NC STATE UNIVERSITY

Department of Mathematics & Statistics

COVID-19, Climate and Socioeconomic Status in the United States:

A Bayesian Analysis of County-Level Case Data

Jordan Bramble, Frederick Donahey, Charlie Frazier, Liam Hanson, and Aaron Marshall

DRUMS

NSA and NSF for support through:
NSA: H98230-20-1-0259 and H98230-21-1-0014
NSF: DMS#2051010

We model COVID-19 spread in county *j* during week *t* via an SIR model:

$$S_j(t)\ \text{(Susceptible)} \xrightarrow{\beta_j(t)\frac{S_j(t)I_j(t)}{N_j(t)}} I_j(t)\ \text{(Infected)} \xrightarrow{\gamma_j} R_j(t)\ \text{(Removed)}$$

$\beta_j(t)$ is the county-dependent, time-varying transmission rate and γ_j is the county-dependent removal rate. Transmission rates are assumed to evolve temporally according to an log AR(1) process:

$$\begin{cases} b_j(t) = \log(\beta_j(t)) \\ b_j(t) \sim \text{Normal}(\mu_j, 1),\ t = 1 \\ b_j(t) \sim \text{Normal}(\mu_j + \rho_j(b_j(t-1)-\mu_j), \tau_j^2),\ t > 1 \end{cases}$$

The process is governed by the mean μ_j, variance τ_j^2, and autocorrelation coefficient ρ_j.

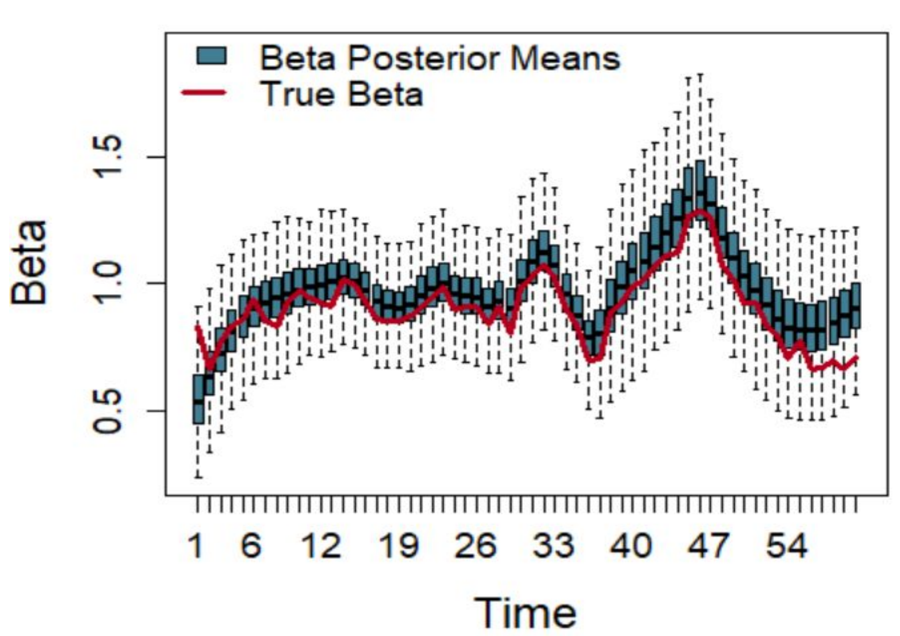
Let $C_{j,\,true}(t) = |\,S_j(t-1) - S_j(t)\,|$ and $C_{j,\,obs}(t)$ be the true and observed numbers of new cases in county j during week t. We assume the following data model:

