Spatial Data Modelling

Ciencia de Datos para la toma de decisiones en Economía

Ignacio Sarmiento-Barbieri

Universidad de los Andes

Recap

- Types of Spatial Data
- Reading and Mapping spatial data in R
- Projections
- Creating Spatial Objects
- Measuring Distances

Agenda

Types of Spatial Data

Spatial data comes in many “shapes” and “sizes”, the most common types of spatial data are:

- Points are the most basic form of spatial data. Denotes a single point location, such as cities, a GPS reading or any other discrete object defined in space.
- Lines are a set of ordered points, connected by straight line segments
- Polygons denote an area, and can be thought as a sequence of connected points, where the first point is the same as the last
- Grid (Raster) are a collection of points or rectangular cells, organized in a regular lattice

Reading and Mapping spatial data in R

- Spatial data in various formats.
- One of the most used format are `shapefiles`
- This type of files stores non topological geometry and attribute information for the spatial features in a data set
- Files
  - Main file: file.shp
  - Index file: file.shx
  - dBASE table: file.dbf
  - Other files
    - file.prj
    - file.sbn , file.sbx
    - file.shp.xml

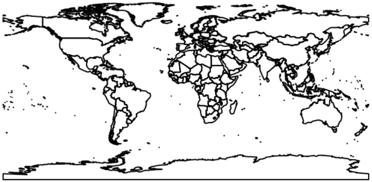
Reading and Mapping spatial data in R

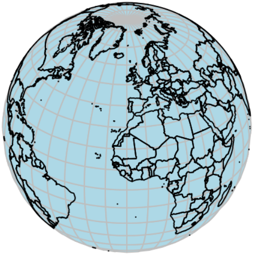
- Another one of the most used format are `json` o `geojson`

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [-74.066391, 4.601590]
  },
  "properties": {
    "name": "Universidad de Los Andes"
  }
}
```

The earth ain't flat

- The world is an irregularly shaped ellipsoid, but plotting devices are flat
- But if you want to show it on a flat map you need a map projection,
- This will determine how to transform and distort latitudes and longitudes to preserve some of the map properties: area, shape, distance, direction or bearing





The earth ain't flat

- For example, sailors use Mercator projection where meridians and parallels cross each other always at the same 90 degrees angle.
- It allows to easy locate your self on the line showing direction in which you sail
- But the projection do not preserve distances

Source: https://www.geoawesomeness.com/all-map-projections-in-compared-and-visualized/

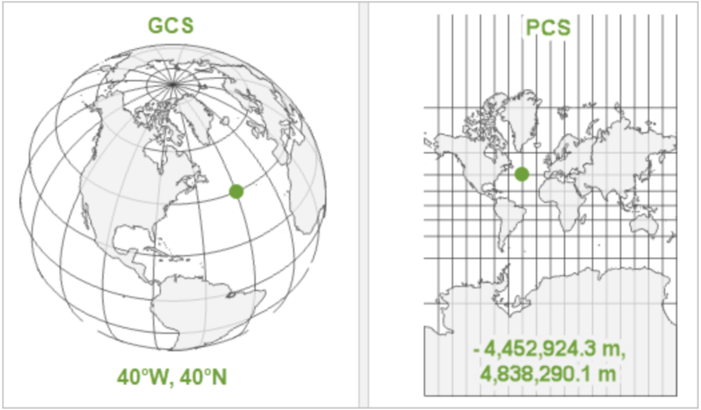
Which projection should I choose?

- "There exist no all-purpose projections, all involve distortion when far from the center of the specified frame" (Bivand, Pebesma, and Gómez-Rubio 2013)
- In some cases, it is not something that we are free to decide: "often the choice of projection is made by a public mapping agency" (Bivand, Pebesma, and Gómez-Rubio 2013).
- This means that when working with local data sources, it is likely preferable to work with the CRS in which the data was provided.

Which projection should I choose?

- Geographic coordinate systems: coordinate systems that span the entire globe (e.g. latitude / longitude).
  - For geographic CRSs, the answer is often WGS84
  - WGS84 is the most common CRS in the world, EPSG code: 4326.
  - For Bogotá the IGAC promotes the adoption of MAGNA-SIRGAS. EPSG code: 4626
- Projected coordinate systems: coordinate systems that are localized to minimize visual distortion in a particular region (e.g. Robinson, UTM, State Plane)

Which projection should I choose?

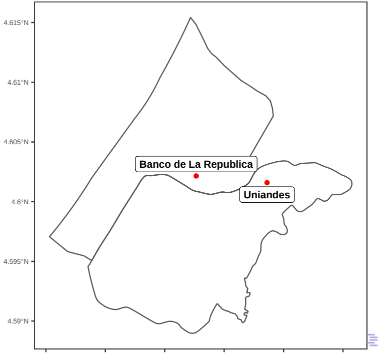


Source: https://www.geoawesomeness.com/all-map-projections-in-compared-and-visualized/

Creating Spatial Objects

```
db<-data.frame(place=c("Uniandes","Banco de La Republica"),
    lat=c(4.601590,4.602151),
    long=c(-74.066391,-74.072350),
    nudge_y=c(-0.001,0.001))
db<-db %>% mutate(latp=lat,longp=long)
db<-st_as_sf(db,coords=c('longp','latp'),crs=4326)
```

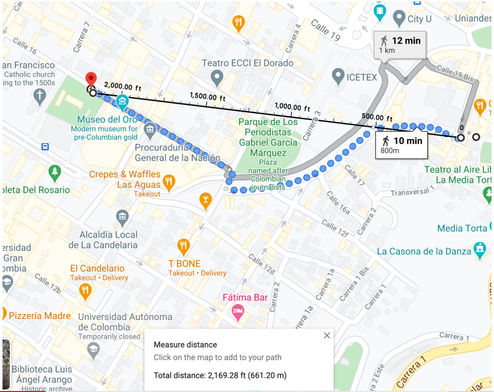
```
ggplot()+
  geom_sf(data=upla
  %>% filter(UPlNombre
  %in%c("LA CANDELARIA","LAS NIEVES")), fill = NA) +
  geom_sf(data=db, col="red") +
  geom_label(data = db, aes(x = long, y = lat,
        label = place),
        size = 3, col = "black", fontface = "bold",
        nudge_y =db$nudge_y) +
  theme_bw() +
  theme(axis.title =element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.text = element_text(size=6))
```



Measuring Distances

