Spatial Data Models

Yusep Suparman

Introduction

Table 2.1 Focal problems in spatial statistics and spatial econometrics

<table>
<tr><th>Spatial statistics</th><th>Spatial econometrics</th></tr>
<tr><td colspan="2">Super population perspective (i.e., realizations from a theoretical population): model-based inference</td></tr>
<tr><td colspan="2">Properties of estimators</td></tr>
<tr><td colspan="2">Specification of geographic neighbourhood structure</td></tr>
<tr><td colspan="2">Modifiable areal unit problem (MAUP)</td></tr>
<tr><td colspan="2">Quantifying spatial autocorrelation</td></tr>
<tr><td colspan="2">Variable transformations: Box-Cox, Box-Tidwell</td></tr>
<tr><td colspan="2">Spatially adjusted statistical techniques</td></tr>
<tr><td colspan="2">Cluster detection: hot and cold spots; LISA statistics</td></tr>
<tr><td colspan="2">Distance as a covariate</td></tr>
<tr><td colspan="2">Bayesian hierarchical models</td></tr>
<tr><td colspan="2">Exploratory spatial data analysis</td></tr>
<tr><td colspan="2">Space-time modelling</td></tr>
<tr><td>Sampling network structure: design-based inference</td><td>Constrained parameter estimation</td></tr>
<tr><td>Ecological fallacy</td><td>Optimization models</td></tr>
<tr><td>Map generalization: spatial interpolation</td><td>Endogenous versus exogenous variables</td></tr>
<tr><td>Missing spatial data imputation</td><td>Spatial complexity and spatial regimes</td></tr>
<tr><td>Auto- model specification: normal, Poisson, binomial</td><td></td></tr>
<tr><td>Spatial structure as a covariate (spatial filtering)</td><td></td></tr>
<tr><td>Bayesian smoothing of map values</td><td></td></tr>
<tr><td>Error propagation</td><td></td></tr>
</table>

Introduction

- Spatial econometrics is a subfield of econometrics dealing with spatial effects among spatial unit (Elhorst, 2014).
- Spatial data are data that have a spatial component; it means that data are connected to a place in the Earth. Data usually collected with reference to location:
  - Location-administrative spatial units (states, districts, counties, etc.)
  - Functional regions (e.g. labour market regions)
  - Points in space (e.g. cities, municipalities, plants)

Introduction

- The distinction between mainstream econometrics and spatial econometrics is an existence of spatial interaction effects.
- Spatial effects have two form (Anselin, 1988):
  - Spatial Dependence (Spatial Autocorrelation)
  - Spatial Heterogeneity (Spatial Structure)
    - non-constant error variances (spatial heteroskedasticity)
    - model coefficients (variable coefficients, spatial regimes).

Introduction

- ***Spatial dependence*** reflects a situation where values observed at one location or region, say observation *i*, depend on the values of *neighboring* observations at nearby locations (LeSage & Pace, 2009).

- **Tobler's first law of geography**:

  "Everything is related to everything else, but near things are more related than distant things"

- **Spatial dependence** is associated with the notion of relative space (location):
  - Neighboring regions are expected to be more alike than arbitrary regions,
  - Spatial dependence is expected to diminish with increasing distance

Introduction

- **Spatial error dependence** is often interpreted as a nuisance., which caused by:
  - Measurement error
  - Omitted variables
- In regression analysis, measurement error and omitted variables are accommodated in error term ε. The error term in region *i* and *j* are no more independent but correlate
  - $Cov(\varepsilon_i,\varepsilon_j) = E(\varepsilon_i,\varepsilon_j)\ ;\ i \neq j \quad (1)$

Introduction

- The identically independently normal distribution of error term ε, can be wrote in matrix as:
  - $\varepsilon \sim IIDN(\mathbf{0}, \sigma^2 \mathbf{I})$
- with $\mathbf{0}$ is a (p x 1) vector mean, σ^2 is variance error, and I is (n x n) identity matrix.
- When the error term are correlated the distribution of ε is
  - $\varepsilon \sim N(\mathbf{0}, \mathbf{\Omega})$
- with Ω is a covariance matrix, which modeled with the aid of spatial weights, defined by distance or contiguity measures

