Vrije Universiteit Amsterdam

Syllabus Spatial Econometrics

Thomas de Graaff

September 23, 2019

> "Everything is related to everything else, but near things are more related than distant things." Tobler (1970)

1. Introduction

The quote above is also known as the *first law of geography* and points to the empirical fact that socio-economic phenomena, such as poverty, GDP and unemployment, often display spatially correlated patterns—or, in other words, are clustered in space. For example, people in Amsterdam who live at the three largest canals (Herengracht, Keizersgracht and Prinsengracht) are usually richer than people who live in the western part of Amsterdam. And most countries in sub-sahara Africa are performing economically worse than most countries in Western Europa. Such examples of spatially related phenomena can be found for almost *every* spatial detail, agent, and phenomenon.

Of course, most spatial patterns, such as the spatial distribution of wealth, are the result of historical processes, but if you think about it, the concept of spatial relationships is actually rather common in everyday life. For example, most would-be students choose an university close to home, most workers search locally for a new job in the labour market, and you usually catch a new disease (like the flu) at work or school.

Thus, (cor)relations over space are rather common. Still, mainstream economics does not regard space as something that truly matters. This perception led to the criticism of one of the founders of regional science, Walter Isard, that there was an "Anglo-Saxon bias" which repudiates the factor of space and compresses everything within the economy to a point, so that all spatial resistance disappears. Maintaining this confines economic theory to "a wonderland of no spatial dimensions."

In the last three decades, however, there seems to be a renewed interest in the concept of space and methods to account for space properly—not in the last place because of the increasing availability of large spatial databases derived from remote sensing, satellite data and mobile devices Another cause is the 'death of distance' debate (see the work of Cairncross, 1997; Friedman, 2005), where it was argued that due to the emergence of information and communication technology the role of distance in trade and labor and housing markets would diminished severely. Maybe surprisingly, we have actually seen the opposite in the last two decades; clustering of most economic activity has only become stronger in recent decades (and especially the most important of all—the growth of cities). The role of distance indeed seems to have become stronger (and we do not *really* understand at the fundamental level *why*, although we have many theories—see as well the very readable and comprehensive overview of Proost and Thisse (2019)).

Partly, the renewed scientific theoretical interest in space may be traced back to the introduction of the monopolistic competition model by Dixit and Stiglitz (1977) and the resulting emergence of the 'new economic geography' by Paul Krugman (see, e.g., Fujita et al., 1999). The renewed empirical interest in spatial methods in econometrics started more or less with the seminal contribution of Anselin (1988) and seems to be finding its way to mainstream econometrics in the last 20 years (see, e.g., Baltagi et al., 2003; Kelejian and Prucha, 1998; Kelejian and Prucha, 2004).[1] Most interest nowadays is (*i*) on the focus of combining time-series data with spatial data (see, e.g. Elhorst, 2001; Baltagi et al., 2003), as a more complex form of panel data methods and (*ii*) on estimating larger spatial systems—the big data phenomenom (see for a good introduction to the problem LeSage and Pace, 2009).

The toolbox that deals with spatial patterns and processes is now commonly referred to as spatial econometrics, a term coined by the Belgian professor Jean Paelinck in the 1970s. Although most concepts seem at first quite similar to time-series econometrics, there are some fundamental differences. This syllabus continues first to explain the most fundamental differences between spatial and time-series econometrics, where we explicitly look at potential dependence structures. We continue with looking directly in possibilities to model space explicitly in a regression framework. Thereafter, we spend some attention on why it actually is important to correct for spatial relationships and, subsequently, we give some possibilities to test for the spatial dependence and the validity of the various spatial models given before. We end this syllabus with an empirical application, where we apply the various techniques to the determinants of crime levels in the city of Columbus, Ohio.

2. THE CHARACTERISTICS OF SPACE

2.1. WHAT IS THE PROBLEM?

Figure 2.1 gives the percentage of the adult population that is obese across the United States. When observing Figure 2.1 one may infer two observations:

[1] Actually, before 1988 there were already quite some good volumes and articles on spatial statistics, including the seminal work of Cliff and Ord (1981) who introduced the concept of spatial autocorrelation. However, they were only used in geography or in a small research niche in the field of regional science.

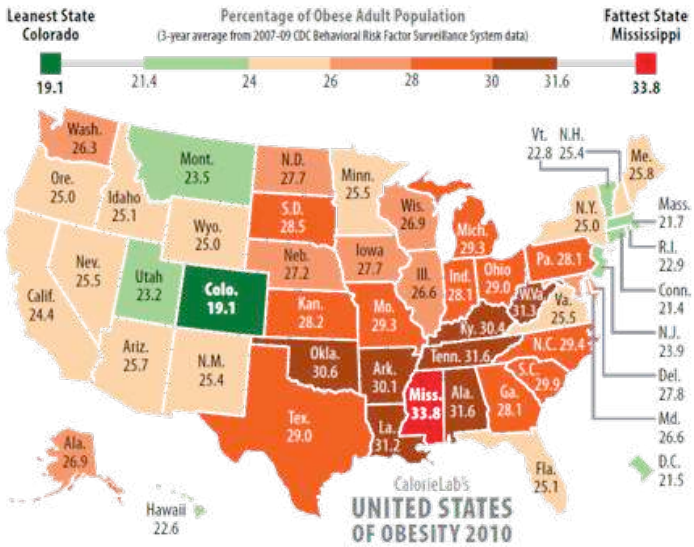

Figure 2.1: Obesity prevalence across the United States

1. The obesity prevalence in the United States is unequal across states, ranging from 19.1% to 33.8%. This phenomenon is called **spatial heterogeneity**; socio-economic variables are unequally distributed over space.

2. The obesity prevalence seems to be clustered over space, being the least in New-England and the west coast and the most in the southern states. This phenomenon is called **spatial dependence** (we will give a formal definition below).

Spatial heterogeneity and spatial dependence are very much related and difficult to discern from each other. For that we need specifically tailored statistical tests.

Most mainstream economics models do not take spatial heterogeneity or spatial dependence into account and are therefore called “topologically invariant”—that is, the phenomenon that is studied exhibit constant characteristics over space. There is one exception to this, and that is the nowadays common practice to incorporate spatial fixed effects for countries, regions and zipcodes. Note that this will tackle spatial heterogeneity as far as the levels are concerned (only the constants will then vary over space), but not deal with spatial dependence.

2.2. MODELLING SPATIAL RELATIONS

So, we know there are spatial effects, whether it is spatial dependence or spatial heterogeneity, but then what? How can we incorporate a spatial system within an econometric or statistical framework?

Let us start with a country with four regions, named A, B, C, and D. Then all possible relations between those four regions can be depicted as in diagram (2.1).

$$\begin{tikzcd} A \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] \arrow[dr, leftrightarrow] & B \arrow[d, leftrightarrow] \arrow[dl, leftrightarrow] \\ C \arrow[r, leftrightarrow] & D \end{tikzcd} \tag{2.1}$$

If we compare diagram (2.1) with a time-series of four observations, such as in the following diagram:

$$\begin{tikzcd} Y_{t-3} \arrow[r] & Y_{t-2} \arrow[r] & Y_{t-1} \arrow[r] & Y_t \end{tikzcd} \tag{2.2}$$

then it becomes clear that there are two main differences between the spatial and the temporal case, namely:

1. Spatial relations are multi-directional. One spatial unit, say a region, can affect several other spatial units directly.

