Daniela Gumprecht

Spatial methods in econometrics

Thesis

WIRTSCHAFTS UNIVERSITÄT WIEN

Titel der Dissertation:

Spatial Methods in Econometrics

Dissertation zur Erlangung des akademischen Grades

einer Doktorin/eines Doktors

der Sozial- und Wirtschaftswissenschaften an der Wirtschaftsuniversität Wien

eingereicht bei

1. Beurteilerin/1. Beurteiler: **Univ.Prof. Mag.Dr. Werner G. Müller**

2. Beurteilerin/2. Beurteiler: **Univ.Prof. Dr. Wilfried Grossmann**

von **Mag. Daniela Gumprecht**

Fachgebiet: **Ökonometrie, Spezielle Statistik**

Wien, im **Mai 2007**

Spatial Methods in Econometrics

Mag. Daniela Gumprecht

May 2007

Abstract

This thesis deals with the appropriate handling of spatial data in general, and in particular in the framework of economic sciences. An overview of well known methods from the field of spatial statistics and spatial econometrics is given. Furthermore a special class of spatial objects is described, namely objects that are that far apart from all other observations in the dataset, that they are not connected to them anymore. Different treatments of such objects are suggested and their influence on the Moran's $\mathcal{I}$ test for spatial autocorrelation is analysed in more detail. After this theoretical part some adequate spatial methods are applied to the well-known problem of R&D spillovers. The corresponding dataset is not obviously spatial, nevertheless spatial methods can be used. The spatial contiguity matrix is based on an economic distance measure instead of the commonly used geographic distances. Finally, optimal design theory and spatial analysis are combined via a new criterion. This criterion was developed to be able to take a potential spatial dependency of the data points into account. The aim is to find the best design points that show the same spatial dependence structure as the true population.

Abstract

Diese Dissertation beschäftigt sich allgemein mit geeigneten Analysemethoden für räumliche Daten, wobei speziell auf Anforderungen der Wirtschaftswissenschaften eingegangen wird. Zunächst wird ein Überblick über die bekanntesten Methoden der räumlichen Statistik und der räumlichen Ökonometrie gegeben. Des Weiteren wird eine spezielle Klasse von räumlichen Objekten behandelt. Es handelt sich hierbei um Objekte, die so weit von allen übrigen Beobachtungen des Datensatzes entfernt sind, dass sie keinen Bezug mehr zu diesen haben. Es werden verschiedene Möglichkeiten der Behandlung solcher Objekte erörtert und der Einfluss dieser Verfahrensformen auf den Moran's $\mathcal{I}$ Test für räumliche Autokorrelation analysiert. Nach diesem theoretischen Teil werden einige geeignete räumliche Methoden für die Analyse von Forschungs- und Entwicklungs Spillover Effekten verwendet. Das Interessante hier ist, dass der entsprechende Datensatz keine offensichtliche räumliche Komponente enthält, und dass das verwendete Distanzschema auf wirtschaftlichen- anstelle der üblicherweise verwendeten geografischen Distanzen basiert. Im letzten Teil dieser Arbeit werden Optimal Design Theorie und räumliche Analyse über ein neues Kriterium kombiniert. Dieses Kriterium wurde entwickelt, um eine möglicherweise vorhandene räumliche Abhängigkeit der Daten berücksichtigen zu können. Ziel ist, das beste Design zu finden, so dass die wahre räumliche Abhängigkeitsstruktur der Population in den Designdaten abgebildet wird.

Acknowledgments

My special thanks go to my supervisor Prof. Werner G. Müller. I would like to thank him for supervising my thesis, for all his time and energy he spent to support me, and all his patience he had with me, for our regular discussions and his valuable advices and fruitful contributions within all steps of my thesis. This supervision was simply perfect.

I am also grateful to Prof. Wilfried Grossmann, my second supervisor, for providing helpful comments and his time he spent for my thesis.

Finally I would like to thank my family and friends, especially my parents Anita and Franz, my sister Nicole, and my boyfriend Roman, for their support during my whole studies.

Contents

1 Preposition **1**

2 Spatial Data **4**
2.1 Spatial Weight Matrix 5
2.1.1 Contiguity Matrices 6
2.1.2 Distance Matrices 7
2.1.3 Coding Schemes 7
2.1.4 Local Spatial Structures 8
2.2 Spatial Statistics versus Spatial Econometrics 8
2.3 Spatial Dependence and Spatial Autocorrelation 10
2.4 Spatial Stochastic Processes 11

3 Measuring Spatial Dependence **15**
3.1 Moran's $\mathcal{I}$ 15
3.1.1 Moran's Plot 16
3.1.2 Moran's $\mathcal{I}$ Test 16
3.1.3 Local Moran's $\mathcal{I}$ 19
3.2 Getis Statistic 20

4 Treatment of Far-Off Objects in Moran's $\mathcal{I}$ Test **21**
4.1 Motivation 21
4.2 Definitions 22
4.3 Moran's $\mathcal{I}$ 25
4.4 Expected Value of $\mathcal{I}$ 27
4.5 Variance of $\mathcal{I}$ 29
4.6 Standardized Moran's $\mathcal{I}$ 36
4.7 Findings 38

