Modelling Spatial Point Data in R - sample applications

{ Piotr Cwiakowski and Piotr Wojcik } University of Warsaw, Faculty of Economic Sciences

University of Warsaw
Faculty of Economic Sciences

Introduction

- spatial point data should be precisely individually geolocalized,
- methodology of modelling spatial points is in the development stage [1, 2],
- the neighbourhood structure is changing after modification/addition/removal of any observation in the dataset,
- the spatial weights matrix is unstable in the training and testing set which makes prediction very difficult.
- currently available models for individually geolocalized points are computationaly inefficient and time-consuming - there is a need for development new methods and algorithms.

Materials & Methods

The following models were used for analysis:
Durbin Spatial Model (Real estate dataset)

$$y = \alpha + \beta\, X + W\, X\theta + \epsilon$$

Durbin Spatial Model (School dataset)

$$ln(\frac{y_{i,t}}{y_{i,t-1}}) = \alpha + \beta y_{i,t-1} + \rho\, W\, ln(\frac{y_{i,t}}{y_{i,t-1}}) + \\ +\gamma\, W\, y_{i,t-1} + \epsilon_t \quad (1)$$

Exam results in schools

Research framework

- **question**: does the direction and speed of convergence of educational achievements on schoool level depend on neighbouring schools' results (initial value and change) ?
- **data**: average result of the lower secondary school leaving exams (math-science part) on the level of schools (2003–2015),
- **model** Spatial Durbin Model for absolute convergence applied for each year separately with spatial weights matrix W based on knn between 1 and 6 (structure of schools and spatial weights changes with time),
- **R packages**: sp, spdep, rgeos, maps, maptools, ggplot,
- Research framework limited due to time consuming estimation of spatial models in R (10–15 min. for one spatial model based on ca. 6000 observations).

Results for exam results in schools

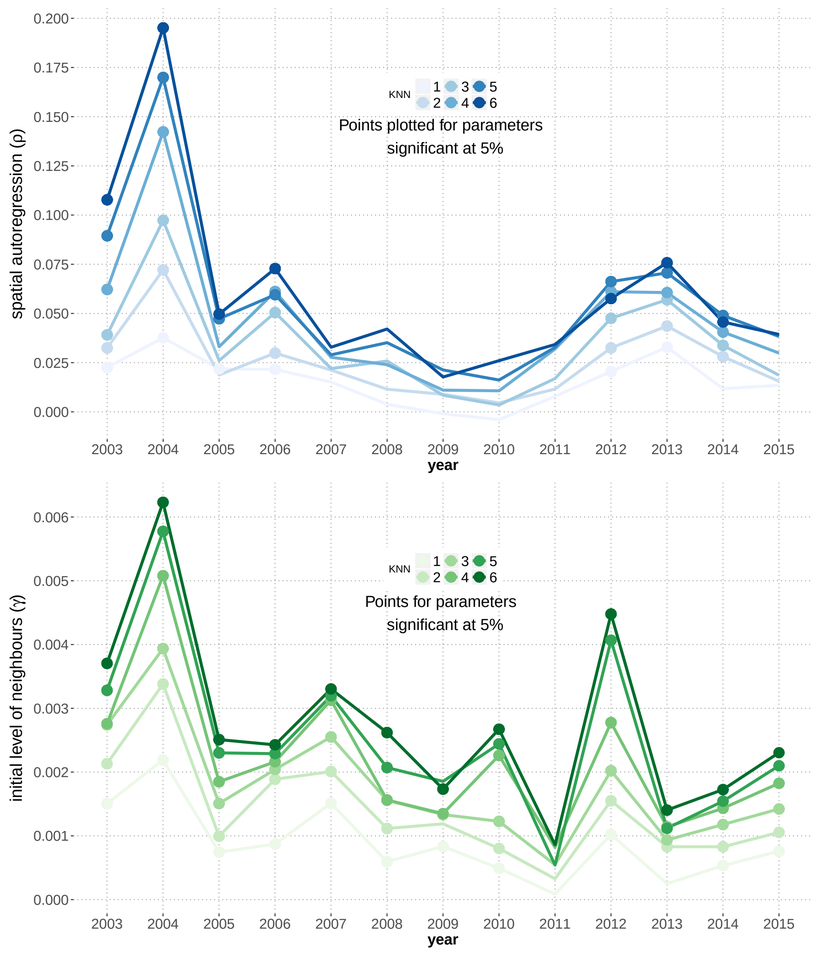

Figure 1: Spatial autoregressive parameter ρ (top) and impact of initial level of neighbours γ (bottom)