```r
st_distance(db)
```

```
## Units: [m]
##           [,1]     [,2]
## [1,]    0.0000 664.1323
## [2,] 664.1323   0.0000
```



Measuring Distances

```
st_distance(db,ciclovias)
Error in st_distance(db, ciclovias) : st_crs(x) == st_crs(y) is not TRUE
st_crs(ciclovias)
```

```
## Coordinate Reference System:
##   User input: 3857
##   wkt:
## PROJCS["WGS 84 / Pseudo-Mercator",
##     GEOGCS["WGS 84",
##         DATUM["WGS_1984",
##             SPHEROID["WGS 84",6378137,298.257223563,
##                 AUTHORITY["EPSG","7030"]],
##             AUTHORITY["EPSG","6326"]],
##         PRIMEM["Greenwich",0,
##             AUTHORITY["EPSG","8901"]],
##         UNIT["degree",0.0174532925199433,
##             AUTHORITY["EPSG","9122"]],
##         AUTHORITY["EPSG","4326"]],
##     PROJECTION["Mercator_1SP"],
##     PARAMETER["central_meridian",0],
##     PARAMETER["scale_factor",1],
##     PARAMETER["false_easting",0],
##     PARAMETER["false_northing",0],
##     UNIT["metre",1,
##         AUTHORITY["EPSG","9001"]],
##     AXIS["X",EAST],
##     AXIS["Y",NORTH],
##     EXTENSION["PROJ4","+proj=merc +a=6378137 +b=6378137 +lat_ts=0.0 +lon_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +units=m +nadgrids=@null +wktext +r
##     AUTHORITY["EPSG","3857"]]
```

Measuring Distances

```r
ciclovias<-st_transform(ciclovias, 4686)
st_crs(ciclovias)
```

```
## Coordinate Reference System:
##   User input: EPSG:4686
##   wkt:
## GEOGCS["MAGNA-SIRGAS",
##     DATUM["Marco_Geocentrico_Nacional_de_Referencia",
##         SPHEROID["GRS 1980",6378137,298.257222101,
##             AUTHORITY["EPSG","7019"]],
##         TOWGS84[0,0,0,0,0,0,0],
##         AUTHORITY["EPSG","6686"]],
##     PRIMEM["Greenwich",0,
##         AUTHORITY["EPSG","8901"]],
##     UNIT["degree",0.0174532925199433,
##         AUTHORITY["EPSG","9122"]],
##     AUTHORITY["EPSG","4686"]]
```

Measuring Distances

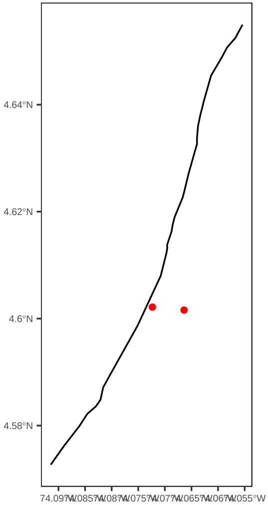
```r
db<-st_transform(db, 4686)
st_distance(db,ciclovias)
```

```
## Units: [m]
##          [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]
## [1,] 9514.617 10789.90 6035.283 12855.90 6025.017 8311.922 4579.450 741.6047
## [2,] 9221.998 10686.39 6143.960 13004.84 5871.073 7656.183 4014.993 116.5939
##           [,9]    [,10]    [,11]    [,12]    [,13]    [,14]
## [1,] 1002.8751 6255.692 2385.125 8402.580 8669.030 3788.265
## [2,]  981.1991 5839.565 2425.508 7738.774 8048.108 3436.819
```

Measuring Distances

```
ciclovias_sp<-ciclovias[8,]

