INSEIKAI Tohoku BootCamp for Economists

Mathematics II

Spatial Econometrics

Quang-Thanh Tran †

Ye Zhang ‡ §

Graduate School of Economics and Management
Tohoku University

July 3, 2024

†tran.quang.thanh.p1@dc.tohoku.ac.jp ("Tedd")

‡zhang.ye.q8@dc.tohoku.ac.jp ("Tony")

§This material is prepared for the Summer Bootcamp 2024. To access other materials, visit:
Camp homepage: https://thanhqtran.github.io/tohoku_bootcamp/
Github: https://github.com/thanhqtran/tohoku_bootcamp/tree/main

Contents

Preface

This course is based on LeSage and Page (2009) textbook ”Introduction to Spatial Econometrics”. We will cover the first two chapters of the book to give you a brief overview of spatial econometrics. For more advanced topics, please refer to later chapters of the book (e.g. Chapters 7 and 8 for panel data).

For coding section, please refer to Burkeyacademy. The complete set of videos, data and commands can be found at the website `https://spatial.burkeyacademy.com/`.

To prepare for coding, it is recommended to download QGIS for maps, and RStudio / Stata for regression analysis. Matlab materials are also available at: `spatial-econometrics.com` and `spatial-statistics.com`.

Professor Luc Anselin is also a prestigious researcher in the field and he has numerous courses on his website and youtube. You may also read his papers and other researchers' papers to get a more comprehensive understanding on empirical spatial econometrics.

Chapter 1

Introduction

Section 1.1 of this chapter introduces the concept of spatial dependence that often arises in cross-sectional spatial data samples.

Section 1.2 sets forth spatial autoregressive data generating processes for spatially dependent sample data along with spatial weight matrices that play an important role in describing the structure of these processes.

Section 1.3 provides a simple example of how congestion effects lead to spatial spillovers that impact neighbouring regions using travel times to the central business district (CBD) region of a metropolitan area.

1.1 Spatial dependence

The data generating process (DGP) for a conventional cross-sectional non- spatial sample of n independent observations y_i, $i = 1, ..., n$ that are linearly related to explanatory variables in a matrix X takes the form in (1.1).

$$y_i = X_i\beta + \varepsilon_i \tag{1.1}$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, ..., n \tag{1.2}$$

In (1.1), X_i represents a $1 \times k$ vector of covariates or explanatory variables, with associated parameters β contained in a $k \times 1$ vector. Each observation has an underlying mean of $X_i\beta$ and a random component ε_i.

An implication of this for situations where the observations i represent regions or points in space is that observed values at one location (or region) are independent of observations made at other locations (or regions). Independent or statistically independent observations imply that $E(\varepsilon_i\varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$. The assumption of independence greatly simplifies models, but in spatial contexts this simplification seems strained.

In contrast, *spatial dependence* reflects a situation where values observed at one location or region, say observation i, depend on the values of neighbouring observations at nearby locations. Suppose we let observations $i = 1$ and $j = 2$ represent neighbours (perhaps regions with borders that touch), then a data generating process might take the form shown in (1.3).

$$\begin{aligned}
y_i &= \alpha_i y_j + X_i\beta + \varepsilon_i \qquad (1.3)\\
y_j &= \alpha_j y_i + X_j\beta + \varepsilon_j \\
\varepsilon_i &\sim N(0, \sigma^2) \quad i = 1 \\
\varepsilon_j &\sim N(0, \sigma^2) \quad j = 2
\end{aligned}$$

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

As a concrete example, consider the set of seven regions shown in Figure 1.1, which represent three regions to the west and three to the east of a central business district (CBD).

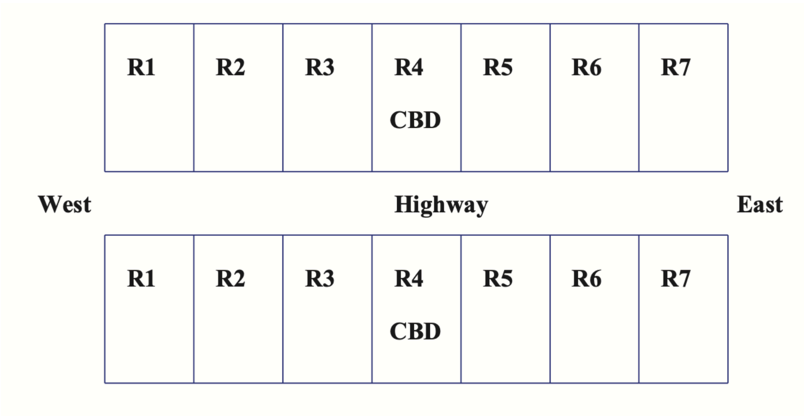


Fig. 1.1. Regions west and east of the CBD

We might observe the following set of sample data for these regions that relates travel times to the CBD (in minutes) contained in the dependent variable vector y to distance (in miles) and population density (population per square block) of the regions in the two columns of the matrix X.

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

The pattern of longer travel times for more distant regions R1 and R7 versus nearer regions R3 and R5 found in the vector y seems to clearly violate independence, since travel times appear similar for neighbouring regions.

However, we might suppose that this pattern is explained by the model variables Distance and Density associated with each region, since these also appear similar for neighbouring regions.

