# COFFEE: COVID-19 Forecasts using Fast Evaluations and Estimation

Lauren Castro[1,2], Geoffrey Fairchild[1], Isaac Michaud[3], and Dave Osthus[3]

[1]Information Systems and Modeling Group, Los Alamos National Laboratory
[2]Center for Nonlinear Studies, Los Alamos National Laboratory
[3]Statistical Sciences Group, Los Alamos National Laboratory

## Abstract

This document details the methodology of the Los Alamos National Laboratory COVID-19 forecasting model, COFFEE (COVID-19 Forecasts using Fast Evaluations and Estimation).[1]

## COFFEE Methodology

COFFEE is a probabilistic model that forecasts daily reported cases and deaths of COVID-19. COFFEE is fit to geographic regions independently, facilitating parallelization for fast computations.

### Notation

- $t$ indexes time, where $t$ is the number of days from a reference starting date (index)
- $T$ is the day of the last observation (index)
- $K$ is the forecast window size, in days (index)
- $y_{c,t}$ is the number of reported cases of COVID-19 on day $t$ as reported on the COVID-19 Dashboard by the Centers for Systems Science Engineering (CSSE) at Johns Hopkins University (JHU) (observable)
- $\ddot{y}_{c,t} = \sum_{j=1}^{t} y_{c,j}$ is the cumulative number of reported cases of COVID-19 through day $t$ as reported by CSSE at JHU (observable)
- $y_{d,t}$ is the number of reported deaths of COVID-19 on day $t$ as reported by CSSE at JHU (observable)

---

[1]Approved for unlimited release and assigned number LA-UR-20-28630

- $\ddot{y}_{d,t} = \sum_{j=1}^{t} y_{d,j}$ is the cumulative number of reported deaths of COVID-19 through day $t$ as reported by CSSE at JHU (observable)
- $\delta_{c,t}$ is the underlying number of reported cases on day $t$ (unobservable)
- $\ddot{\delta}_{c,t} = \sum_{j=1}^{t} \delta_{c,j}$ is the underlying number of cumulative reported cases through day $t$ (unobservable)
- $\delta_{d,t}$ is the underlying number of reported deaths on day $t$ (unobservable)
- $\ddot{\delta}_{d,t} = \sum_{j=1}^{t} \delta_{d,j}$ is the underlying number of cumulative reported deaths through day $t$ (unobservable)
- $\delta_{s,0}$ is the underlying number of susceptible individuals at the start of the pandemic (unobservable)
- $\delta_{s,t} = \delta_{s,0} - \ddot{\delta}_{c,t}$ is the underlying number of susceptible individuals on day $t$ (unobservable)

We use the convention that bolded quantities are vectors and unbolded quantities are scalars. For concreteness, $y_{c,t}$ is a scalar while $\boldsymbol{y}_{c,1:t} = (y_{c,1}, y_{c,2}, \ldots, y_{c,t})'$ is a $t \times 1$ vector.

## Cases Model

Let

$$y_{c,t}|\delta_{c,t}, \alpha \sim \text{NB}\left(\delta_{c,t}, \frac{\delta_{c,t}}{\alpha}\right) \tag{1}$$

where NB(a,b) is a Negative-Binomial model with mean parameter a $> 0$ and size parameter b $> 0$ where

$$\text{E}(y_{c,t}|\delta_{c,t}, \alpha) = \delta_{c,t} \tag{2}$$

$$\text{Var}(y_{c,t}|\delta_{c,t}, \alpha) = \delta_{c,t}(1 + \alpha). \tag{3}$$

Figure 1 shows the daily reported cases for New Mexico, the United States (US), and France. All three regions have gone through rising and declining periods of cases with various levels of noise in the reported cases.

In what follows, we outline the steps COFFEE takes to produce forecasts of reported cases.

### Step 1: Identify and Adjust Outliers

COFFEE automatically identifies and adjusts outliers [2]. It runs five different outlier detection algorithms on the reported data, taking into account possible day-of-week (DOW) effects. A datum is declared an outlier if three or more of the five detection algorithms identify that datum as an outlier. The outliers are not removed, but rather adjusted to ensure all values are non-negative. Figure 2 shows the result of this process on daily cases for New Mexico, the US, and France. All subsequent modeling steps are conducted with outlier adjusted data.
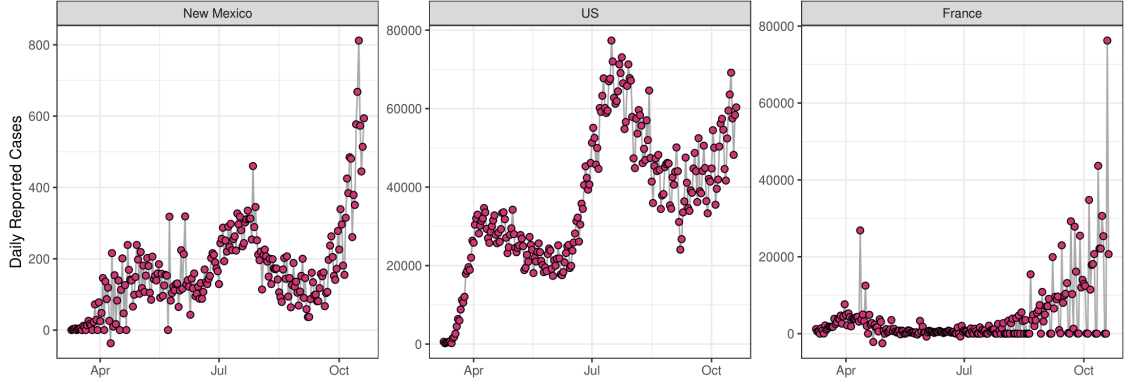
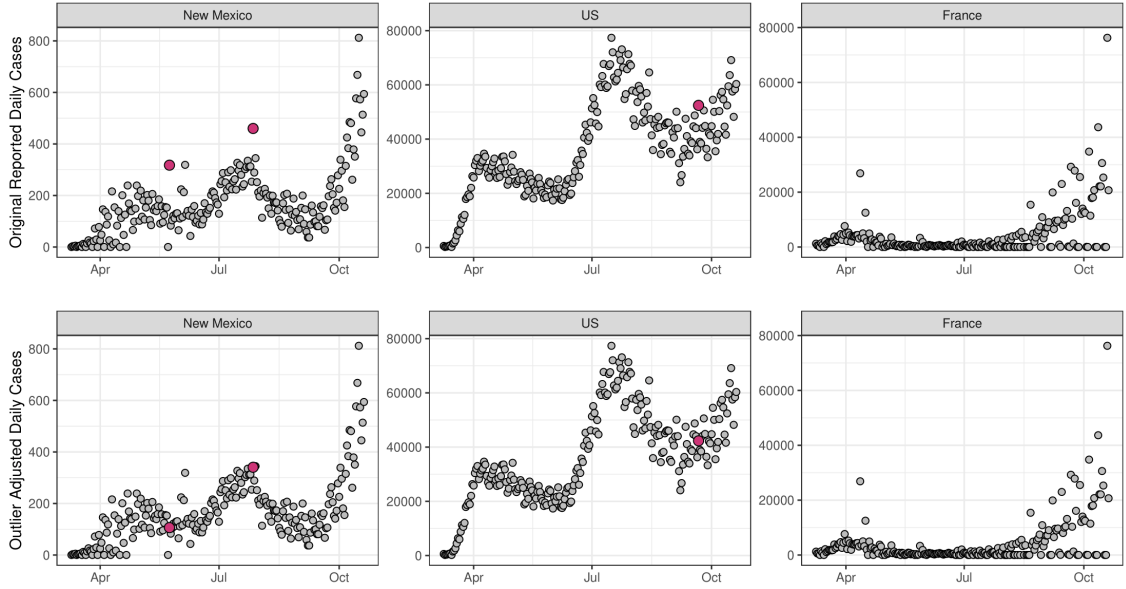Figure 1: The daily reported cases of COVID-19 for New Mexico, the US, and France.



Figure 2: The reported daily cases of COVID-19 (top) and the outlier adjusted daily cases (bottom). (Top) Magenta points were identified as outliers. (Bottom) Magenta points are the adjusted outliers.

