The effect of spatial scale on bias in regression models with spatial confounding

Chris Paciorek
Department of Biostatistics
Harvard School of Public Health

November 14, 2007

Outline

1. Framework
   - Correlated Residuals
   - Spatial Confounding
2. Results
   - Analytic
   - Simulation
3. Extensions

Themes

- Intuition about residual correlation can be deceptive.
- Scales of spatial correlation are critical.
- Accounting for spatial correlation may help reduce bias from confounding in **some** situations.

Uncertainty and Correlated Residuals

- Variance of regression estimates, Var($\hat{\beta}$):
    - naive OLS variance is incorrect
    - GLS is the minimum variance estimator: $\hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$
        - lower variance than OLS with corrected variance estimate
- Question: How does residual correlation affect variance?
-

Uncertainty and Correlated Residuals

- Variance of regression estimates, Var($\hat{\beta}$):
  - naive OLS variance is incorrect
  - GLS is the minimum variance estimator
    - lower variance than OLS with corrected variance estimate
- Question: How does residual correlation affect variance?
- Conventional wisdom: Correlated residuals reduce the effective sample size, so their presence adds uncertainty.

Uncertainty and Correlated Residuals (2)

- Reality:
  - Correlated residuals offer an **opportunity** to improve precision by systematically explaining a portion of the residual variability.
  - Equivalent models

$$\text{GLS:} \quad Y \sim \mathcal{N}(X\beta, \sigma_r^2 R + \tau^2 I)$$

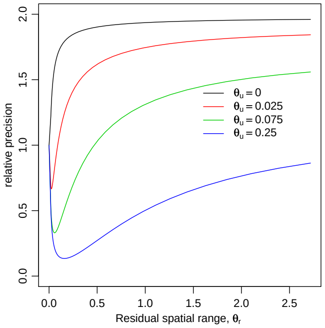
$$\text{GAM:} \quad Y \sim \mathcal{N}(X\beta + g, \tau^2 I)$$
$$g \sim \mathcal{N}(0, \sigma_r^2 R)$$

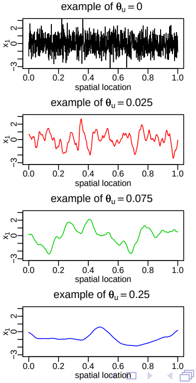
  - Heuristic is that fitting either a GLS or GAM model allows one to attribute residual variability to the spatial component of the residual, reducing the unexplained variability in the model and decreasing $\text{Var}(\hat{\beta})$.

Precision with Correlated Residuals

$$E(Var(\hat{\beta})^{-1}) = E(X_1^T\Sigma^{-1}X_1) = \mathrm{tr}(\Sigma^{-1}\sigma_u^2 R(\theta_u)) = \frac{\sigma_u^2}{\tau^2}\mathrm{tr}((I + \frac{\sigma_r^2}{\tau^2}R(\theta_r))^{-1}R(\theta_u)).$$

Results depend on the scales of the correlation in X_1 and the residual.





Key Question

- We know that attributing variability to a spatial component in the residual can reduce variance.
- Can it alleviate bias from an unmeasured, but spatially-correlated, confounder?
  - Potential mechanism: attribute variability from confounder to the spatial residual (or to a spatial term in the mean).
- Conventional Wisdom?
  - Accounting for spatial correlation in the residual can account for a spatial confounding and reduce (eliminate?) bias.
- Reality:
  - It depends on the spatial scales involved.
  - Dominici et al. (2004, JASA): control for spatial structure at large scales to eliminate confounding at that scale.
  - Goal is to assess association based on nearby observations, which share the same large-scale spatial effect.

Thought Experiment

- Suppose pollution varies smoothly in space. Also, suppose that (unmeasured) SES varies smoothly in space.
- If we analyze a health outcome as a function of pollution, the residuals will be correlated because of SES.
- There is a fundamental non-identifiability in the model

$$Y_i = X(s_i)\beta + g(s_i) + \epsilon_i$$

which we could re-express as

$$Y_i = g^*(s_i) + \epsilon_i.$$

That is, how do we separate the pollution effect from the spatial effect (spatial confounder) if the pollution effect is just another form of spatial effect.

- Questions:
  - How does the model attribute variation between $X(s)\beta$ and $g(s)$?
  - What aspects of $X(s)$ are used to estimate β?