Now, consider that our set of observed travel times represent measurements taken on a particular day, so we have travel times to the CBD averaged over a 24 hour period. In this case, some of the observed pattern might be explained by congestion effects that arise from the shared highway. It seems plausible that longer travel times in one region should lead to longer travel times in neighbouring regions on any given day. This is because commuters pass from one region to another as they travel along the highway to the CBD. Slower times in R3 on a particular day should produce slower times for this day in regions R2 and R1. Congestion effects represent one type of **spatial spillover**, which do not occur simultaneously, but require some time for the traffic delay to arise. From a modelling viewpoint, congestion effects such as these will not be explained by the model variables Distance and Density. These are dynamic feedback effects from travel time on a particular day that impact travel times of neighbouring regions in the short time interval required for the traffic delay to occur. Since the explanatory variable distance would not change from day to day, and population density would change very slowly on a daily time scale, these variables would not be capable of explaining daily delay phenomena. Observed daily variation in travel times would be better explained by relying on travel times from neighbouring regions on that day. This is the situation depicted in (1.3), where we rely on travel time from a neighbouring observation y_j as an explanatory variable for travel time in region i, y_i. Similarly we use y_i to explain region j travel time, y_j.

1.2 Spatial autoregressive process

We could continue in the fashion of (1.3) to generate a larger set of observations as shown in (1.4).

$$\begin{aligned}
y_i &= \alpha_{i,j} y_j + \alpha_{i,k} y_k + X_i \beta + \varepsilon_i \\
y_j &= \alpha_{j,i} y_i + \alpha_{j,k} y_k + X_j \beta + \varepsilon_j \\
y_k &= \alpha_{k,i} y_i + \alpha_{k,j} y_k + X_k \beta + \varepsilon_k \\
\varepsilon_i &\sim N(0, \sigma^2) \quad i = 1 \\
\varepsilon_j &\sim N(0, \sigma^2) \quad j = 2 \\
\varepsilon_k &\sim N(0, \sigma^2) \quad k = 3
\end{aligned} \tag{1.4}$$

It is easy to see that this would be of little practical usefulness, since it would result in a system with many more parameters than observations.

The solution to the over-parameterisation problem is to impose structure on the spatial dependence relations. This structure gives rise to a data generating process known as a *spatial autoregressive process* (1.5).

$$\begin{aligned}
y_i &= \rho \sum_{j=1}^{n} W_{ij} y_j + \varepsilon_i \\
\varepsilon_i &\sim N(0, \sigma^2) \quad i = 1, ..., n
\end{aligned} \tag{1.5}$$

Where we eliminate an intercept term by assuming that the vector of observations on the variable y is in deviations from means form.

The term $\sum_{j=1}^{n} W_{ij} y_j$ is called a *spatial lag*, since it represents a linear combination of values of the variable y constructed from observations/regions that neighbour observation i. This is accomplished by placing elements W_{ij} in the $n \times n$ spatial weight matrix W , such that $\sum_{j=1}^{n} W_{ij} y_j$results in a scalar that represents a linear combination of values taken by neighbouring observations.

We can write a matrix version of the spatial autoregressive process as in (1.6), where we use $N(0, \sigma^2 I_n)$ to denote a zero mean disturbance process that exhibits constant variance σ^2, and zero covariance between observations. This results in the diagonal variance-covariance matrix σI_n, where I_n represents an n-dimensional identity matrix. Expression (1.6) makes it clear that we are describing a relation between the vector y and the vector Wy representing a linear combination of neighbouring values to each observation.

$$
\begin{aligned}
y &= \rho W y + \varepsilon \\
\varepsilon &\sim N(0, \sigma^2 I_n)
\end{aligned}
\tag{1.6}
$$

To illustrate this, we form a 7×7 spatial weight matrix W using the first-order contiguity relations for the seven regions shown in Figure 1.1. We begin by forming a first-order contiguity matrix C shown in (1.7). For row 1 we place a value of 1 in column 2, reflecting the fact that region R2 is first-order contiguous to region R1. All other elements of row 1 receive values of zero. Similarly, for each row we place a 1 in columns associated with first-order contiguous neighbours, resulting in the matrix C shown in (1.7).

$$
C = \begin{array}{cc}
 & \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\
\begin{array}{c} R1 \\ R2 \\ R3 \\ R4 \\ R5 \\ R6 \\ R7 \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
\tag{1.7}
$$

Note that the diagonal elements of the matrix C are zero, so regions are not considered neighbours to themselves. For the purpose of forming a spatial lag or linear combination of values from neighbouring observations, we can normalise the matrix C to have row sums of unity.

This *row-stochastic* matrix which we label W is shown in (1.8), where the term row-stochastic refers to a non-negative matrix having row sums normalised so they equal one.

$$
W = \begin{array}{cc}
 & \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\
\begin{array}{c} R1 \\ R2 \\ R3 \\ R4 \\ R5 \\ R6 \\ R7 \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
\tag{1.8}
$$

The 7×7 matrix W can be multiplied with a 7×1 vector y of values taken by each region to produce a spatial lag vector of the dependent variable vector taking the form Wy. The matrix product Wy works to produce a 7×1 vector representing the value of the spatial lag vector for each observation i, $i = 1, ..., 7$.

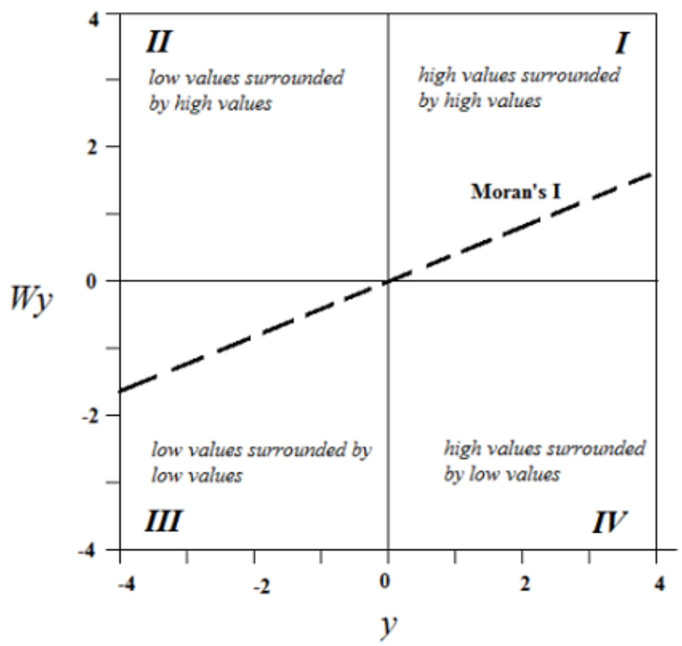
The matrix multiplication process is shown in (1.9), along with the resulting spatial lag vector Wy.

$$Wy = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} \tag{1.9}$$

$$= \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix}$$

The scalar parameter ρ in (1.6) describes the strength of spatial dependence in the sample of observations.