ggplot()+
  geom_sf(data=ciclovias[8,], fill = NA) +
  geom_sf(data=db, col="red") +
  theme_bw() +
  theme(axis.title =element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.text = element_text(size=6))
```



Motivation

Cross-sectional iid non-spatial data

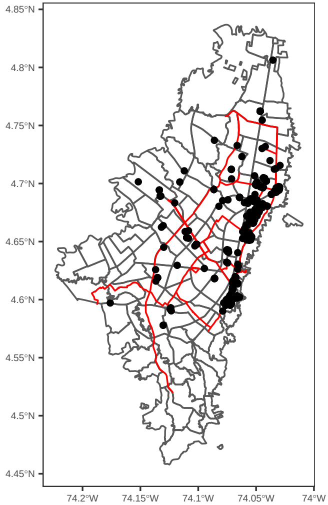
Standard cross-sectional models

$$y_i = X_i\beta + \epsilon_i \tag{1}$$

$$i = 1, \ldots, n \tag{2}$$

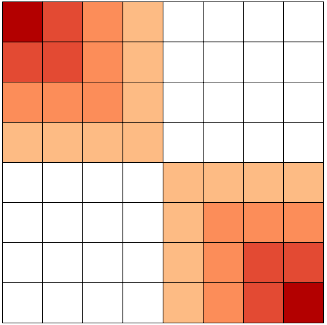
Motivation

- Independence assumption between observation is no longer valid
- Attributes of observation i may influence the attributes of observation j.
- We will consider various alternatives to model spatial dependence



Spatial Dependence

- We now take a closer look at spatial dependence, or to be more precise on it's weaker expression spatial (auto)correlation.
- Spatial autocorrelation measures the degree to which a phenomenon of interest is correlated to itself in space (Cliff and Ord (1973)).
- For example, positive spatial correlation arises when units that are *close* to one another are more similar than units that are far apart



Spatial Dependence

- Anselin and Bera (1998) tell us that we can express the existence of spatial autocorrelation with the following moment condition:

$$Cov(y_i, y_j) \neq 0 \quad for \quad i \neq j \tag{3}$$

  were y_i and y_j are observations on a random variable at locations i and j.

- The problem here is that we need to estimate N by N covariance terms directly for N observations.

- To overcome this problem we impose restrictions on the nature of the interactions.

Closeness

"Everything is related to everything else, but close things are more related than things that are far apart" (Tobler, 1979).

- One of the major differences between standard econometrics and standard spatial econometrics lies, in the fact that, in order to treat spatial data, we need to use two different sets of information

  1. Observed values of the variables
  2. Particular location where those variables are observed and to the various links of proximity between all spatial observations

Weights Matrix

- At the heart of traditional spatial econometrics is the definition of the *weights matrix*:

$$W = \begin{pmatrix} w_{11} & \dots & \dots & w_{n1} \\ \vdots & w_{ij} & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{n_1} & \dots & \dots & w_{nn} \end{pmatrix}_{n\times n} \tag{4}$$

with generic element:

$$w_{ij} = \begin{cases} 1 & \mathit{if}\ j \in N(i) \\ 0 & o.w \end{cases} \tag{5}$$

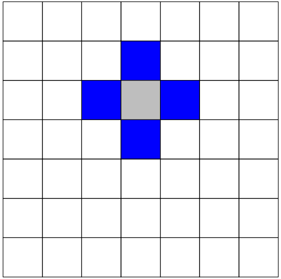
$N(i)$ being the set of neighbors of location j. By convention, the diagonal elements are set to zero, i.e. $w_{ii} = 0$.

Weights Matrix

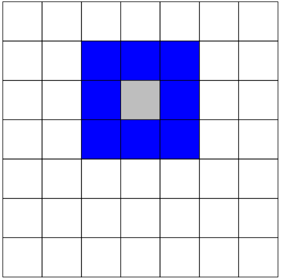
- The specification of the neighboring set ($N(i)$) is quite arbitrary and there's a wide range of suggestions in the literature.
  - Adjacency
    - Rook criterion
    - Queen criterion
  - Two observations are neighbors if they are within a certain distance, i.e., $j \in N(j)$ if $d_{ij} < d_{max}$ where d is the distance between location i and j.
  - Closest neighbor, ties can be solved randomly
  - More general matrices can also be specified by considering entries of w_{ij} as functions of geographical, economic or social distances between areas rather than simply characterized by dichotomous entries

Weights Matrix

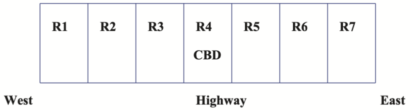
Rook criterion: two units are close to one another if they share a side



Queen criterion: two units are close if they share a side or an edge.



Some Examples of Weights Matrices

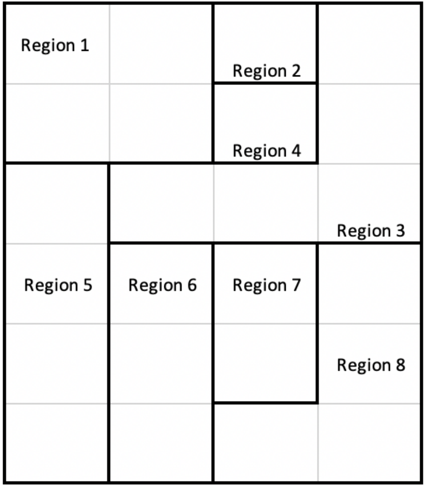


Fuente: LeSage & Pace (2009)

Adjacency Criterion

$$
W = \begin{array}{c|ccccccc}
 & R1 & R2 & R3 & R4 & R5 & R6 & R7 \\
\hline
R1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
R2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
R3 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
R4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
R5 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
R6 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
R7 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array} \tag{6}
$$

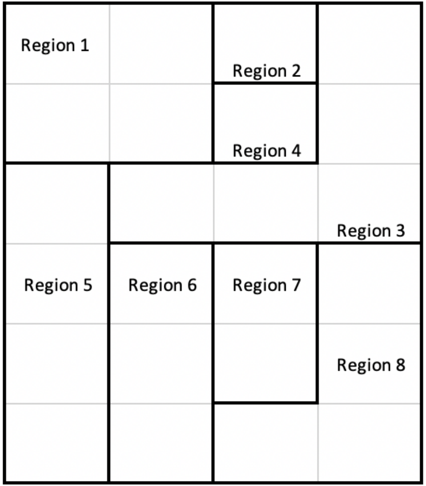
Some Examples of Weights Matrices



Adjacency Criterion

$$W = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}_{8\times 8}$$

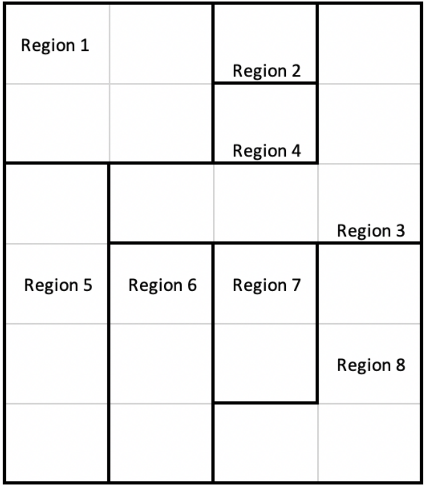
Some Examples of Weights Matrices



Nearest Neighbor

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{8\times 8}$$

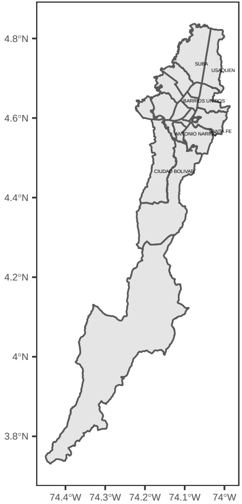
Some Examples of Weights Matrices



Distance < 2

$$W = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{8\times 8}$$

Some Examples of Weights Matrices



Some Examples of Weights Matrices

	ANTONIO NARIÑO	TUNJUELITO	RAFAEL URIBE URIBE	CANDELARIA	BARRIOS UNIDOS	TEUSAQUILLO	PUENTE ARANDA	LOS MARTIRES	SUMAPAZ	USAQUEN	CHAPINERO	SANTA FE	SAN CRISTOBAL
ANTONIO NARIÑO	0	1	1	0	0	0	1	1	0	0	0	1	1
TUNJUELITO	1	0	1	0	0	0	1	0	0	0	0	0	0
RAFAEL URIBE URIBE	1	1	0	0	0	0	0	0	0	0	0	0	1
CANDELARIA	0	0	0	0	0	0	0	0	0	0	0	1	0
BARRIOS UNIDOS	0	0	0	0	0	1	0	0	0	1	1	0	0
TEUSAQUILLO	0	0	0	0	1	0	1	1	0	0	1	1	0
PUENTE ARANDA	1	1	0	0	0	1	0	1	0	0	0	0	0
LOS MARTIRES	1	0	0	0	0	1	1	0	0	0	0	1	0
SUMAPAZ	0	0	0	0	0	0	0	0	0	0	0	0	0
USAQUEN	0	0	0	0	1	0	0	0	0	0	1	0	0
CHAPINERO	0	0	0	0	1	1	0	0	0	1	0	1	0
SANTA FE	1	0	0	1	0	1	0	1	0	0	1	0	1
SAN CRISTOBAL	1	0	1	0	0	0	0	0	0	0	0	1	0
USME	0	1	1	0	0	0	0	0	1	0	0	0	1
CIUDAD BOLIVAR	0	1	0	0	0	0	0	0	0	0	0	0	0
BOSA	0	0	0	0	0	0	0	0	0	0	0	0	0
KENNEDY	0	1	0	0	0	0	1	0	0	0	0	0	0
FONTIBON	0	0	0	0	0	1	1	0	0	0	0	0	0
ENGATIVA	0	0	0	0	1	1	0	0	0	0	0	0	0
SUBA	0	0	0	0	1	0	0	0	0	1	1	0	0

Some Examples of Weights Matrices

Quite often the W matrices are standardized to sum to one in each row

$$w^*_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \tag{7}$$

This can be quite useful since $L(y) = W^* y$ in which each single element is equal to

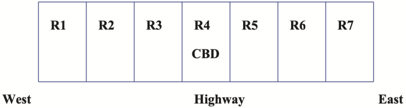
$$L(y_i) = \sum_{j=1}^{n} w^*_{ij} y_j \tag{8}$$

$$= \sum_{j=1}^{n} \frac{w_{ij} y_j}{\sum_{j=1}^{n} w_{ij}}$$

$$= \frac{\sum_{j \in N(i)} y_j}{\#N(i)} \tag{9}$$

Some Examples of Weights Matrices

Quite often the W matrices are standardized to sum to one in each row