2. Spatial relations are reciprocal. Spatial unit A affects B and B affects A at the same time.

Of course, one might argue that there is always a temporal dimension (just as there is always a spatial dimension), but for reasons of simplicitly we refrain from that possibility and assume that a spatial relation occurs instantaneously (again, for space-time models see, e.g, Elhorst, 2001).

But how to measure such a spatial relationship? Consider the map of GDP growth across the world (data are obtained from the Mankiw et al., 1992)) in Figure 2.2. Each country can be connected to another with some kind of function. The two most used functions in the literature are:

1. **Contiguity based**. For so-called first-order contiguity based relations, two areas have a relation if and only if they share a common border. Otherwise the relation is zero. In this respect, relations between direct adjacent neighbours are called first order contiguity, between areas that share a common neighbour second order contiguity, and so on. If regions or countries are isolated (e.g., with islands) this might pose a problem.

2. **Distance based**. Here the relation between two areas is measured by some notion of distance of travel time. A commonly used metric is the inverse of the distance between the two areas, such as $1/d$. Sometimes a more general power function is used, such as $d^{-\alpha}$, to capture a non-linear relationship. If one feels that a relationship only extends over a certain distance, say x kilometres, then one can also use a cut-off function, indicating that the relationship is proportional to $1/d$ within x kilometres, and 0 after x kilometres.

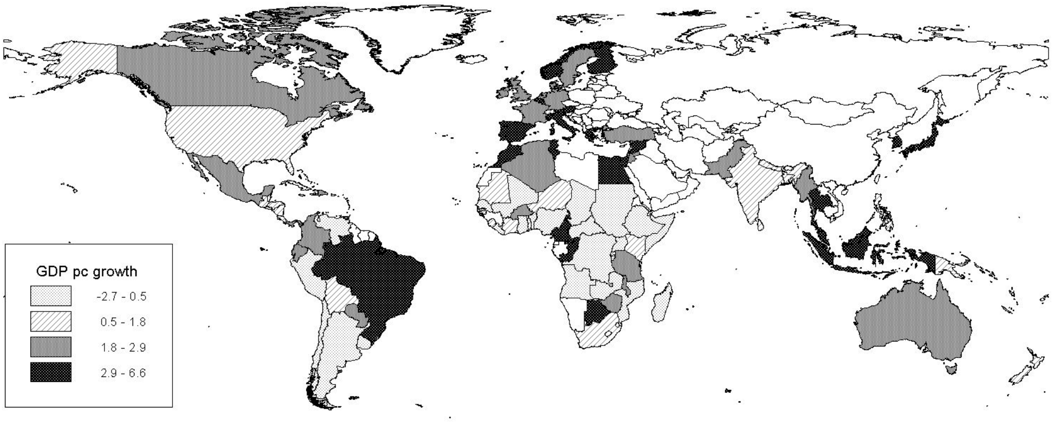

Figure 2.2: GDP growth across countries

We would like to stipulate here that, alhough often used, distance does not have to be based on a Euclid distance metric. It is just as valid to base your distance measure on, e.g., social or information networks. That is, *the distance* can be expressed like that. Note, however, that geographically based distances, such as kilometers, can truly be regarded as exogenous, while metrics based on social interactions are usually correlated with the phenomenon to be explained. The latter introduces then another kind of endogeneity, which further complicates matters when applying this in a regression framework.

So, we have some kind of function that describes how units, such as areas, regions, countries and humans, relate to each other in space, but how can we depict such a relationship for the whole spatial system? Consider again a country with four regions, named A, B, C, and D, where the relationships between the regions are as depicted in diagram (2.3)

$$\begin{tikzcd} A \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & B \arrow[dl, leftrightarrow] \\ C \arrow[r, leftrightarrow] & D \end{tikzcd} \tag{2.3}$$

Now, the way to go forward is to use the information from diagram (2.3) and display this in a so-called *spatial weight matrix*, usually denoted as $\mathbf{W}$.[2] If we use a first-order contiguity spatial weight matrix, then (where rows denote always the origin and the columns the destination; so, the 1 in the first row and the second column denotes from A to B):

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.4}$$

and if we use an inverse distance spatial weight matrix, where d_{ij} denotes some kind of distance measure between spatial unit i and j $(i, j \in \{A, B, C, D\})$:

$$\mathbf{W} = \begin{bmatrix} 0 & 1/d_{AB} & 1/d_{AC} & 1/d_{AD} \\ 1/d_{BA} & 0 & 1/d_{BC} & 1/d_{BD} \\ 1/d_{CA} & 1/d_{CB} & 0 & 1/d_{CD} \\ 1/d_{DA} & 1/d_{DB} & 1/d_{DC} & 0 \end{bmatrix}. \tag{2.5}$$

Three things become immediately clear when looking at the two types of spatial weight matrices. First, they are *symmetric*, reflecting the reciprocal nature of spatial relationships. Asymmetric spatial weight matrices are sometimes used as well, but make things considerably more complex. Secondly, the *diagonal* is always zero. Thus, a spatial unit does not affect itself *directly*. Thirdly, first order-contiguity spatial weight matrices exhibit quite a lot of zeros when spatial systems become larger, and are therefore also called *sparse* matrices. This feature is attractive because is has some direct computational advantages (basically, it is a lot faster to store such matrices in computer memory; see, e.g., LeSage and Pace, 2009). Distance based weight matrices are usually full weight matrices (each entry except those on the diagonal is non-zero).

For ease of computation and estimation spatial weight matrices are typically *row-standardized.* This mean that all entries are divided by their respective row-sums, or for a typical four by four matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & \frac{a_{12}}{\sum_j a_{1j}} & \frac{a_{13}}{\sum_j a_{1j}} & \frac{a_{14}}{\sum_j a_{1j}} \\ \frac{a_{21}}{\sum_j a_{2j}} & 0 & \frac{a_{23}}{\sum_j a_{2j}} & \frac{a_{24}}{\sum_j a_{2j}} \\ \frac{a_{31}}{\sum_j a_{3j}} & \frac{a_{32}}{\sum_j a_{3j}} & 0 & \frac{a_{34}}{\sum_j a_{3j}} \\ \frac{a_{41}}{\sum_j a_{4j}} & \frac{a_{42}}{\sum_j a_{4j}} & \frac{a_{43}}{\sum_j a_{4j}} & 0 \end{bmatrix}. \tag{2.6}$$

Note that with this procedure all entries on a role sum up to 1 and that spatial weight matrices that initially were symmetrical loose their symmetry.

[2] Because of notational ease, spatial econometrics usually is done in matrix notation. For those not accustomed to matrix notation and algebra, appendix A gives a quick recap.

