Modeling Spatial Dependence in Child Poverty Rates: A Spatial Regression Analysis of North Carolina Counties

Zach Ginder and Jo Brennan

Introduction

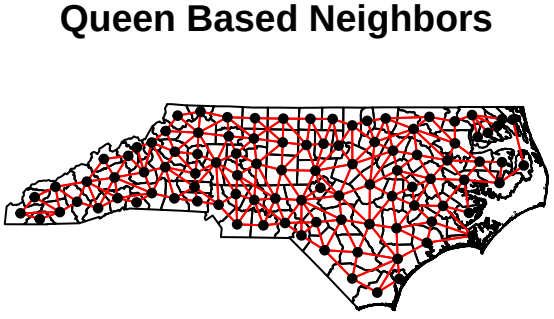
Each year the North Carolina Association of County Commissioners (NCACC) releases a County Map Book detailing data on demographics, economics, environment, education, health, and taxation in the state but on a county level. This project will evaluate a subset of the data from a spatial statistics perspective, evaluating whether spatial dependence is evident in some of the data, and then constructing a regression model to attempt to account for the spatial dependence of a response variable of interest. If the residuals of the model are found to exhibit spatial dependence, then a construction of Spatial Autoregressive (SAR) and Conditional Autoregressive (CAR) areal models will be considered to account for the spatial confounding evident.

Spatial Dependence Testing

Of interest in this project was to evaluate variables that were in relation to the topic of poverty in North Carolina. As shown in recent public perception polling by Gallup, a plurality of Americans view economic problems as the most important problem facing the country today. Therefore, it was of interest to see what predictors may be useful in modeling aspects of economic status in North Carolina.

Choosing the Proximity Matrix

In testing for spatial dependence it is important to determine to proximity matrix for determining neighboring counties. The analysis used to determine the proximity matrix involved using the queen based neighborhood structure with style "W" which represents row standardization (sums over all links to n). A plot of the queen based neighborhood structure is as follows:



The three variables of interest that were tested to determine potential spatial dependence were as follows: Gross Domestic Product (2022), Percent of Children in Poverty, and Percent of Uninsured Residents (2021). To determine potential spatial dependence, evaluation through the use of the Geary's C test for spatial autocorrelation. The results are as follows:

Variable	Geary's C Test Statistic	P-Value
Gross Domestic Product (2022)	0.9546642	0.2544
Percent of Children in Poverty	0.5750683	2.936e-10
Percent of Uninsured Residents (2021)	0.6793242	1.476e-06

The Geary's C test statistics ranges from zero to two, and reflects the degree of spatial autocorrelation. A value closer to zero reflects perfect positive spatial autocorrelation, while a value closer to two reflects perfect negative spatial autocorrelation. A value that is close to one reflects little to no spatial autocorrelation.

We see from the tests conducted on the three variables that Gross Domestic Product (2022) appears very close to one, and with a p-value that is not significant, we can conclude that there appears to be little to no spatial autocorrelation. In relation to the Percent of Children in Poverty, the Geary's C test statistic appears to be closer to zero than two and appears significant according to the p-value, thus showing there exists positive spatial autocorrelation. The same can be said for the Percent of Uninsured Residents (2021) which appears to exhibit positive spatial autocorrelation, though less than that of the Percent of Children in Poverty.

Spatial Areal Regression Analysis

As the variable associated with the Percent of Children in Poverty seems to be the most spatial dependent of the three variables of initial interest, further spatial area regression analysis will be considered. Of interest in this study is creating a regression model to attempt to account for the spatial dependence of a response variable of interest. The predictors that will be considered include: Gross Domestic Product (2022), Adults Without a High School Diploma, Taxable Property Valuation Per Capita, and Overdose Deaths Per 100,000. A simple least squares can be computed to evaluate which of these predictors are significant in predicting the Percent of Children in Poverty. When such a model is performed, it is found that the predictors that are significant (alpha=0.01) in predicting the Percent of Children in Poverty are as follows: Adults Without a High School Diploma and Taxable Property Valuation Per Capita.

It is possible that such a model accounts for the spatial dependence of the Percent of Children in Poverty, and thus a spatial areal regression model is not needed to be considered. To evaluate this it is important to perform a Geary's C test on the standardized residuals of the model using the statistically significant predictors. The outcome of that Geary's C test on the standardized residuals of the linear model is as follows:

Model	Geary's C Test Statistic	P-Value
Simple Linear Regression of Percent of Children in Poverty regressed on Adults Without a High School Diploma and Taxable Property Valuation Per Capita.	0.658162147	3.136e-07

The results of the Geary's C test for spatial autocorrelation for the residuals shows a test statistic closer to one than when the test was performed on the Percent of Children in Poverty alone. This suggest that there is less spatial dependence in the residuals than in the response variable. However, there still exists some positive spatial dependence as the value of the Geary's C test statistic is not equal to one. Thus construction of Spatial Autoregressive (SAR) and Conditional Autoregressive (CAR) areal models are appropriate.

