Supplement to The Evolving Roles of Partisanship and Vulnerability in the COVID-19 Pandemic

Introduction

These supplemental materials present a conditional auto-regressive (CAR) Poisson model of COVID-19 county death rates in the United States. The inclusion of a conditional auto-regressive spatial component in this model is intended to address the question of how robust the relationships are between observed covariates and COVID-19 death rates given the expected correlations in rates between counties which neighbor one another.

Methods

The implementation of our model is based on the Stan Case Study, *Exact Sparse CAR Models in Stan* (Joseph 2016). Spatial models including CAR models and improvements on the Besag-York-Mollié model have often been used in current epidemiology and disease risk mapping applications to distinguish spatially structured effects from the effects of observed covariates [Lee, Rushworth, and Sahu (2014); Wakefield (2007); Morris et al. (2019)].

To describe our model, we write the deaths observed as $y_1, y_2, ..., y_{2683}$ for each of the 2,683 counties which have neighboring counties and for which all covariates were available and hence were considered in our main manuscript. Letting X_i for i in $1 \ldots 2{,}683$ represent the vector of observed covariates for the ith county and similarly P_i represent the population of the ith county, we write that

$$y_i \sim \text{Poisson}(\exp(X_i\beta + \phi_i + \log(P_i))),$$

where β is a vector of the estimated coefficients for the covariates and ϕ_i is the the spatial component of the model. See Joseph (2016) for the details of the prior distributions on β and ϕ. This model is fit twice with data from period 2 and period 3 separately.

Given the computational complexity in fitting these models with over 2600 county observations and a large number of covariates, we opted to only include a selection of the variables which had the highest measures of feature importance in the LASSO and spatial linear models from our main manuscript. In order to include parameters parsimoniously, we chose to include parameters which appeared in the top three most important features from the LASSO and spatial linear models for periods 2 and 3. Since we only modeled the probability of counties being seeded during period 1 and did not model death rates, we have only calibrated the spatial model presented here to the deaths data from periods 2 and 3.

All analyses in these supplemental materials were conducted using R version 4.0.2 (R Core Team 2020) and the model analyses were conducted in the Bayesian statistical computing and modeling framework Stan using the No-U-Turns Hamiltonian Monte Carlo Sampler (Stan Development Team 2021; Homan, Matthew D. and Gelman, Andrew 2014).

Results

Parameter Estimates

We found that the results of fitting a spatial sparse CAR Poisson model were consistent with our findings from the main analyses for periods 2 and 3.

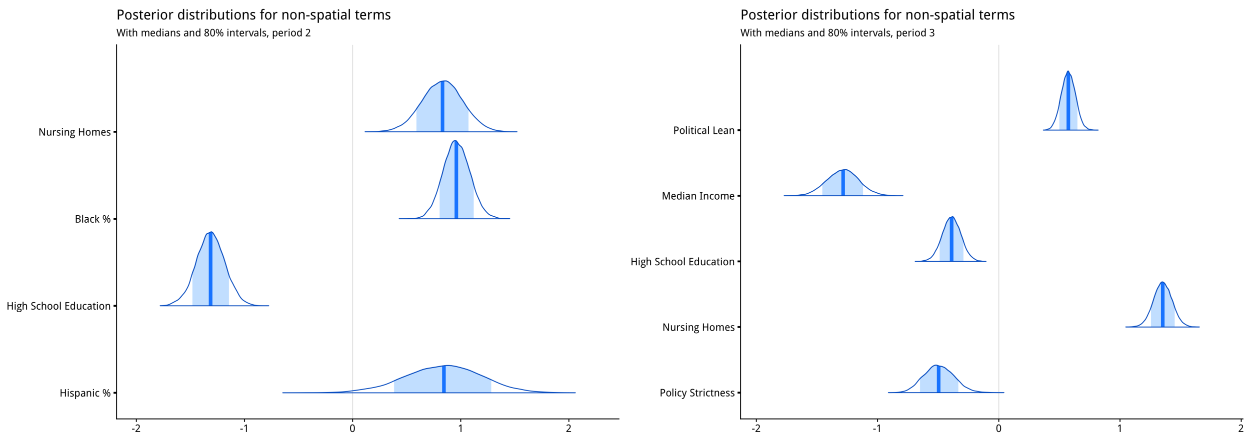

Figure 1: Non-spatial parameter estimates for periods 2 and 3

These model estimates reflect the associations between county-level covariates and COVID-19 death rates after accounting for the estimated spatial correlation structure included in the model.