To give an example, consider again the countries in Figure 2.2. The darker the country code areas the higher economic growth. But how to implement this in a statistical framework?

We proceed as follows. In this case, we measured the distance between each country as the crow flies (why not take a first contiguity approach here?), giving us the distance d_{ij} between each country i and j. Using these distances, we construct a distance based spatial weight matrix, **W**, where the entry for each row i and column j is formed by $1/d_{ij}$. Note that there about 97 countries in this dataset, thus we end up with a matrix **W** with size 97 × 97, or 9,409 entries in total. Now, we know for each country the growth of GDP per worker well, denoted with the vector y, then a measure for GDP weighted by distance can be denoted as

$$\tilde{y} = \mathbf{W}y, \tag{2.7}$$

where the i-th element of $\tilde{y}$ gives an weighted average of GDP, where we weight by distance.

.

3. SPATIAL ECONOMETRIC MODELS

3.1. SPATIAL DEPENDENCE

The basic concept in spatial econometrics is spatial autocorrelation or spatial dependence. Fundamentally, these two concepts are not the same, but usually they are treated similarly, just as we do in this chapter. To proceed, first note that independence between two stochastic variables can be formalized as follows:

$$\Pr(X_i = x_i) = \Pr(X_i = x_i | X_j = x_j), \tag{3.1}$$

where i and j are two spatial units. Equality (3.1) basically states that the probability that x_i occurs in spatial unit i is not related with the probability that x_j occurs in spatial unit j. But what if spatial units i and j are related to each other, then equality (3.1) becomes an inequality. If we now assume that there is a set of J neighbours around i that exert an influence on i – via, e.g., a first order contiguity matrix –, then the whole system of spatial dependence may be denoted as:

$$J | \left\{ Pr(X_i = x_i) \neq \Pr(X_i = x_i | X_j = x_j) \right\}. \tag{3.2}$$

So, what does equation (3.2) actually say? Loosely speaking, it says that the spatial system of dependence around i consists of all neighbours $j \in J$ that have a *statistical* relation with i. In order words, i and j are not independent.

This shows that spatial dependence is basically a statistical concept, which does not say anything about causality, but merely about (cor)relations. Thus, phenomena occurring in spatial units i and j may be correlated because there is a fundamental process working or because some important variables have been left out of the model that influence both the phenomenon in spatial unit i as well as in spatial unit j. The latter we denote with the term unobserved spatial heterogeneity.

3.2. A BASIC TAXONOMY OF SPATIAL ECONOMETRIC MODELS

In a regression framework we may denote now the following general spatial model, using the concept of the spatial weight matrix as explained above and in matrix form:

$$\begin{aligned} y &= \rho \mathbf{W}_1 y + \mathbf{W}_2 \mathbf{Z}\gamma + \mathbf{X}\beta + \epsilon \\ \epsilon &= \lambda \mathbf{W}_3 \epsilon + \mu, \end{aligned} \tag{3.3}$$

where y is a vector of endogenous variables, $\mathbf{X}$ is a matrix of exogenous variables, $\{\rho, \gamma, \beta, \lambda\}$ is a vector of parameters, and μ is a vector of i.i.d. distributed residuals (usually assumed to be normally distributed). Note that the spatial weight matrices do not have to be the same. $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ may differ, but are usually equal to each other. To facilitate estimation and testing in a regression framework, these weight matrices are almost always *row-standardized.* So again, this means that every element in these matrices are divided by the sum of all elements in their specific rows ($w_{ij}/\sum_i w_{ij}$). Spatial processes work through these spatial weight matrices and are measured by their corresponding parameters, ρ, γ and λ. So, spatial dependence is captured in only a few parameters to be estimated, which is rather efficient. Attempts have been made to estimate the whole spatial weight matrix parametrically, but this does not seem to improve matters much.

Note that both $\mathbf{Z}$ and $\mathbf{X}$ are both two sets of exogenous variables, which may be identical to each other (but not necessarily so). Moreover, if $\rho = \gamma = \lambda = 0$ model (3.3) simplifies to the multivariate ordinary regression model:

$$y = \mathbf{X}\beta + \mu, \tag{3.4}$$

Model (3.3) gives the general expression of the most simple form of modelling space in a regression framework. This model can be decomposed in the following three separate *spatial* models.

The spatial lag model ($\rho \neq 0,\ \gamma = 0,\ \lambda = 0$) This leads – in matrix notation – to the following expression:

$$y = \rho \mathbf{W}_1 y + \mathbf{X}\beta + \mu. \tag{3.5}$$

Note that for each region r and all its possible neighbours r' yields: $y_r = \rho \sum_{r'} w_{rr'} y_r + \beta x_r + \mu_r$, which, for example, collapses with only two regions (r and r') and weights $w_{rr'}$ set at 1 and 0 (for a contiguity matrix), to: $y_r = \rho y_{r'} + \beta x_r + \mu_r$, which bears again close resemblance to the time-series model with autocorrelation.

We can rewrite the spatial lag model (3.5) as: $(\mathbf{I} - \rho \mathbf{W}_1) y = \mathbf{X}\beta + \mu$, where $\mathbf{I}$ stands for the identity matrix. Rewriting leaves us with: $y = (\mathbf{I} - \rho \mathbf{W}_1)^{-1}(\mathbf{X}\beta + \mu)$.[3] Thus a change in y causes a changes throughout the whole system, because some kind of feedback system is introduced. To see this, consider the following mathematical equality:

$$(\mathbf{I} - \rho \mathbf{W}_1)^{-1} = \mathbf{I} + \rho \mathbf{W}_1 + \rho^2 \mathbf{W}_1 \mathbf{W}_1 + \rho^3 \mathbf{W}_1 \mathbf{W}_1 \mathbf{W}_1 + \ldots \tag{3.6}$$

[3]The trick here is to multiply both sides with $(\mathbf{I} - \rho \mathbf{W}_1)^{-1}$ and to remember that $(\mathbf{I} - \rho \mathbf{W}_1)^{-1}(\mathbf{I} - \rho \mathbf{W}_1) = \mathbf{I}$.

which bears close resemble with the mathematical formulation of conventional input-output models. Namely, when y changes this has an effect on itself (through the identity matrix), if affects its neighbours (through the second term), those neighbours affect their neighbours again (through the third term), and so forth. The nice thing about this model is that it has an interesting *theoretical* interpretation. For example, it can—at least theoretically—model the transfer of knowledge, the spread of disease, or agglomeration effects.

The spatial cross-regressive model ($\rho = 0,\ \gamma \neq 0,\ \lambda = 0$) This leads to the following expression:

$$y = \gamma \mathbf{W}_2 \mathbf{Z} + \mathbf{X}\beta + \mu. \tag{3.7}$$

Note that now for a region r and say its neighbour r' yields: $y_r = z_{r'}\gamma + x_r\beta + \mu_r$ when the spatial weight connection between r and r' is set again at 1. Thus, some kind of phenomenon, say regional growth, in r depends on some exogenous variables, say the growth of human capital, in region r and in its neighbouring region r'.