We can check the spatial relationship by plotting y against Wy. Such a scatter plot is called a *Moran scatter plot* and the coefficient is *Moran's I*.



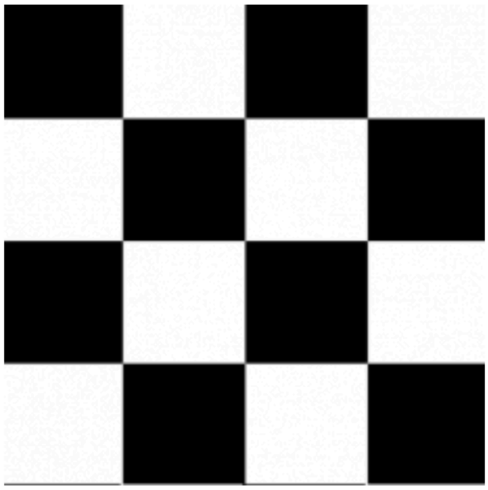
Global Moran's I can be calculated as:

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

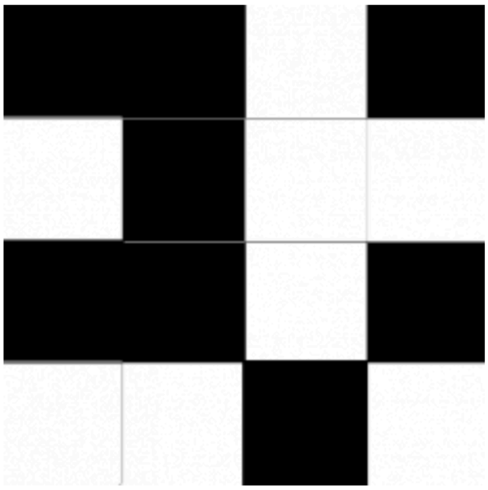
The expected value of Moran's I under the null hypothesis of no spatial autocorrelation is

$$E(I) = \frac{-1}{N-1}$$

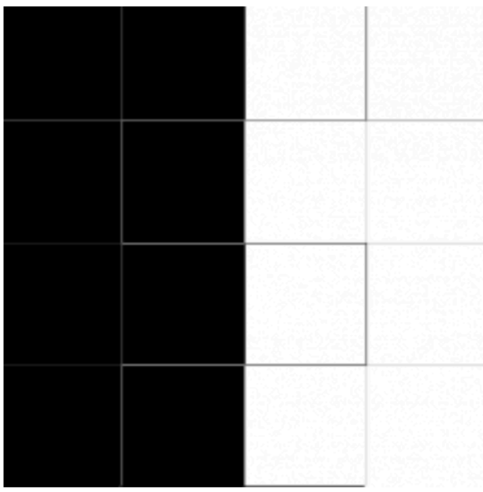
When Moran's I is -1, we have perfect dispersion:



When Moran's I is 0, we have perfect randomness:



When Moran's I is +1, we have perfect clustering:



In this simple example, ρ is similar to Moran's I. Note that it is tempting to interpret

ρ as a conventional correlation coefficient. However, this should be avoided as we will see in the next chapter.

1.2.1 Spatial autoregressive data generating process

The spatial autoregressive process is shown in (1.10) using matrix notation, and the implied data generating process for this type of process is in (1.11). We introduce a constant term vector of ones ι_n, and associated parameter α to accommodate situations where the vector y does not have a mean value of zero.

$$
\begin{aligned}
y &= \alpha\iota_n + \rho W y + \varepsilon && (1.10)\\
(I_n - \rho W)y &= \alpha\iota_n + \varepsilon \\
y &= (I_n - \rho W)^{-1}\iota_n\alpha + (I_n - \rho W)^{-1}\varepsilon && (1.11)\\
\varepsilon &\sim N(0, \sigma^2 I_n)
\end{aligned}
$$

The data generating process statement in (1.11) expresses the simultaneous nature of the spatial autoregressive process. To further explore the nature of this, we can use the following infinite series to express the inverse (Binomial expansion):

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + ... \tag{1.12}$$

where we assume that $|\rho| < 1$.This leads to a spatial autoregressive data generating process for a variable vector y:

$$
\begin{aligned}
y &= (I_n - \rho W)^{-1}\iota_n\alpha + (I_n - \rho W)^{-1}\varepsilon \\
y &= \alpha\iota_n + \rho W\iota_n\alpha + \rho^2 W^2\iota_n\alpha + ... \\
&\quad + \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + ... && (1.13)
\end{aligned}
$$

Expression (1.13) can be simplified since the first part of the infinite series converges to $(1-\rho)^{-1}\iota_n\alpha$. By definition, $W\iota_n = \iota_n$ and therefore $W(W\iota_n) = \iota_n$. This allows us to write:

$$y = \frac{1}{1-\rho}\iota_n\alpha + \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + ... \tag{1.14}$$

To further explore the nature of this data generating process, we consider powers of the row-stochastic spatial weight matrices W^2, W^3, ... that appear in (1.14). For illustration, we present W^2 in (1.15):

$$
W^2 = \begin{pmatrix}
0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0.75 & 0 & 0.25 & 0 & 0 & 0 \\
0.25 & 0 & 0.5 & 0 & 0.25 & 0 & 0 \\
0 & 0.25 & 0 & 0.5 & 0 & 0.25 & 0 \\
0 & 0 & 0.25 & 0 & 0.5 & 0 & 0.25 \\
0 & 0 & 0 & 0.25 & 0 & 0.75 & 0 \\
0 & 0 & 0 & 0 & 0.5 & 0 & 0.5
\end{pmatrix} \tag{1.15}
$$

It can be seen that the diagonal elements are non-zero so that $W^2\varepsilon$ will extract observations from the vector ε that point back to the observation i itself. This is in contrast with our initial independence relation in (1.1), where the Gauss-Markov assumptions rule out dependence of ε_i on other observations j, by assuming zero covariance between observations i and j in the data generating process.