We found that higher percentages of residents in nursing homes, Black non-Hispanic population percentages, and Hispanic population percentages at the county level were associated with higher COVID-19 death rates during period 2.Counties with higher high school graduation rates were associated with having lower COVID-19 death rates in period 2 after accounting for spatial correlations. During period 3, counties where the population voted more Republican (positive political lean), and counties with higher percentages of the population living in nursing homes were associated with higher COVID-19 death rates. Counties with higher median income, greater high school graduation rates, and policy strictness during period 3 were found to be have lower COVID-19 death rates.

Spatial Components

We present the spatial model component below visualized as $\exp(\phi_i + \log(P_i))$, the expected COVID-19 deaths per capita in each period conditioning out the effects from covariates.

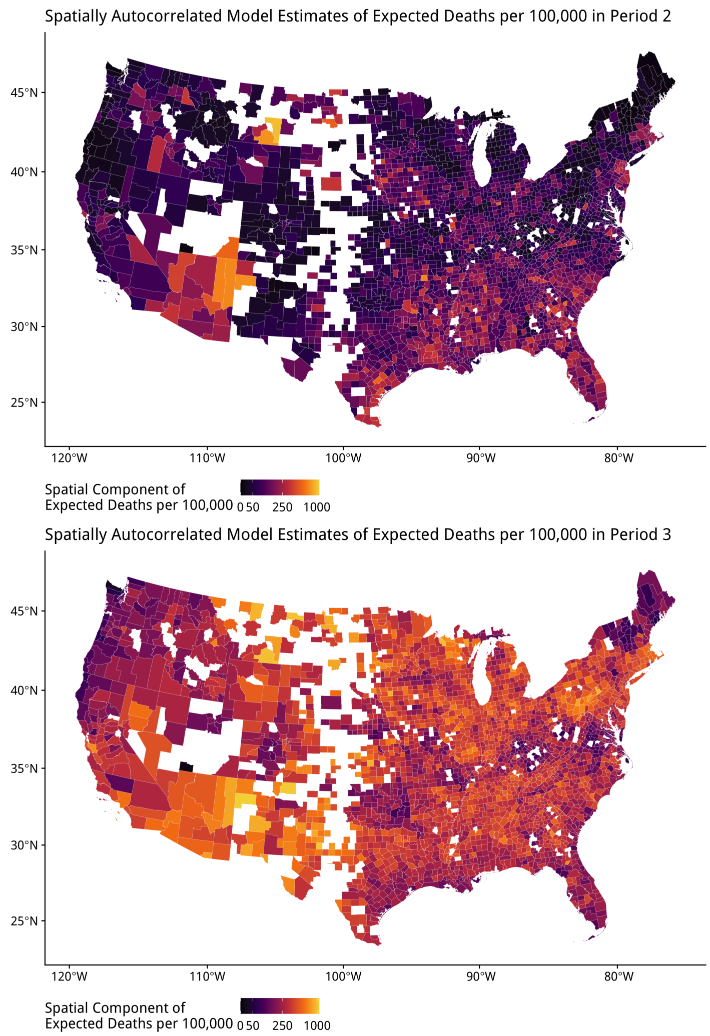

Figure 2: Spatial model component during periods 2 and 3

Supplemental Figure 2 allows us to visualize the estimated spatial correlation structure and how neighboring counties tend to be correlated with one another. In essence, we expect to see that counties which have high rates are surrounded by counties that also have high rates and vice-versa for low rate counties.