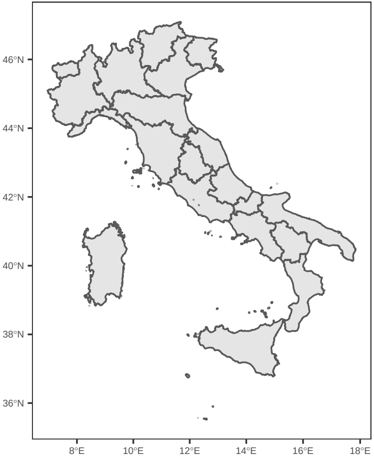
Fuente: LeSage & Pace (2009)

$$W = \begin{array}{c} \\ R1 \\ R2 \\ R3 \\ R4 \\ R5 \\ R6 \\ R7 \end{array} \begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \tag{10}$$

Some Examples of Weights Matrices

	ANTONIO NARIÑO	TUNJUELITO	RAFAEL URIBE URIBE	CANDELARIA	BARRIOS UNIDOS	TEUSAQUILLO	PUENTE ARANDA	LOS MARTIRES	SUMAPAZ	USAQUEN	CHAPINERO	SANTA FE	SAN CRISTOBAL	
ANTONIO NARIÑO	0.0000000	0.1666667	0.1666667	0.0000000	0.0000000	0.0000000	0.1666667	0.1666667	0.0	0.00	0.0000000	0.1666667	0.1666667	0
TUNJUELITO	0.1666667	0.0000000	0.1666667	0.0000000	0.0000000	0.0000000	0.1666667	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
RAFAEL URIBE URIBE	0.2500000	0.2500000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.2500000	0
CANDELARIA	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	1.0000000	0.0000000	0
BARRIOS UNIDOS	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2000000	0.0000000	0.0000000	0.0	0.20	0.2000000	0.0000000	0.0000000	0
TEUSAQUILLO	0.0000000	0.0000000	0.0000000	0.0000000	0.1428571	0.0000000	0.1428571	0.1428571	0.0	0.00	0.1428571	0.1428571	0.0000000	0
PUENTE ARANDA	0.1666667	0.1666667	0.0000000	0.0000000	0.0000000	0.1666667	0.0000000	0.1666667	0.0	0.00	0.0000000	0.0000000	0.0000000	0
LOS MARTIRES	0.2500000	0.0000000	0.0000000	0.0000000	0.0000000	0.2500000	0.2500000	0.0000000	0.0	0.00	0.0000000	0.2500000	0.0000000	0
SUMAPAZ	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	1
USAQUEN	0.0000000	0.0000000	0.0000000	0.0000000	0.3333333	0.0000000	0.0000000	0.0000000	0.0	0.00	0.3333333	0.0000000	0.0000000	0
CHAPINERO	0.0000000	0.0000000	0.0000000	0.0000000	0.2000000	0.2000000	0.0000000	0.0000000	0.0	0.20	0.0000000	0.2000000	0.0000000	0
SANTA FE	0.1666667	0.0000000	0.0000000	0.1666667	0.0000000	0.1666667	0.0000000	0.1666667	0.0	0.00	0.1666667	0.0000000	0.1666667	0
SAN CRISTOBAL	0.2500000	0.0000000	0.2500000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.2500000	0.0000000	0
USME	0.0000000	0.2000000	0.2000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2	0.00	0.0000000	0.0000000	0.2000000	0
CIUDAD BOLIVAR	0.0000000	0.2500000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
BOSA	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
KENNEDY	0.0000000	0.2000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
FONTIBON	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2500000	0.2500000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
ENGATIVA	0.0000000	0.0000000	0.0000000	0.0000000	0.2500000	0.2500000	0.0000000	0.0000000	0.0	0.00	0.0000000	0.0000000	0.0000000	0
SUBA	0.0000000	0.0000000	0.0000000	0.0000000	0.2500000	0.0000000	0.0000000	0.0000000	0.0	0.25	0.2500000	0.0000000	0.0000000	0

Some Examples of Weights Matrices



Some Examples of Weights Matrices

	Piemonte	Valle D'Aosta	Lombardia	Trentino-Alto Adige	Veneto	Friuli Venezia Giulia	Liguria	Emilia-Romagna	Toscana	Umbria	Marche
Piemonte	0.0000000	0.25	0.2500000	0.00	0.0000000	0.00	0.2500000	0.2500000	0.0000000	0.0000000	0.0000000
Valle D'Aosta	1.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Lombardia	0.2500000	0.00	0.0000000	0.25	0.2500000	0.00	0.0000000	0.2500000	0.0000000	0.0000000	0.0000000
Trentino-Alto Adige	0.0000000	0.00	0.5000000	0.00	0.5000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Veneto	0.0000000	0.00	0.2500000	0.25	0.0000000	0.25	0.0000000	0.2500000	0.0000000	0.0000000	0.0000000
Friuli Venezia Giulia	0.0000000	0.00	0.0000000	0.00	1.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Liguria	0.3333333	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.3333333	0.3333333	0.0000000	0.0000000
Emilia-Romagna	0.1666667	0.00	0.1666667	0.00	0.1666667	0.00	0.1666667	0.0000000	0.1666667	0.0000000	0.1666667
Toscana	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.2000000	0.2000000	0.0000000	0.2000000	0.2000000
Umbria	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.3333333	0.0000000	0.3333333
Marche	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.2000000	0.2000000	0.2000000	0.0000000
Lazio	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.1666667	0.1666667	0.1666667
Abruzzo	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.3333333
Molise	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Campania	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Puglia	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Basilicata	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Calabria	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Sicilia	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Sardegna	0.0000000	0.00	0.0000000	0.00	0.0000000	0.00	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
Piemonte	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Valle D'Aosta	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Lombardia	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Trentino-Alto Adige	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Veneto	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Friuli Venezia Giulia	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Liguria	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Emilia-Romagna	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Toscana	0.2000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Umbria	0.3333333	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Marche	0.2000000	0.2000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Lazio	0.0000000	0.1666667	0.1666667	0.1666667	0.0000000	0.0000000	0.0000000	0	0
Abruzzo	0.3333333	0.0000000	0.3333333	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Molise	0.2500000	0.2500000	0.0000000	0.2500000	0.2500000	0.0000000	0.0000000	0	0
Campania	0.2500000	0.0000000	0.2500000	0.0000000	0.2500000	0.2500000	0.0000000	0	0
Puglia	0.0000000	0.0000000	0.3333333	0.3333333	0.0000000	0.3333333	0.0000000	0	0
Basilicata	0.0000000	0.0000000	0.0000000	0.3333333	0.3333333	0.0000000	0.3333333	0	0
Calabria	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	1.0000000	0.0000000	0	0
Sicilia	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0
Sardegna	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0

Weights Matrix in R