### Step 2: Compute the Empirical Growth Rate, $\hat{\kappa}_t$

The model for the underlying number of reported daily cases, $\delta_{c,t}$, is a dynamic susceptible-infectious (SI) model [1], where

$$\delta_{s,t} = \delta_{s,t-1} - \delta_{c,t} \tag{4}$$

$$\ddot{\delta}_{c,t} = \ddot{\delta}_{c,t-1} + \delta_{c,t}, \tag{5}$$

3

and

$$\delta_{c,t} = \kappa_t \frac{\delta_{s,t-1}}{\delta_{s,0}} \ddot{\delta}_{c,t-1}. \tag{6}$$

The quantity $\ddot{\delta}_{c,t-1}$ is the cumulative number of underlying cases on day $t-1$, $\frac{\delta_{s,t-1}}{\delta_{s,0}}$ is the proportion of the population still susceptible at time $t-1$, and $\kappa_t$ is the growth rate on day $t$.

When $\frac{\delta_{s,t-1}}{\delta_{s,0}} \approx 1$ (when most of the susceptible population is still susceptible), we can rearrange Equations 4, 5, and 6 to identify a crude estimator of $\kappa_t$:

$$\hat{\kappa}_t \approx \left( \frac{\ddot{y}_{c,t}}{\ddot{y}_{c,t-1}} - 1 \right). \tag{7}$$

Estimates for $\kappa_t$ are shown in Figure 3. It is clear that $\hat{\kappa}_t$ is dynamic and changes over time. This is what makes forecasting COVID-19 so challenging; parameters of epidemiologically-motivated models are dynamic and *forecasting with them requires anticipating how these dynamic parameters will change in the future, not just tracking where they have been in the past.* In what follows, we describe how we forecast $\hat{\kappa}_t$.
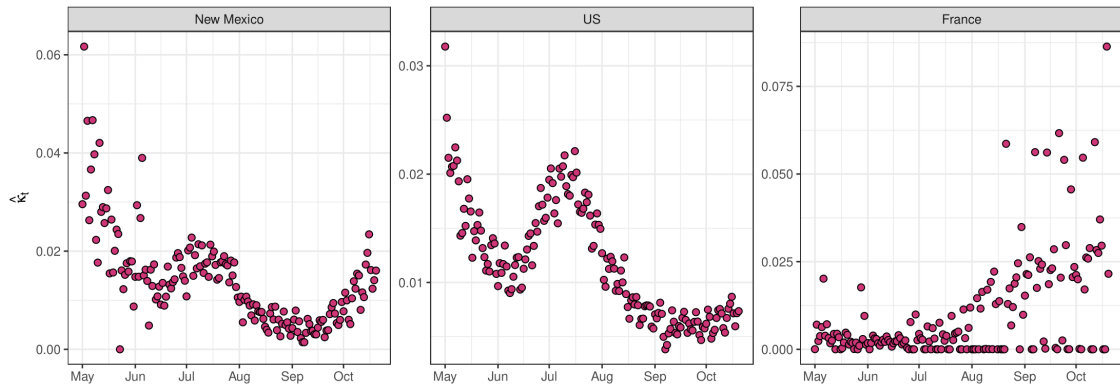


Figure 3: $\hat{\kappa}_t$ starting in May for New Mexico, the US, and France.

4

**Step 3: Compute $\hat{\kappa}_t^* = \text{logit}(\hat{\kappa}_t)$**

After the initial portion of the outbreak, $\hat{\kappa}_t$ is almost always between 0 and 1. COFFEE logit transforms $\hat{\kappa}_t$, where $\text{logit}(p) = \log(p/(1-p))$ for $p \in (0,1)$. For all days with no reported cases, $\hat{\kappa}_t = 0$, an incompatible value with the logit transform. Thus, we compute $\hat{\kappa}_t^* = \text{logit}(\hat{\kappa}_t)$ as follows:

$$\hat{\kappa}_t^* = \begin{cases} \text{logit}(\hat{\kappa}_t) & \text{if } \hat{\kappa}_t > \tau_c \text{ and } \hat{\kappa}_t < 1 - \tau_c \\ \text{logit}(\tau_c) & \text{if } \hat{\kappa}_t \leq \tau_c \\ \text{logit}(1 - \tau_c) & \text{if } \hat{\kappa}_t \geq 1 - \tau_c, \end{cases} \tag{8}$$

where $\tau_c = 0.95 * \min(\{\hat{\kappa}_t | \hat{\kappa}_t > 0\})$, ensuring that if $\hat{\kappa}_t \geq \hat{\kappa}_{t'}$, then $\hat{\kappa}_t^* \geq \hat{\kappa}_{t'}^*$. The logit transformed $\hat{\kappa}_t$ are shown in Figure 4.
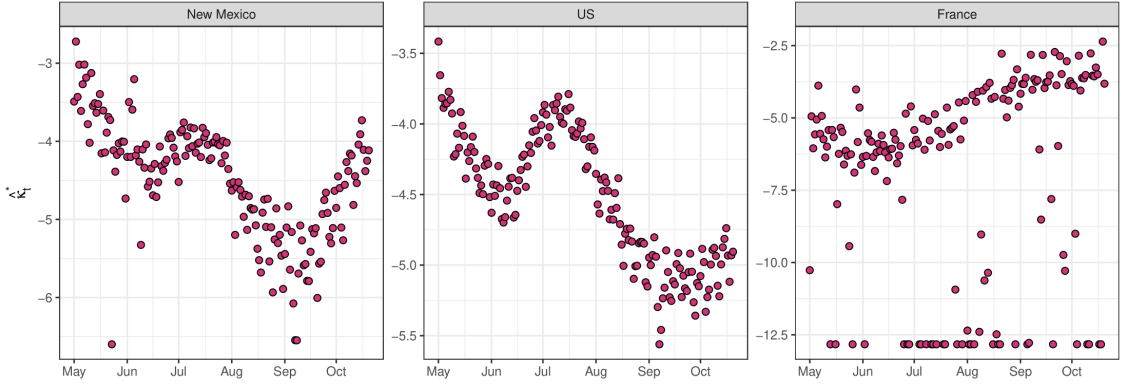


Figure 4: The quantities $\hat{\kappa}_t^*$ starting in May for New Mexico, the US, and France.

**Step 4: Split Data into Training and Testing Sets**

Let $T$ be the last observed day. We only consider the last 42 days of data when fitting a model for $\hat{\kappa}_t^*$ and split those days into training and testing data [3]. Days $T - 41$ through $T - 14$ constitute the training data, while $T - 13$ through $T$ constitutes the testing data. We will denote the last day of the training data by $T^{\text{train}} = T - 14$. The splits are shown for New Mexico, the US, and France in Figure 5.

**Step 5: Compute $\hat{\kappa}_t^{\text{trend}}$**

We fit a weighted regression to the training data, where we downweight influential points that could have an outsized influence on the regression using the inverse of Cook's distance. The regression has a linear trend over time and a DOW effect:

$$\hat{\kappa}_t^* = \beta_0 + \beta_1 t + \beta_2 \text{I}(t = \text{Monday}) + \beta_3 \text{I}(t = \text{Tuesday}) + \ldots + \beta_7 \text{I}(t = \text{Saturday}). \tag{9}$$
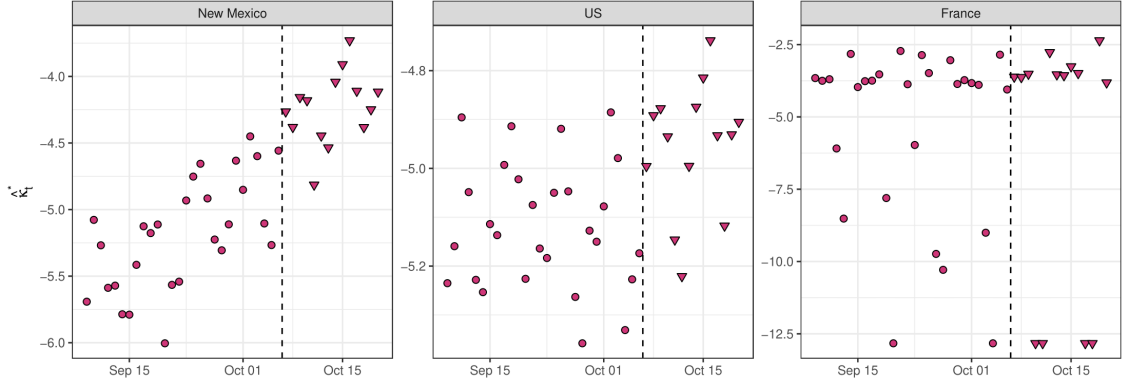
5

Figure 5: The quantities $\hat{\kappa}_t^*$ for New Mexico, the US, and France. Circles are training data, left of the vertical dashed line, while testing data are the triangles to the right of the dashed vertical line.