Scale Matters

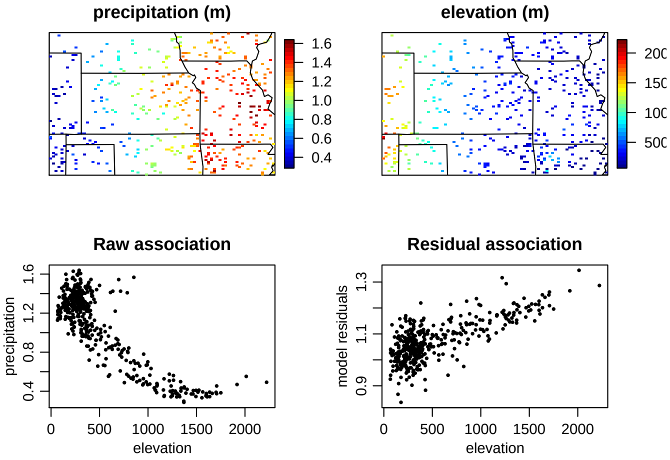
- A non-health example: how does elevation affect precipitation in the central United States?
- At large scale, precipitation increases with decreasing elevation as topography slopes gently downwards from the Rockies to the Mississippi River.
  - Elevation is not the causal effect.
- At smaller scale, precipitation increases with increasing elevation.
- A spatial model here can account for confounding from other factors that vary smoothly west to east, and isolate the elevation effect to the effect of elevation at small scales.

$$\text{GLS:} \quad Y \sim \mathcal{N}(X\beta, \sigma_r^2 R + \tau^2 I)$$

$$\text{GAM:} \quad Y \sim \mathcal{N}(X\beta + g, \tau^2 I)$$

In the GAM, roughness in g is penalized with a penalty parameter estimated by an analog of generalized cross-validation.

Association of Elevation and Precipitation



Dominici et al. approach

- Model

$$y_t = \beta x_t + g(t) + \epsilon_t$$
$$x_t = x_c(t) + x_{u,t}$$

  - The paper explores the effects of modeling the temporal variability in $g(t)$ with orthogonal basis functions.

- Results:
  - If $g(t)$ is modeled with sufficient basis functions to fully capture the temporal variation in $x_c(t)$, then:
    1. if $x_c(t)$ is smoother than $g(t)$, $\hat{\beta}$ is asymptotically unbiased.
    2. if $x_c(t)$ is rougher than $g(t)$, $\hat{\beta}$ is unbiased.

Building on the approach

- Key insight: the spatial model should account for correlation in the covariate, not in the outcome/residuals.
- Unresolved issues:
  - What happens if the unconfounded portion of the covariate, $x_{u,t}$, is spatially correlated?
    - How do the relative spatial scales affect bias and precision?
  - What is the bias when one fits a standard GLS model or GAM for the covariate, accounting for spatial correlation?
  - The model doesn't have correlation between $g(t)$ and $x_c(t)$.

A Simple Model

- We can explore bias by starting with a simple generative model:

$$y_i = \beta_1 x_1(s_i) + \beta_2 x_2(s_i) + \epsilon_i$$

  Let $x_1(s)$ and $x_2(s)$ be Gaussian processes, with $\text{Cor}(x_1(s_i), x_2(s_i)) = \rho$.

- If x_2 is unmeasured, we arrive at the GLS model

$$y_i = \beta_1 x_1(s_i) + \epsilon_i^*$$
$$\text{Cov}(\epsilon^*) = \Sigma = \sigma_r^2 R(\theta_r) + \tau^2 I$$

  where $\sigma_r^2 = \beta_2^2 \text{Var}(x_2)$.

Bias in the simple model

$$E(Y|x_1) = \beta_1 x_1(s) + \epsilon^*$$
$$\text{Cov}(Y|x_1) = \text{Cov}(\epsilon^*) = \Sigma = \sigma_r^2 R(\theta_r) + \tau^2 I$$

Bias comes from fitting models under the assumption that ϵ^* is uncorrelated with x_1.

- In calculating $E(Y|x_1)$ and $\text{Cov}(Y|x_1)$ in the GLS model above, we have used the marginal, $P(X_2)$ instead of the conditional, $P(X_2|X_1)$.
- The GLS model and its GAM analog match what practicioners do when they fit regressions with spatial structure.

Known parameters, single scale

- Suppose $x_1(s)$ and $x_2(s)$ share the same range of spatial correlation, but may be scaled differently in magnitude, namely, $\text{Cov}(x_1) = \sigma_c^2 R(\theta_r)$ and $\text{Cov}(x_2) = \sigma_2^2 R(\theta_r)$, then

$$\begin{aligned} E(\hat{\beta}_1|x_1) &= \beta_1 + (x_1^T \Sigma^{-1} x_1)^{-1} x_1^T \Sigma^{-1} E(x_2|x_1)\beta_2 \\ &= \beta_1 + \rho \frac{\sigma_2}{\sigma_c} \beta_2 \end{aligned}$$

  because $E(x_2|x_1) = \rho\sigma_2\sigma_c R(\theta_r)\sigma_c^{-2} R(\theta_r)^{-1} x_1$.
- The resulting bias, $\rho\frac{\sigma_2}{\sigma_u}\beta_2$, is the same as if the covariates were not spatially structured.
- Heuristically, the model attributes variability from the confounder to the covariate of interest.