5 Handling of Spatial Data **42**
5.1 Spatial Filtering . . . 42
5.2 Estimation in Spatial Econometrics . . . 43
5.2.1 Spatial Two Stage Least Square Estimation . . . 44
5.2.2 Maximum Likelihood Estimation . . . 44
5.3 Measuring and Estimating Spatial Relations in Geostatistics . 45
5.3.1 Covariogram and Correlogram . . . 46
5.3.2 Theoretical Variogram . . . 46
5.3.3 Variogram Cloud and Sample Variogram . . . 47
5.3.4 Regional Variogram . . . 48
5.3.5 Variogram Models . . . 48
5.3.6 Upper-Austria SO_2 Monitoring Network . . . 49
5.3.7 Kriging . . . 50

6 R&D Spillovers - An Application in Spatial Econometrics **51**
6.1 R&D Spillovers Data . . . 51
6.1.1 'Non-Spatial' Analysis and Results . . . 52
6.1.2 Other Non-Spatial Approaches . . . 54
6.2 A Spatial Approach for the Analysis of R&D Spillovers . . . 56
6.2.1 Results of a Spatial Analysis . . . 58
6.3 Findings . . . 60

7 Optimal Design for Spatial Data **64**
7.1 Motivation . . . 65
7.1.1 Spatial Link Matrices . . . 67
7.1.2 Models . . . 67
7.2 Moran's $\mathcal{I}$ . . . 68
7.2.1 Status of the World: Spaceless . . . 69
7.2.2 Status of the World: Spatial Dependence . . . 70
7.3 Optimal Design Considerations . . . 71
7.3.1 A Criterion . . . 71
7.3.2 Design Algorithms . . . 72
7.4 Examples . . . 76
7.4.1 Artificial Dataset . . . 76
7.4.2 Columbus Crime Data . . . 79
7.4.3 Upper-Austria SO_2 Monitoring Network . . . 80
7.4.4 R&D Spillovers . . . 82
7.5 Findings . . . 83

8 Conclusions **84**

References **85**

List of Figures

3.1 Moran scatterplot for 49 observations 17

4.1 Variances of $\mathcal{I}$ and differences between the variances within the extreme cases 39
4.2 $z(\mathcal{I})$ for $\mathcal{I} = 0.01$, and differences within the extreme cases . . 40
4.3 $z(\mathcal{I})$ for $\mathcal{I} = 0.10$, and differences within the extreme cases . . 40
4.4 $z(\mathcal{I})$ for $\mathcal{I} = 0.50$, and differences within the extreme cases . . 41
4.5 $z(\mathcal{I})$ for $\mathcal{I} = 0.99$, and differences within the extreme cases . . 41

5.1 Upper-Austria SO_2 monitoring network 50

6.1 Landscape based on trade-intensities between the countries . . 57

7.1 Neighbourhoods in Columbus, Ohio 66
7.2 Correlation for a 25 point grid 77
7.3 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices 78
7.4 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices 79
7.5 Left: Simple search. Right: Fedorov exchange algorithm. . . . 80
7.6 Columbus crime data, simple search & exchange algorithm, contiguity matrix 81
7.7 Upper-Austria SO_2 monitoring network 82
7.8 Economic landscape with optimal design points 83

List of Tables

6.1 Moran's $\mathcal{I}$ for residuals of the Fixed Effects model with independent variable $logS_{it}$; for all periods $\mathrm{E}[\mathcal{I}] = -0.0476$ and $\mathrm{Var}[\mathcal{I}] = 0.0053$. 61
6.2 Results for R&D Spillovers: Static Fixed Effects Model. 61
6.3 Results for R&D Spillovers: Dynamic Fixed Effects Model. . . 62
6.4 Results for R&D Spillovers: Static Random Coefficients Model 62
6.5 Results for R&D Spillovers: Dynamic Random Coefficients Model 63

Chapter 1

Preposition

The topic of this thesis is the appropriate handling of spatial data in general, and in particular in the framework of economic sciences. The field of spatial econometrics is comparatively new, it arose in the 1970ies and although it gained a high level of acceptance in the meanwhile, it is not for a long time completely worked out yet, and there are still many questions left open.

The three general aims of this thesis are: Firstly, giving an overview over spatial data and different appropriate estimation techniques that stem from different disciplines, namely the field of spatial statistics and the field of spatial econometrics. Secondly, use some of these spatial methods to analyse real economic data, which are not obviously 'spatial', to show that spatial analysis is much more powerful than it might appear at first glance. Thirdly, combine the fields of optimal design theory and spatial analysis via a new criterion.

The thesis is structured in the following way:

Chapter 2: A general introduction to the field of spatial analysis is given. In general, spatial data are characterized, and the differences to non-spatial data are shown, and in particular different types of spatial data are described and several approaches for the analysis are shown. This introduction also gives an answer to the question why and when spatial methods are needed, and what are the advantages compared with well known and well understood 'standard' (non-spatial) methods of statistics and econometrics. The similarities and differences between theses aspects are explained and singled out,

and also the distinction between spatial statistics and spatial econometrics is worked out.

Chapter 3: This chapter deals with the first question that has to be answered whenever spatial data are of interest, namely how to measure and test a potentially existent spatial effect in a dataset. The focus is put on the probably most famous statistic for spatial dependence, the Moran's $\mathcal{I}$. Moran's $\mathcal{I}$ can be used for global as well as local spatial structures. Furthermore, the moments of $\mathcal{I}$, which are needed for a quite simple asymptotic test about the level of autocorrelation, are given. Close related to $\mathcal{I}$ is the Moran's plot, which is a diagnostic tool to visualize spatial autocorrelation. Beside $\mathcal{I}$ the Getis statistic, which is an other measurement for spatial autocorrelation is mentioned.

