Spatial Confounding Adjustment for Count Data

Sam Herold, *Dr. Kayleigh Keller*
College of Statistics

Background and Significance

Controlling for confounders is a classic problem in statistics.

- Research focus: **Unobservable confounders in spatial**
- **data** (data with information about location).
- Unobserved spatial variables influence both dependent and independent variables
- Ex. If relationship of interest is air quality on health, income level may be a confounder.

Research Goal:
Combat spatial confounding in linear and count data.

Methods

This Plate Regression Splines combat unknown spatial confounding.

- **TPRS use a mathematical process to create many spatial surfaces with varying structure, which are added to the model.**
- TPRS try to spatially represent any structure not represented by the independent variable, and can isolate the desired relationship.

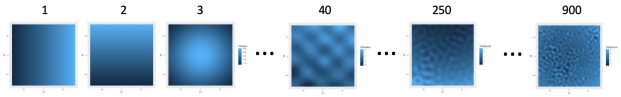

Figure 1. Different Thin Plate Regression Splines. The early TRPS only capture general structure, while later TPRS try to capture detail.

- Any amount of TPRS can be added to a model (TPRS degreed of freedom).
- TPRS become more complex and detailed as degrees of freedom increases.

Experiment Setup

Simulation was used to explore the effectiveness of TPRS for linear and count data. Linearly related data was examined first:

- True Model: $\boldsymbol{y(s) = \beta_0 + \beta_1 x(s) + c(s) + \varepsilon}$, where $c(s)$ is an unknown confounder
- SLR Model: $\boldsymbol{y(s) = \beta_0 + \beta_1 x(s) + \varepsilon}$
- SLR w/ TPRS: $\boldsymbol{y(s) = \beta_0 + \beta_1 x(s) + f(s)_i + \varepsilon}$, where $\boldsymbol{f(s)_i}$ are i Thin Plate Regression Splines

In this experiment $\boldsymbol{\beta_1 = 1}$, but $\widehat{\boldsymbol{\beta_1}}$ is artificially inflated when there is an unobserved confounder. TPRS hope to bring $\widehat{\boldsymbol{\beta_1}}$ to 1.

Results confirmed the benefits of TPRS for linearly related data. This was known, but helpful to understand the problem and set up the next test: TPRS was investigated for Poisson distributed count data. Again, $\boldsymbol{\beta_1 = 1}$, but $\widehat{\boldsymbol{\beta_1}}$ is inflated.

- True Model: $\boldsymbol{log(\mu) = \beta_0 + \beta_1 x(s) + c(s) + \varepsilon}$,
- SLR Model: $\boldsymbol{log(\mu) = \beta_0 + \beta_1 x(s) + \varepsilon}$
- w/ TPRS: $\boldsymbol{log(\mu) = \beta_0 + \beta_1 x(s) + f(s)_i + \varepsilon}$, where $\boldsymbol{f(s)_i}$ are i Thin Plate Regression Splines

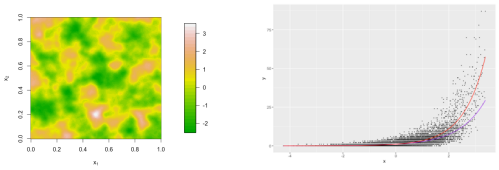

Fig 2:
(Left) Example of simulated spatial surface $\boldsymbol{x(s)}$. $\boldsymbol{y(s)}$ is a function of $\boldsymbol{x(s)}$ and $\boldsymbol{c(s)}$, which is another one of these surfaces.
(Right) Plotting simulated $\boldsymbol{x}$ and $\boldsymbol{y}$. The red curve visualizes the inflated $\boldsymbol{\beta_1}$ caused by the confounder, while the purple curve represents the desired $\boldsymbol{\beta_1}$. TPRS hope to represent the confounder and isolate the desired relationship.

Results

The results of this experiment suggest that TPRS are ineffective for Poisson spatial data.

- For almost all simulation situations, $\widehat{\boldsymbol{\beta_1}} > \boldsymbol{\beta_1}$.
- The estimate are larger (worse) than it would be if we didn't add TPRS's.
- The more TPRS added, the worse the performance.

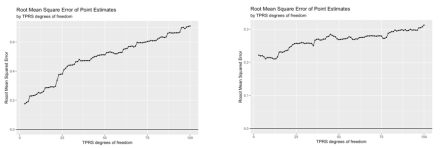

Figure 3. RMSE of $\widehat{\boldsymbol{\beta_1}}$ by number of TPRS added. RMSE, bias, standard error and standard deviation all increase similarly.

Conclusion: Thin Plate Regression Splines are not suitable for combatting spatial confounding in count data. It is suspected that the continuous nature of TPRS cannot handle the log transform and the discrete nature of the response variable.

Other Experiments

- Negative Binomial GLM (rather than Poisson Regression)
- Quasi-Poisson GLM (rather than Poisson Regression)
- Two predictions (with Poisson Regression)
- Non-square data (with Poisson Regression)

TPRS also performed poorly in these experiments for the same reasons.

Acknowledgements

Huge thank you to Dr. Keller! Not only for the guidance on the project, but for inspiring me and helping me with grad school.

[1] K. Keller, A. Szpiro. Selecting a Scale for Spatial Confounding Adjustment. 2020.