Known parameters, multi-scale

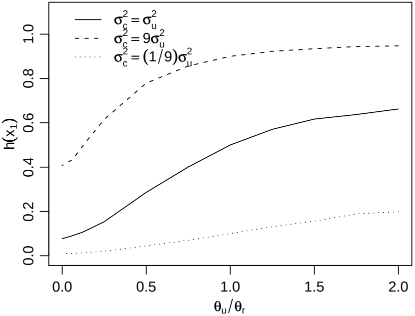
Let $x_1(s) = x_c(s) + x_u(s)$ with $\text{Cov}(x_1) = \sigma_c^2 R(\theta_r) + \sigma_u^2 R(\theta_u)$.
Let $\text{Cov}(x_2) = \sigma_2^2 R(\theta_r)$ and $\text{Cor}(x_c(s_i), x_2(s_i)) = \rho$.

$$E(\hat{\beta}_1 | x_1)$$

$$= \beta_1 + (x_1^T \Sigma^{-1} x_1)^{-1} x_1^T \Sigma^{-1} E(x_2|x_1)\beta_2$$

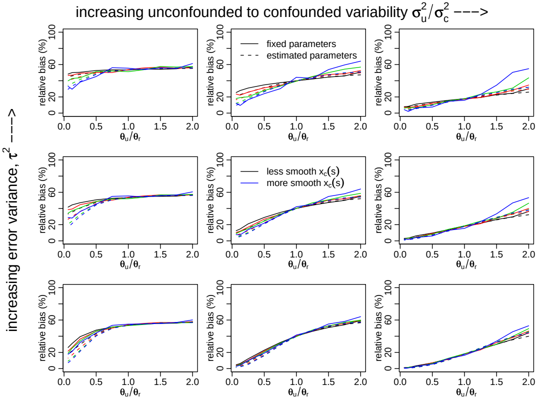
$$= \beta_1 + \frac{x_1^T \Sigma^{-1}(I + \frac{\sigma_u^2}{\sigma_c^2} R(\theta_u) R(\theta_r)^{-1})^{-1} x_1}{(x_1^T \Sigma^{-1} x_1)} \rho \frac{\sigma_2}{\sigma_c} \beta_2$$

$$= \beta_1 + h(x_1)\rho \frac{\sigma_2}{\sigma_c} \beta_2$$



Unknown parameters

Simulation results indicate that bias when estimating parameters in a GLS framework (or also in a GAM framework) is similar to that with known parameters.



Heuristics

- Reducing bias requires the covariate of interest to have a spatial scale at which it is unconfounded, and that scale must be smaller than the scale at which confounding operates.
- We would like the covariate to have as much variation at the unconfounded scale and as little at the confounded scale as possible.
- Other results are straightforward and match the non-spatial setting for confounding. We want:
  - the magnitude of variation in the confounder (or its effect on the outcome) to be small.
  - the correlation between confounder and covariate to be small.

Ongoing work

- Analysis of precision and MSE
- Simulations for non-linear settings
- Effects of choosing incorrect parameter values to minimize bias
    - Using fixed df to model the residual correlation (a la Dominici et al. 2004)
- Areal data settings
- Implications of measurement error in x_1
- Is there related work in spatial econometrics?
    - Regression discontinuity in spatial settings?

Areally-aggregated Data

- Aggregated data in areal units such as zip codes, census tracts and counties are often the finest resolution data available for disease mapping analyses.
- Spatial confounding may be an issue in spatial regression models for aggregated data.
- Conditional auto-regressive (CAR) models are often used; these models smooth based on weighted averaging of neighboring units.
- Two key issues in areal models:
  1. Aggregation smooths over fine-scale heterogeneity.
  2. CAR models (by using local averaging) do not model large-scale spatial patterns.
- Both of these issues suggest that bias could be substantial in CAR-type models based on the results presented here.

Measurement Error

- Classical error:
  - Preliminary work suggests that under classical error, the model attributes variability in the outcome to the spatial residual, not to the error-contaminated covariate of interest.
  - Model attenuates the effect estimate because the spatial residual is a well-measured surrogate that can stand in for the covariate.
- Berkson error/regression calibration:
  - Gryparis, Paciorek, and Coull (under revision) argue that spatial smoothing models are a form of regression calibration that induce Berkson type error when using predictions
  - Under Berkson error, we should be in the framework discussed here, except that smoothing done to make predictions will reduce fine-scale heterogeneity, decreasing our ability to reduce bias.