Chapter 4: Treatment of far-off objects in Moran's $\mathcal{I}$ test. Here a question, that arose during the work on this thesis is discussed. It is about a special class of objects in a spatial dataset, namely objects that are far apart from all other observations in the sense that they are not spatially connected to others, but still belong to the dataset. The questions to be answered are, how to deal with such objects, and how do different treatments of such observations influence the Moran's $\mathcal{I}$ test.

Chapter 5: This chapter deals with the appropriate estimation of parameters in spatial analysis. There are two different ideas behind the estimation of spatial models, and therefore two different approaches. One possibility is to filter out the spatial effect and use standard statistical methods for further analysis of the filtered (spaceless) data. The other alternative is to use special spatial estimation techniques where the spatial dependence is explicitly included in the estimation. Depending whether the methods stem from the field of spatial statistics or spatial econometrics, there are differences in the procedures.

Chapter 6: It contains an application of some of the previously explained methods from spatial econometrics to empirical data, namely an analysis of R&D spillovers. For the R&D spillovers data, the question to be answered is: Do R&D spillover effects exist? This topic is discussed in many papers and as non-spatial analysis lead to different conclusions, methods from the field of spatial econometrics are applied.

Chapter 7: The last chapter deals with methods from the field of Optimal Design. The aim is to find the best design for an experiment to detect spatial dependence if it is existent and at the same time to avoid to spuriously detect a spatial structure if it is not inherent in the population. Therefore a new design criterion is presented and different algorithms how to compute a design are discussed. The procedure is illustrated by applying it to different datasets: an artificial dataset, a well-known case study in spatial analysis, the R&D dataset from chapter 6, and an other real dataset from the field of geostatistics presented in chapter 5.

Chapter 2

Spatial Data

The main characteristic of spatial data is, that they contain beside the attribute information also information about the location of the object, therefore they are also called geo-referenced data. Whenever data contain information about their location, the question rises whether this information can be used to improve statistical conclusions drawn from this data. In recent years a growing number of spatial data can be observed, mainly through the dispersion of geographic information systems (GIS) and spatial data analysis software.

The locational information of a spatial object can represent either single points, or areas in a certain study region. Commonly it is assumed that the study area is a two dimensional surface or map. The pattern of points on a two dimensional surface can be divided into three classes:

(i) Regular pattern: The distances between the points are equal, the objects form a regular grid.

(ii) Random pattern: The placements of the points vary randomly across the whole region, and do not influence each other.

(iii) Aggregated or clustered pattern: The points tend to occur in clusters.

see e.g. Davis (2002). Alternatively to points, the surface can be partitioned into areas. These areas are always assumed to be mutually exclusive and two adjacent regions always share a common bound. Such a partition is then

called spatial tessellation of the region into n tiles, see e.g. Tiefelsdorf (2000). The areas are not restricted in any way, except of the limitations given above. They can occur in various forms, e.g. the bounds can be straight lines or cleft edges, the sizes can be the same or diverse. The areas can represent natural entities like political or economic districts, or artificial entities created by a scientist. Spatial tessellation can easily be reduced to point patterns, simply using only a single point instead of the whole region, usually the centroid is used to represent an area, but also capitals of political districts can be used.

Spatial relationships are modelled via the connectivity of the single observations. There are many different approaches known, how to quantify the connectivity or proximity. There is a distinction between two general proposals, (1) measuring the similarities e.g. via neighbourships, and (2) measuring the dissimilarities e.g. via the distances between the spatial objects. For a strict and general mathematical definition of relations between spatial objects see e.g. Tiefelsdorf (2000, p. 25).

2.1 Spatial Weight Matrix

However the spatial relationships are quantified, they are represented in spatial weight matrices (also called spatial link matrices). In general, a spatial link matrix $\mathbf{U} = [u_{ij}]$ is a fixed (non-stochastic) n by n matrix (n is the number of observations in the dataset) with the following properties:

(i) $u_{ij} = 0$ if i and j are not spatially connected and if $i = j$ by definition, i.e. all elements on the main diagonal are zero, this means an object is not spatially connected with itself.

(ii) $u_{ij} \neq 0$ if i and j are spatially connected, and usually these values are greater than zero.

If similarities are measured the spatial link matrix is called contiguity matrix, if dissimilarities are measured it is called distance matrix. Similarity and dissimilarity matrices are inversely related - the higher the connectivity, the smaller the distance and vice versa.

2.1.1 Contiguity Matrices

Similarity can be quantified in many different ways, hence there are many different spatial link matrices in use. **Neighbourhood matrices** are symmetric, binary, n by n spatial link matrices with $u_{ij} = 1$ if two observations are neighbours and $u_{ij} = 0$ if not and if $i = j$ by definition, i.e. an object is not neighbour of itself. The concrete form of the matrix depends on the definition of the neighbourship. The most commonly used definitions of neighbourhood are:

(i) Rook's criterion: Adjacent areas are neighbours if they share nonzero-length boundaries.