Introduction

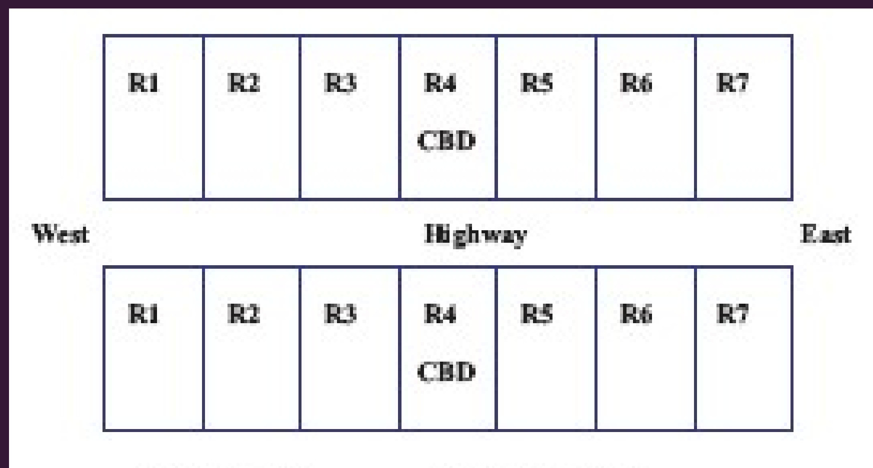
- **Substantive spatial** dependence reflects the existence of spatial interaction effect. The complex pattern of the interaction and dependence of socio-demographic and economic activity on the regional level to be a problem in modeling. In regional science, location and distance are important variables to be considered at work in human geography and market activity.
- The formal way to modeling the substantive spatial dependence the geo-referenced variable to be explained Y, is subject to the moment condition.

  - $Cov(y_i, y_j) \neq 0; \ i \neq j$ (1.2)

- **Geo-referenced variable Y** is values of Y are measured for spatial units (regions, districts, etc.)

Introduction

- Example



$$y = \begin{pmatrix} \text{Travel times} \\ 42 \\ 37 \\ 30 \\ 26 \\ 30 \\ 37 \\ 42 \end{pmatrix} \qquad X = \begin{pmatrix} \text{Density} & \text{Distance} \\ 10 & 30 \\ 20 & 20 \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix} \begin{array}{ll} & \\ \text{ex-urban areas} & R1 \\ \text{far suburbs} & R2 \\ \text{near suburbs} & R3 \\ \text{CBD} & R4 \\ \text{near suburbs} & R5 \\ \text{far suburbs} & R6 \\ \text{ex-urban areas} & R7 \end{array}$$

Spatial Weight Matrix

A spatial weights matrix is an $n \times n$ positive symmetric matrix $\mathbf{W}$ with element w_{ij} at location i, j for n locations. The values of w_{ij} or the weights for each pair of locations are assigned by some preset rules which define the spatial relations among locations and, therefore, determine the spatial autocorrelation statistics. By convention, $w_{ii} = 0$ for the diagonal elements (Zhou X and Lin H, 2008)

- **Contiguity (neighborhood)**
  - Contiguity reflects the relative location of one spatial unit to other regions in space. Neighbourhoodships of spatial units are usually established from a map.

- **Distance**
  - The location in space represented by latitude and longitude is one source of information. This information allows to calculate distance between points in space. In regional science points in space may represent centers or cities of regions.

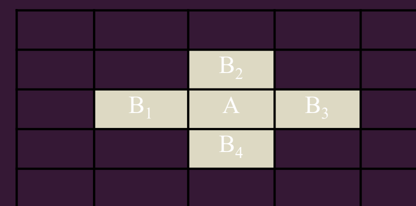
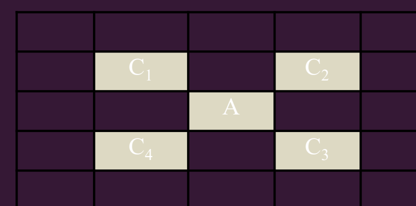
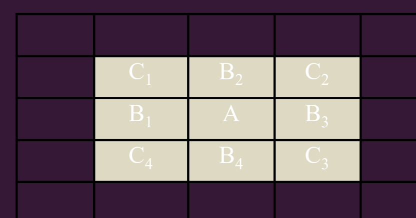
Spatial weight matrix