Variable selection is performed, potentially resulting in a subset of the model parameters in Equation 9. We refer to the fits and predictions from this linear model as $\hat{\kappa}_t^{\text{trend}}$ where

$$\hat{\kappa}_t^{\text{trend}} = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 I(t = \text{Monday}) + \hat{\beta}_3 I(t = \text{Tuesday}) + \ldots + \hat{\beta}_7 I(t = \text{Saturday}) \qquad (10)$$

where, if a variable was removed during the variable selection phase, then the corresponding $\hat{\beta}$ is set equal to 0. Figure 6 shows $\hat{\kappa}_t^{\text{trend}}$ for the training and testing windows.
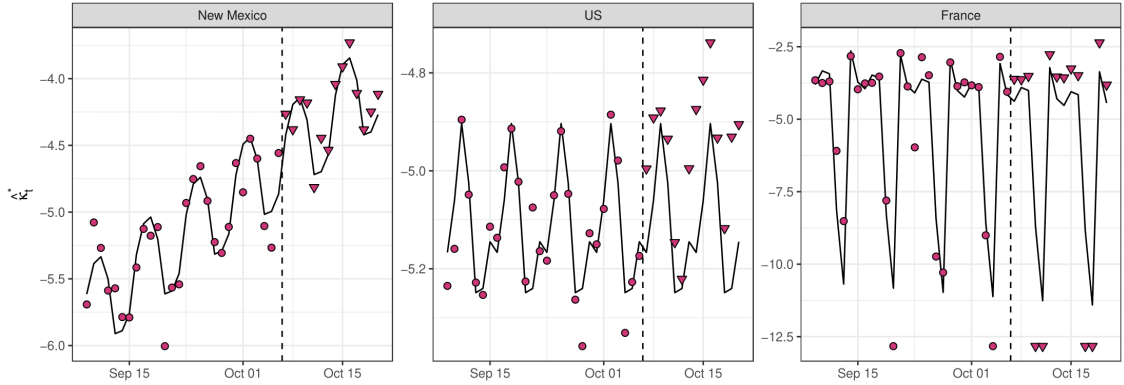


Figure 6: The quantities $\hat{\kappa}_t^*$ for New Mexico, the US, and France. Circles are training data, left of the vertical dashed line, while testing data are the triangles to the right of the dashed vertical line. Solid line represents the fits (train) and predictions (test) of $\hat{\kappa}_t^{\text{trend}}$ based on the regression.

**Step 6: Compute $\hat{\kappa}_t^{\text{constant}}$**

There is a $\hat{\kappa}_t$ trajectory corresponding to a constant new number of cases day over day. Let

$$\bar{y}_{c,T^{\text{train}}} = \frac{1}{7} \sum_{t=T^{\text{train}}-6}^{T^{\text{train}}} y_{c,t} \tag{11}$$

be the average number of daily reported cases over the last week of the training window.

Then

$$\hat{\kappa}_t^{\text{constant}} = \text{logit}\left( \bar{y}_{c,T^{\text{train}}} \left[ \left( \frac{\delta_{s,0} - \ddot{y}_{c,t-1}}{\delta_{s,0}} \right) \ddot{y}_{c,t-1} \right]^{-1} \right) \tag{12}$$

where we set $\delta_{s,0} = 0.55N$, where $N$ is the population of the forecasted region and 0.55 is a nominal attack rate for COVID-19. Figure 7 shows $\hat{\kappa}_t^{\text{constant}}$ for $t = T^{\text{train}} + k$ and $k \in 1, 2, \ldots, 14$.
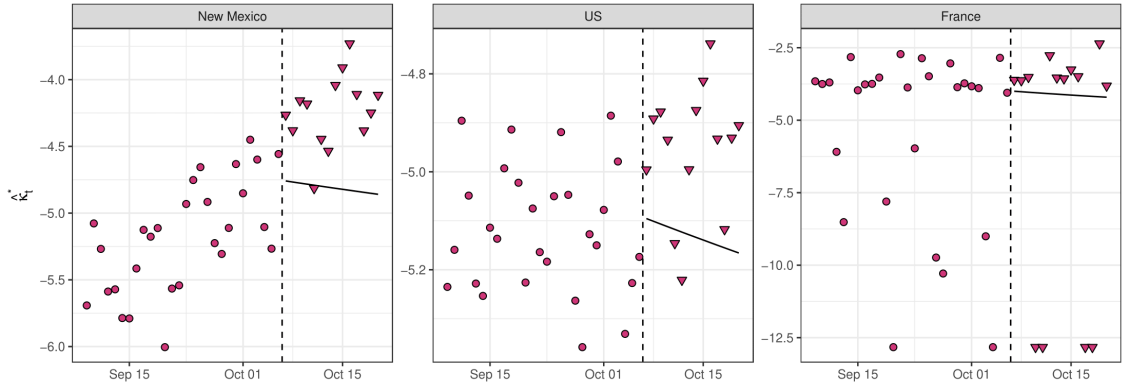


Figure 7: The quantities $\hat{\kappa}_t^*$ for New Mexico, the US, and France. Circles are training data, left of the vertical dashed line, while testing data are the triangles to the right of the dashed vertical line. Solid line represents the values of $\hat{\kappa}_t^{\text{constant}}$, the trajectory corresponding to a constant number of new reported cases, equal to $\bar{y}_{c,T^{\text{train}}}$.

**Step 7: Compute a Joint Probability Distribution over Tuning Parameters**

The form of the forecasting model for $\hat{\kappa}_t^{\text{forecast}}$ is

$$\hat{\kappa}_t^{\text{forecast}}(\eta, \omega, \phi) = \lambda_t(\phi)[w_t \min(\eta^*, \hat{\kappa}_t^{\text{trend}}) + (1 - w_t)\hat{\kappa}_t^{\text{constant+DOW}}] \tag{13}$$

for $t = T^{\text{train}} + k$ and $k \in 1, 2, \ldots, 14$. The objective is to produce forecasts for $\hat{\kappa}_t$. The way COFFEE does this is be creating a blended combination of $\hat{\kappa}_t^{\text{trend}}$ and $\hat{\kappa}_t^{\text{constant+DOW}}$ where

$$\hat{\kappa}_t^{\text{constant+DOW}} = \hat{\kappa}_t^{\text{constant}} +$$
$$\hat{\beta}_2 \text{I}(t = \text{Monday}) + \hat{\beta}_3 \text{I}(t = \text{Tuesday}) + \ldots + \hat{\beta}_7 \text{I}(t = \text{Saturday}), \tag{14}$$

which is $\hat{\kappa}_t^{\text{constant}}$ with the DOW effects estimated in Equation 10 added.

There are three tuning parameters in Equation 13, each playing a role in controlling the form of $\hat{\kappa}_t^{\text{forecast}}$.

The first tuning parameter $\eta$ puts a cap on how large $\hat{\kappa}_t^{\text{trend}}$ can get. A forecast can blow up if $\hat{\kappa}_t^{\text{trend}}$ is growing in an unmitigated fashion. The parameter $\eta$ is a safeguard against this unmitigated growth. We set

$$\eta^* = \text{median}(\{\hat{\kappa}_{T^{\text{train}}-6}^*, \hat{\kappa}_{T^{\text{train}}-5}^*, \ldots, \hat{\kappa}_{T^{\text{train}}}^*\})\eta \tag{15}$$

for $\eta \in [0,1]$.