Spatial Autoregressive (SAR) and Conditional Autoregressive (CAR) Models

Since the Geary's C test for independence in the residuals found that the errors were still dependent, a spatial areal regression model needs to be formulate to account for the dependent errors. A SAR model to account for these dependent errors as well as a CAR model were considered.

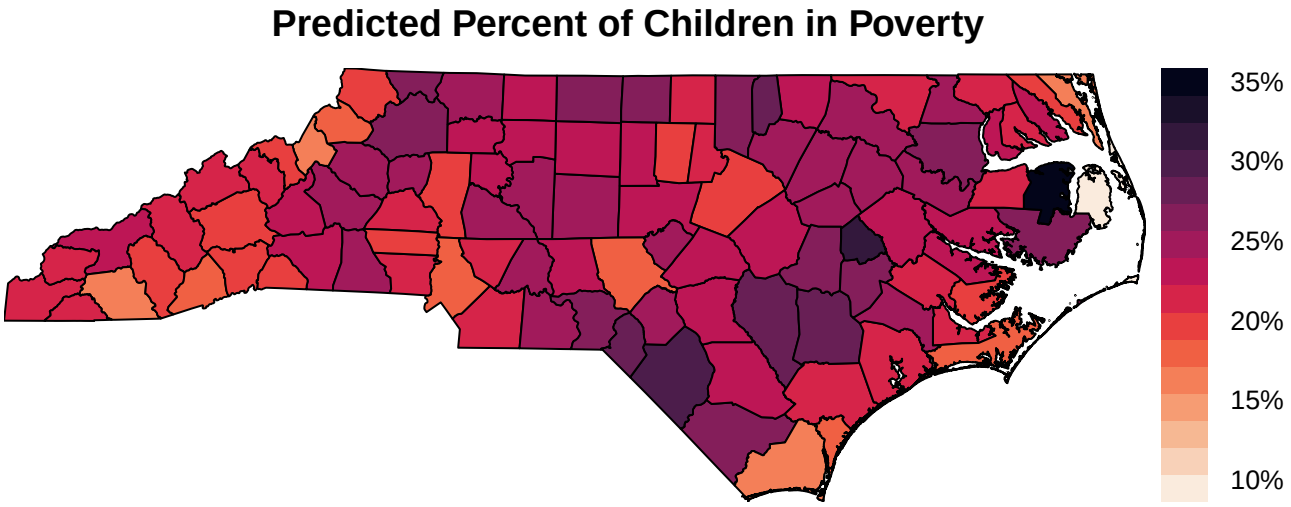
Model	AIC
SAR	-283.9900
CAR	-206.4569

Comparing the models off of the Akaike Information Criterion (AIC), showed that the SAR model had the lower AIC value indicating a better fit. Thus, the optimal spatial areal regression model is the SAR model and the parameter estimates are as follows:

Parameter	Estimate (for the SAR model)	P-Value
Intercept	2.232e-01	2e-16
Adults.Without.a.High.School.Diploma	4.569e-01	3.28e-06
Taxable.Property.Valuation.Per.Capita	-3.190e-07	0.000101

SAR Spatial Covariance Coefficient Estimates:

Parameter	Estimate (for the SAR model)
Dependent Error Variance	0.001998
Parameter Controlling Spatial Dependence (ρ)	0.628089



Discussion

The spatial analysis emphasizes the significant spatial dependence present in child poverty rates across North Carolina. While the proportion of adults without a high school diploma and taxable property valuation per capita emerged as statistically significant predictors, they were insufficient to fully explain the spatial autocorrelation observed. This suggests that child poverty in one county is strongly influenced by poverty levels in adjacent counties, highlighting the need for regional, rather than isolated, policy interventions. These findings support the idea that

area-wide strategies may be more effective in addressing persistent poverty. However, the analysis is limited by the complexity of factors contributing to child poverty. Additional data and broader variable inclusion are necessary to more accurately capture and predict poverty dynamics across counties especially in North Carolina.

References

County data and information - North Carolina Association of County Commissioners : North Carolina Association of County Commissioners. (n.d.-a). https://www.ncacc.org/research-and-publications/research/county-data-and-information/

Gallup. (2024, October 16). *Most important problem*. Gallup.com. https://news.gallup.com/poll/1675/most-important-problem.aspx