(ii) Bishop's criterion: Adjacent areas are neighbours if they share zero-length boundaries.

(iii) Queen's criterion : Adjacent areas are neighbours if they share zero-length or nonzero-length boundaries.

Corresponding to the diverse criterions, the neighbours (N) of an object (O) can be seen in the following pictures:

Rook's criterion

	N	
N	O	N
	N	

Bishop's criterion

N		N
	O	
N		N

Queen's criterion

N	N	N
N	O	N
N	N	N

To build neighbourhood matrices one can use first-, second-, ... order neighbours. Neighbours can also be defined via a certain distance r, all objects which lie within this distance (radius) are neighbours of an observation.

Spatial **connectivity matrices** are similar to neighbourhood matrices, but they are non-binary. They are also originally symmetric n by n matrices, and the elements u_{ij} measure the intensity of the contiguousness by some adequate function, e.g. simply using the inverse distances between the locations.

A quite common practice to define a spatial weight matrix **U** is the use of a distance decay function, that means observations which are geographically further distant are downweighted, the degree of the weighting is controlled by a locality parameter δ. The distance decay can e.g. be modelled by a negative exponential function,

$$u_{ij} = \exp(-\delta d_{ij}) \tag{2.1}$$

with d_{ij} denoting the geographical distances between region i and j, see e.g. Badinger et al. (2002).

2.1.2 Distance Matrices

Similar to these connectivity matrices are distance matrices, which are again symmetric n by n matrices, here the elements u_{ij} measure the distance between locations. Again the elements on the main diagonal are zero by definition, i.e. the distance between an object and itself is always zero. In practice the geographic distances between two observations are often used. If the location of an observation is a whole area, the centers of the regions can be used to measure the distances.

In spatial econometrics usually contiguity matrices are used, whereas in spatial statistics the analysis are typically based on distance matrices. It is worth to notice, that neither distance- nor contiguity matrices are restricted to geographic space, one can also use some other kind of characteristics to formulate an adequate distance or connectivity measurement. In chapter 6 such 'non-geographic' distances and connectivity measurements are used to analyse empirical data.

2.1.3 Coding Schemes

The original symmetric spatial link matrices are often converted by using coding schemes to cope with the heterogeneity which is induced by the different linkage degrees of the spatial objects. Tiefelsdorf (2000) defines the linkage degree of a spatial object i by the total sum of its interconnections with all other spatial objects, that is $d_i = \sum_{j=1}^{n} u_{ij}$. There are mainly three different coding schemes used: (1) the globally standardized C-coding scheme, (2) the row-sum standardized W-coding scheme, and (3) the variance stabilizing S-coding scheme. In all these three schemes the overall sum of the elements is

n. Standardized spatial weight matrices are non-binary and often also non-symmetric. The most important, or most commonly used coding scheme is the row-sum standardized W-coding scheme. In such a row standardized spatial link matrix the sum of each row is equal to one, the elements are simply calculated by $w_{ij} = \frac{u_{ij}}{\sum_{j=1}^{n} u_{ij}}$.

Beside these global spatial link matrices, which represent the connectivity between all observations, there exist also local spatial weight matrices, which show the spatial connectivity for a single object.

2.1.4 Local Spatial Structures

For each spatial object i a local spatial link matrix can be constructed, by simply using the i-th row and the i-th column of the global spatial link matrix and setting all other elements zero. This gives n star-shaped local spatial matrices $\mathbf{U}_i$:

$$\mathbf{U}_i = \begin{pmatrix} 0 & \cdots & 0 & u_{1i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{i-1,i} & 0 & \cdots & 0 \\ u_{i1} & \cdots & u_{i,i-1} & 0 & u_{i,i+1} & \cdots & u_{i,n} \\ 0 & \cdots & 0 & u_{i+1,i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{ni} & 0 & \cdots & 0 \end{pmatrix}. \tag{2.2}$$

The global spatial link matrix $\mathbf{U}$ in turn yields to be $\mathbf{U} = \frac{1}{2}\sum_{i=1}^{n} \mathbf{U}_i$ for symmetric link matrices. For non-symmetric link matrices $\mathbf{V}$ it holds that $\frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}') = \sum_{i=1}^{n} \mathbf{V}_i$, see e.g. Tiefelsdorf (2000).

2.2 Spatial Statistics versus Spatial Econometrics

The difference between spatial statistics or geostatistics and spatial econometrics is the same as in statistics and econometrics in general. In spatial econometrics a theoretical model plays the central role, whereas in geostatistics the data are in the focus. Thus, in spatial statistics the approach is data

based, in spatial econometrics the approach is model based.

The original problem in geostatistics, summarized by Clark and Harper (2000), is on one hand

> to characterise and interpret the behavior of the existing sample data

and on the other hand

> to produce 'maps' of the values at unsampled locations. That is, to estimate unknown values at locations which have not been sampled.

A definition of the main subject in spatial econometrics is e.g. given in Anselin (1999):

> Spatial econometrics is a subfield of econometrics that deals with the treatment of spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) in regression models for cross-sectional and panel data.

Based on the different approaches and requirements of spatial statistics and spatial econometrics, there are different concepts used, and their importance is weighted differently, although both fields deal more or less with the same tenor, namely data which are somehow related over space. The theoretical background are for both subjects spatial stochastic processes, as spatial data are assumed to exhibit some kind of structure reflecting the spatial dependence, and some additional random effect.