```r
require("sf")
require("spdep")
require("dplyr")
```

```r
chi.poly<-read_sf("foreclosures/foreclosures.shp")
st_crs(chi.poly) #doesn't have a projection
```

```
## Coordinate Reference System: NA
```

```r
st_crs(chi.poly)<-4326 #WGS84 set it in the map
```

Weights Matrix in R

```r
chi.poly<-st_transform(chi.poly,26916) #reproject planarly
#NAD83 UTM Zone 16N
st_crs(chi.poly)
```

```
## Coordinate Reference System:
##   User input: EPSG:26916
##   wkt:
## PROJCS["NAD83 / UTM zone 16N",
##     GEOGCS["NAD83",
##         DATUM["North_American_Datum_1983",
##             SPHEROID["GRS 1980",6378137,298.257222101,
##                 AUTHORITY["EPSG","7019"]],
##             TOWGS84[0,0,0,0,0,0,0],
##             AUTHORITY["EPSG","6269"]],
##         PRIMEM["Greenwich",0,
##             AUTHORITY["EPSG","8901"]],
##         UNIT["degree",0.0174532925199433,
##             AUTHORITY["EPSG","9122"]],
##         AUTHORITY["EPSG","4269"]],
##     PROJECTION["Transverse_Mercator"],
##     PARAMETER["latitude_of_origin",0],
##     PARAMETER["central_meridian",-87],
##     PARAMETER["scale_factor",0.9996],
##     PARAMETER["false_easting",500000],
##     PARAMETER["false_northing",0],
##     UNIT["metre",1,
##         AUTHORITY["EPSG","9001"]],
##     AXIS["Easting",EAST],
##     AXIS["Northing",NORTH],
##     AUTHORITY["EPSG","26916"]]
```

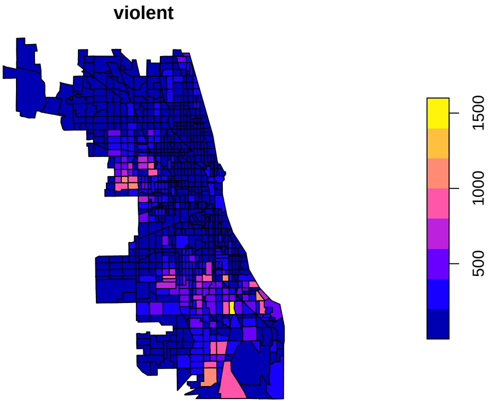
Weights Matrix in R