This model is *econometrically* the least interesting of the three. Essentially it is a transformation on (a subset of) the exogenous variables ($\mathbf{Z}$), and therefore reduces to a basic regression framework. If you think a variable needs a spatial transformation, then that transformation can be performed on that variable without transforming the other variables or changing the model specification.

The spatial (autoregressive) error model ($\rho = 0,\ \gamma = 0,\ \lambda \neq 0$) This leads to the following expression:

$$\begin{aligned} y &= \mathbf{X}\beta + \epsilon \\ \epsilon &= \lambda \mathbf{W}_3 \epsilon + \mu. \end{aligned} \tag{3.8}$$

where for one region r and it neighbour r' the expression for the *residual* of r becomes: $\epsilon_r = \lambda\epsilon_{r'} + \mu_r$, with the weight between r and r' set at 1. This shows that the residual is basically a random effect (μ_r) in combination with a part of the residual in its neighbouring region.

Now note that $\epsilon - \lambda\mathbf{W}_3\epsilon = \mu$, solving for ϵ leads to $(\mathbf{I} - \lambda\mathbf{W}_3)\epsilon = \mu$ and thus $\epsilon = (\mathbf{I} - \lambda\mathbf{W}_3)^{-1}\mu$, which basically reduces the spatial error model to $y = \mathbf{X}\beta + (\mathbf{I} - \lambda\mathbf{W}_3)^{-1}\mu$. This last expression clearly shows that in theory spatial dependence only shows up in the obtained regression residuals through the same effect as in equation (3.6). The spatial error model is *theoretically* therefore more difficult to interpret than the spatial lag model.

Basically, the three models above are the ones most commonly used, but other models and combinations of the above three exist as well, but for reasons of simplicity we will restrain ourselves to these ones.

3.3. DOES IT MATTER?

A question that naturally arises is whether it actually matters when spatial dependence is present. Can we still apply OLS on those regression models? We answer this question in the light of the three models above. For this purpose, we need the concepts of statistical bias and efficiency. Recall that statistical bias refers to the fact that an estimator of a parameter ($\hat{\theta}$) does not produce the correct true parameter ($\mathbb{E}(\hat{\theta}) \neq \theta$). And an efficient estimator $\hat{\theta}$ indicates that there is no estimator $\hat{\theta}_1$ that has a variance smaller than the variance of $\hat{\theta}$. For instance, not accounting for heteroskedasticity by not applying robust standard errors usually lead to inefficient standard errors (but not to a *biased* estimator).

The spatial lag model ($y = \rho \mathbf{W}_1 y + \mathbf{X}\beta + \mu$) When this is the true model, OLS is not only inefficient, but it always produces *biased* results.[4]The reason for this may be clear. There are endogenous variables on the right-hand side, which leads to $\mathbb{E}[\mu|X] \neq 0$ or biased results when simply applying OLS. Actually, this spatial bias can be proven rather easily. Without loss of generality we may simplify the model to:

$$y = \rho \mathbf{W}_1 y + \mu. \tag{3.9}$$

Now remember that the OLS estimate for ρ is:

$$\hat{\rho} = [(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]^{-1}(\mathbf{W}_1 y)' y. \tag{3.10}$$

with $'$ denoting the transpose of a matrix.[5]

Now substitute the expression for y from equation (3.9) into equation (3.10) and we finally get:

$$\begin{aligned}
\hat{\rho} &= [(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]^{-1}(\mathbf{W}_1 y)'(\rho \mathbf{W}_1 y + \mu) && (3.11)\\
&= [(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]^{-1}[(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]\rho + [(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]^{-1}(\mathbf{W}_1 y)'\mu) && (3.12)\\
&= \rho + [(\mathbf{W}_1 y)'(\mathbf{W}_1 y)]^{-1}(\mathbf{W}_1 y)'\mu. && (3.13)
\end{aligned}$$

The second term in (3.13) is usually not equal to zero (because μ is correlated with y) and therefore OLS does not produce a consistent estimator of ρ (remember that consistency means loosely that if the number of observations become very large, the estimator convergences in probability to the parameter that is estimating). Including the β coefficients in this proof is straightforward.

[4]Recall that bias indicates that the expected value of the difference between and estimator and the parameter that is is estimating is non-zero. Or, if $\hat{\mu}_Y$ is an estimator of μ_Y then the size of the bias is $E(\hat{\mu}_Y) - \mu_Y$ and we say that the estimator is biased (Stock and Watson, 2015).

[5]Here $\mathbf{W}_1 y$ is used instead of the normal matrix $\mathbf{X}$. Inserting $\mathbf{X}$ for $\mathbf{W}_1 y$ leaves us again with the familiar result: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$.

The spatial cross-regressive model ($y = \gamma\mathsf{W}_2\mathsf{Z} + \mathsf{X}\beta + \mu$) Because the exogenous parameters are only (spatially) transformed, this model can still be consistently and efficiently estimated by OLS.

The spatial (autoregressive) error model ($y = \mathsf{X}\beta + (\mathbf{I} - \lambda\mathsf{W}_3)^{-1}\mu$) As seen above, this model produces a spatial dependence structure in the residuals, indicating that the residuals are not identically and independently distributed anymore. Fortunately, this does not lead to biased results, but it does create inefficiency. Thus, there are better estimation procedures than OLS out there that produce consistent and efficient results. Note that OLS here gives inefficient results with sample sizes smaller than infinitely (and how often does that happen). This means that when correcting for autoregressive errors the standard errors might turn out to be rather different than the standard errors coefficients resulting from an OLS regression (and with small samples this might affect the β coefficients as well).

4. TESTING FOR SPATIAL EFFECTS

Before starting to estimate all kinds of specifications, the econometrician would like to have some clue about which model specification would be more or less the correct one. Fortunately, several test-statistics have been devised to test for the presence and type of spatial dependence. In general there are two type of tests. Unfocused or general misspecification tests, which test for the presence of spatial dependence in what ever form present. And focused or specific misspecification tests that test for the presence of a specific type of spatial dependence (i.e., spatial lag or spatial error dependence). In this section we will only give a few of these tests, without going into the mathematical background of these tests. The *null*-hypothesis is always no spatial dependence. In other words, the null model is the ordinary regression model.

$$y = \mathbf{X}\beta + \mu \tag{4.1}$$

We always assume that the econometrician has estimated the null model and kept the realisations of the error term, the residuals, which we call u.

4.1. GENERAL TESTS FOR SPATIAL DEPENDENCE

Now, the oldest and most common general misspefication test is the so-called Moran's I:

$$I = \frac{R}{S_0} \times \frac{u'\mathbf{W}u}{u'u}, \tag{4.2}$$

where R is the number of spatial units, S_0 is the sum of all elements of the spatial weight matrix, and $\mathbf{W}$ and u are as specified before. Note that when the spatial weight matrix is row-standardized all rows will sum up to 1 and S_0 to R, thus the term $\frac{R}{S_0}$ becomes equal to 1. The advantage of this test-statistic is that is has power against all kinds of spatial dependence processes. The disadvantage is that we do not know against which kind of spatial process.

