Can Conventional Measures Identify Geographically Varying Mixed Regression Relationships? A Simulation-based Analysis

Aaron Swoboda

Carleton College

[aswoboda@carleton.edu](mailto:aswoboda@carleton.edu)

Economics Department

One North College St

Northfield, MN 55057

Regional scientists are increasingly utilizing data analysis techniques that allow for spatial heterogeneity of various forms. Regression specifications like Geographically Weighted Regression and Locally Weighted Regression allow the regression parameters to vary over space thereby reflecting spatially non-stationary relationships. Such spatial heterogeneity is appealing in a regional science context in which location is assumed to matter in various ways. Recent work has begun to estimate "mixed" relationships in which some variables exert non-stationary effects on the dependent variable, while others exhibit a constant effect over space. With so many possibilities available, researchers face a daunting task determining which variables to estimate in a (non-)stationary fashion. This paper uses Monte Carlo simulations to generate mixed regression relationships to determine whether commonly used metrics (for example Leave One Out Cross Validation and the Akaike Information Criterion) can adequately distinguish between the different possibilities.

Experimental data are generated according to a spatially explicit linear model containing two independent variables and an intercept, which allows for a mix of spatially non-stationary relations for one variable while retaining spatial stationarity for others. We generate 675 different specifications allowing for varying degrees of spatial non-stationarity in the model coefficients, different sample sizes, and multiple levels of model error. Each combination is replicated 100 times for an overall sample size of 67,500 model iterations. In each iteration we estimate multiple locally weighted regression models and choose among them based on commonly used cross-validation and information metrics (Leave One Out, Generalized Cross Validation, Row Standardized Cross Validation and the Akaike Information Criterion).

Results suggest that the Leave One Out Cross Validation Metric is least susceptible to false positives (choosing models that suggest spatial non-stationarity when there is none). Interestingly, we also find that the most accurate coefficient estimates are not typically associated with the correct underlying regression model, but instead are associated with models that include additional non-stationarity beyond the original data generation process.

Keywords: Spatial Econometrics, Geocomputation, Spatial Heterogeneity, Monte Carlo Simulation