The second tuning parameter is $\omega$. The basic form of the forecasting model is to transition from a forecast that relies on the current trend $\hat{\kappa}_t^{\text{trend}}$ to a forecast that relies on $\hat{\kappa}_t^{\text{constant+DOW}}$. If $\hat{\kappa}_t^{\text{trend}}$ is trending up, this transition keeps the forecasts from blowing up. If $\hat{\kappa}_t^{\text{trend}}$ is trending down, this transition keeps the forecasts from flat-lining at 0 new cases. The assumption behind this modeling choice is that, as cases are going up, people will take action to curb the growth of the pandemic, either through independent choices of personal responsibility or governmental policies. As cases are going down, however, we assume policies will be relaxed or people will become more comfortable engaging in activities that will increase transmission pathways. The tuning parameter $\omega \geq 1$ determines how quickly the forecast transitions from $\hat{\kappa}_t^{\text{trend}}$ to $\hat{\kappa}_t^{\text{constant+DOW}}$. The closer $\omega$ is to 0, the quicker the transition occurs.

$$w_{T^{\text{train}}+k} = \begin{cases} 1 - \left(\frac{k-1}{\omega}\right)^2 & \text{if } k \leq \omega + 1 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $k$ is a positive integer. Figure 8 shows weight trajectories $w_t$ for various choices of $\omega$. When $w_{T^{\text{train}}+k} = 1$, all weight is on $\hat{\kappa}_t^{\text{trend}}$; when $w_{T^{\text{train}}+k} = 0$, all weight is on $\hat{\kappa}_t^{\text{constant+DOW}}$.

The third tuning parameter is $\phi$. The trajectory $\lambda_t$ is defined as

$$\lambda_{T^{\text{train}}+k} = 1 + k\frac{\phi - 1}{30}, \tag{17}$$

a linear trend starting at 1 when $k = 0$. The tuning parameter $\phi > 0$ determines whether $\lambda_{T^{\text{train}}+k}$ trends up ($\phi > 1$) or down ($\phi < 1$). Examples of $\lambda_{T^{\text{train}}+k}$ are shown in Figure 9.

For a combination of $\eta$, $\omega$, and $\phi$, we compute $\hat{\kappa}_t^{\text{forecast}}(\eta, \omega, \phi)$ and compute the inverse-distance between the inverse-logit of $\hat{\kappa}_t^{\text{forecast}}(\eta, \omega, \phi)$ and $\hat{\kappa}_t$ over the test period:

$$d^{-1}(\eta, \omega, \phi) = \left(\sum_{t=T^{\text{train}}+1}^{T^{\text{train}}+14} \left[\text{logit}^{-1}\left(\hat{\kappa}_t^{\text{forecast}}(\eta, \omega, \phi)\right) - \hat{\kappa}_t\right]^2\right)^{-1}. \tag{18}$$

Finally we compute a joint probability distribution over $\eta$, $\omega$, and $\phi$ as the normalized inverse-distance. The probability distributions for New Mexico, the US, and France are shown in Figure 10.

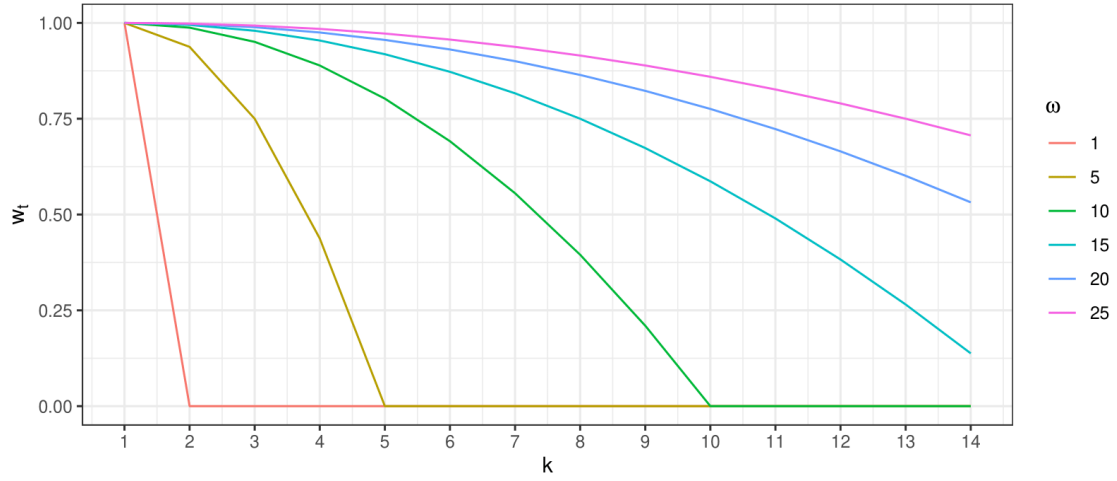Figure 8: Weight trajectories $w_{T^{\mathrm{train}}+k}$ for different values of $\omega$. The smaller $\omega$ is, the quicker $w_{T^{\mathrm{train}}+k}$ transitions from 1 where all weight is assigned to $\hat{\kappa}^{\mathrm{trend}}$ to 0 where all weight is assigned to $\hat{\kappa}_t^{\mathrm{constant+DOW}}$.
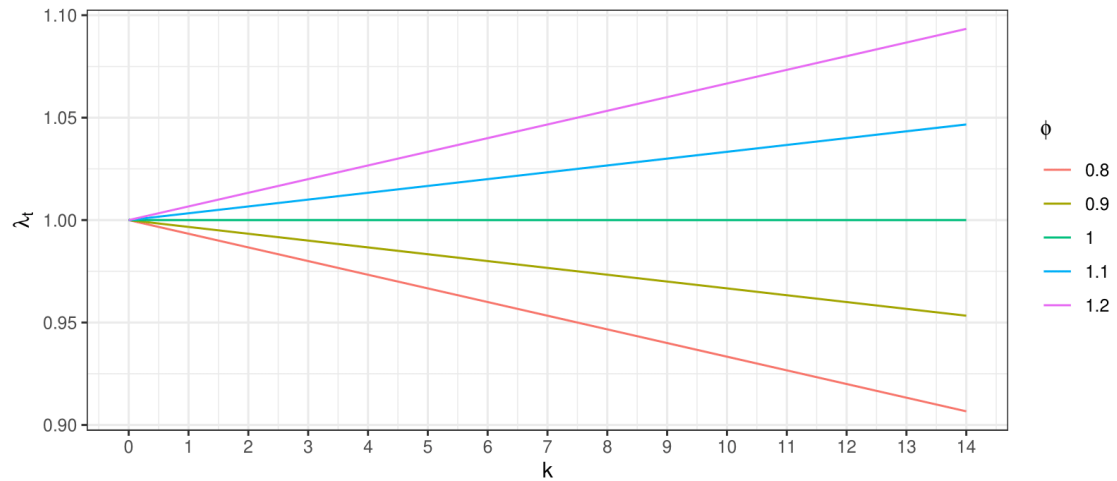


Figure 9: The trajectories $\lambda_{T^{\mathrm{train}}+k}$ for different values of $\phi$. For $\phi$ less than 1, $\lambda_{T^{\mathrm{train}}+k}$ tilts the blended combination of $\hat{\kappa}_t^{\mathrm{trend}}$ and $\hat{\kappa}_t^{\mathrm{constant+DOW}}$ up. When $\phi$ is greater than 1, $\lambda_{T^{\mathrm{train}}+k}$ tilts it down.

### Step 8: Produce the Reported Cases Forecast

The final step is to simulate reported cases. The purpose of the previous steps was to get a joint probability distribution over the tuning parameters that can be used to sample from.