Given that $|\rho| < 1$, the data generating process assigns less influence to disturbance terms associated with higher-order neighbours, with a geometric decay of influence as the order rises.

The dependence of each observation y_i on disturbances associated with neighbouring observations as well as higher-order neighbours suggests a mean and variance-covariance structure for the observations in the vector y that depend in a complicated way on other observations.

It is instructive to consider the mean of the variable y that arises from the spatial autoregressive data generating process in (1.13). Note that we assume the spatial weight matrix is exogenous, or fixed in repeated sampling, so that:

$$\begin{aligned} E(y) &= \frac{1}{1-\rho}\alpha\iota_n + E(\varepsilon) + \rho W E(\varepsilon) + \rho^2 W^2 E(\varepsilon) + ... \\ &= \frac{1}{1-\rho}\alpha\iota_n \end{aligned} \tag{1.16}$$

As demonstrated the conventional regression procedures do not work with spatial models. Therefore we need adopt other techniques to solve for the parameters. This will be Chapter 3 of the textbook but it is beyond the scope of this course. For people interested, please read more on Chapter 3 Maximum Likelihood Estimation.

1.3 Illustration of spatial spillovers

The spatial autoregressive structure can be combined with a conventional regression model to produce a spatial extension of the standard regression model shown in (1.17), with the implied data generating process in (1.18). We will refer to this as the spatial autoregressive model (SAR).

$$y = \rho W y + X\beta + \varepsilon \tag{1.17}$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\varepsilon \tag{1.18}$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

In this model, the parameters to be estimated are the usual regression parameters β, σ and the additional parameter ρ. It is noteworthy that if the scalar parameter ρ takes a value of zero so there is no spatial dependence in the vector of cross-sectional observations y, this yields the least-squares regression model as a special case of the SAR model.

To provide an illustration of how the spatial regression model can be used to quantify spatial spillovers, we reuse the earlier example of travel times to the CBD from the seven regions shown in Figure 1.1. We consider the impact of a change in population density for a single region on travel times to the CBD for all seven regions. Specifically, we double the population density in region R2 and make a prediction of the impact on travel times to the CBD for all seven regions.

For this example we use parameter estimates $\hat{\beta}' = [0.135\ 0.561]$ and $\hat{\rho} = 0.642$. The estimated value of ρ indicates positive spatial dependence in commuting times. Predictions from the model based on the explanatory variables matrix X would take the form:

$$\hat{y}^{(1)} = (I_n - \hat{\rho} W)^{-1} X\hat{\beta}$$

where $r\hat{h}o$, $\hat{\beta}$ are maximum likelihood estimates.

A comparison of predictions $\hat{y}^{(1)}$ from the model with explanatory variables from X and $\hat{y}^{(2)}$ from the model based on $\tilde{X}$ shown in (1.19) is used to illustrate how the model generates spatial spillovers when the population density of a single region changes. The matrix $\tilde{X}$ reflects a doubling of the population density of region R2.

$$\tilde{X} = \begin{pmatrix} 10 & 30 \\ 20 & \mathbf{40} \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix} \tag{1.19}$$

The two sets of predictions $\hat{y}^{(1)}$, $\hat{y}^{(2)}$ are shown in Table 1.1, where we see that the change in region R2 population density has a direct effect that increases the commuting times for residents of region R2 by 4 minutes. It also has an indirect or spillover effect that produces an increase in commuting times for the other six regions. The increase in commuting times for neighbouring regions to the east and west (regions R1 and R3) are the greatest and these spillovers decline as we move to regions in the sample that are located farther away from region R2 where the change in population density occurred.

Table 1.1: Spatial spillovers from changes in Region R2 population density

Regions / Scenario	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)}$ - $\hat{y}^{(1)}$
R1:	42.01	44.58	2.57
R2:	37.06	41.06	4.00
R3:	29.94	31.39	1.45
R4: CBD	26.00	26.54	0.53
R5:	29.94	30.14	0.20
R6:	37.06	37.14	0.07
R7:	42.01	42.06	0.05

Ex. 1.1. I think the $\hat{y}^{(2)}$ values in the book are wrong. Please try calculating yourself and see (But the main findings are still the same).

It is also of interest that the cumulative indirect impacts (spillovers) can be found by adding up the increased commuting times across all other regions (excluding the own-region change in commuting time). This equals $2.57 + 1.45 + 0.53 + 0.20 + 0.07 + 0.05 = 4.87$ minutes, which is larger than the direct (own-region) impact of 4 minutes. The total impact on all residents of the seven region metropolitan area from the change in population density of region R2 is the sum of the direct and indirect effects, or 8.87 minutes increase in travel times to the CBD.

The model literally suggests that the change in population density of region R2 would immediately lead to increases in the observed daily commuting times for all regions. A more palatable interpretation would be that the change in population density would lead over time to a new equilibrium steady state in the relation between daily commuting times and the distance and density variables. The predictions of the direct impacts arising from the change in density reflect $\partial y_i/\partial X_{i2}$, where X_{i2} refers to the ith observation of the second explanatory variable in the model. The cross-partial derivatives $\partial y_j/\partial X_{i2}$ represent indirect effects associated with this change.