1. Spatial Contiguity Matrix

The spatial contiguity matrix W* is a binary n x n matrix which elements of the matrix w^*_{ij} are 0 or 1. The element w^*_{ij} is equal to one if the regions i and j are neighbors and 0 otherwise.

$$w^*_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are neighbours} \\ 0 \text{ otherwise} \end{cases} \quad (2.4)$$

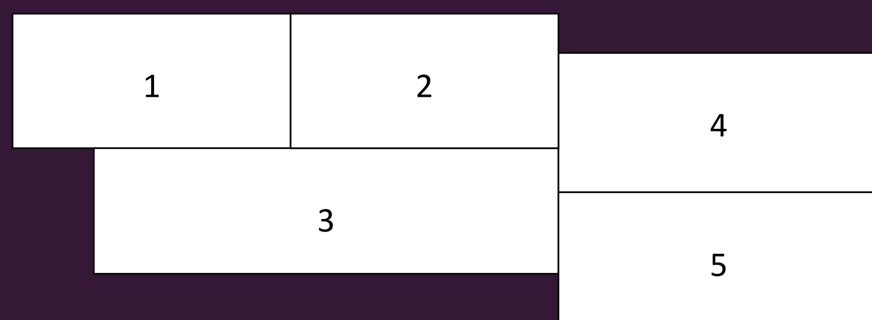
In a regular grid, the contiguity matrix can be distinguished in three different types with the game of chess analogy, i.e Rook contiguity, bishop contiguity, and queen contiguity

Types	Form
Rock contiguity A spatial unit is a neigbhour of another unit if both areas share a common edge (side). Unit B_1, B_2, B_3 and B_4 are neighbor of unit A	
Bishop contiguity A spatial unit is a neighbor of another unit if both areas share a common vertex. Unit C_1, C_2, C_3 and C_4 are neighbor of unit A	
Queen contiguity: A spatial unit is a neighbor of another unit if both areas share a common edge or vertex. Unit B_1, B_2, B_3, B_4, C_1, C_2, C_3, and C_4 are neighbor of unit A	

Spatial Weight Matrix

A common border (not vertex) is usually used to define the neighborhood in irregular grid.

Figure: Irregular arrangement of spatial unit



Contiguity matrix

$$\underset{5x1}{\boldsymbol{W}^*} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Spatial weight matrix

Standardized Weight Matrix

Row-standardization

For ease of interpretation, the weight matrix is often standardized such that the element of a row sum to one. The elements of a row-standardized weights matrix thus equal to:

$$w_{ij} = \frac{w_{ij}^*}{\sum_{j=1}^{n} w_{ij}^*} \qquad (2.5)$$

Standardized contiguity matrix for the irregular grid:

$$W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

Spatial Weight Matrix

This ensure that all weight are between 0 and 1 and facilitates the interpretation of operations with the weight matrix as an averaging of the neighboring value. It also ensures that the spatial parameter in many spatial stochastic process are comparable between models. For a row-standardized weight matrix, the largest Eigen value is always +1, which facilitates the interpretation of the autoregressive coefficient as "Correlation"

Effect of row-standardization with X is geo-referenced variable

$$\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} (x_2 + x_3)/2 \\ (x_1 + x_3 + x_4)/3 \\ (x_1 + x_2 + x_4 + x_5)/4 \\ (x_2 + x_3 + x_5)/3 \\ (x_3 + x_5)/2 \end{bmatrix} = \begin{bmatrix} \bar{x}_1^N \\ \bar{x}_2^N \\ \bar{x}_3^N \\ \bar{x}_4^N \\ \bar{x}_5^N \end{bmatrix}$$

Spatial weight matrix

2. Distance-based spatial weight matrix

It is assumed that spatial interaction will decline increasing distance due to increasing geographical impediments. Nearer regions have a greater potential influence.