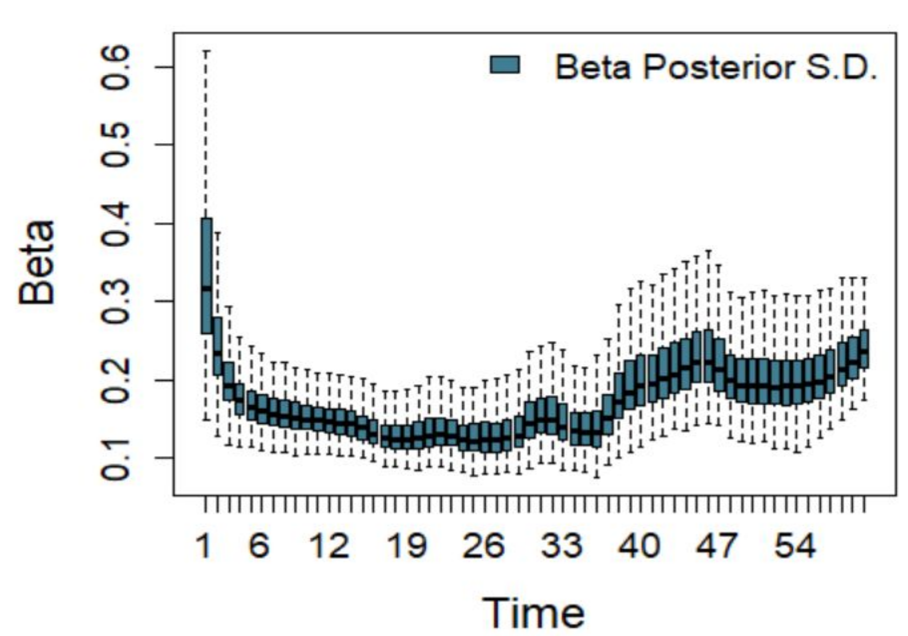
$$C_{j,\,true}(t) \mid C_{j,\,obs}(t) \sim \text{Poisson}(C_{j,\,true}(t-\ell)),\quad \ell = \text{Reporting Lag}$$

Finally, we assume gamma priors for $I_j(0)$, $R_j(0)$, and τ_j; beta priors for γ_j and ρ_j; and a normal prior for μ_j.

- 7 parameter sets, 500 data sets each
- b(t) generated by sampling from AR(1) processes with differing ρ values
- Obtained posterior means and standard deviations for β(t)
- Calculated coverage, bias, and MSE
- Posterior means corresponding to lower case-counts exhibited greater variability
- Coverage generally high, bias and MSE generally low

Rho (ρ)	Tau² (τ²)	Cov.	S.D.
0.001	0.001	0.999	0.068
0.001	0.1	0.991	0.126
0.2	0.1	0.951	0.136
0.4	0.1	0.993	0.152
0.6	0.1	0.715	0.157
0.8	0.1	0.953	0.178
0.999	0.1	0.821	0.168





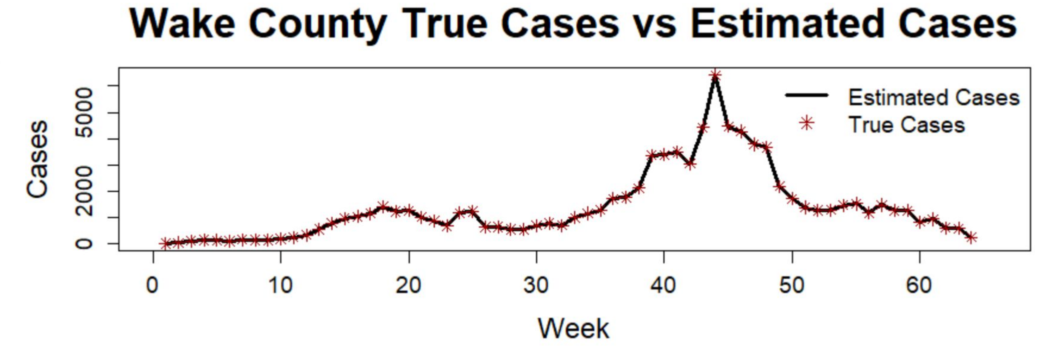
Data Prep

- To remedy daily case count anomalies, data is aggregated by week.
- Data is trimmed to exclude leading weeks with 0 cumulative cases.
- This results in 30-70 timepoints per county.
- Fit via MCMC/JAGS on the henry2 cluster [7].

