Model for the i^{th} county at time point t:

$$R(t)_i = R_{0i} \times S_i(t)^{z_i} \times \prod_{k=1}^{6} d_{ik}^{\beta_k} \times \prod_{j=1}^{3} c_{ij}(dj)^{\alpha_j(t-dj)} \tag{1}$$

Annotations:

- $R(t)_i$: effective reproductive number at time t for the i^{th} county
- R_{0i}: basic reproductive number at the beginning for the i^{th} county
- $S_i(t)$: proportion of susceptible at time t in the county i: $\frac{S_i(0)-h(t)_i}{S(0)_i}$
- $S_i(0)$: susceptible at the beginning in the county i (assuming all population)
- $h(t)_i$: number of cases up to time t-1
- $d_{i1}...d_{i6}$: 6 fixed variable/effects for the i^{th} county: % of essential workers, humidity, poverty, population density, education, and % over 60 years old
- $c_{i1}...c_{i3}$: 3 time-varying variables/effects for i^{th} county: # of tests per 1000 people, social distancing policy, and mask policy

Taking the log on both side

$$\begin{aligned} log(R(t)_i) =& log(R_{0i}) + z_i log S_i(t) + \sum_{k=1}^{6} \beta_k log(d_{ik}) + \underbrace{\sum_{j=1}^{3} \alpha_j(t-d_j) log(c_{ij}(d_j))} \\ =& log(R_{0i}) + z_i log S_i(t) + \sum_{k=1}^{6} \beta_k log(d_{ik}) + \sum_{j=1}^{3} \sum_{d=t-b_j}^{t} \alpha_j(t-d_j) log(c_{ij}(d_j)) \end{aligned} \tag{2}$$

Interpretation:

- $log(R_{0i})$: county specific intercept
- b_j: # of days we will look back for variable c_{ij}, this will define the time window when the variable could affect $R(t)_i$
- z_i: how quickly susceptibles deplete
- $c_{ij}(t)$: the exposure value for covariate j for county i at time t

We will estimate each $\alpha_j(0), \alpha_j(1)...\alpha_j(d_j)$, j = 1, 2, and 3 with a restricted cubic spline approach with q=5 knots (approximately 3 months, each 2 weeks will be an interval)

$$\sum_{d=t-b_j}^{t} \alpha_j(t-d) log(c_{ij}(d)) = \sum_{d=t-b_j}^{t} \{[\phi_1 B_{-1}(t-d) + ... + \phi_Q B_{Q-2}(t-d)] log(c_{ij}(d))\}$$

Where Q is the number of knots (Q=5)

Step 1: Estimate R_0 for each county in the US
"J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B Biol. Sci. 274, 599–604 (2007)."

$$R(t) = R(today) = \sum_{s=today}^{today+stop} \frac{b(s)g(s-today)}{\sum_{a=0}^{stop} b(s-a)g(a)} \tag{3}$$

Annotation:

- "The reproductive number to the time t at could be find by summing over all possible times t"
- $\sum_{s=t}^{t+stop}$: look ahead starting today through t+stop day
- $R(t)$: the effective reproductive number today (t)
- $b(t)$: incidence number at time t
- $g(t)$: value for general interval distribution at time t (the generation interval distribution is the probability distribution function for the time from infection of an individual to the infection of a secondary case by that individual)
- stop: max generation interval as first day that captures > 99% of the density

Step 2: Estimate time-varying effect by ML using restricted cubic spline
"Wang, Molin, et al. "Quantifying Risk over the Life Course - Latency, Age-Related Susceptibility, and Other Time-Varying Exposure Metrics." Statistics in Medicine, vol. 35, no. 13, Oct. 2016, pp. 2283–2295., doi:10.1002/sim.6864."
This is a function of d_1

$$\alpha_\phi(t-d_1) = \phi_1 B_{-1}(t-d_1) + \phi_2 B_0(t-d_1) + \phi_3 B_1(t-d_1) + \phi_4 B_2(t-d_1) + \phi_5 B_3(t-d_1)$$

This is a function of d_2

$$\alpha_\pi(t-d_2) = \pi_1 B_{-1}(t-d_2) + \pi_2 B_0(t-d_2) + \pi_3 B_1(t-d_2) + \pi_4 B_2(t-d_2) + \pi_5 B_3(t-d_2)$$

This is a function of d_3

$$\alpha_\psi(t-d_3) = \psi_1 B_{-1}(t-d_3) + \psi_2 B_0(t-d_3) + \psi_3 B_1(t-d_3) + \psi_4 B_2(t-d_3) + \psi_5 B_3(t-d_3)$$

Step 3: Estimate fixed effect by OLS

Step 4: Estimate random intercept by BLUP