Extending the CAR Model to Account for General Temporal Neighbourhood Structures

Eric Kalendra

SAMSI: Spatial Epidemiology
2009-12-08

1 Introduction

Preterm birth refers to the birth of a baby with a gestational age less than 37 weeks. Preterm birth is the most common cause of prematurity and is the major cause of neonatal mortality in developed countries. Infants who are born premature are at greater risk of many short term and long term adverse health complications. Understanding preterm birth is an important step in understanding premature birth.

I have already preformed an extensive analysis using birth outcome data from New York City. The analysis showed that there is spatial variation, even after allowing each race category to independently vary spatially. Another complicating factor is that the data is from 1995 to 2003, 9 years of data. A reasonable assumption is that New York City is changing over time and that the spatial pattern for the risk of preterm birth from 1995 is not the same as the spatial pattern in 2003. After allowing for a flexible temporal structure, we investigate if there is any evidence that the spatial pattern of preterm birth changes over time.

The question of interest is investigated using the birth outcome data from New York City for the years 1995 to 2003. Because of time constrains in running the model for the project, only a subset of the original dataset was used. To create the subset of data, we restrict the data to include mothers who are white and live in Bronx or New York counties. The Bronx and New York counties together contain 650 census tracts. Also, the data is only for pregnancies which resulted in a single birth. This results in a total of 65,628 mothers included in the analysis, of which 3,442 give birth preterm. We use additional individual level data about each mother, which includes the census tract in which they live, age, education level, and country of birth. The country of birth is used to determine whether or not the mother is foreign born. Lastly, we have information about each baby. We know the gestational age in weeks, the sex, and the year and month of delivery. Instead of using the continuous gestation age, we use an indicator of if the baby is born preterm.

2 Model

To study the questions of interest, we fit a probit regression model expanded to include a spatial-temporal intercept. The model is

$$\Phi^{-1}(E(y_i)) = Z_i\gamma + \beta(s_i, t_i), \tag{1}$$

where y_i is indicator of whether mother i gives birth prematurely, Z_i is the vector of confounders (individual level characteristics), γ are the corresponding fixed effects, β is the spatial-temporal intercept having a spatial-temporal conditional autoregressive (STCAR) prior, s_i is the census tract where mother i lives, and t_i is the time period when mother i gives birth, month and year.

The STCAR model is a conditional autoregressive model that includes a neighbour structure for both space and time. Because the STCAR model allows for a general temporal neighbourhood structure, it is more general than the dynamic linear model. It is called the STCAR instead of a spatio-temporal autoregressive (STAR), because we are using the conditional aspect. The STCAR model is defined as:

$$\beta \sim MVN\left(0, [\tau\left(\alpha(M_s - \rho_s D_s) + (1-\alpha)(M_t - \rho_t D_t)\right)]^{-1}\right), \tag{2}$$

where τ is a precision component that controls the variance, M_s and M_t are diagonal matrices listing the number of neighbours for space and time respectively, ρ_s and ρ_t control the smoothing, D_s and D_t are the neighbourhood matrices for space and time respectively, and α controls the amount of smoothing coming from spaces versus time. If $\alpha- > 0$ then there is only smoothing across time, and if $\alpha- > 1$ then there is only smoothing across space. The corresponding full conditional is

$$\beta(s,t)|\beta(-(s,t)) \sim N\left(\mu(s,t), per = \tau\left(\alpha M_{s=s} + (1-\alpha)M_{t=t}\right)\right) \tag{3}$$

$$\mu(s,t) = \frac{\alpha\rho_s M_{s=s}\bar{\beta}_s + (1-\alpha)\rho_t M_{t=t}\bar{\beta}_t}{\alpha M_{s=s} + (1-\alpha)M_{t=t}} \tag{4}$$

where $\mu(s,t)$ is between the average of the spatial neighbours and temporal neighbours (based on the number of neighbours and α), the precision is between τ times the number of spatial neighbours and τ times the number of temporal neighbours, and we assume the weights to be equal to 1 if two tracts are neighbours and 0 if the two tracts are not neighbours.

3 Example Analysis

The example analysis is performed using the New York data described above. Due to time constraints a complete model was not fit. In order to speed up the time in which the model could be fit, only three time periods are used. They are 1995 - 1997, 1998 - 2000, and 2001 - 2003. Three years in each category is reasonable to evaluate the performance of the model.

The fixed effects (presented in the presentation) are given first, which are very similar to the fixed effects found when analysing the full dataset. Because there are a small number of mothers in the ”$<$ 18 years old category and elementary school category, the standard errors are much greater than the analysis with the full dataset. The spatial results (also presented in the presentation) are given as the difference between the time category 2001-2003 and 1995-1997. These plots suggest that there is some temporal pattern to the data, but the results are not statistically significant. This lack of significance may be due to the small sample size or the small spatial region (2 counties out of 5). This idea is also suggested when looking at the posterior of α, which appears to slightly favour spatial smoothing over temporal smoothing.

4 Further Work

The results presented here are the beginnings of work to investigate the use of a STCAR model. In order to further the research, we plan to compare the STCAR with the typical STAR and dynamic linear models. Because of the large number of correlated parameters, a model comparison method such as DIC will be used to help provide comparisons between the models. Also, we plan to implement models that allow α to be non-stationary. We don't expect the smoothness of the surface to remain constant across time or space, so the model should allow for the possibility. Finally, in order to evaluate the sensitivity of the model to the space time covariance structure, we will implement other covariance structures which allow for directionality in the temporal neighbours, larger temporal neighbourhoods, and more temporal categories.