If more than 14 of the last 28 days had zero reported cases, we take independent and identically distributed (iid) samples of future reported cases from the empirical distribution of outlier adjusted reported cases over the last 28 days. If no cases were reported over the
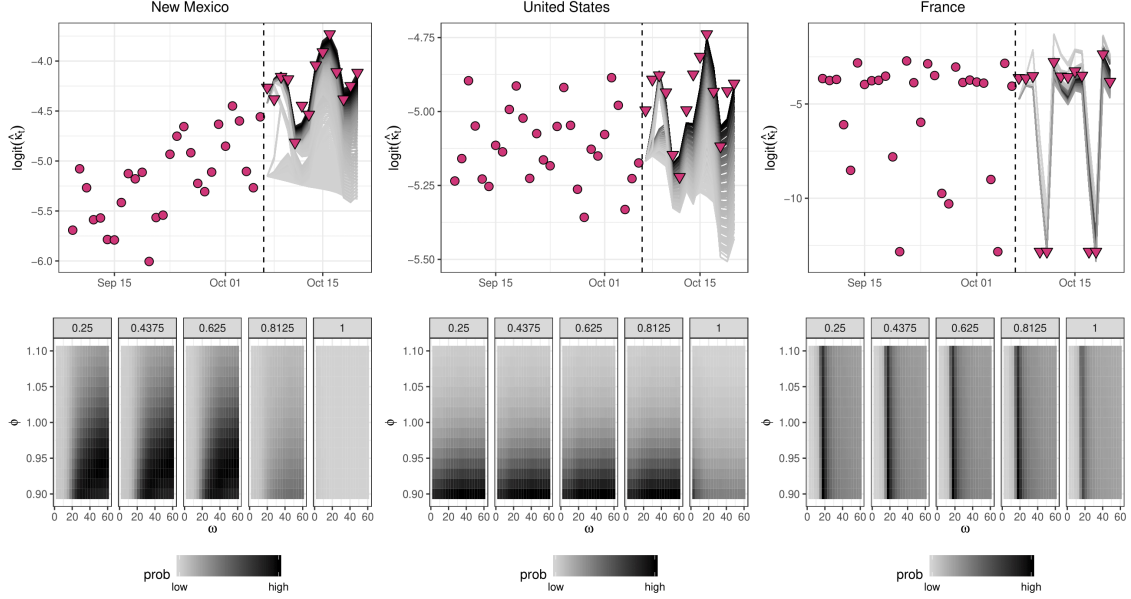
Figure 10: (Top) $\hat{\kappa}_t^*$ for the training period (circles) and testing period (triangles). Lines in the testing period are $\hat{\kappa}_t^{\text{forecast}}$. Each line corresponds to a combination of $\eta$, $\omega$, and $\phi$. The color of the line is proportional to $d^{-1}(\eta, \omega, \phi)$ with darker lines corresponding to better agreement between $\hat{\kappa}_t^{\text{forecast}}$ and $\hat{\kappa}_t^*$ in the test set. (Bottom) The normalized inverse-distance values for $\omega$ (x-axis), $\phi$ (y-axis), and $\eta$ (panels). Darker tiles correspond to larger inverse-distances.

last 28 days, we sample future reported cases as iid Bernoulli draws with success probability equal to $1/29$. If 14 or more of the last 28 days observed at least 1 reported case, we simulate $\hat{\kappa}_{T+k}^{\text{forecast}}$ for $k \in 1, 2, ..., K$ by doing the following:

1. Do Step 5, treating the training data as days $T$ to $T - 27$, resulting in a fitted linear model in the form of Equation 9. Use this to compute $\hat{\kappa}_{T+k}^{\text{trend}}$.

2. Do Step 6 to compute $\hat{\kappa}_{T+k}^{\text{constant+DOW}}$, replacing $T^{\text{train}}$ with $T$ in Equations 11 and 12.

3. Draw a vector of $(\omega, \phi, \eta)$ from the joint distribution computed in Step 7.

4. Compute $\hat{\kappa}_{T+k}^{\text{forecast}}$ following Equation 13.

5. Compute $\text{logit}^{-1}(\hat{\kappa}_{T+k}^{\text{forecast}})$.

6. Draw an attack rate $p \sim \text{Uniform}(0.4, 0.7)$ and set $\delta_{s,0}^{\text{forecast}} = pN$.

7. Set $\ddot{\delta}_{c,T} = \ddot{y}_{c,T}$ and $\delta_{s,T} = \delta_{s,0}^{\text{forecast}} - \ddot{y}_{c,T}$

8. For $k = 1, 2, \ldots, K$, compute

   (a) $\delta_{c,T+k}^{\text{forecast}} = \text{logit}^{-1}(\hat{\kappa}_{T+k}^{\text{forecast}}) \left( \frac{\delta_{s,T+k-1}^{\text{forecast}}}{\delta_{s,0}^{\text{forecast}}} \right) \ddot{\delta}_{c,T+k-1}^{\text{forecast}}$

   (b) $\ddot{\delta}_{c,T+k}^{\text{forecast}} = \ddot{\delta}_{c,T+k-1}^{\text{forecast}} + \delta_{c,T+k}^{\text{forecast}}$

   (c) $\delta_{s,T+k}^{\text{forecast}} = \delta_{s,T+k-1}^{\text{forecast}} - \delta_{c,T+k}^{\text{forecast}}$

9. Draw $y_{c,t}^{\text{forecast}} | \delta_{c,T+k}^{\text{forecast}}, \hat{\alpha} \sim \text{NB}\left( \delta_{c,T+k}^{\text{forecast}}, \frac{\delta_{c,T+k}^{\text{forecast}}}{\hat{\alpha}} \right)$ where $\hat{\alpha}$ is the maximum likelihood estimate.

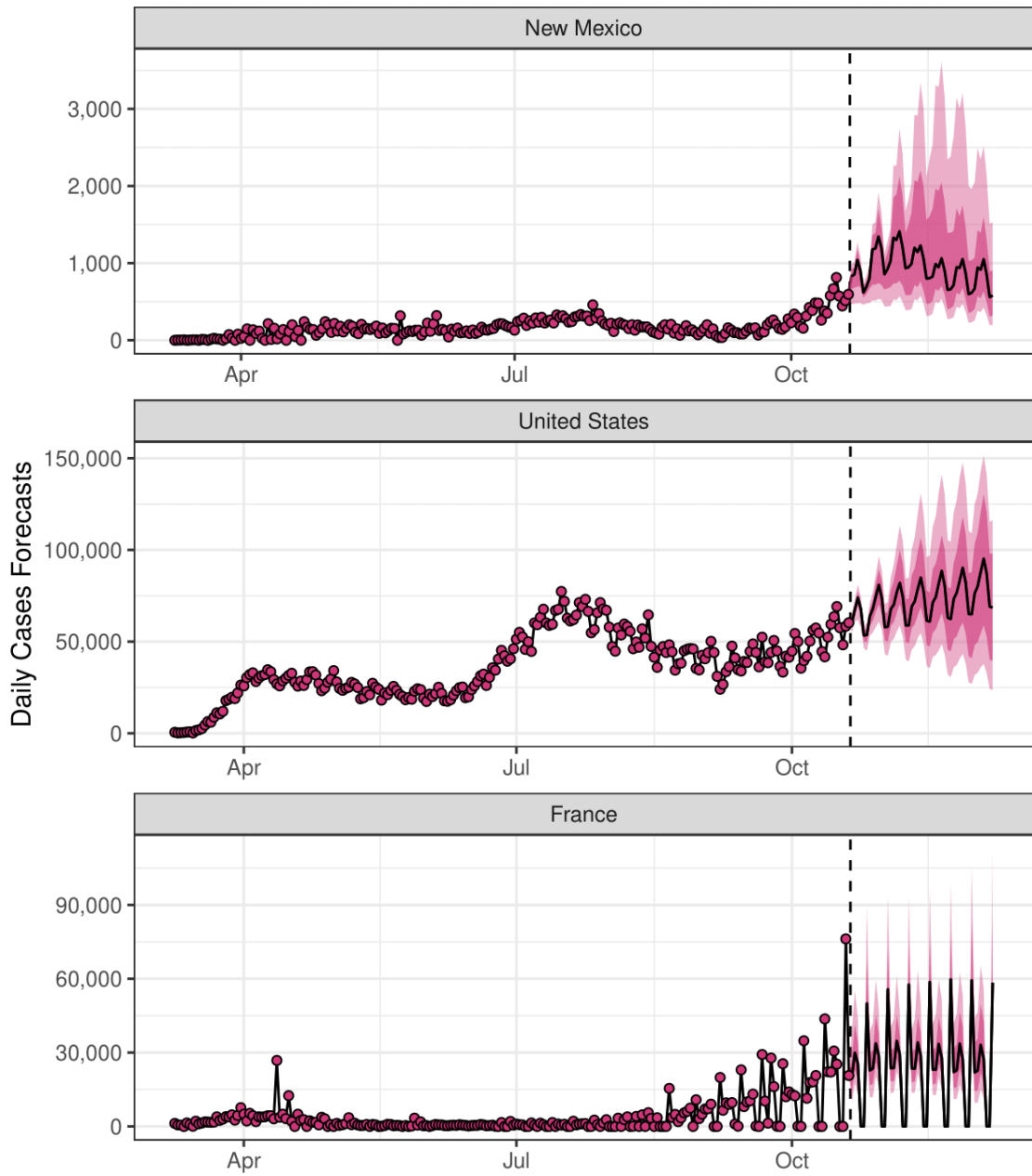Figure 11 shows the forecasts for New Mexico, the US, and France.