The main question, both in the field of spatial statistics and the field of spatial econometrics, is whether there is a significant regional effect inherent in the data or not. If data are spatially independent, there is no need for a special treatment, standard methods can be used, whereas if data are spatially connected, special techniques have to be adopted because standard assumptions, like uncorrelated errors, are violated.

2.3 Spatial Dependence and Spatial Autocorrelation

The first question that rises, if one has to deal with spatial data, is: Are data spatially dependent? If data are spatially dependent, this dependence can be exploited to improve statistical conclusions. In this case special methods are needed. If data are not regional dependent, there is no information spillover in the dataset that can be used, and standard methods can be executed. Many different explanations of spatial dependence and spatial autocorrelation can be found in literature. Fotheringham et al. (2002) e.g. give the following specification of spatial dependence:

> It is the extent to which the value of an attribute in one location depends on the values of the attribute in nearby locations.

And Griffith (2003) says about spatial autocorrelation:

> It (...) is the correlation among values of a single variable strictly attributable to the proximity of those values in geographic space (...).

Both of these explanations are related to geographic space, nevertheless this is no general restriction, all kinds of measurement that are modelled in a spatial link matrix can be used.

The subject of spatial dependence and autocorrelation is related to the well known problem of autocorrelation in time series. The main difference between spatial- and time autocorrelation is the dimension and the direction of the dependence, in time series there is only one dimension and only one direction of influence possible, by contrast in the field of spatial dependence there are two dimensions and therefore also two directions of dependence. This is, what makes the whole field more complex. In general, the meaning of autocorrelation is the same:

(i) Positive spatial autocorrelation: Nearby values of a variable tend to be similar, high values tend to occur near high values, medium values near medium values, and low values near low values.

(ii) Negative spatial autocorrelation: Nearby values of a variable tend to be dissimilar, high values tend to occur near low ones, medium values near medium ones, and low values near high ones.

(iii) No spatial autocorrelation: there is no pattern identifiable, high, medium and low values appear in a random formation.

2.4 Spatial Stochastic Processes

A stochastic (or random) process is a collection of random variables $\{Z(\mathbf{t})|\mathbf{t} \in T\}$ indexed by a set T, where T is typically assumed to be a subset of the real numbers or $[0, \infty)$. This is standard practice in time series literature. However, the set can also be a more general index set D, e.g. pairs of integers which can represent coordinates on a surface, or a subset of the plane. The big difference to time series is, that spatial index sets do not have a natural ordering like time indices (Ripley, 1981). The process $\{Z(\mathbf{x})|\mathbf{x} \in D\}$ is often called a (random) process if $D \subseteq \mathbb{R}$ and it is often called a (random) field if $D \subseteq \mathbb{R}^n$ for $n \geq 2$ (Rao, 1979). The concrete form of the stochastic process is then given by the joint distribution function $F_{Z_1,...,Z_n}(\cdot, ..., \cdot)$. The covariance and the correlation between two observations $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$ in X are given by

$$\mathrm{VC}(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{E}[\{Z(\mathbf{x}_i) - \mathrm{E}(Z(\mathbf{x}_i))\}\{Z(\mathbf{x}_j) - \mathrm{E}(Z(\mathbf{x}_j))\}] \tag{2.3}$$

and

$$\mathrm{Corr}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathrm{VC}(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{\mathrm{Var}(\mathbf{x}_i)\mathrm{Var}(\mathbf{x}_j)}} \ . \tag{2.4}$$

There are different kinds of stochastic processes, depending on some special properties they own. A stochastic process is called

(i) *Stationary*, if the distributions are invariant under an arbitrary translation of the points by a vector $\mathbf{h}$:

$$\mathrm{P}(Z(\mathbf{x}_1) < z_1, ..., Z(\mathbf{x}_k) < z_k) = \mathrm{P}(Z(\mathbf{x}_1+\mathbf{h}) < z_1, ..., Z(\mathbf{x}_k+\mathbf{h}) < z_k)$$

A stationary process is homogeneous in space.

(ii) *Weakly Stationary* (or second-order stationary), if the moments of the stochastic process are invariant under translations. This means for the first moment, that $\mathrm{E}[Z(\mathbf{x})] = \mathrm{E}[Z(\mathbf{x}+\mathbf{h})] = m$, and for the covariance function, that $\mathrm{E}[\{Z(\mathbf{x}) - \mathrm{E}(Z(\mathbf{x}))\}\{Z(\mathbf{x}+\mathbf{h}) - \mathrm{E}(Z(\mathbf{x}+\mathbf{h}))\}] = \mathrm{VC}(\mathbf{h})$. The mean is constant and the covariance only depends on the separation $\mathbf{h}$.

(iii) *Isotropic*, if its covariance function only depends on the length $|\mathbf{h}|$ of the vector $\mathbf{h}$ and not on its orientation.

(iv) *Intrinsic*, if for every vector $\mathbf{h}$ the increment $Y_{\mathbf{h}}(\mathbf{x}) = Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$ is a weakly stationary random process in $\mathbf{x}$.

(v) *Gaussian*, if its finite-dimensional distributions are multivariate Gaussian.

These definitions can be found e.g. in Chilès and Delfiner (1999).