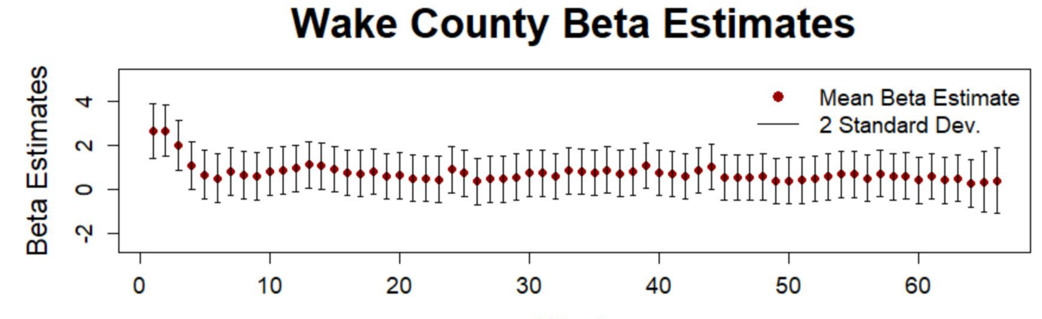
Output

- The model is fit separately to each County, and Bayesian analysis returns Estimates and uncertainty measures for $b_j(t)$.

Wake County True Cases vs Estimated Cases



Wake County Beta Estimates



Results

Good fits were obtained for medium to large counties with weekly case numbers in the high hundreds and upwards. In smaller counties with at most 100 cases per week, poorer fits were obtained at both the starts and ends of outbreaks, but otherwise performance was otherwise acceptable. This is consistent with the model's underperformance for low case numbers in the simulation study.

Background

Differing COVID-19 transmission rates across the U.S. may be partially explained by climate and demographic factors. COVID-19 is transmitted via droplets and possibly aerosols, which may be influenced by temperature and relative humidity [1,2]. Healthcare access, crowding level, and employment type - which relate to poverty rate, percent non-white ethnic groups, and number of ICU beds - may also influence COVID-19 transmission [3,4,5]. Furthermore, impoverished individuals are more likely to have greater environmental exposure due to their occupations [6]. Our goal is to examine these factors in relation to COVID-19 transmission and determine if climate variables have varying effects across different social groups.

Formulate Bayesian SIR Model
↓
Perform Simulation Study
↓
Fit Bayesian SIR Model to County-Level Case Data
↓
Formulate Bayesian Regression Model
↓
Perform Regression Analysis

Conclusions

- All effects, barring interactions between ICU beds & temperature and ICU beds & relative humidity, were significant.
- All significant interaction effects were positive.
- Counties w/ high African-American or Hispanic proportions, high median age, high population density, or low poverty rate had higher transmission rates.
- Counties w/ low African-American or Hispanic proportions, low median age, low population density, or low poverty rate were more strongly affected by temperature changes.

Limitations & Looking Forward

- Issues with parameter convergence prevented use of higher-dimensional compartmental models such as SEIR.
- While state and month indicators may crudely capture large-scale dynamics, state policy and mobility data would more accurately capture these trends.
- We plan to incorporate neighboring county and air pollution effects in the future.