Figure 11: The median (black line) and 50% and 80% prediction intervals (ribbons) for New Mexico, the US, and France for daily reported cases.

# Deaths Model

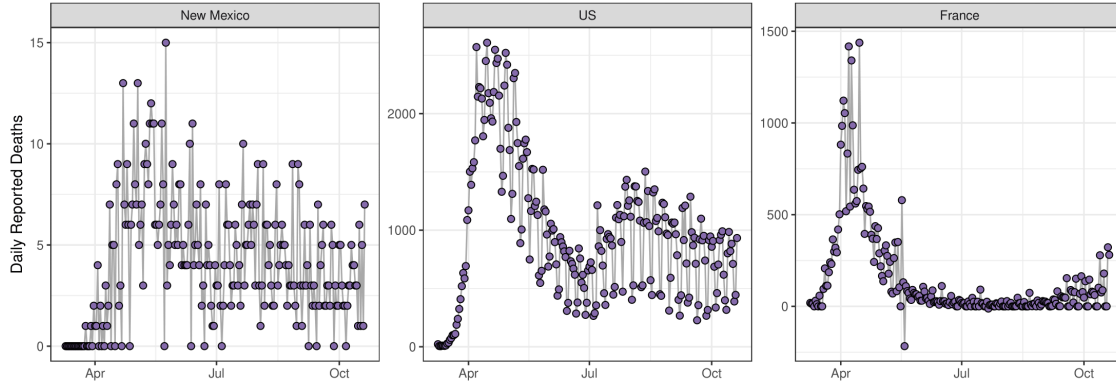Figure 12 shows the daily deaths for New Mexico, the US, and France.



Figure 12: Daily reported deaths for New Mexico, the US, and France.

The COFFEE deaths model is

$$\delta_{d,t} = \gamma_t f(\boldsymbol{\delta}_{c,1:t}, \nu) \tag{19}$$

where $\gamma_t$ is the case fatality ratio and $f(\boldsymbol{\delta}_{c,1:t}, \nu)$ is a moving average of $\boldsymbol{\delta}_{c,1:t}$ with window size equal to $\nu$:

$$f(\boldsymbol{\delta}_{c,1:t}, \nu) = \frac{1}{\nu} \sum_{j=t-\nu+1}^{t} \delta_{c,j}. \tag{20}$$

The deaths model proceeds with the following steps.

## Step 1: Identify and Adjust Outliers

COFFEE uses the same outlier identification and adjustment routine as with cases, resulting in outlier adjusted deaths which are used for all subsequent forecasting steps. The outlier adjusted deaths are shown in Figure 13.

## Step 2: Compute the Case Fatality Ratio, $\hat{\gamma}_t$

We estimate $\gamma_t$ by rearranging Equation 19 and replacing $\delta_{d,t}$ with $y_{d,t}$ and $\boldsymbol{\delta}_{c,1:t}$ with $\boldsymbol{y}_{c,1:t}$ for $t \leq T$:

$$\hat{\gamma}_t = y_{d,t}/f(\boldsymbol{y}_{c,1:t}, \nu). \tag{21}$$

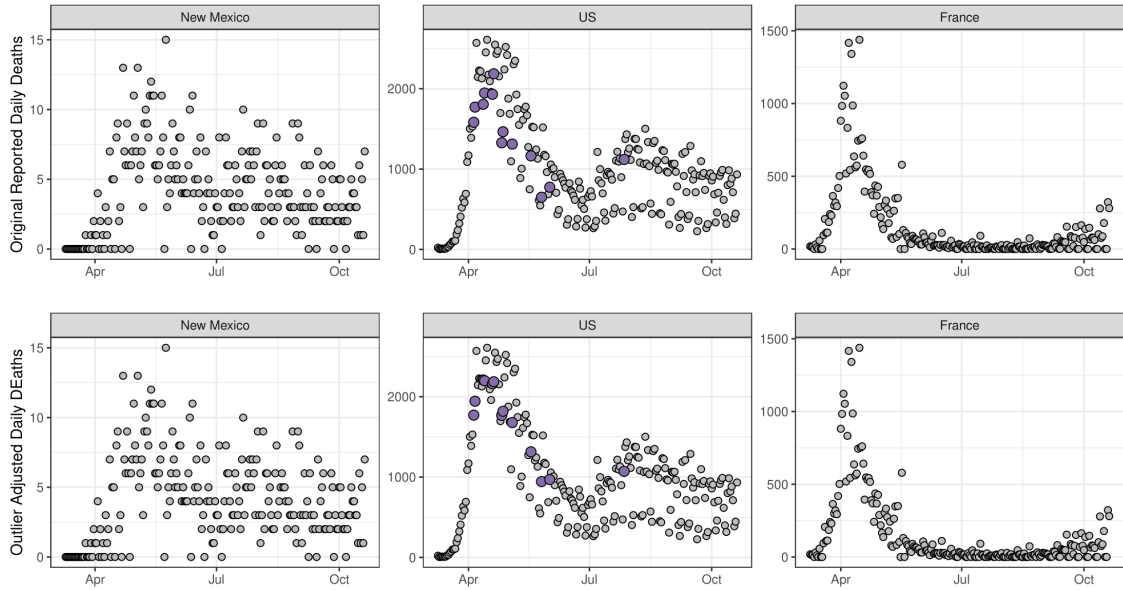Figure 14 displays $\hat{\gamma}_t$ for $\nu \in 7, 14, 21, 28, 35$.

Figure 13: The originally reported daily deaths of COVID-19 (top) and the outlier adjusted daily deaths (bottom). (Top) Purple points were identified as outliers. (Bottom) Purple points are the adjusted outliers.
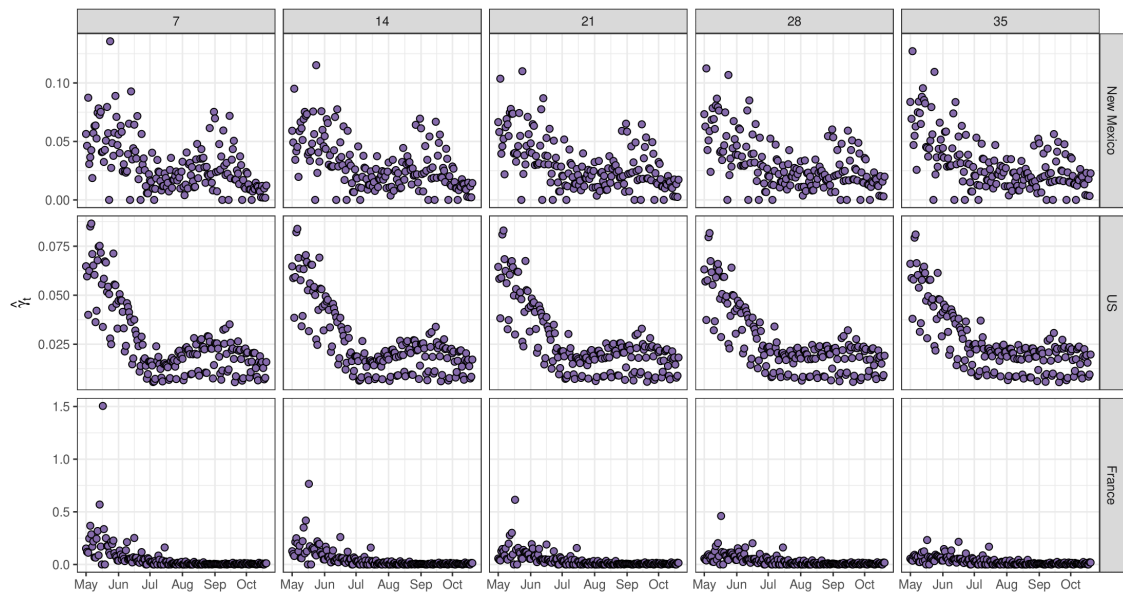


Figure 14: The values $\hat{\gamma}_t$ for New Mexico, the US, and France (rows) for different moving average window sizes of $\nu$ (columns).