- in 2003–2004 higher impact of neighbours but also large variation of estimation results depending on knn,
- since 2005 results more consistent across different knns,
- **positive** but **very weak spatial impact** of progress in neighbouring schools' results (top figure) – non-significant in 2007–2011 and 2015,
- **higher initial level** of educational achievements in neighboring schools results in **consistently significant higher progress** of exam results (bottom figure),
- **further step 1**: development of time efficient estimation of spatial models for large datasets (with the use of Rcpp package),
- **further step 2**:development of methodology of estimation of spatial models robust to changes in the neighborhood structure (due to changing sample of points),

Warsaw estate market

Research framework

- **objective**: application of different spatial models and Voronoi polygons to build a predictive model for real estate prices in Warsaw (apartments' prices),
- **data**: market transactions between 2005–2015 (apartments),
- **model:** several models currently used for real estate market modelling will be considered: Spatial Durbin Model, Spatial Autoregressive Model, General Spatial Model, Geographically Weighted Regression, Spatial Expansion Model.
- **R packages**: sp, spdep, rgdal, deldir, maptools, ggplot, spatstat.

Results for Warsaw real estate market

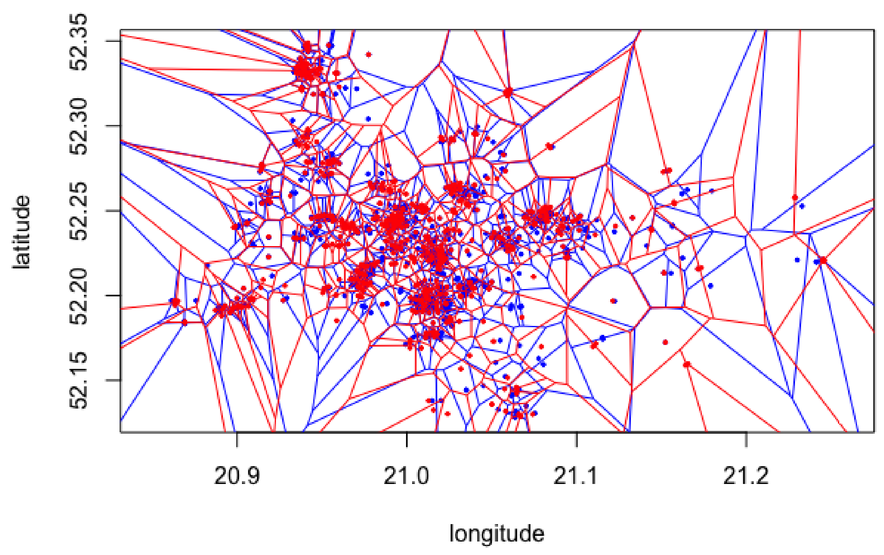

Figure 2: Comparison of regions estimated on two random subsamples derived from tessalation algorithm

- Empirical strategy assumes:
  1. development of sampling method for estimation of robust Voronoi polygons in spatial models,
  2. investigation of the impact of a particular subsample characteristics and definition of neighbourhood on estimation results
  3. examination of sampling methods in different spatial models.
- very early results using Spatial Durbin Model suggests that parameteres of the model are highly unstable and depends on number of nearest neighbours and subsample of localisations.

Contact Information

Email pwojcik@wne.uw.edu.pl
Email pcwiakowski@wne.uw.edu.pl

References

[1] Adrian Baddeley. Analysing spatial point patterns in 'r'. 2009.

[2] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial point patterns: methodology and applications with R*. CRC Press, 2015.