp. 1–18. arXiv: `2003.00122`. URL: `http://arxiv.org/abs/2003.00122`.

Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press. ISBN: 978-0521321969.

Kristensen, K., A. Nielsen, C.W. Berg, H.J. Skaug, and B.M. Bell (2016). "TMB: Automatic Differentiation and Laplace Approximation". In: *Journal of Statistical Software* 70, pp. 1–21. DOI: `doi:10.18637/jss.v070.i05`.

Nielsen, Anders and Casper W. Berg (2014). "Estimation of time-varying selectivity in stock assessments using state-space models". In: *Fish. Res.* 158, pp. 96–101. ISSN: 01657836. DOI: `10.1016/j.fishres.2014.01.014`. URL: `http://dx.doi.org/10.1016/j.fishres.2014.01.014`.

Sibert, John (2017). "Assessing of a portion of the Pacific Thunnus albacares stock : Ahi in the Main Hawaiian Islands". In: *arxiv.org* arXiv:1702. arXiv: `arXiv:1702.01217v1`.

Skaug, Hans J and David A Fournier (2006). "Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models". In: *Comput. Stat. Data Anal.* 51.2, pp. 699–709. ISSN: 01679473. DOI: `10.1016/j.csda.2006.03.005`.