Spatial Regression in GeoDa

GeoDa is a software package developed by the Spatial Analysis Lab (Luc Anselin) at the University of Illinois. It is a collection of software tools designed for exploratory spatial data analysis (ESDA). GeoDa adheres to ESRI's shape files as the standard for storing the information. GeoDa is highly interactive based on dynamically linked windows. For example, you can select cases from any of several windows (histogram, scatter plot) and see those cases highlighted in the other active windows.

The software can be downloaded for free at https://www.geoda.uiuc.edu/downloadin.php
Extended documentations, tutorials, and support are also available at https://www.geoda.uiuc.edu.

Example 1: Homicide rates in the southern U.S.
(Baller et al., 2001)

Variables

HR: Homicide rate per 100,000

RD: Resource Deprivation/Affluence Component

(prinicipal component composed of percent black, log of median family income, gini index of family income inequality, percent of families female headed (percent of families single parent for 1960), and percent of families below poverty (percent of families below $3,000 for 1960)

PS: Population Structure Component

(principal component composed of the log of population and the log of population density)

MA: Median age

DV: Percent of males 14 and over who are divorced (aged 15 and over for 1980 and 1990)

UE: Percent of civilian labor force that is unemployed

Example 1

south.shp
FIPSNO

1412 counties in the southern U.S.

☆ **Exercises**

1. Conduct ESDA to explore the response variable HR60 and relationships between HR60 and the structural variables.

2. Calculate the Global Moran's I statistic to see if there is evidence of spatial autocorrelation in the dependent variable HR60.

3. Construct a cumulative 1st and 2nd order rook weights matrix called `southrk12.gal`.

Ordinary Least Squares (OLS) Regression

Be sure to give your files meaningful names. A second report with the same name does not append, it overwrites.

- `Regress >`

  `Report Title:OLS Regression`
  `Output File: OLS_Regression.OLS`

  `Check box "Moran's I z-value"`

  `Dependent variable: HR60`
  `Independent variables: RD60, PS60, MA60, DV60, UE60`
  `Weight matrix: southrk12.gal`
  `Models: Classic`

  `Run`

  `Save` → `Predicted Value` and `Residual`

The *.OLS regression output files can be opened in WordPad.

- `Map > St Dev` → `OLS_RESIDU`
- `Space > Univariate Moran` → Select OLS_RESIDU and weights matrix `southrk12.gal`

Discussion and Other Tools

A Moran's I plot for residuals is purely *descriptive*. It is not appropriate to use the permutation approach to assess significant autocorrelation for residuals.

1. Understanding the output

Fit Statistics

R^2: Coefficient of determination
F-statistic: omnibus measure of model fit

Model Comparison Statistics

Log-likelihood: the larger the better
Akaike Information Criterion: the smaller the better
Schwarz Criterion: the smaller the better

Assumption Diagnostics

Multicollinearity: scores >30 indicate a problem
Jarque-Bera: small p-value suggests normality assumption is questionable
Breusch-Pagan, Koenker-Bassett, White } small p-value suggests non-constant error variance

Spatial Autocorrelation Diagnostics

Moran's I for residuals (cannot assess significance)

Lagrange Multiplier (LM) Test Statistics

LM (lag): small p-value suggests spatial lag model
LM (error): small p-value suggests spatial error model

Robust LM (lag): small p-value suggests spatial lag model
Robust LM (error): small p-value suggests spatial error model

The Robust LM statistics are *only* used if both the LM (lag) and LM (error) tests are significant.

2. The residual standard deviation plot indicates where the model is under-fitting or over-fitting.

3. The Lagrange Multiplier tests provide some guidance as to what spatial regression model might be appropriate for the data.

Spatial Lag Regression

- `Regress >`

  `Report Title: SpLag Regression`
  `Output File: SPLAG_REGRESSION.OLS`

  `Check box "Moran I's z-value"`

  `Dependent variable: HR60`
  `Independent variables: RD60, PS60, MA60, DV60, UE60`
  `Weights matrix: southrk12.gal`
  `Models: Lag`

  `Run`

  `Save` → `Predicted Value` and `Residual`

Usually the absolute value of the coefficients in the spatial lag model will be smaller than in the OLS model. This indicates the effects of the independent variables are partially due to the lag variable.

Exercises

1. Compare the summary statistics of the OLS model and the spatial lag model.

2. Check whether or not the lag variable W_HR60 is significant.

3. Compare the individual coefficients of the OLS model and the spatial lag model.

4. Use model comparison statistics to compare the OLS and spatial lag models.

5. Repeat the analysis using data from 1990: HR90 (dependent variable), RD90, PS90, MA90, DV90, and UE90 (independent variables). What is the best spatial regression model to use in this case?