Spatial Autocorrelation in Areal Data

**Handed out:** Thursday, April 23, 2020

**Return date:** Sunday, May 3, 2020 (no extensions can be granted)

**Grading:** This lab counts 12 % towards your final grade

**Other Dates:**

* The lab will be returned on Tuesday May 5, 2020
* An online Q&A session will be held on Tuesday May5, 2020, from 4:00 pm to 5:00 pm
* The final exam will be written online on Thursday, May 7, 2020, from 4:00 to 7:00 pm.

**Objectives:** The first part introduces you to different model structures and the concept of spatial spillovers. The second part explores properties of regression based global spatial statistics, models and their identification, while the last part explores local spatial statistics and their impact on global statistics.

**Format of answer:** Your answers (statistical figures and verbal description) should be submitted as ***hardcopy***. Add a running title with the following information: Lab05, your name and page numbers. You may use this document as template. Copy the requested statistical figures into your document. No trial and error answers are permitted. Label each answer properly with the bold task headings. You are expected to hand in professionally formatted answers: use a fixed pitch font, like **Courier New**, for any R code the use mathematical type-setting when equations are required. Copy and paste figures into your document. Make sure that each figure has a proper ***caption*** describing its content.

# Part I: Spatial spillovers (4 points)

**Task 1 (4 points):** Study the paper by Golpher , A. B., and P. R. Voss. How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects. *Spatial Demography* (2016) 4:175-205.

Note: In this lab the terminology of the Golpher’s paper is used.

Assuming that we just deal with a simple bivariate base *SLM* model with with spatial observations and .

**[a]** How to interpret the entries in the derivatives matrix below with regards to changes in the independent variable at location ? (1 point)

[b] Explain, based on the derivative matrix , what the *direct*, *indirect* and *total* effects are as well as how to interpret these effects? (1 point)

[c] Why does the spatial error model *SEM* have only direct effect predictions ? (0.5 points)

[d] The spatial error model *SEM* can be rewritten as (see page 179)

Argue why the estimates and of ordinary least squares and this general least squares model are the same. Thus, the *OLS* estimates are unbiased, i.e., and . (0.5 points)

[e] Discuss for Table 2 and the models SAR & SAC with and at how and why the spillover effects distribute around the central region 7 and the peripheral region 3. (1 point)

You may also use for your argument the property (see page 180). Note that consists of 2-steps links and predominately of 3-steps links.

# Part II: Global autocorrelation (4 points)

Throughout Parts II and III you will be working with the **tractShp** dataset in the package **TexMix**. Make sure to ***exclude*** the Love Field and the DFW airport tracts

**Task 2 (1 points):** Calculate the nearest neighbor graphs for the census tracts and the standard adjacency graph. (see Bivand *et al.* page 269)

Map and interpret both graphs.

Why are binary nearest neighbor graphs not necessarily symmetric?

**Task 3 (1 points):** Based the neighbor’s adjacency link structure in the *W*-coding scheme evaluate the standard deviate of the global Moran’s statistic for the residuals of the sequence of regression models:

|  |  |
| --- | --- |
| Model | Moran’s I standard deviate |
| PCTNOHINS~1 |  |
| PCTNOHINS~PCTHISPAN |  |
| PCTNOHINS~PCTHISPAN+PCTUNIVDEG |  |
| PCTNOHINS~PCTHISPAN+PCTUNIVDEG+PCTFAMPOV |  |

Explain why usually the autocorrelation level shrinks as additional exogeneous variables are added to the model.

Also interpret the correlogram up to lag 10 for the last model.

**Task 4 (2 points):** Calibrate a SAR model with **errorsarlm( )** and a LAG model with **lagsarlm( )**.

Compare the estimated regression coefficients of both models with those of the OLS model.

Test the residuals of the selected spatial autocorrelation model with the function **moran.mc( )**.

Why are the residuals expected to be spatially uncorrelated for a ***properly*** specified model?

# Part III: Local autocorrelation (4 points)

**Task 5 (2 points):** For the model with **PCTNOHINS~PCTHISPAN+PCTUNIVDEG+PCTFAMPOV** under the assumption of spatial independence calculated the cumulative probabilities of local Moran’s  for all areas with the saddlepoint approximation. Use the neighbor’s adjacency link structure in the *W*-coding scheme.

Map the residuals with a bipolar map theme for the sub-region of the ***north-east quadrant*** of Dallas county and superimpose map symbols for the cumulative probabilities of local Moran’s .

Interpret your map by looking for significant clusters and/or hot spots.

**Task 6 (1 points):** Calculate for the model in task 5 the variances of local Moran’s  in the *C*-, *S*-, and *W*-coding schemes and map these variances with a gradient color scheme.

Does a specific pattern emerge?

**Task 7 (1 point):** Discuss how the variance of local Moran’s  in the different coding schemes influences the contribution of individual areas on global Moran’s . Recall . See also Golgher *et al.* (2016 page 191).