For those familiar with time-series, notice the close resemblence with the Durbin-Watson statistic.[6] Statistical inference can be based on the standardised or z-value of Moran's I, as follows:

$$z_I = \frac{I - \mathbb{E}(I)}{\sqrt{\mathbb{V}ar(I)}},$$

Unfortunately, the expectation and the variance have rather long technical expressions, but can be derived analytically based on the assumption of a normal distribution (Anselin, 1988). Fortunately, computer programs will do this now (one can resort as well to simulations). Like Moran's I, other general test statistics have been developed, such as Geary's c and Getis & Ord's G. However, the latter two tests do now seem to perform as well as Moran's I, and in most applications nowadays only Moran's I is reported.

4.2. FOCUSED TEST FOR SPATIAL DEPENDENCE

There are test as well that directly test for the presence of a spatial lag or spatial error model. There are no tests for spatial cross-regressive models, mainly because they do not affect the residuals, u, and are not considered as misspecifications. These focused tests are so-called Lagrange Multiplier (LM) tests and are given here to complete the picture. Unfortunately, these tests are somewhat complex but because every preprogrammed estimation procedure gives these as output and every paper reports them, they are needed to get a good overview. We start with the simplest, the LM$_\lambda$-test for the presence of a spatial error component model, which is basically a scaled Moran's I test (see also Florax and Nijkamp, 2005).

$$LM_\lambda = \frac{1}{T} \times \left(\frac{u'\mathbf{W}u}{s^2}\right)^2 \tag{4.3}$$

where s^2 is the maximum-likelihood variance $u'u/R$ and where T is somewhat more complicated. It is the *trace* of a quadatric expression of the weight matrix: $T = \text{tr}(\mathbf{W}'\mathbf{W} + \mathbf{WW})$. The trace is here a new matrix operation. It is essentially the sum of all the diagonal elements of a matrix ($\text{tr}(\mathbf{W}) = \sum_{i=1}^{R} w_{ii}$). Fortunately, this test statistics follows nicely a χ^2 distribution with one degree of freedom.

The LM$_\rho$ has the same distribution and looks similar:

$$LM_\rho = \frac{1}{RJ} \times \left(\frac{u'\mathbf{W}y}{s^2}\right)^2 \tag{4.4}$$

where J has a rather ugly expression, namely: $J = [(\mathbf{WX}\hat{\beta})'\mathbf{M}(\mathbf{WX}\hat{\beta}) + Ts^2]/Rs^2$. Here $\hat{\beta}$ are the estimated OLS coefficients and $\mathbf{M}$ is the projection matrix $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. There is a lot

[6]This test statistic, d, measures whether temporal residuals are autocorrelated by

$$d = \frac{\sum_{t=2}^{T}(u_t - u_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

where a value of d that differs significantly from 2 suggests temporal autocorrelation.

more to say about tests for spatial dependence and there are quite some more, but these are the most common and the ones you'll find reported in almost every table of results.

These tests have as disadvantage that they do not distinguish properly between spatial error and spatial lag processes. Therefore, so-called *robust* LM tests have been constructed. These tests are called robust because they account for the *potential* presence of a spatial lag of spatial error model when testing for a spatial error or spatial lag model, respectively. The test for a spatial error robust to the presence of a spatial lag model is as follows:

$$LM^*_\lambda = \frac{1}{T - T^2(RJ)^{-1}} \times \left(\frac{u'\mathbf{W}u}{s^2} - T(RJ)^{-1}\frac{u'\mathbf{W}y}{s^2}\right)^2, \tag{4.5}$$

with all notation as above. This expression looks (and is) ugly, but if you look closer, you see that there is some of subtraction present (equation (4.3) - equation (4.4) that accounts for the local misspecifcation of a spatial lag process.

Alternatively, the test for spatial lag robust to the presence of a spatial error model is:

$$LM^*_\rho = \frac{1}{RJ - T} \times \left(\frac{u'\mathbf{W}y}{s^2} - \frac{u'\mathbf{W}u}{s^2}\right)^2. \tag{4.6}$$

Both tests follow nicely a χ^2 distribution with one degree of freedom. Most common practice is to choose and estimate that model, which robust test-statistic is (most) significant. Usually only one test-statistic is significant. If none of these test-statistics are significant, there is no misspecification in our basic regression model (4.1) and we may confidently apply OLS. If both robust test statistics are significant, one can nowadays use the generalised spatial 2SLS estimator of Kelejian and Prucha (1998) and Kelejian and Prucha (2004).

4.3. LOCAL MORAN AND MORAN SCATTERPLOTS

Although the test statistics above seem to perform fairly well in detecting spatial dependence, they are not particularly suitable in detecting outliers or localised clusters of spatial dependence (hotspots). Therefore, so-called local indicators of spatial association (LISA) have been developed in the 1990s. They are basically a graphical tool to detect spatial dependence and help in identifying the correct spatial pattern. Here, we will only discuss the local Moran statistic and the Moran scatterplot, which are the ones most commonly used.

For a standardized weight matrix, the local Moran for spatial unit r is given by:

$$I_r = \frac{(x_r - \bar{x})\sum_{r=1}^{R} w_{rr'}(x_{r'} - \bar{x})}{\sum_{r=1}^{R}(x_r - \bar{x})^2/R}, \tag{4.7}$$

where R is again the total number of spatial units and the $\bar{x}$ denote the average of x, and x may stand for every vector of variable of interest, including the residuals. Calculating the test-statistic (4.7) for every spatial unit, gives a vector of indicators for spatial dependence. In this way, outliers can be quite easily identified. Note that $\sum_r I_r$ gives again our global Moran's I.

Looking more closely to (4.7) reveals another pattern. Namely, it is not difficult to see that we can transform x_r to z_r by:

$$z_r = \frac{x_r - \bar{x}}{\sqrt{\sum_{r=1}^{R}(x_r - \bar{x})^2/R}} = \frac{x_r - \mathbb{E}(x)}{\sqrt{\mathbb{V}ar(x)}},$$

which is the *standardized* version of x. Now, (4.7) can also be rewritten as $I_r = z_r \sum_{r=1}^{R} w_{rr'} z_{r'}$. In other words, depicting z against $\mathbf{W}z$ in a scatterplot, is another way of presenting the information behind (4.7). Moreover, the slope of the regression through the scatterplot gives then the aggregate Moran's I again! The reasoning behind this is that the Moran's I can be interpreted as the correlation between a phenomenon z and the occurrence of that phenomenon in neigbouring regions.

Because z is standardized, this slope of the regression runs always through the origin (remember: standardization means correction for the expectation, resulting in deviations around zero) and divides the scatterplot in four quadrants: namely positive-positive, positive-negative, negative-negative, and negative-position (for those of you who want a visual example, Figure 6.2 gives such a scatterplot. If there is positive spatial dependence then most observations will lie in the positive-positive and negative-negative quadrants. This basically means that values of z are closely correlated with values of z in neighbouring regions. Negative spatial dependence will cause most observations to be located in the negative-positive and the positve-negative quadrant.