These classifications of stochastic processes are commonly used and quite important in geostatistics, in the field of spatial econometrics this distinction does not play such an important role and therefore can not be found in econometric literature that often, except of the specification of a Gaussian process. Hence, it is specified in more detail.

Gaussian Spatial Process

A random process depends on its internal structure and on the input of the spatial process. If the input of the spatial process is a white noise, i.e. it is a random vector where the elements are independently identically normally distributed, the random process is called *Gaussian*, see Tiefelsdorf (2000). For a regression model with normally distributed regression disturbances $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, all finite-dimensional distributions are multivariate Gaussian, that is:

$$f(z_1, ..., z_n) = \frac{1}{\sqrt[n]{(2\pi)^2}} |\boldsymbol{\Omega}(\rho)|^{-\frac{1}{2}} \exp\left(\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\rho)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right),$$

where $\boldsymbol{\Omega}(\rho)$ denotes the covariance matrix of the disturbances, it depends on a spatial autocorrelation parameter ρ, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\rho))$. The spatial parameter ρ gives the magnitude of the spatial dependence, it can be positive or negative, if ρ is zero, the spatial structure is irrelevant (Tiefelsdorf, 2000).

Spatial Gaussian processes are further divided into sub-divisions, namely the autoregressive (AR) processes and the moving average (MA) processes.

Autoregressive Spatial Process

The **Simultaneous Autoregressive Process** is the most commonly used AR process in spatial econometrics. In the common econometric notation of a regression model, it is expressed by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \text{ with } \mathbf{u} = \rho\mathbf{V}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{2.5}$$

where ρ is the spatial autoregressive parameter, $\mathbf{V}$ is a spatial link matrix, now denoted by $\mathbf{V}$ to give consideration to the fact that it can be standardized and asymmetric, and $\boldsymbol{\varepsilon}$ are i.i.d. $\sim N(\mathbf{0}, \sigma^2\mathbf{I})$ errors. This model is called **spatial autoregressive (SAR) error model** and it can be easily transformed into the following formula, which is called the reduced form of (2.5), see Anselin (1999):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{V}\mathbf{y} - \mathbf{V}\mathbf{X}\rho\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.6}$$

The dependent variable $\mathbf{y}$ is influenced by $\mathbf{X}\boldsymbol{\beta}$ (like in standard regression models), by the spatially dependent endogenous variable $\rho\mathbf{V}\mathbf{y}$, and by the spatial trend component $-\mathbf{V}\mathbf{X}\rho\boldsymbol{\beta}$, where the components of the coefficient $\rho\boldsymbol{\beta}$ are not directly identifiable. It is required that the estimated coefficient $\widehat{\rho\boldsymbol{\beta}}$ fulfills the so called common factor constraint: $\widehat{\rho\boldsymbol{\beta}} = \hat{\rho}\hat{\boldsymbol{\beta}}$ (Tiefelsdorf, 2000).
The covariance matrix $\boldsymbol{\Omega}(\rho)$ between the error terms specifies the kind of spatial process, for the SAR process it is

$$\boldsymbol{\Omega}(\rho) = \sigma^2[(\mathbf{I} - \rho\mathbf{V})'(\mathbf{I} - \rho\mathbf{V})]^{-1}. \tag{2.7}$$

As $\boldsymbol{\Omega}(\rho)$ has to be positive definite, ρ is restricted to the interval $]\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}[$, where λ denotes the eigenvalues of the spatial link matrix $\mathbf{V}$, the smallest and the biggest one.

Assuming that there is no spatial effect of $\mathbf{X}\boldsymbol{\beta}$, i.e. leaving out the spatial trend component $-\mathbf{V}\mathbf{X}\rho\boldsymbol{\beta}$, leads to a **spatial lag model**:

$$\mathbf{y} = \rho\mathbf{V}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.8}$$

with $\boldsymbol{\varepsilon}$ being i.i.d. error terms. In this case the spatial dependence is strictly attributable to the endogenous variable $\mathbf{V}\mathbf{y}$. The term $\mathbf{V}\mathbf{y}$ is called *spatial lag* or *spatial smoother*, if the spatial link matrix $\mathbf{V}$ is row-standardized, it

can be interpreted as a weighted average of spatially connected observations. The spatial lag corresponds to the shift in time series. The main difference between the spatial lag and the time shift is, that a spatial lag does not have a 'natural' direction of the influence like a time shift.

The **Conditional Autoregressive (CAR) Process** is very seldom used in practice because the covariance matrix is given by

$$\mathbf{\Omega}(\rho) = \sigma^2(\mathbf{I} - \rho\mathbf{V})^{-1}$$

and is therefore very restrictive, the spatial link matrix has to be symmetric, which forbids the application of the S- and W-coding scheme. For a CAR process, values of ρ are also restricted to the interval $]\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}[$, see Tiefelsdorf (2000).

Moving Average Process

The Spatial Moving Average Process is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \text{ with } \mathbf{u} = \rho\mathbf{V}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon},$$

ρ stands again for the spatial parameter, here it is the moving average parameter, and it can take values in the interval $]-\frac{1}{\lambda_{max}}, -\frac{1}{\lambda_{min}}[$, and $\boldsymbol{\varepsilon}$ are again i.i.d. errors with $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\mathbf{I})$. The covariance matrix between the error terms is given by

$$\mathbf{\Omega}(\rho) = \sigma^2(\mathbf{I} + \rho\mathbf{V})(\mathbf{I} + \rho\mathbf{V})'$$

see Tiefelsdorf (2000).