```r
str(chi.poly)
```

```
## tibble [897 x 17] (S3: sf/tbl_df/tbl/data.frame)
##  $ SP_ID     : chr [1:897] "1" "2" "3" "4" ...
##  $ fips      : chr [1:897] "17031010100" "17031010200" "17031010300" "17031010400" ...
##  $ est_fcs   : int [1:897] 43 129 55 21 64 56 107 43 7 51 ...
##  $ est_mtgs  : int [1:897] 904 2122 1151 574 1427 1241 1959 830 208 928 ...
##  $ est_fcs_rt: num [1:897] 4.76 6.08 4.78 3.66 4.48 4.51 5.46 5.18 3.37 5.5 ...
##  $ res_addr  : int [1:897] 2530 3947 3204 2306 5485 2994 3701 1694 443 1552 ...
##  $ est_90d_va: num [1:897] 12.61 12.36 10.46 5.03 8.44 ...
##  $ bls_unemp : num [1:897] 8.16 8.16 8.16 8.16 8.16 8.16 8.16 8.16 8.16 8.16 ...
##  $ county    : chr [1:897] "Cook County" "Cook County" "Cook County" "Cook County" ...
##  $ fips_num  : num [1:897] 1.7e+10 1.7e+10 1.7e+10 1.7e+10 1.7e+10 ...
##  $ totpop    : int [1:897] 5391 10706 6649 5325 10944 7178 10799 5403 1089 3634 ...
##  $ tothu     : int [1:897] 2557 3981 3281 2464 5843 3136 3875 1768 453 1555 ...
##  $ huage     : int [1:897] 61 53 56 60 54 58 48 57 61 48 ...
##  $ oomedval  : int [1:897] 169900 147000 119800 151500 143600 145900 153400 170500 215900 114700 ...
##  $ property  : num [1:897] 646 914 478 509 641 612 678 332 147 351 ...
##  $ violent   : num [1:897] 433 421 235 159 240 266 272 146 78 84 ...
##  $ geometry  :sfc_POLYGON of length 897; first list element: List of 1
##   ..$ : num [1:15, 1:2] 443923 444329 444814 444839 444935 ...
##   ..- attr(*, "class")= chr [1:3] "XY" "POLYGON" "sfg"
##  - attr(*, "sf_column")= chr "geometry"
##  - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",..: NA NA NA NA NA NA NA NA NA NA ...
##   ..- attr(*, "names")= chr [1:16] "SP_ID" "fips" "est_fcs" "est_mtgs" ...
```

Weights Matrix in R

```r
plot(chi.poly['violent'])
```



Weights Matrix in R

```
list.queen<-poly2nb(chi.poly, queen=TRUE)
W<-nb2listw(list.queen, style="W", zero.policy=TRUE)
W
```

```
## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 897
## Number of nonzero links: 6140
## Percentage nonzero weights: 0.7631036
## Average number of links: 6.845039
##
## Weights style: W
## Weights constants summary:
##     n     nn  S0       S1       S2
## W 897 804609 897 274.4893 3640.864
```

Weights Matrix in R

```
plot(W,st_geometry(st_centroid(chi.poly)))
```

Weights Matrix in R

```
coords <- st_centroid(st_geometry(chi.poly), of_largest_polygon=TRUE)
W_dist<-dnearneigh(coords,0,1000)
W_dist
```

```
## Neighbour list object:
## Number of regions: 897
## Number of nonzero links: 5448
## Percentage nonzero weights: 0.6770991
## Average number of links: 6.073579
## 55 regions with no links:
## 141 142 143 145 153 154 155 158 462 631 637 638 642 643 644 645 655 656 657 658 659 758 759 769 820 821 822 823 824 855 856 857 861 862 8
```

```
plot(W_dist, coords)
```

Weights Matrix in R

```
W_dist<-dnearneigh(coords,0,4300)
W_dist
```

```
## Neighbour list object:
## Number of regions: 897
## Number of nonzero links: 87988
## Percentage nonzero weights: 10.9355
## Average number of links: 98.09142
```

```
plot(W_dist, coords)
```

Testing for Spatial Dependence

$$y = X\beta + \epsilon$$

```
chi.ols<-lm(violent~est_fcs_rt+bls_unemp, data=chi.poly)
```

```
summary(chi.ols)
```

```
##
## Call:
## lm(formula = violent ~ est_fcs_rt + bls_unemp, data = chi.poly)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -892.02  -77.02  -23.73   41.90 1238.22
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -18.627     45.366  -0.411    0.681
## est_fcs_rt    28.298      1.435  19.720   <2e-16 ***
## bls_unemp     -0.308      5.770  -0.053    0.957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Testing for Spatial Dependence

- We can use the OLS reiduals to test for spatial correlation.
- The most basic one is Moran's I test (1950), a test statistics for the null of uncorrelation among regression residuals.

$$I = \left(\frac{e'We}{e'e}\right) \tag{11}$$

- where $e = y - X\beta$ is a vector of OLS residuals $\beta = (X'X)^{-1}X'y$, W is the row standardized spatial weights matrix
- Moran's I test was originally developed as a two-dimensional analog of Durbin-Watson's test

Testing for Spatial Dependence

```
moran.lm<-lm.morantest(chi.ols, W, alternative="two.sided")
print(moran.lm)
```

```
##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = violent ~ est_fcs_rt + bls_unemp, data =
## chi.poly)
## weights: W
##
## Moran I statistic standard deviate = 11.785, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation         Variance
##     0.2142252370    -0.0020099108     0.0003366648
```

Spatial Lag Model

- We can think of situations where values observed at one location or region, say observation i, depend on the values of neighboring observations at nearby locations.

$$y_i = \rho_i y_j + X_i \beta + \epsilon_i \tag{12}$$
$$y_j = \rho_j y_i + X_j \beta + \epsilon_j \tag{13}$$

- This situation suggests a simultaneous data generating process, where the value taken by y_i depends on that of y_j and vice versa.