5. ESTIMATION OF SPATIAL MODELS

Thus, for the spatial lag (and to a lesser extent for the spatial error) OLS is not feasible anymore. Which alternatives are there? The two mostly used estimation strategies are:

1. A maximum likelihood approach. This basically implies that the researcher should assume a stochastic distribution for the error term (typically normal) and then searches for that parameter combination that fits the data best. Besides complex and computationally intensive, the biggest drawback is that one has to assume a stochastic distribution (see for some technical details Anselin and Hudak, 1992).

2. A spatial instrumental variable approach. Usually the instruments are formed by spatially transforming the exogeneous variables (thus $\tilde{\mathbf{X}} = \mathbf{WX}$). The main benefit of this approach is that it is computationally much faster than the maximum likelihood approach (and it does not requires restrictive distributional assumptions). However, the exogeneity of the instruments is often doubtful and should therefore be very carefully adopted. In a seminal series of articles Kelejian and Prucha (1998) and Kelejian and Prucha (2004) have expanded this technique with a generalized method of moment estimator and have shown that the estimator is unbiased and efficient under a reasonably set of mild assumptions. Moreover, their generalised spatial two stage least squares (GS2SLS) estimator is able to simultaneously estimate a model in the presence of both a spatial lag and a spatial error.

Fortunately, computer applications and packages for most statistical software is now widely available. The most often used are listed below:

GeoDa Open source application written by Luc Anselin Anselin (2000). Although somewhat limited in terms of models it can estimate, this application has a nice integration with GIS techniques (it can create maps) and is able to deal with larger (> 20,000 geographical observation points) datasets. Figure 6.1, for example, is created with GeoDa.

Stata There are some user-written packages in Stata that allow for the creation of spatial weights matrices and the estimation of spatial econometric models. The estimations in Section 6, e.g., have been produced by Stata. Moreover, Ingmar Prucha gives on his website [7] Stata code as well for the estimations in Kelejian and Prucha (1998) and Kelejian and Prucha (2004).

MatLab Kelly Pace (Pace, 1997; Pace and Barry, 1997; Pace et al., 1998) maintains a well good organised webpage[8] for spatial statistic routines in Matlab just as Ingmar Prucha does on his website.

R R is now the environment that offer the most in terms of software for spatial econometric analysis (just as it provides a nice integration with GIS routines). R's CRAN task view: Analysis of Spatial Data[9] maintained by Roger Bivand (Bivand, 2002) gives a very thorough overview of all possibilities. It is open source, so free. That is in beers not in time (mind the *learning curve*).

ArcGis ArcGis is obviously mostly used to handle and display spatial data, but can as well, to a certain extent, perform some spatial econometric exercises uses the spatial analyst toolbox[10]. Be careful though for the burden on your wallet and on your computer memory.

Before we turn to an empirical example of these models, we first look into possibilities to test a priori whether there is spatial dependence present, and if so, which model specification is most likely to prevail.

6. SPATIAL ECONOMETRICS IN PRACTICE: THE DETERMINANTS OF CRIME

We have seen above that OLS is not correct anymore when spatial dependence is present.[11] So, for the spatial lag model, conventional regression techniques are not appropriate anymore and we have to turn to more complex methods, like maximum likelihood or method of moments. For the spatial error model we might apply instrumental variables (IV) techniques *if* we have some kind of consistent estimate of λ. Usually, that is not the case and we have

[7] http://econweb.umd.edu/~prucha/Research_Prog.htm.
[8] http://www.spatial-statistics.com/software_index.htm.
[9] https://cran.r-project.org/web/views/Spatial.html.
[10] http://www.esri.com/software/arcgis/extensions/spatialanalyst.
[11] Except for the spatial cross-regressive model, but basically that is a normal regression and we do not look into this model any further.

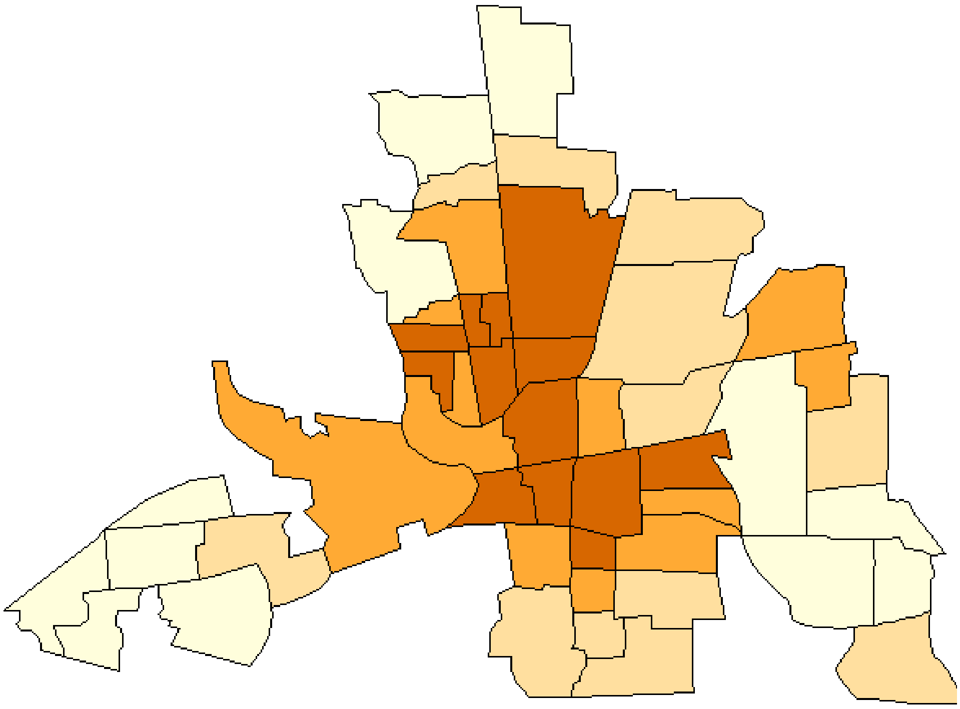

Figure 6.1: Crime rates, in quartiles, in the neighbourhoods of Columbus, Ohio.

to rely again on maximum likelihood or method of moment methods. Fortunately, as we have seen statistical softare packages are readily available now for the estimation of spatial dependence. For our application we use Luc Anselin's Geoda[12], which is very useful for making nice spatial maps as well. Moreover, we use the user-written plug-in for Stata, the so-called 'sg162' package[13], which is able to deal with spatial datasets, spatial tests and statistics and the estimation of spatial dependence in regression frameworks.

A long-lasting and still somewhat unsettled scientific literature concerns the determinants of crime. Already since the 19th century sociologists discovered that crime rates were correlated with poverty rates. We will look into this issue using the often used Columbus, Ohio, crime dataset. First look at Figure 6.1.

Here, the darker the color, the larger the crime rate, with high crime rates downtown and smaller crime rates uptown. We are interested in the fact whether crime rates are correlated with poverty and whether there is spatial dependence present in these data. In other words, do crime rates exibit some kind of spatial pattern and may this even be generated by a particular spatial process (e.g., by contagion). To answer these questions, we first have to

[12] See `https://www.geoda.uiuc.edu/`.