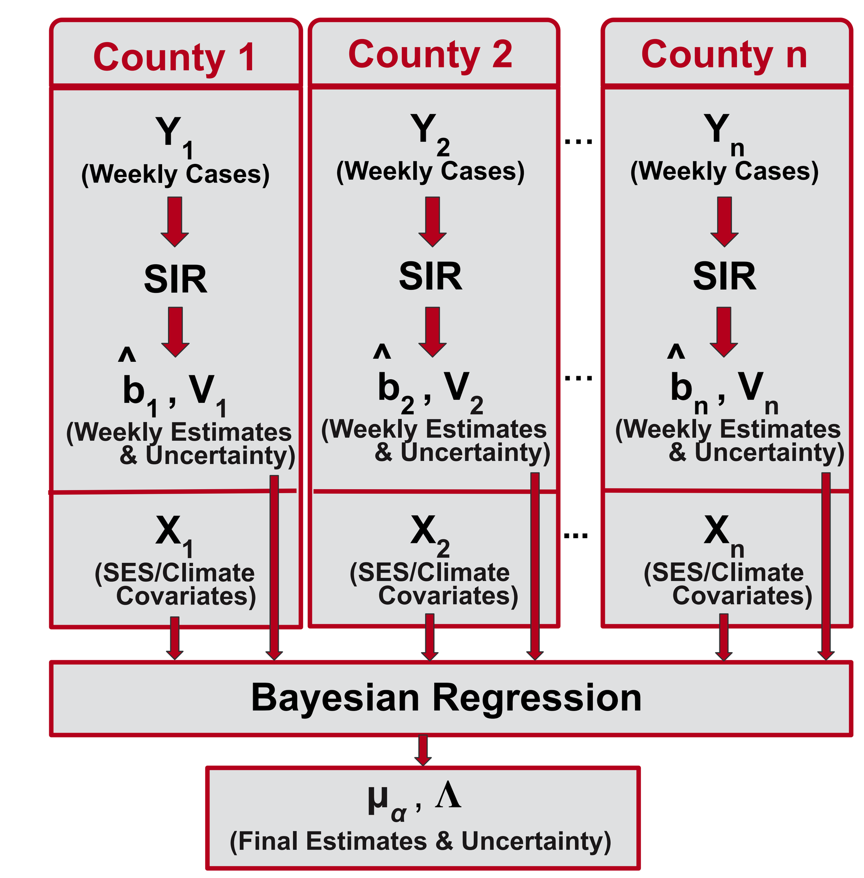
In the next stage, we combine the posterior means $\hat{b}_j = (\hat{b}_j(1),...,\hat{b}_j(n_t))$ and covariance V_j and combine these in a Bayesian Linear Regression Model