Spatial Lag Model

- Spatial lag dependence in a regression setting can be modeled similar to an autoregressive process in time series. Formally,

$$y = \rho W y + X\beta + u$$

- Wy induces a nonzero correlation with the error term, similar to the presence of an endogenous variable.
- Unlike to time series, Wy_i is always correlated with u
- OLS estimates in the non spatial model will be biased and inconsistent. (Anselin and Bera, 1998)
- In `R` the function `lagsarlm` uses MLE

Spatial Lag Model

```r
sar.chi<-lagsarlm(violent~est_fcs_rt+bls_unemp, data=chi.poly, W)
summary(sar.chi)
```

```
##
## Call:
## lagsarlm(formula = violent ~ est_fcs_rt + bls_unemp, data = chi.poly,
##     listw = W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -519.127  -65.003  -15.226   36.423 1184.193
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -93.7885    41.3162  -2.270  0.02321
## est_fcs_rt   15.6822     1.5600  10.053  < 2e-16
## bls_unemp     8.8949     5.2447   1.696  0.08989
##
## Rho: 0.49037, LR test value: 141.33, p-value: < 2.22e-16
## Asymptotic standard error: 0.039524
##     z-value: 12.407, p-value: < 2.22e-16
## Wald statistic: 153.93, p-value: < 2.22e-16
##
## Log likelihood: -5738.047 for lag model
## ML residual variance (sigma squared): 20200, (sigma: 142.13)
## Number of observations: 897
## Number of parameters estimated: 5
## AIC: 11486, (AIC for lm: 11625)
## LM test for residual autocorrelation
```

Spatial Lag Model

	Dependent variable:	
	Violent Crime	
	OLS	*SAR*
	(1)	(2)
Foreclosures	28.298***	15.682***
	(1.435)	(1.560)
Unemployment	−0.308	8.895*
	(5.770)	(5.245)
Constant	−18.627	−93.789**
	(45.366)	(41.316)
Observations	897	897
Note:	*p<0.1; **p<0.05; ***p<0.01	

Spatial Lag Model

The model is then

$$y = \rho W y + X\beta + u$$

with $|\rho| < 1$, we also assume that W is exogenous

If W is row standardized:

- Guarantees $|\rho| < 1$ (Anselin, 1982)
- $[0, 1]$ Weights
- Wy Average of neighboring values
- W is no longer symmetric $\sum_j w_{ij} \neq \sum_i w_{ji}$ (complicates computation)

Spatial Lag Model
Maximum Likelihood Estimator

Note that we can write

$$(I - \rho W)y = X\beta + u$$

- We can think this model as a way to correct for loss of information coming from spatial dependence.
- $(1 - \rho W)y$ is a spatially filtered dependent variable, i.e., the effect of spatial autocorrelation taken out

Spatial Lag Model

- In this case, endogeneity emerges because the spatially lagged value of y is correlated with the stochastic disturbance.
- One solution that emerged in the literature is MLE
- We need an extra assumption, i.e., $u \sim_{iid} N(0, \sigma^2 I)$.

Spatial Lag Model

Maximum Likelihood Estimator

The associated likelihood function is then

$$\mathcal{L}\left(\sigma^2, \rho, y\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\sigma^2\Omega|^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(y-(I-\rho W)^{-1}X\beta)'\Omega^{-1}(y-(I-\rho W)^{-1}X\beta)\right\}$$

the log likelihood

$$l\left(\sigma^2, \rho, y\right) = constant - \frac{1}{2}ln|\sigma^2\Omega| - \frac{1}{2\sigma^2}(y-(I-\rho W)^{-1}X\beta)'\Omega^{-1}(y-(I-\rho W)^{-1}X\beta)$$

so returning to the log likelihood we have that the log likelihood is

$$l\left(\sigma^2,\rho,y\right) = constant - \frac{n}{2}ln\left(\sigma^2\right) + ln\left(\left|(I-\rho W)\right|\right) - \frac{1}{2\sigma^2}(y-(I-\rho W)^{-1}X\beta)'(I-\rho W)'(I-\rho W)(y-(I-\rho W)^{-1}X\beta) \tag{14}$$

then

$$\begin{aligned} l\left(\sigma^2,\rho,y\right) = {} & constant - \frac{n}{2}ln\left(\sigma^2\right) \\ & - \frac{1}{2\sigma^2}((I-\rho W)y-X\beta)'((I-\rho W)-X\beta) \\ & + ln\left(\left|(I-\rho W)\right|\right) \end{aligned} \tag{15}$$

Spatial Lag Model

Maximum Likelihood Estimator

- The determinant $|(I - \rho W)|$ is quite complicated because in contrast to the time series, where it is a triangular matrix, here it is a full matrix.
- However, Ord (1975) showed that it can be expressed as a function of the eigenvalues ω_i

$$|(I - \rho W)| = \Pi_{i=1}^{n}(1 - \rho\omega_i)$$

So the log likelihood is simplified to

$$\begin{aligned} l\left(\sigma^2, \rho, y\right) = constant - \frac{n}{2} ln\left(\sigma^2\right) \\ - \frac{1}{2\sigma^2}((I - \rho W)y - X\beta)'((I - \rho W)y - X\beta) \\ + \sum ln(1 - \rho\omega_i) \end{aligned} \tag{16}$$

Applying FOC, the ML estimates for β and σ^2 are:

$$\beta_{MLE} = (X'X)^{-1}X'(I - \rho W)y$$

$$\sigma^2_{MLE} = \frac{1}{n}(y - \rho Xy - X\beta_{MLE})'(y - \rho Xy - X\beta_{MLE})$$

- Conditional on ρ these estimates are simply OLS applied to the spatially filtered dependent variable and explanatory variables X.

Spatial Lag Model

Maximum Likelihood Estimator

- Substituting these in the log likelihood we have a concentrated log-likelihood as a nonlinear function of a single parameter ρ

$$l(\rho) = -\frac{n}{2} ln \left(\frac{1}{n} (e_0 - \rho e_L)'(e_0 - \rho e_L) \right) + \sum ln(1 - \rho \omega_i) \tag{17}$$

- where e_0 are the residuals in a regression of y on X and
- e_L of a regression of Wy on X.

- This expression can be maximized numerically to obtain the estimators for the unknown parameters ρ.