Chapter 3

Measuring Spatial Dependence

The first question to be answered whenever analysts deal with spatial data is, whether there is a regional effect inherent or not. If not, i.e. if the observations are spatially independent, standard methods can be used. If there exists a significant spatial structure in the data, standard methods lead to biased and/or inconsistent results. There are different methods for measuring and testing a potential spatial effect in a dataset.

3.1 Moran's $\mathcal{I}$

The most famous procedure for measuring and testing spatial autocorrelation is based on Moran's $\mathcal{I}$ statistic developed by Moran (1948), (1950a) and (1950b). Moran's $\mathcal{I}$ is a measure for the intensity of spatial autocorrelation in a spatial stochastic process. It was originally developed for binary continguity matrices (neighbourhood matrices) but it is not restricted to this case, one can also use standardized continuous spatial link matrices.

For a standard regression model, normally one is interested whether the residuals are spatially dependent or not. If they are not, one can use standard estimation methods. If they are, one has to use special methods to take care of the spatial dependence because autocorrelation in the error term leads to biased estimates of the residual variance and inefficient estimates of the regression coefficients when the OLS estimation method is applied, see e.g. Cliff and Ord (1981). Tiefelsdorf (2000) focuses on spatial econometrics, and therefore gives the following explanation of Moran's $\mathcal{I}$: It is a scale invariant

ratio of quadratic forms in the normally distributed regression residuals $\hat{\boldsymbol{\varepsilon}}$.

$$\mathcal{I} = \frac{\hat{\boldsymbol{\varepsilon}}' \frac{1}{2}(\mathbf{V} + \mathbf{V}')\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}, \tag{3.1}$$

where $\mathbf{V}$ is a standardized spatial link matrix (generated by using the C-, S- or W-coding scheme) and $\hat{\boldsymbol{\varepsilon}}$ are usually the normally distributed OLS residuals of the regression model. If the spatial dependence of a single variable is of interest, one simply uses a regression on only an intercept and takes the corresponding residuals for the calculation of $\mathcal{I}$. This form can also be found in Cliff & Ord (1981).

3.1.1 Moran's Plot

Moran's plot is a diagnostic tool to visualize spatial autocorrelation. It is close related to Moran's $\mathcal{I}$. The Moran's plot is a scatterplot, and it gives a graphical representation of the relationship between a variable $\mathbf{y}$ and its spatial lag $\mathbf{Vy}$, typically the row-standardized spatial weight matrix is used. For such a scatterplot first the variable $\mathbf{y}$ is standardized to z-scores $\mathbf{z}_Y$, second the spatial lag of $\mathbf{z}_Y$ is constructed by multiplying it with the spatial weight matrix $\mathbf{V}$ to get $\mathbf{Vz}_Y$. Moran's plot is the scatterplot of $\mathbf{Vz}_Y$ against $\mathbf{z}_Y$. The slope of a no-intercept regression model of $\mathbf{Vz}_Y$ on $\mathbf{z}_Y$ is Moran's $\mathcal{I}$. $\mathcal{I}$ can be interpreted as the spatial autocorrelation, i.e. the correlation between $\mathbf{z}_Y$ and its spatial lag $\mathbf{Vz}_Y$. In Figure 3.1 an example for a Moran scatterplot is given, Moran's $\mathcal{I}$ is exactly the slope of the regression line. Data for this plot stem from the classical Columbus Crime dataset from Anselin (1988) which includes observations for 49 contiguous Planning Neighbourhoods in Columbus, Ohio. Variable $\mathbf{y}$ is a measure of crime, it includes residential burglaries and vehicle thefts per thousand households in a region, and the spatial weight matrix $\mathbf{V}$ is the row-standardized neighbourhood matrix.

3.1.2 Moran's $\mathcal{I}$ Test

The Moran's $\mathcal{I}$ test is used for parametric hypotheses about the spatial autocorrelation level ρ, i.e. $H_0 : \rho = 0$ against $H_A : \rho > 0$ for positive spatial autocorrelation; or $H_0 : \rho = 0$ against $H_A : \rho < 0$ for negative spatial autocorrelation. Tests for positive correlation are much more relevant in practice, because negative spatial autocorrelation very seldomly appears in the real world.