To elaborate on this, we note that the DGP for the SAR model can be written as in

(1.20), where the subscript r denotes explanatory variable r,

$$y = \sum_{r=1}^{k} S_r(W)X_r + (I_n - \rho W)^{-1}\varepsilon \tag{1.20}$$

$$E(y) = \sum_{r=1}^{k} S_r(W)X_r \tag{1.21}$$

where $S_r(W) = (I_n - \rho W)^{-1}\beta_r$ acts a multiplier matrix that applies higher-order neighbouring relations to X_r. It follows from (1.21) that:

$$\frac{\partial E(y_i)}{\partial X_{jr}} = S_r(W)_{ij} \tag{1.22}$$

As expression (1.22) indicates, the standard regression interpretation of coefficient estimates as partial derivatives: $\hat{\beta}_r = \partial y/\partial X_r$, no longer holds.

If the DGP for our observed daily travel times is that of the SAR model, least-squares estimates will be biased and inconsistent, since they ignore the spatial lag of the dependent variable.

To see this, note that the estimates for $\hat{\beta}$ from the SAR model takes the form $\hat{\beta} = (X'X)^{-1}X'(I_n - \hat{\rho}W)y$. For our simple illustration where all values of y and X are positive, and the spatial dependence parameter is also positive, this suggests an upward bias in the least-squares estimates:

$$\hat{\beta} = (X'X)^{-1}X'(I_n - \hat{\rho}W)y$$
$$\hat{\beta} = \hat{\beta}_0 - \hat{\rho}(X'X)^{-1}X'Wy$$
$$\hat{\beta}_0 = \hat{\beta} + \hat{\rho}(X'X)^{-1}X'Wy$$

For illustration, the OLS estimates are $\hat{\beta}_0' = [0.55\ 1.25]$, which show upward bias relative to the spatial autoregressive model estimates $\hat{\beta}' = [0.135\ 0.561]$.

Least-squares predictions based on the matrices X and $\tilde{X}$ are presented in Table 1.2. We see that no spatial spillovers arise from this model, since only the travel time to the CBD for region R2 is affected by the change in population density of region R2. We also see the impact of the upward bias in the least-squares estimates, which produce an inflated prediction of travel time change that would arise from the change in population density.

Table 1.2: Non-spatial predictions for changes in Region R2 population density

Regions / Scenario	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)}$ - $\hat{y}^{(1)}$
R1:	42.98	42.98	0.00
R2:	36.00	47.03	11.02
R3:	29.02	29.02	0.00
R4: CBD	27.56	27.56	0.00
R5:	29.02	29.02	0.00
R6:	36.00	36.00	0.00
R7:	42.98	42.98	0.00

Ex. 1.2. You can find the Stata data and Do-file in Github that give the estimates of the SAR model. Try to reproduce the OLS estimates (in any software of your choice or manually using the formula).

Chapter 2

Spatial Econometric Models

Section 2.1 briefly introduces a family of conventional spatial regression models that have appeared in the empirical literature.

Section 2.2 is devoted to a discussion of interpreting the parameter estimates from spatial regression models.

2.1 Spatial regression models

We start with a full model including all possible spatial correlations (The Manski model):

$$
\begin{aligned}
y &= \rho W y + X\beta + W X\theta + u \qquad (2.1)\\
u &= \lambda W u + \varepsilon
\end{aligned}
$$

In this model, all three (dependent variable, independent variables and error term) are potentially spatially-correlated. This considers everything but is rarely used empirically because it is not or only weakly identifiable.

The case $\rho = 0$ reduces the model to a Spatial Durbin Error Model (SDEM) but again it is less frequently used.

When $\theta = 0$, the model becomes the Kelejian-Prucha model or Spatial Autoregressive Confused (SAC) model.

$$
\begin{aligned}
y &= \rho W y + X\beta + u \qquad (2.2)\\
u &= \lambda W u + \varepsilon
\end{aligned}
$$

The estimators of β are flawed in that they are biased and not convergent when the real model includes exogenous interactions WX. In addition, Le Gallo 2002 emphasises that choosing the same neighbourhood matrix W for this model results in weak parameter identification.

If we assume that the model is such that $\lambda = 0$, we have what is known as the Spatial Durbin Model (SDM). The estimators will be unbiased even if, in reality, we are in the presence of spatially auto-correlated errors. This model is therefore more robust in the face of a poor specification choice.

$$
y = \rho W y + X\beta + W X\theta + \varepsilon \qquad (2.3)
$$

The SDM then includes sub-models.

If $\rho = 0$, this becomes the Spatially Lagged X (SLX) model.

$$y = X\beta + WX\theta + \varepsilon \tag{2.4}$$

If $\theta = 0$, this degenerates to the familiar SAR model.