[13] See package sg162 from `http://www.stata.com/stb/stb60` developed by Maurizio Pisati, University of Milano Bicocca, Italy.

collect data.[14] The data we have are a first order contigous spatial weight matrix, indicating for 49 neighbourhoods whether they share a border, the variable 'crime' which measures the residential burglaries and vehicle theft per thousand households in a neighbourhood, the variable 'income' which is the average income in thousand dollars and the variable 'hoval' which is the average house value in thousand dollars. All data pertain to 1980.

To start with our analysis, we first run an OLS regression on 'crime'. Table 6.1 gives the STATA results.

Table 6.1: OLS estimation results: crime

Variable	**Coefficient**	**Std. Err.**
hoval	-0.274†	0.163
income	-1.597**	0.461
Intercept	68.619**	4.233
N	49	
R^2	0.552	
$F_{(2,46)}$	45.466	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Clearly, crime is in this specification negatively related with house values and income. For example, if average income increase with a 1,000 dollars in a neighbourhood, the amount of crime incidences per thousand household decreases with more than 1.5. But is there spatial dependence present in this dataset? We therefore first look at the relation between 'crime' and the level of 'crime' in the neighbours. To to so, we standardize the 'crime' variable and depict this against the spatial lag of 'crime' ($\mathbf{W}\times$ 'crime') in a moran scatterplot. This scatterplot is depicted in Figure 6.2. Obviously, there is a positive relation between those two variables. Crime is high when crime levels in the neighbourhoods are high and low when crime levels in the neighbourhoods are low. And of course, the steeper the regression slope, the more spatial dependence and the higher the global Moran's I.

Thus, the residuals of the regression most likely show spatial dependence as well. To investigate these residuals, we run our test-statistics by invoking the STATA command 'spatdiag': Table 6.2 displays the test-statistics. First, the Moran's I clearly shows that some kind of spatial dependence is present in the residuals. If we now look at the Lagrange multiplier test, then we see that both tests for spatial error and lag models are significant. However, we do not know which model to choose. Therefore, we look at the Robust LM tests, which clearly indicates that the spatial lag model (even when corrected for the presence of an error model) is to preferred above the spatial error model. Note that the Robust LM test is now only marginally significant. Running the spatial lag model with STATA, by using the 'spatreg' command gives the output as displayed in Table 6.3:

Clearly the spatial dependence parameter, $\hat{\rho}$, is significant and rather large. Namely, $\hat{\rho} = 0.43$ means that 43% of the crime rate in a neighbourhood is related with the weighted average

[14] these data can be dowloaded as well from `http://www.stata.com/stb/stb60` and are called columbusdata.dta and columbusswm.dta for the data and spatial weight matrix, respectively.

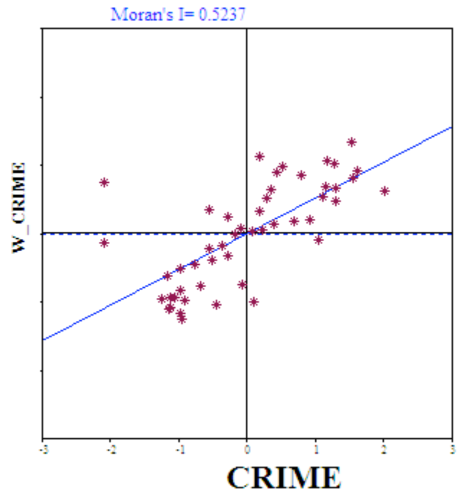

Figure 6.2: Moran scatterplot of 'crime' in Columbus, Ohio.

Table 6.2: Test statistics for spatial dependence

Test	Statistic	p-value
Moran's *I*	2.955	0.003
Spatial error:		
Lagrange multiplier	5.723	0.017
Robust Lagrange multiplier	0.079	0.778
Spatial lag:		
Lagrange multiplier	9.364	0.002
Robust Lagrange multiplier	3.720	0.054

Table 6.3: Estimation results : spatial lag

Variable	Coefficient	Std. Err.
hoval	-0.266^{**}	0.088
income	-1.032^{**}	0.328
Intercept	45.079^{**}	7.871
Spatial dependence parameter		
$\hat{\rho}$	0.431^{**}	0.124
Estimation of standard deviation		
$\hat{\sigma}$	9.772^{**}	0.998
N	49	
Log-likelihood	-182.39	
$\chi^2_{(1)}$	9.974	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

crime rate of its neighbours. Further, we see here that spatial lag processes indeed induces a bias in OLS regressions. To see this, compare the coefficient of income in Table 6.3 and Table 6.1. The coefficient has dropped with almost a third, indicating that income levels in a neighbourhood are not as important as previously believed when correcting for spatial dependence. So, does this mean that crime rates are *caused* by crime rates in neighborhoods nearby (copy-catting, contagion or peer effects?). Not necessarily so. It could as well be that there are unobserved spatial variables (socio-economic class, education, percentage broken families) that relate to both income and crime rates. So, this result signifies the importance of correcting for spatial dependence—especially in cross-sections, but points as well to the fact that one should be careful to interpretate these results directly.

7. CONCLUDING REMARKS

The use of spatial econometrics is now almost common practice—at least in the fields of regional science and urban economics. With the advent of increasing larger spatial datasets, there is also a need for spatial dependence correction. Besides, more and more computer software is available for estimating spatial regression models. Further, spatial dependence—be it as a spatial lag, spatially correlated exogenous variables or a spatial error model—seems to be omnipresent in cross-sectional data.

However, we would like to make a caveat here. The presence of spatial dependence does not necessarily mean that there is some sort of fancy spatial process—such as spillover processes between firms or network externalities—at work (see for a similar critique Gibbons and Overman, 2012). Usually, it indicates that the researcher has omitted an important spatially correlated variable or that the important spatially correlated variable can not be easily observed, like ambition and intelligence. Unobserved spatial heterogeneity is more rule than exception. The problem is though that if you are interested in finding a causal effect, spatial econometrics if not of much help as it does not aid identification of the underlying causal mechanism. What it can do is tacking spatial unobserved heterogeneity if one can not use spatial fixed effects as in a cross-section. Moreover, the marginal effects from a spatial econometric regression are difficult to assess. Therefore, and that is what Gibbons and Overman (2012) argue, the best way to incorporate spatial variables is via spatially correlated exogenous variables (the Durbin model). These are easy to interpret and do not require any additional assumptions. Still note that this not mean that you have found a causal effect.

That does not render spatial econometrics meaningless. On the contrary, controlling for spatial dependence might have a large impact in the size of the other coefficients and removes at least one particular and important source of bias in the data, even though a direct interpretation is often cumbersome. And if you have a (structural) model[15] that neatly describes the interaction then spatial econometrics give the right tools to empirical validate such a model.

[15] One can think here about models describing epidemics, congestion and social interaction effects in networks. Especially the latter now receives more and more attention in main stream economics and econometrics, but note that the causality here still needs attention (König et al., 2017).