1. Power function

$$w_{ij}^{*} = d_{ij}^{-\alpha} \qquad (2.6)$$

Where
α: power parameter
α=1: inverse distance
α=2: quadratic inverse distance (gravity model of spatial interaction)
For spatial units outside a critical distance cut-off d_{max} the weights may be set equal to 0. The distance d_{ij} are usually measured between the centers of the region. Using the latitude And longitude coordinates the shortest distance between two centers are given by the great circle distance.

$$d = r * arcos[sin(lat_1) * sin(lat_2) + cos(lat_1) * cos(lat_2) * cos(lat_2 - lat_1)]$$

radios of the earth: r = 6387.7 kilometers

Spatial weight matrix

2. Negative exponential function

The weight formulation of negative exponential function is

$$w_{ij}^{*} = e^{-\beta d_{ij}} \quad (2.7)$$

$$\gamma \bar{d} = 1 - e^{-\beta \bar{d}} \quad (2.8)$$

: average distance between immediate neighboring regions over the whole cross-section.

γ : transformed distance decay parameter

It is assumed that spatial interaction such as commuting, migration or interregional trade is exposed to the frictional effects of geographical distance. With increasing distance from region i these geographical impediments gain in strength, so that the decline of spatial effects become more and more pronounced.

Distance corresponding to a decrease of spatial effects of γ100%

$$d_{\gamma} = \frac{\ln(1-\gamma)}{\beta}; \quad 0 < \gamma < 1 \quad (2.9)$$

The Spatial Econometrics taxonomy

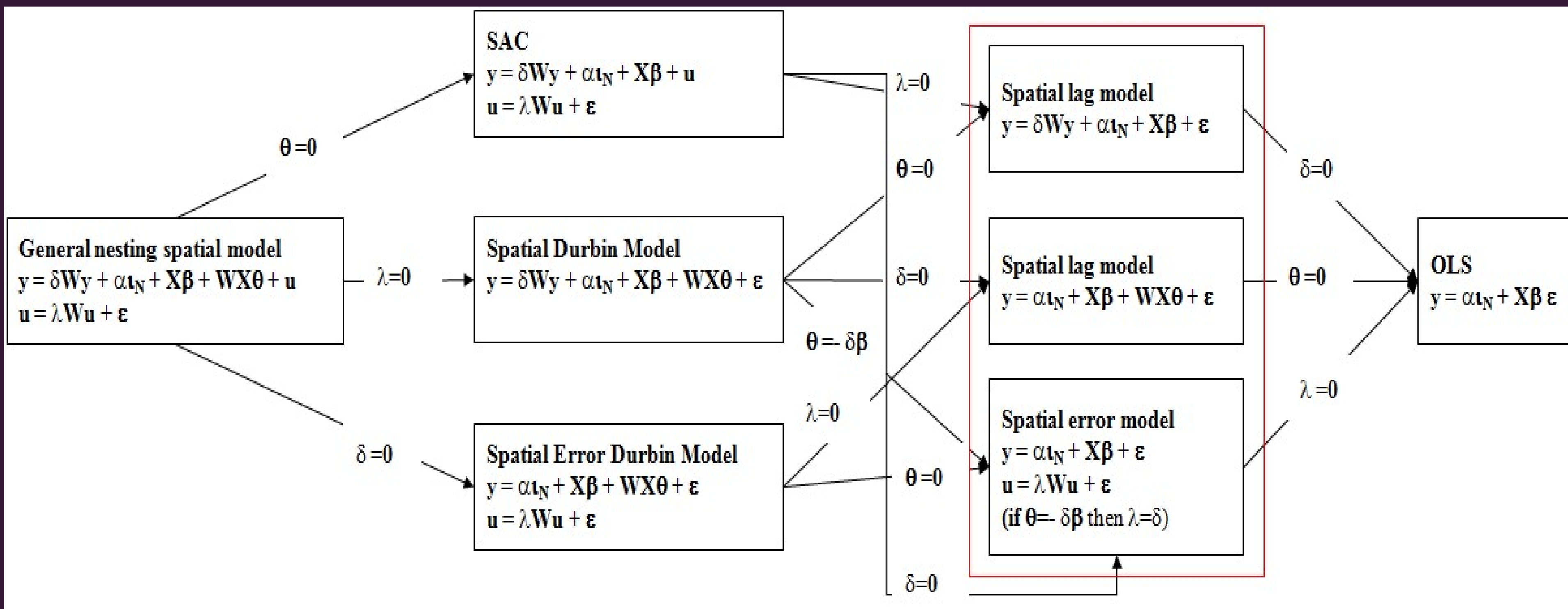


Figure 5.1 The relationships between different spatial dependence models for cross-section data (source Halleck Vega and Elhorst 2012)