$$y = \rho Wy + X\beta + \varepsilon \tag{2.5}$$

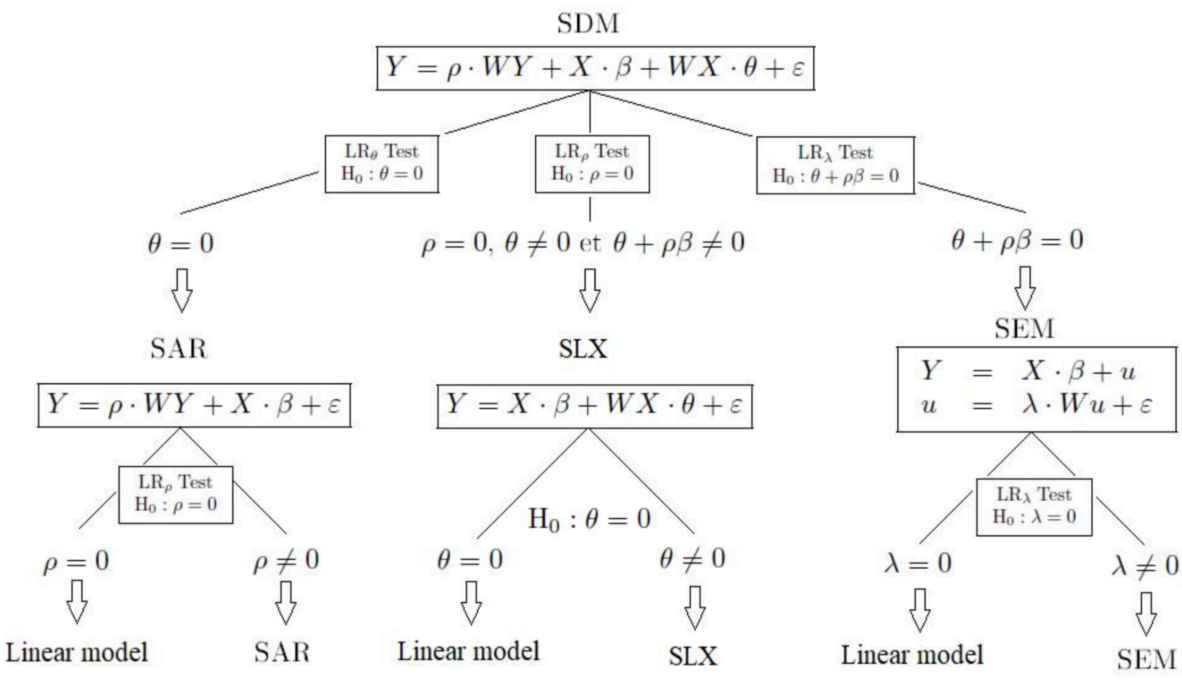
If $\theta = -\rho\beta$, this simplifies to the Spatial Error Model (SEM).

$$\begin{aligned} y &= \rho Wy + X\beta + WX(-\rho\beta) + \varepsilon \\ y &= X\beta + \rho W(y - X\beta) + \varepsilon \end{aligned}$$

By noting $u = y - X\beta$:

$$\begin{aligned} y &= X\beta + u \\ u &= \lambda Wu + \varepsilon \end{aligned} \tag{2.6}$$

Nowadays model selection is done with the 'top-down' approach. It starts from the SDM and reduces based on tests of the likelihood ratio:



2.2 Interpreting parameter estimates

2.2.1 Direct and indirect impacts in theory

In models containing spatial lags of the explanatory or dependent variables, interpretation of the parameters becomes richer and more complicated than conventional OLS.

In essence, spatial regression models expand the information set to include information from neighbouring regions/observations. To see the effect of this, consider the SDM model which we have re-written in (2.7).

$$
\begin{aligned}
(I_n - \rho W)y &= X\beta + WX\theta + \iota_n\alpha + \varepsilon \\
y &= \sum_{r=1}^{k} S_r(W)x_r + V(W)\iota_n\alpha + V(W)\varepsilon \qquad (2.7) \\
S_r(W) &= V(W)(I_n\beta_r + W\theta_r) \\
V(W) &= (I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + ...
\end{aligned}
$$

To illustrate the role of $S_r(W)$, consider the expansion of the data generating process as shown in (2.8)

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{r=1}^{k} \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \dots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & & \\ \vdots & \vdots & \ddots & \\ S_r(W)_{n1} & S_r(W)_{n2} & \dots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} \qquad (2.8)
$$
$$
+ V(W)\iota_n\alpha + V(W)\varepsilon
$$

The case of a single dependent variable observation in (2.9) makes the role of the matrix $S_r(W)$ more transparent.

$$
y_i = \sum_{r=1}^{k} [S_r(W)_{i1}x_{1r} + S_r(W)_{i2}x_{2r} + ... + S_r(W)_{in}x_{nr}] \qquad (2.9)
$$
$$
+ V(W)_i\iota_n\alpha + V(W)_i\varepsilon
$$

It can be seen that the derivative of y_i with respect to x_{jr} does not equal to β_r as in least-squares.

$$
\frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij} \qquad (2.10)
$$

The own derivative results in an expression $S_r(W)_{ii}$ that measures the impact on the dependent variable observation i from a change in x_{ir}. This impact includes the effect of feedback loops where observation i affects observation j and observation j also affects observation i as well as longer paths which might go from observation i to j to k and back to i.

$$
\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \qquad (2.11)
$$

To measure the impact of changes in an explanatory variable, LeSage and Pace (2006) suggest a summary measure:

1. *Average Direct Impact*

   An average of the diagonal of matrix $S_r(W)$, which equals $n^{-1}tr(S_r(W))$.

2. *Average Total Impact **to** an Observation*

   An average of row sums of matrix $S_r(W)$, which equals $n^{-1}\iota_n' c_r$ where $c_r = S_r(W)\iota_n$ is the column vector for the row sums.

3. *Average Total Impact* ***from*** *an Observation*

   An average of column sums of matrix $S_r(W)$, which equals $n^{-1}r_r\iota_n$ where $r_r = \iota_n' S_r(W)$ is the row vector for the column sums.

It is easy to see that the numerical values of the summary measures for the two forms of average total impacts set forth in 2) and 3) above are equal. However, these two measures allow for different interpretative viewpoints, despite their numerical equality.

The 'from' an observation view expressed in 3) above relates how changes in a single observation j influences all observations. In contrast, the 'to' an observation view expressed in 2) above considers how changes in all observations influence a single observation i.