$$\hat{b}_j \sim \text{Normal}(X_j\alpha, V_j)$$

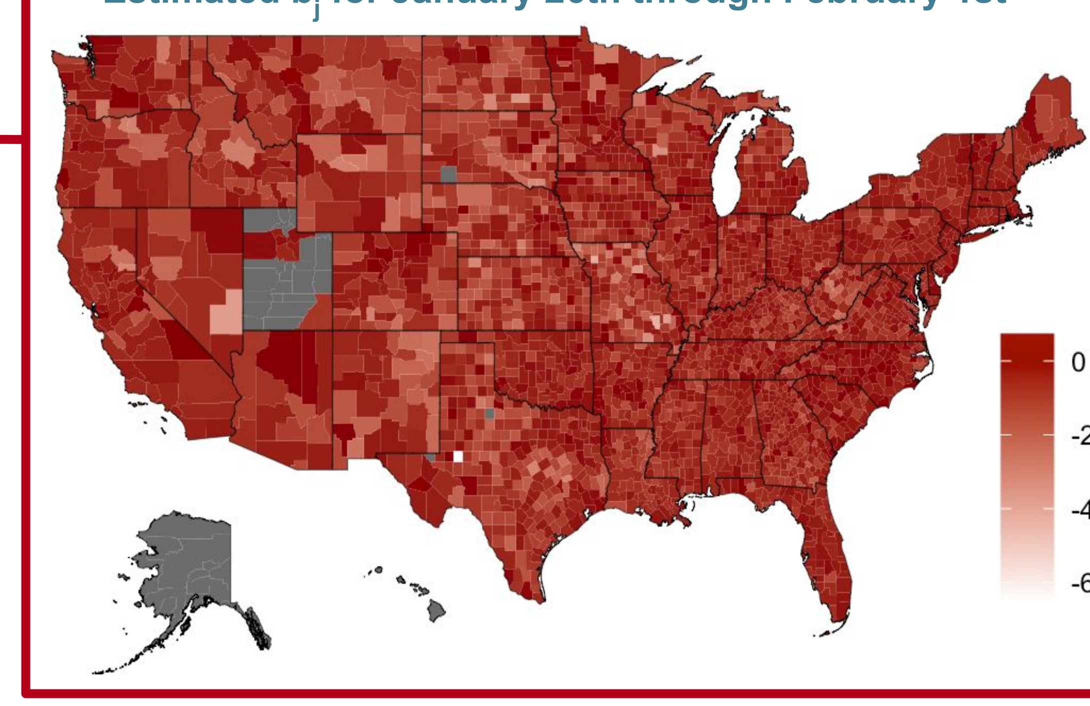
and prior

$$\alpha \sim \text{Normal}(0, 100 \cdot I_p)$$

For covariates X_j, we include socioeconomic and climate variables (listed below) as well as state and month indicators to capture large scale spatiotemporal trends. We obtain the posterior means, standard deviations, and 95% credible intervals for **α**.



Estimated $\hat{b}_j$ for January 26th through February 1st



Variables

$\hat{b}_j$	Vector of Posterior Means of $b_j(t)$ for ea. Week in County j
V_j	Covariance Matrix of b_j
n_t	Total Number of Weeks
X_j	Covariate Matrix of County j
n_c	Total Number of Counties
α	Vector of Regression Coefficients (Constant Across Counties)

Model Selection

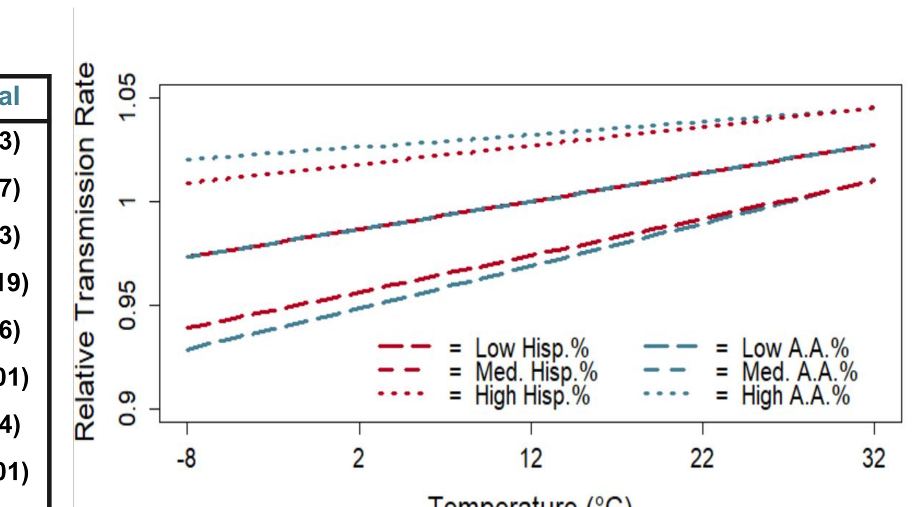
Model	ΔBIC
Baseline	0
Baseline + SES	-39,092
Baseline + CLIM	-16,813
Baseline + SES*CLIM	-87,452
Baseline + SES + CLIM	-55,162
Baseline + SES + SES*CLIM	-144,173
Baseline + CLIM + SES*CLIM	-89,160
*Baseline + SES + CLIM + SES*CLIM*	*-146,961*