A. MATRIX ALGEBRA

Matrix algebra is slightly different than normal algebra (with just numbers), especially in terms of multiplication and division. At first, it might seem very outlandish if you are not used to it. Therefore, this small appendix is provided which gives the most important characteristics of matrices and rules of matrix algebra.

A.1. WHAT IS A MATRIX?

A matrix $\mathbf{X}$ can be seen as a rectangular box filled with numbers x_{ij}, where x denotes a number and i and j are indices that run from 1 to I or J, respectively. Specifically, i stands for the i-th row of $\mathbf{X}$ and j for the j-th column (thus $x_{row,column}$). So, in general we have

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ij} & \cdots & x_{IJ} \end{bmatrix}. \tag{A.1}$$

A.2. IDENTITY MATRIX

And identity, $\mathbf{I}$, matrix is a matrix with zeros on the off-diagional and ones on the diagonal, so:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}. \tag{A.2}$$

multiplying a matrix or a vector with the identity matrix always gives back the same matrix of vector, so: $\mathbf{I}y = y$.

A.3. MATRIX ADDITION AND SUBTRACTION

Matrix addition and subtraction is fortunately rather easy. For matrix addition:

$$\mathbf{X} + \mathbf{Y} = \mathbf{Z}, \tag{A.3}$$

where each z_{ij} in matrix $\mathbf{Z}$ is calculated by $x_{ij} + y_{ij}$. Likewise for subtraction. So Matrix addition and subtraction is element wise, but only goes for matrices of the same size.

A.4. MATRIX MULTIPLICATION

For matrix multiplication we now use the so-called inner products of vectors (usually denoted by a $'$). Say, $\mathbf{x} = [x_1 \quad x_2 \quad x_3]$ and $\mathbf{y} = [y_1 \quad y_2 \quad y_3]$, then $\mathbf{x}'\mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3$. Assume that we have the following matrix multiplication

$$\mathbf{XY} = \mathbf{Z}, \tag{A.4}$$

then the number on the i-th row and j-th column of $\mathbf{Z}$ is calculated by the inner product of the i-th row of $\mathbf{X}$ and the j-th column of $\mathbf{Y}$. For example:

$$\mathbf{XY} = \begin{bmatrix} 1 & 3 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 1 \times 2 + 3 \times 1 & 1 \times 4 + 3 \times 6 \\ 4 \times 2 + 4 \times 1 & 4 \times 4 + 4 \times 6 \end{bmatrix} = \begin{bmatrix} 5 & 22 \\ 12 & 40 \end{bmatrix} = \mathbf{Z} \tag{A.5}$$

A.5. MATRIX DIVISION

Matrix division is the tough, both in concept and calculation (therefore, let a computer do all the work). Obviously, $1/\mathbf{X}$ is a strange concept. Therefore, the inverse, $\mathbf{X}^{-1}$, is invented and defined as follows (only for symmetrical matrices with equal number of rows and columns):

$$\mathbf{X}^{-1}\mathbf{X} = \mathbf{X}\mathbf{X}^{-1} = \mathbf{I} \tag{A.6}$$

So multiplying a matrix with its inverse, and vice versa, gives the identity matrix.

REFERENCES

Anselin, L (1988). *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers (cit. on pp. 2, 12).

— (2000). “Computing Environments for Spatial Data Analysis”. In: *Journal of Geographical Systems* 2, pp. 201–220 (cit. on p. 15).

Anselin, L and S Hudak (1992). “Spatial Econometrics in Practice”. In: *Regional Science and Urban Economics* 22, pp. 509–536 (cit. on p. 14).

Baltagi, B. H., S. H. Song, and W Koh (2003). “Testing Panel Data Regression Models with Spatial Error Correlation”. In: *Journal of Econometrics* 117, pp. 123–150 (cit. on p. 2).

Bivand, R. (2002). “Spatial econometrics functions in R: Classes and methods”. In: *Journal of geographical systems* 4.4, pp. 405–421 (cit. on p. 15).

Cairncross, F (1997). *The death of distance: how the communications revolution will change our lives.* London: Orion Publishing (cit. on p. 2).

Cliff, A and J. K. Ord (1981). *Spatial Processes: Models and Applications.* London: Pion (cit. on p. 2).

Dixit, A. K. and J. E. Stiglitz (1977). “Monopolistic Competition and Optimum Product Diversity”. In: *American Economic Review* 67, pp. 297–308 (cit. on p. 2).

Elhorst, J. P. (2001). “Dynamic Models in Space and Time”. In: *Geographical Analysis* 33.2, pp. 120–140 (cit. on pp. 2, 4).

Florax, R. and P Nijkamp (2005). “Misspecification in Linear Spatial Regression Models”. In: *Encyclopedia of Social Measurement* 2, pp. 695–707 (cit. on p. 12).

Friedman, T. (2005). *The world is flat: a brief history of the twenty-first century*. New York: Farrar, Straus and Giroux (cit. on p. 2).

Fujita, M, P Krugman, and A. J. Venables (1999). *The Spatial Economy: Cities, Regions, and Economic Trade*. Cambridge: MIT Press (cit. on p. 2).

Gibbons, S. and H. G. Overman (2012). “Mostly Pointless Spatial Econometrics?” In: *Journal of Regional Science* 52.2, pp. 172–191 (cit. on p. 19).

Kelejian, H. H. and I. R. Prucha (1998). “A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances”. In: *Journal of Real Estate Finance and Economics* 17.1, pp. 99–121 (cit. on pp. 2, 13–15).

— (2004). “Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations”. In: *Journal of Econometrics* 118, pp. 27–50 (cit. on pp. 2, 13–15).

König, M. D. et al. (2017). “Networks in conflict: Theory and evidence from the great war of africa”. In: *Econometrica* 85.4, pp. 1093–1132 (cit. on p. 19).

LeSage, J. P. and R. K. Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton: Chapman \& Hall/CRC (cit. on pp. 2, 6).

Mankiw, N. G., D Romer, and D. N. Weil (1992). “A Contribution to the Empiricis of Economic Growth”. In: *Quarterly Journal of Economics* 107, pp. 407–437 (cit. on p. 4).

Pace, R. K. (1997). “Performing Large Spatial Regressions and Autoregressions”. In: *Economics Letters* 54, pp. 283–291 (cit. on p. 15).

Pace, R. K. and R Barry (1997). “Sparse Spatial Autoregressions”. In: *Statistics & Probability Letters* 33, pp. 291–297 (cit. on p. 15).

Pace, R. K. et al. (1998). “Spatiotemporal Autoregressive Models of Neighborhood Models”. In: *Journal of Real Estate Finance and Economics* 17.1, pp. 15–33 (cit. on p. 15).

Proost, S. and J.-F. Thisse (2019). “What Can Be Learned from Spatial Economics?” In: *Journal of Economic Literature* (cit. on p. 2).

Stock, J. H. and M. W. Watson (2015). “Introduction to Econometrics (3rd Updated Edition)”. In: (cit. on p. 10).

Tobler, W (1970). “A Computer Movie Simulating Urban Growth in the Detroit Region”. In: *Economic Geography* 46, pp. 234–240 (cit. on p. 1).