Averaging over all n of the total impacts, whether taking the from an observation or to an observation approaches, leads to the same numerical result. Therefore, the average total impact is the average of all derivatives of y_i with respect to x_{jr} for any i, j.

In contrast, the average direct impact is the average of all own derivatives. Consequently, the average of all derivatives (average total impact) less the average own derivative (average direct impact) equals the average cross derivative (average indirect impact).

Formally, we define the three impacts below:

$$\bar{M}(r)_{direct} = n^{-1}tr(S_r(W)) \tag{2.12}$$

$$\bar{M}(r)_{total} = n^{-1}\iota_n' S_r(W)\iota_n \tag{2.13}$$

$$\bar{M}(r)_{indirect} = \bar{M}(r)_{total} - \bar{M}(r)_{direct} \tag{2.14}$$

Note that in practice, it is computationally inefficient to calculate the summary impact estimates using the definitions above. For approximation method, please refer to Chapter 4 of the textbook.

Example 2.1. Derive the average total impacts for SAR model.

Recall that the SAR model takes the form:

$$\begin{aligned}
(I_n - \rho W)y &= X\beta + \iota_n\alpha + \varepsilon \\
y &= \sum_{r=1}^{k} S_r(W)x_r + V(W)\iota_n\alpha + V(W)\varepsilon \qquad (2.15) \\
S_r(W) &= V(W)I_n\beta_r \\
V(W) &= (I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + ...
\end{aligned}$$

The average total impacts for this model then take the form in (2.16) for row-stochastic W.

$$\begin{aligned}
n^{-1}\iota_n' S_r(W)\iota_n &= n^{-1}\iota_n'(I_n - \rho W)^{-1}\beta_r\iota_n \\
&= (1-\rho)^{-1}\beta_r \qquad (2.16)
\end{aligned}$$

Ex. 2.1. Find the Average Total Impacts for SAC model and compare your result with that of the SAR model.

2.2.2 Partitioning the impacts by order of neighbours

Recall from the numerical example in Chapter 1 that impacts arising from a change in the explanatory variables will influence low-order neighbours more than higher-order neighbours.

Since the impacts are a function of $S_r(W)$, these can be expanded as a linear combination of powers of the weight matrix W using the infinite series expansion of $(I_n - \rho W)^{-1}$. Applying this to (2.12) and (2.13) for the SAR model allows us to observe the impact associated with each power of W. These powers correspond to the observations themselves (zero-order), immediate neighbours (first-order), neighbours of neighbours (second-order), and so on.

$$S_r(W) \approx (I_n + \rho W + \rho^2 W^2 + ... + \rho^q W^q)\beta_r \tag{2.17}$$

As an example, Table 2.1 shows both the cumulative and marginal or spatially partitioned direct, indirect and total impacts associated with orders 0 to 9 for the case of a SAR model where $\beta_r = 0.5$ and $\rho = 0.7$.

Table 2.1: Spatial partitioning of direct, indirect and total impacts

	Cumulative Effects		
	Mean	Std. dev	t-statistic
Direct effect X_r	0.5860	0.0148	39.6106
Indirect effect X_r	1.0841	0.0587	18.4745
Total effect X_r	1.6700	0.0735	22.7302
	Spatially Partitioned Effects		
W-order	Total	Direct	Indirect
W^0	0.5000	0.5000	0
W^1	0.3500	0	0.3500
W^2	0.2452	0.0407	0.2045
W^3	0.1718	0.0144	0.1574
W^4	0.1204	0.0114	0.1090
W^5	0.0844	0.0066	0.0778
W^6	0.0591	0.0044	0.0547
W^7	0.0415	0.0028	0.0386
W^8	0.0291	0.0019	0.0272
W^9	0.0204	0.0012	0.0191
$\sum_{q=0}^{9} W^q$	1.6220	0.5834	1.0386

From the table we see a cumulative direct effect equal to 0.586, which given the coefficient of 0.5 indicates that there is feedback equal to 0.086 arising from each region impacting neighbours that in turn impacts neighbours to neighbours and so on. In this case these feedback effects account for the difference between the coefficient value of $\beta_r = 0.5$ and the cumulative direct effect of 0.586.

The cumulative indirect effects equal to 1.0841 are nearly twice the magnitude of the cumulative direct effects of 0.5860. Based on the t-statistics calculated from a set of 5,000 simulated parameter values, all three effects are significantly different from zero.

The spatial partitioning of the direct effect shows that by the time we reach 9th-order neighbours we have accounted for 0.5834 of the 0.5860 cumulative direct effect. Of note is the fact that for W^0 there is no indirect effect, only direct effect, and for W^1 there is no direct effect, only indirect.

While cumulative indirect effects having larger magnitudes than the direct effects might seem counterintuitive, the marginal or partitioned impacts make it clear that individual indirect effects falling on first-order, second-order and higher-order neighbouring regions are smaller than the average direct effect of 0.5 falling on the own-region. Cumulating these effects however leads to a larger indirect effect which represents smaller impacts spread over many regions.

We see the direct effects die down quickly as we move to higher-order neighbours, whereas the indirect or spatial spillover effects decay more slowly as we move to higher-order neighbours.

Ex. 2.2. Reproduce Table 2.1 above using the spatial weight matrix and given parameters and explain why for W^0 there is only direct effect, and for W^1 there is only indirect effect (Note that your answers may not coincide exactly with all numbers in the table so just check the pattern).

Chapter 3

Project

In this course's project, we will do the followings to get an overview of spatial econometrics programming:

1. Use QGIS to create map graphs
2. Use Stata to conduct basic calculations

Please download the dataset from Github for later use.