Model Convergence Diagnostics

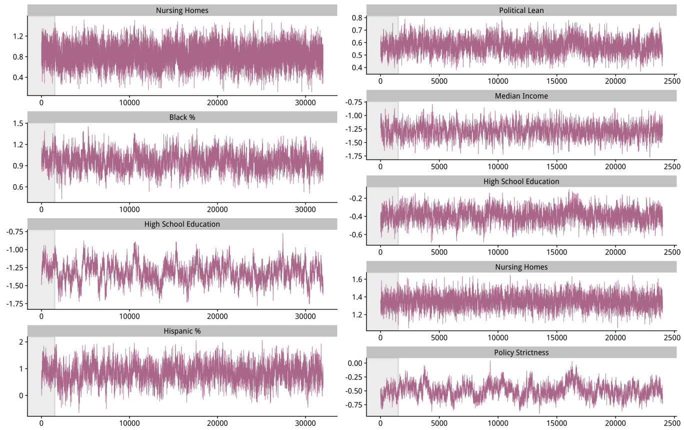

Figure 3: Traceplots for non-spatial parameters

Contemporary advice recommends that Bayesian models should be considered to have converged only if the Markov chains have convergence diagnostics of $\hat{R} < 1.05$ (Stan Development Team 2020; Vehtari et al. 2020). In Supplemental Figure 4 we present the $\hat{R}$ convergence diagnostics for our non-spatial and spatial effects for both periods 2 and 3 which are all below 1.05.

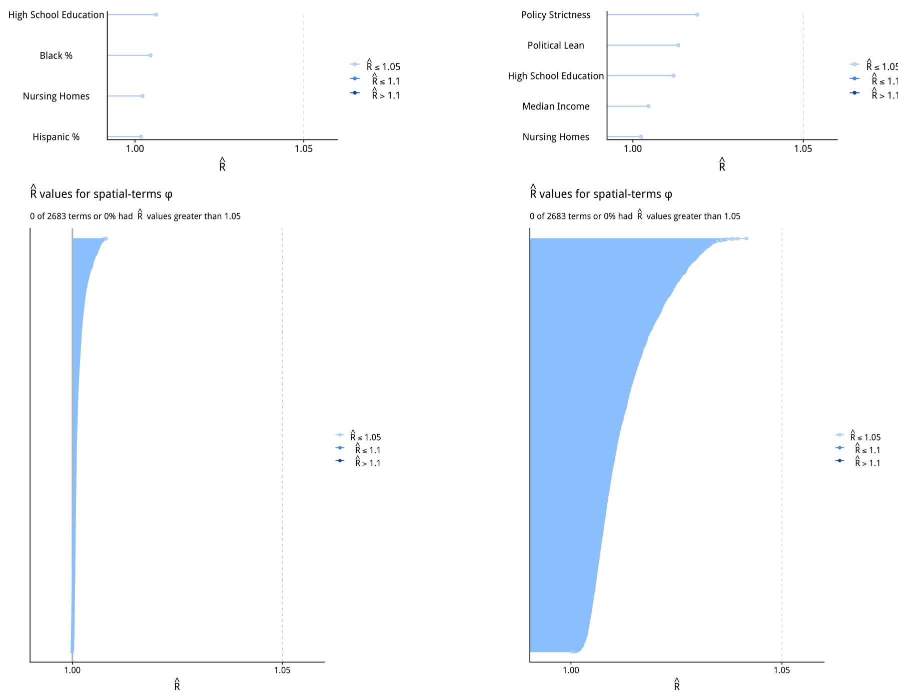

Figure 4: Convergence diagnostics for all parameters

References

Homan, Matthew D. and Gelman, Andrew. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *The Journal of Machine Learning Research* 15 (January). https://dl.acm.org/doi/10.5555/2627435.2638586.

Joseph, Maxwell. 2016. *Exact Sparse CAR Models in Stan. Stan Case Studies*. Vol. 3.

Lee, Duncan, Alastair Rushworth, and Sujit K. Sahu. 2014. "A Bayesian Localized Conditional Autoregressive Model for Estimating the Health Effects of Air Pollution." *Biometrics* 70 (2): 419–29. https://doi.org/10.1111/biom.12156.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stan Development Team. 2020. *Convergence and Efficiency Diagnostics for Markov Chains, Rstan 2.12.2 Documentation*. https://mc-stan.org/rstan/reference/Rhat.html.

———. 2021. *Stan Modeling Language Users Guide and Reference Manual, V2.26*. https://mc-stan.org.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2020. "Rank-Normalization, Folding, and Localization: An Improved $\Widehat{R}$ for Assessing Convergence of MCMC." *Bayesian Analysis*, July. https://doi.org/10.1214/20-BA1221.

Wakefield, Jon. 2007. "Disease Mapping and Spatial Regression with Count Data." *Biostatistics* 8 (2): 158–83. https://doi.org/10.1093/biostatistics/kxl008.