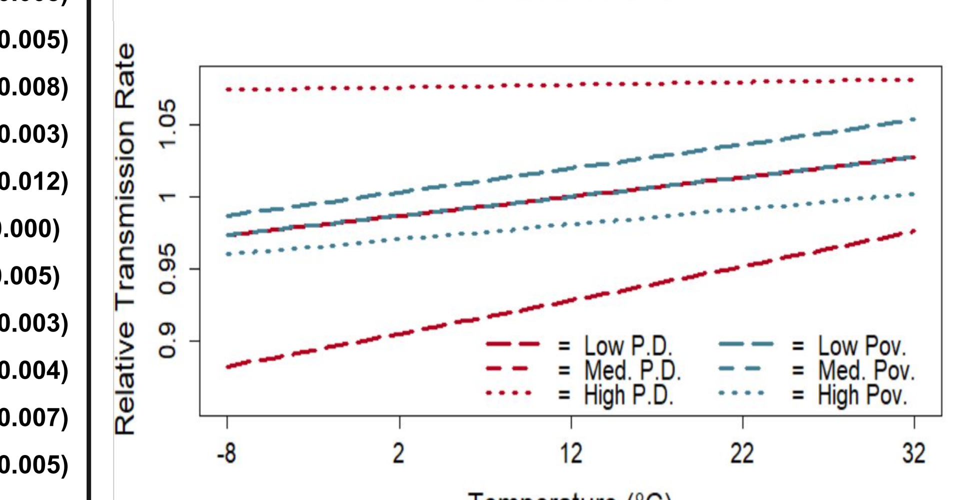
- SES = Socioeconomic Variables
- CLIM = Climate Variables
- Bayesian Information Criterion (BIC) measures fit
- The lower the BIC, the better the fit
- The table to the left gives the difference in BIC for each model and the baseline model with only state and month effects.

Interaction Plots:
For SES variable X, the plot below is the relative risk, RR(X, temp), which is the multiplicative increase in $\beta_j(t)$ compared to the baseline with all other covariates fixed at their mean values.

Alpha Estimates for Best Fitting Model

Covariate	Mean	95% Interval
African-American (A.A.)	0.032	(0.031, 0.033)
Hispanic (Hisp.)	0.026	(0.026, 0.027)
Median Age	0.032	(0.031, 0.033)
Poverty (Pov.)	-0.019	(-0.020, -0.019)
Population Density (P.D.)	0.075	(0.074, 0.076)
ICU Beds	-0.001	(-0.001, -0.001)
Temperature (Temp.)	0.013	(0.013, 0.014)
Relative Humidity (R.H.)	-0.002	(-0.002, -0.001)
A.A.*Temp.	-0.008	(-0.008, -0.008)
Hispanic*Temp.	-0.005	(-0.005, -0.005)
Age*Temp.	-0.008	(-0.008, -0.008)
Poverty*Temp.	-0.003	(-0.003, -0.003)
P.D.*Temp.	-0.012	(-0.012, -0.012)
ICU Beds*Temp.	0.000	(0.000, 0.000)
A.A.*R.H.	0.005	(0.005, 0.005)
Hispanic*R.H.	-0.003	(-0.003, -0.003)
Age*R.H.	-0.004	(-0.005, -0.004)
Poverty*R.H.	-0.008	(-0.008, -0.007)
P.D.*R.H.	-0.005	(-0.005, -0.005)
ICU Beds*R.H.	0.000	(0.000, 0.000)





1. Song Tang, Yixin Mao, Rachael M Jones, Qiyue Tan, John S Ji, Na Li, Jin Shen, Yuebin Lv, Lijun Pan, Pei Ding, et al. Aerosol transmission of sars-cov-2? Evidence, prevention and control
2. CDC. Scientific brief: Sars-cov-2 transmission.
3. Devan Hawkins. Differential occupational risk for covid-19 and other infection exposure according to race and ethnicity.
4. Philip Blumenshine, Arthur Reingold, Susan Egerter, Robin Mockenhaupt, Paula Braveman, and James Marks. Pandemic influenza planning in the united states from a health disparities perspective.
5. Vida Abedi, Oluwaseyi Olulana, Venkatesh Avula, Durgesh Chaudhary, Ayesha Khan, Shima Shahjouei, Jiang Li, and Ramin Zand. Racial, economic, and health inequality and covid-19 infection in the united states.
6. US Bureau of Labor Statistics. Working poor, 2014.
7. https://people.stat.sc.edu/hansont/stat740/jags_user_manual.pdf