1. QGIS

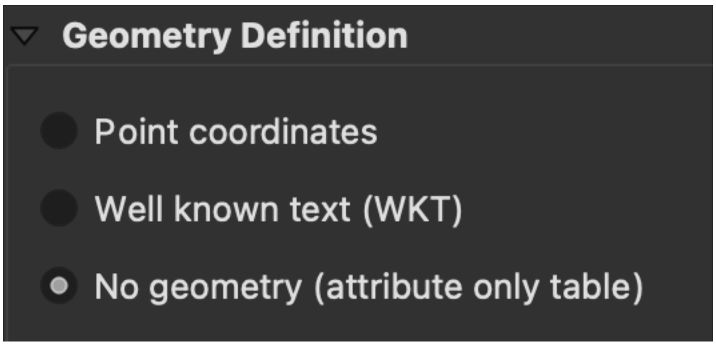
After downloading and installing QGIS, we can use it to draw graphs.

First of all, we need to load the map into QGIS. A map normally has several files as you can see from the folder. I will leave you to explore the properties of each of the extensions but for this project, we will only focus on the .shp files.

You can directly open the `NCVA CO.shp` file by changing your default app to QGIS, or add the vector layer manually in QGIS (`Layer -> Add Layer -> Add Vector Layer`).

This map shows all the states in the US. Each state has an administrative code and this is recorded as "FIPS".

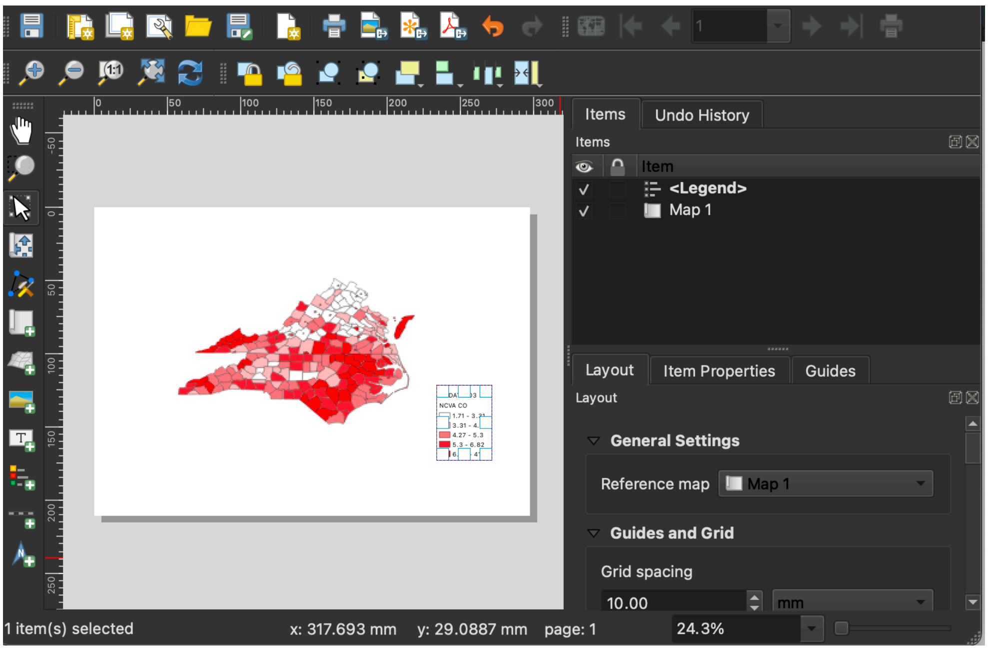
Now suppose we have some economic data stored in a .csv file. We can load this into QGIS by adding a text delimited layer. For this particular exercise, since we know the coordinates of each state centre, we can select the `point coordinates` for geometry definition. Here I will demonstrate the more general case where we do not have all longitude and latitude information (i.e. No geometry).

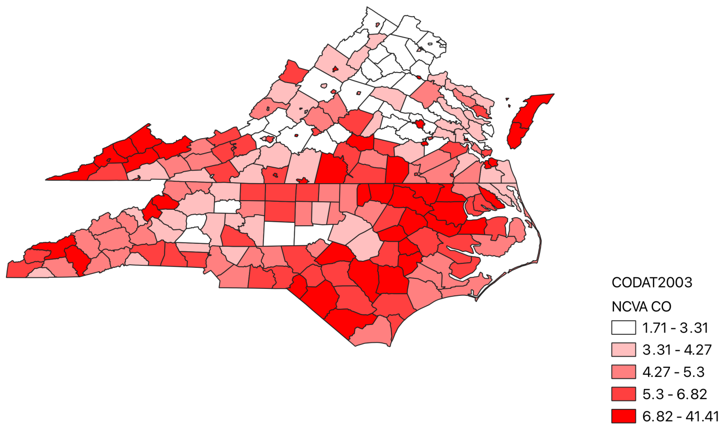


We need to `join` the two files together. To achieve this, right click the vector layer and select `Properties`. Then go to `Joins`, click the "+" button and change both the `join field` and `target field` to "FIPS". Finally apply this setting and you should see all economic data in the map's `Attribute table`.

As a final exercise, we can plot the economic variable by location. Change from the `Joins` window to `Symbology`. At the very top, you should have `Single symbol`. Change this to `Categorized` or `Graduated`, select your variable, classify and apply. Now you should see the new map plotting the variable of your choice by location.

Once the plot is ready, you can then add a legend and export to your preferred format. Go to `Project -> New Print Layout`. On the left hand side in the toolbar, you can see `Add Map` and `Add Legend`. Select the area you want these and then on the top you can export to image, PDF or any format you want.





Example graph of Unemployment by state

Finally, save the whole project (possibly as a .qgz file) so that you can visualise and modify everything you have so far created.

2. Stata

Stata is a powerful tool for spatial analysis and the manual gives a more comprehensive set of steps (https://www.stata.com/manuals/sp.pdf).

Here we will follow the example in Intro 7 using our own data. Please check the Do file in Github for more information.