### Step 3: Compute $\hat{\gamma}_t^* = \text{logit}(\hat{\gamma}_t)$

COFFEE logit transforms $\hat{\gamma}_t$, setting all values of $\hat{\gamma}_t < \tau_d$ equal to $\tau_d$ and all values of $\hat{\gamma}_t > 1 - \tau_d$ equal to $1 - \tau_d$ where $\tau_d = 0.95 * \min(\{\hat{\gamma}_t | \hat{\gamma}_t > 0\})$. The logit transformed $\hat{\gamma}_t$ are
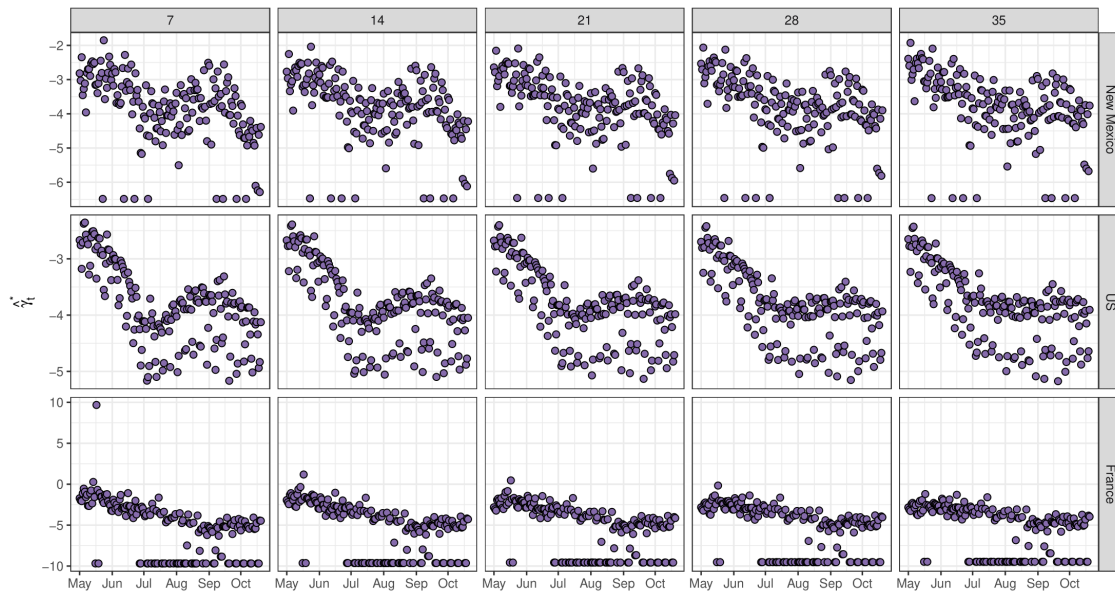
13

shown in Figure 15.



Figure 15: The values of $\hat{\gamma}_t^*$ for New Mexico, the US, and France (rows) for different values of $\nu$ (columns).

### Step 4: Split Data into Training and Testing Sets

Split $\hat{\gamma}_t^*$ in a training and testing data set, same as with the cases model.

### Step 5: Compute $\hat{\gamma}_t^{\mathbf{trend}}$

Fit a regression model with a linear date term and a DOW effect to $\hat{\gamma}_t^*$, analogous to Equation 9. Variable selection is then performed. The fitted regression and predictions ($\hat{\gamma}_t^{\mathrm{trend}}$) are shown in Figure 16.
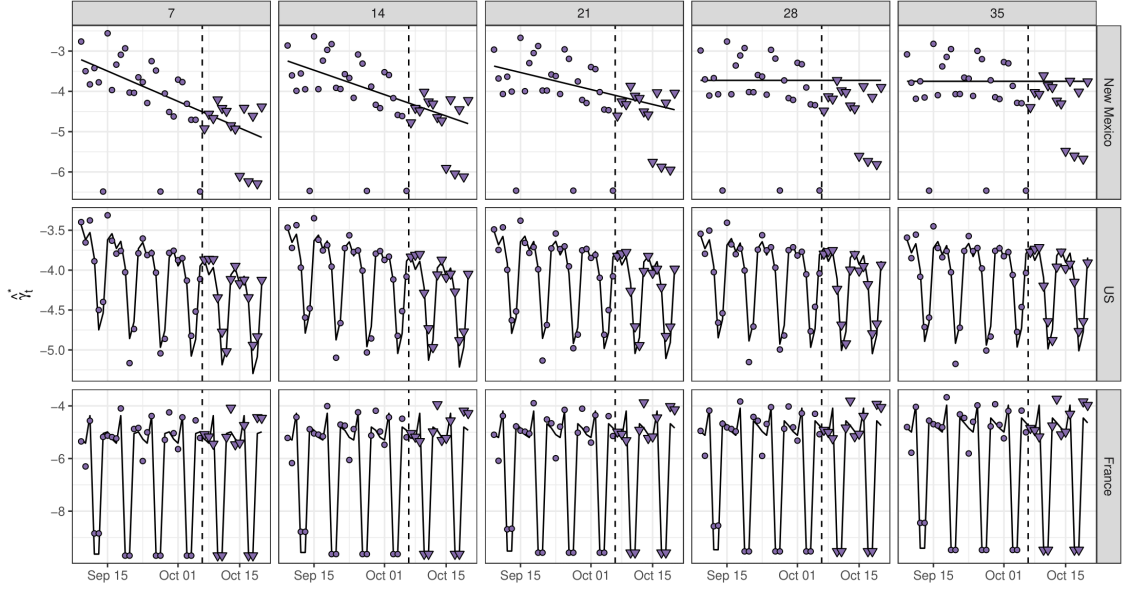
Figure 16: The quantities $\hat{\gamma}_t^*$ for New Mexico, the US, and France. Circles are training data, left of the vertical dashed line, while testing data are the triangles to the right of the dashed vertical line. Solid line represents the fits (train) and predictions (test) of $\hat{\gamma}_t^{\text{trend}}$ based on the regression. The US and France have a DOW effect, while New Mexico had the DOW effect removed in the variable selection phase.

### Step 6: Compute a Joint Probability Distribution over Tuning Parameters

The form of the forecasting model for $\hat{\gamma}_t^{\text{forecast}}$ is

$$\hat{\gamma}_t^{\text{forecast}}(\nu, \theta_{\text{lower}}, \theta_{\text{upper}}) = \begin{cases} \theta_{\text{lower}} & \text{if } \hat{\gamma}_t^{\text{trend}} < \theta_{\text{lower}} \\ \theta_{\text{upper}} & \text{if } \hat{\gamma}_t^{\text{trend}} > \theta_{\text{upper}} \\ \hat{\gamma}_t^{\text{trend}} & \text{otherwise} \end{cases} \tag{22}$$

for $t = T^{\text{train}} + k$ and $k \in 1, 2, \ldots, 14$. The parameters $\theta_{\text{lower}}$ and $\theta_{\text{upper}}$ act as a floor and a ceiling to $\hat{\gamma}_t^{\text{forecast}}$, keeping it from getting too large or too small. We evaluate $\hat{\gamma}_t^{\text{forecast}}$ on a grid over $\nu$, $\theta_{\text{lower}}$, $\theta_{\text{upper}}$ and compute the joint distribution as proportional to the inverse-distance between $\text{logit}^{-1}(\hat{\gamma}_t^{\text{forecast}})$ and $\hat{\gamma}_t$, similar to Equation 18. The estimated joint probability distribution over tuning parameters is shown in Figure 17.

### Step 7: Produce the Reported Deaths Forecast

The final step is to simulate reported deaths. The purpose of the previous steps was to get a joint probability distribution over the tuning parameters that can be used to sample from.