Spatial Lag Model

Two-Stage Least Squares estimators

- An alternative to MLE we can us 2SLS to eliminate endogeneity.
- Key is to identify proper instruments
    - Need to be uncorrelated with the error term
    - Correlated with WY

Spatial Lag Model

Two-Stage Least Squares estimators

Consider the following

$$E(y) = (I - \rho W)^{-1} X\beta$$

now, since $|\rho| < 1$ we can use Neumann series property to expand the inverse matrix as

$$(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + ...$$

hence

$$E(y) = (I + \rho W + \rho^2 W^2 + \rho^3 W^3 + ...)X\beta$$

$$= X\beta + \rho W X\beta + \rho^2 W^2 X\beta + \rho^3 W^3 X\beta + ...$$

so we can express $E(y)$ as a function of X, WX, W^2X,...

Spatial Lag Model

Two-Stage Least Squares estimators

We can use the first three elements of the expansion as instruments. Let's define H as the matrix with our instruments

$$H = [X, WX, W^2X]$$

Interpretation of Parameters

- Consider the following model for the $i-th$ observation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_r x_{ir} + \cdots + \beta_k x_{ik} \;\; i = 1, \ldots, n$$

- Recall that in OLS we have

$$\beta_1 = \frac{\partial y_i}{\partial x_{i1}}$$

- Interpretation is straight forward as long as we take into account units
- In spatial models the interpretation is less immediate and require some clarification

Interpretation of Parameters

- Lets consider the case of a simple Spatial Lag model with a single regressor

$$y_i = \alpha + \beta x_i + \rho \sum w_{ij} y_j + \epsilon_i \tag{18}$$

with $|\rho| < 1$, and

$$\beta \neq \frac{\partial y_i}{\partial x_i}$$

$$\frac{\partial y_i}{\partial x_i} = diag(I - \rho W)^{-1} \beta$$

- The impact depends also on the parameter ρ
- The impact is different in each location

Interpretation of Parameters

More generally consider

$$
\begin{aligned}
y &= \lambda W y + X\beta + u \\
&= (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} u
\end{aligned}
$$

Then

$$E(y) = (I - \lambda W)^{-1} X\beta \tag{19}$$

we define

$$S(W) = (I - \lambda W)^{-1} \beta \tag{20}$$

Interpretation of Parameters

Therefore the impact of *each variable* x on y can be described through the partial derivatives $\frac{\partial E(y)}{\partial x}$ which can be arranged in the following matrix:

$$S(W) = \frac{\partial E(y)}{\partial x} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_1} & \cdots & \frac{\partial E(y_1)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_i)}{\partial x_1} & \cdots & \frac{\partial E(y_i)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_1} & \cdots & \frac{\partial E(y_n)}{\partial x_n} \end{pmatrix} \tag{21}$$

Interpretation of Parameters

On this basis, LeSage and Pace (2009) suggested the following impact measures that can be calculated for each independent variable X_i included in the model

- *Average Direct Impact*: this measure refers to the impact of changes in the $i-th$ observation of x, which we denote x_i, on y_i. This is the average of all diagonal entries in S

$$\begin{aligned} ADI &= \frac{tr(S(W))}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ii} \end{aligned} \tag{22}$$

Interpretation of Parameters

- *Average Total Impact To* an observation: this measure is related to the impact produced on one single observation y_i. For each observation this is calculated as the sum of the $i-th$ row of matrix S

$$
\begin{aligned}
ATIT_j &= \frac{\iota' S(W)}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ij}
\end{aligned}
\tag{23}
$$

Interpretation of Parameters

- *Average Total Impact From* an observation: this measure is related to the total impact on all other observations y_i . For each observation this is calculated as the sum of the $j-th$ column of matrix S

$$
\begin{aligned}
ATIF_i &= \frac{1}{n} S(W) \iota \\
&= \frac{\sum_{j=1}^{n} S(W)_{ij}}{n}
\end{aligned}
\tag{24}
$$

Interpretation of Parameters

- A Global measure of the average impact obtained from the two previous measures.
- It is simply the average of all entries of matrix S

$$ATI = \frac{1}{n}\iota' S(W)\iota = \frac{1}{n}\sum_{i=1}^{n} ATIT_i = \frac{1}{n}\sum_{j=1}^{n} ATIF_i \tag{25}$$

- The numerical values of the summary measures for the two forms of average total impacts are equal.
- The ATIF relates how changes in a single observation j influences all observations.
- In contrast, the ATIT considers how changes in all observations influence a single observation i.

Interpretation of Parameters

- *Average Indirect Impact* obtained as the difference between ATI and ADI

$$AII = ATI - ADI \tag{26}$$

- It is simply the average of all off-diagonal entries of matrix S

Interpretation of Parameters: Example

- We have data on 20 Italian regions on GDP and unemployment.
- We want to estimate the effect of GDP on Unemployment (Okun's Law)

	OLS	Spatial Lag Model
Intercept	10.971***	3.12275***
GDP	-3.326***	-1.13532***
ρ	-	0.7476***
ADI	-	-1.542448
AII	-	-2.95571
ATI	-	-4.498159

Spatial Error Model (SEM)

An OVB motivation

- True GDP

$$y = X\beta + Z\theta$$

- but Z is not observed and $Z \perp X$
- we estimate

$$y = X\beta + \epsilon$$

- if Z has a spatial autoregressive process

$$Z = \rho WZ + r$$

Spatial Error Model (SEM)

An OVB motivation

- Then

$$y = X\beta + (I - \rho W)^{-1}(\theta r)$$

- calling $(\theta r) = u$

$$y = X\beta + (I - \rho W)^{-1}u$$

- β will be unbiased but inefficient

Spatial Error Model (SEM)

The model is now

$$y = X\beta + u$$

with

$$u = \rho X u + \epsilon$$

with $|\rho| < 1$, we also assume that W is exogenous We can estimate this by MLE or FGLS

Further Readings

- Arbia, G. (2014). A primer for spatial econometrics with applications in R. Palgrave Macmillan. (Chapter 2 and 3)
- Anselin, Luc, & Anil K Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." Statistics Textbooks and Monographs 155. MARCEL DEKKER AG: 237–90.
- Sarmiento-Barbieri, I. (2016). An Introduction to Spatial Econometrics in R. http://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html
- Tobler, WR. 1979. "Cellular Geography." In Philosophy in Geography, 379–86. Springer.