If more than 14 of the last 28 days had zero reported deaths, we take iid samples of future reported deaths from the empirical distribution of outlier adjusted reported deaths over the
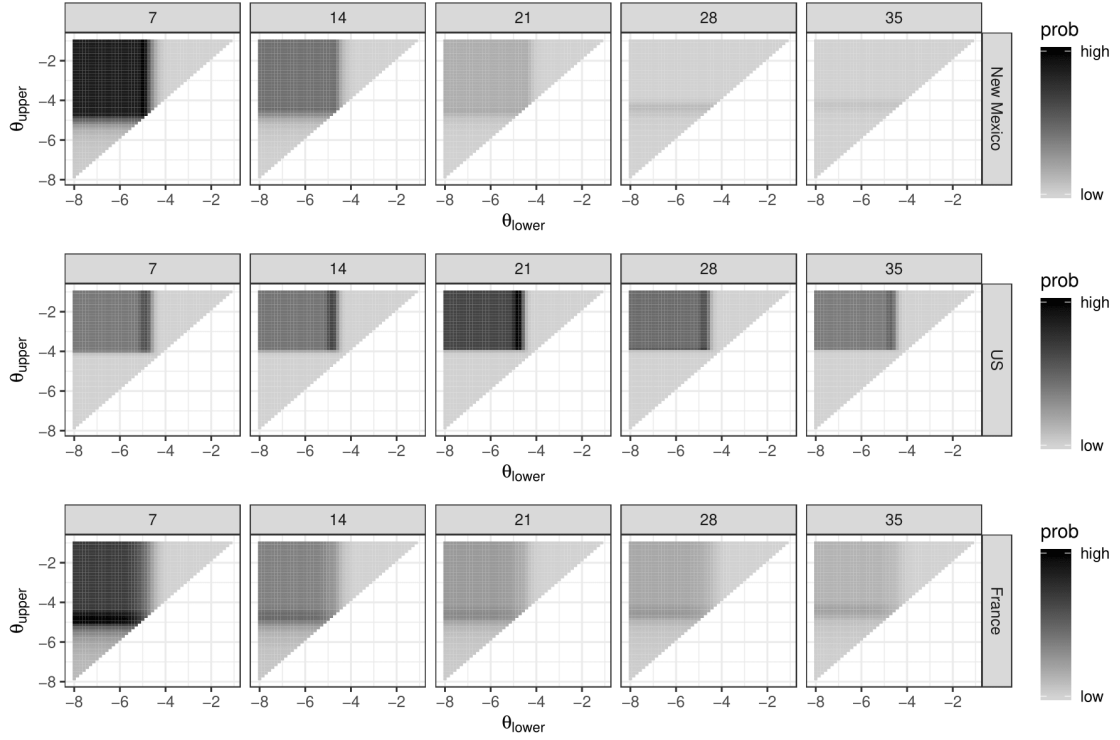
15

Figure 17: Joint probability distributions over tuning parameters $\nu$ (columns), $\theta_{\mathrm{upper}}$, and $\theta_{\mathrm{lower}}$ for New Mexico, the US, and France (rows).

last 28 days. If no deaths were reported over the last 28 days, we sample future reported deaths as iid Bernoulli draws with success probability equal to $1/29$. If 14 or more of the last 28 days observed at least 1 reported death, we simulate $\hat{\gamma}_{T+k}^{\mathrm{forecast}}$ for $k \in 1, 2, ..., K$ by doing the following:

1. Fit the regression outlined in Step 5 to days $T$ to $T - 27$. Use this to compute $\hat{\gamma}_{T+k}^{\mathrm{trend}}$.

2. Draw a vector of $(\nu, \theta_{\mathrm{lower}}, \theta_{\mathrm{upper}})$ from the joint distribution computed in Step 6.

3. Compute $\hat{\gamma}_{T+k}^{\mathrm{forecast}}$ following Equation 22.

4. Compute $\mathrm{logit}^{-1}(\hat{\gamma}_{T+k}^{\mathrm{forecast}})$.

5. For $k = 1, 2, \ldots, K$, compute

   (a) $y_{d,T+k}^{\mathrm{forecast}} = \mathrm{logit}^{-1}(\hat{\gamma}_{T+k}^{\mathrm{forecast}}) f(\boldsymbol{\delta}_{c,1:t}^{\mathrm{forecast}}, \nu)$, where $\delta_{c,t}^{\mathrm{forecast}} = y_{c,t}$ if $k - \nu \leq 0$ and $\delta_{c,t}^{\mathrm{forecast}} = y_{c,t}^{\mathrm{forecast}}$ if $k - \nu > 0$.

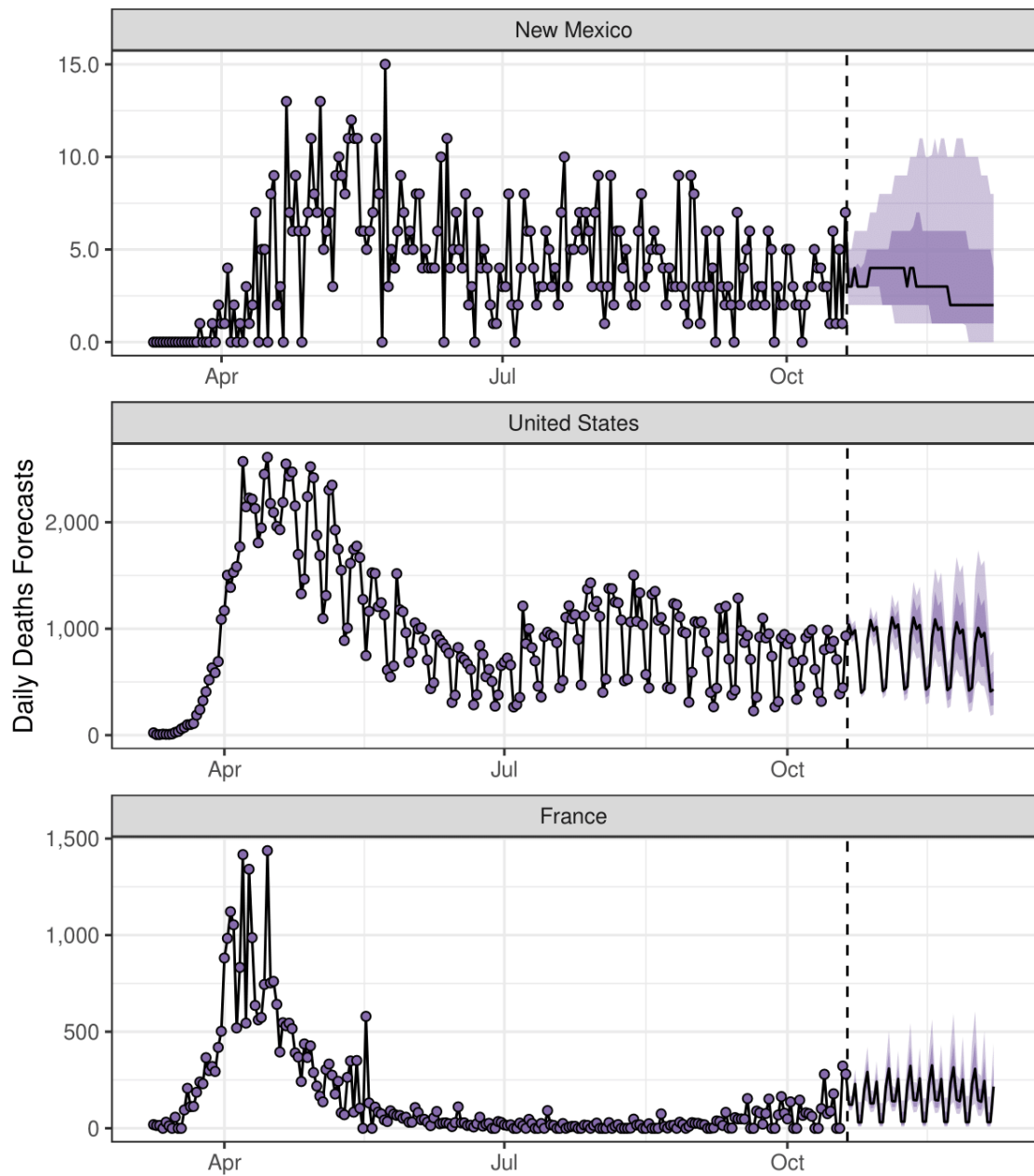Figure 18 shows the daily deaths forecasts for New Mexico, the US, and France.

Figure 18: The median (black line) and 50% and 80% prediction intervals (ribbons) for New Mexico, the US, and France for daily reported deaths.

# References

[1] Jacquez, John A and Simon, Carl P. The stochastic SI model with recruitment and deaths I. Comparison with the closed SIS model. *Mathematical biosciences*, 117(1-2):77–125, 1993.

[2] Javier López-de-Lacalle. *tsoutliers: Detection of Outliers in Time Series*, 2019. R package

version 0.6-8.

[3] Picard, Richard R and Berk, Kenneth N. Data splitting. *The American Statistician*, 44(2):140–147, 1990.