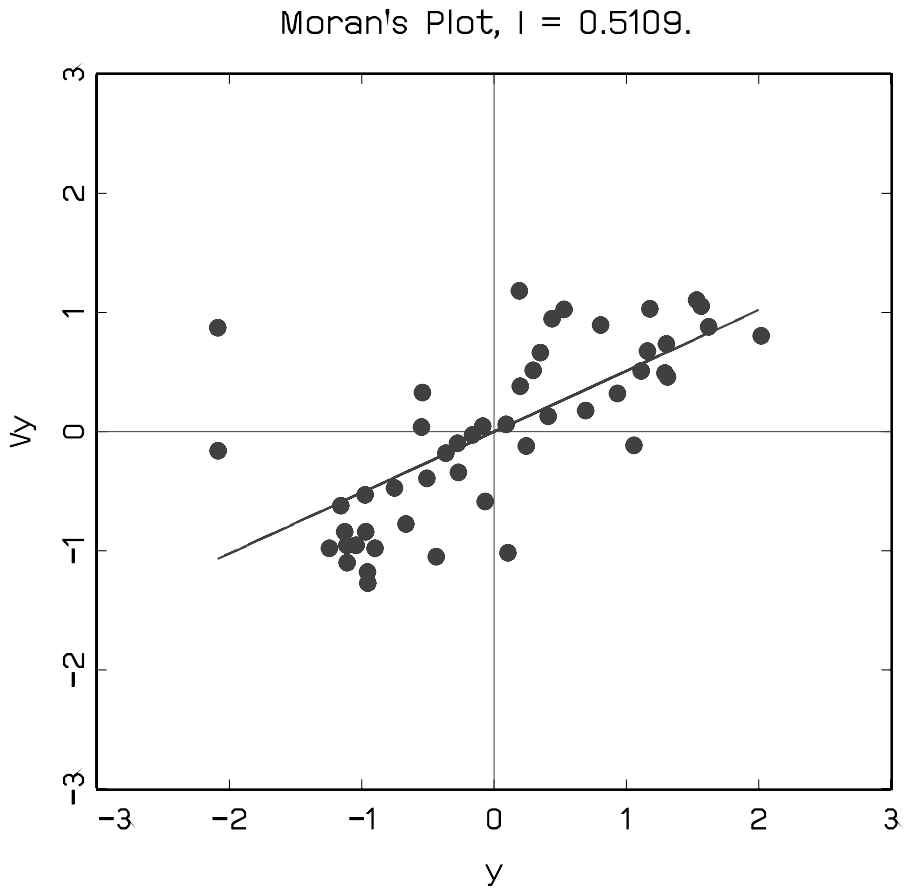


Figure 3.1: Moran scatterplot for 49 observations

Inference for Moran's $\mathcal{I}$ is commonly based on a normal approximation using standardized $z(\mathcal{I})$-values, although an exact test can be constructed. The practical problem, when running the exact test is, that numerical integration is needed.

The moments of Moran's $\mathcal{I}$ under the assumption of spatial independence can can be expressed in terms of the eigenvalues of the matrix $\mathbf{K} = \mathbf{M}\frac{1}{2}(\mathbf{V}+\mathbf{V}')\mathbf{M}$ (Tiefelsdorf, 2000), with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ denoting the general projection matrix. The eigenvalue expression is useful, if the exact distribution

of $\mathcal{I}$ under the null is required. The distribution function of $\mathcal{I}$ is given in Tiefelsdorf (2000). If only the moments are of interest, the evaluation of the eigenvalues can be bypassed by making use of the trace operator tr(.). The expected value of $\mathcal{I}$ under the null is given by

$$\mathrm{E}[\mathcal{I} \mid H_0] = \frac{\mathrm{tr}(\mathbf{K})}{n-k} = \frac{\mathrm{tr}\{\mathbf{M}\frac{1}{2}(\mathbf{V}+\mathbf{V}')\mathbf{M}\}}{n-k} = \frac{\mathrm{tr}(\mathbf{MV})}{n-k}. \tag{3.2}$$

The expected value is independent of the design points, i.e. the location of the observations is not relevant. If the spatial dependence of a single variable is of interest, the regression on only an intercept is used and the corresponding residuals are taken to calculate $\mathcal{I}$. In this case $k = 1$ and $\mathbf{M} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$. Under the assumption that $\mathbf{V}$ is standardized and $\sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij} = n$ holds, the expected value of $\mathcal{I}$ under the null reduces to $\mathrm{E}[\mathcal{I}|H_0] = -\frac{1}{n-1}$, see Cliff & Ord (1981).
The variance of $\mathcal{I}$ under the null hypothesis is given by

$$\begin{aligned}\mathrm{Var}[\mathcal{I} \mid H_0] &= \frac{\mathrm{tr}(\mathbf{MVMV}') + \mathrm{tr}(\mathbf{MV})^2 + \{\mathrm{tr}(\mathbf{MV})\}^2}{(n-k)(n-k+2)} - \{\mathrm{E}[\mathcal{I} \mid H_0]\}^2 \\ &= \frac{2\{(n-k)\mathrm{tr}(\mathbf{K}^2) - \mathrm{tr}(\mathbf{K})^2\}}{(n-k)^2(n-k+2)},\end{aligned} \tag{3.3}$$

see Henshaw (1966). For $k = 1$ and $\mathrm{rank}(\mathbf{V}) = n$, this variance can be given in terms of the eigenvalues γ_i of matrix $\mathbf{K}$: $\mathrm{Var}[\mathcal{I}|H_0] = \frac{2n}{n^2-1}\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2 = \frac{2n}{n^2-1}\sigma^2_\gamma$, given in Cliff & Ord (1981).

An application of the theoretical moments of Moran's $\mathcal{I}$ is the approximation of the exact distribution of Moran's $\mathcal{I}$ by well-known simple distributions, that allow fast assessment of the significance of an observed Moran's $\mathcal{I}$ without numerical evaluation of its exact probability. If the skewness and the kurtosis of Moran's $\mathcal{I}$ (see Tiefelsdorf, 2000) do not differ substantially from their counterparts of the normal distribution, the z-transformation of Moran's $\mathcal{I}$ can be used to obtain the significance of an observed Moran's $\mathcal{I}$. However, if there is a marked difference between the skewness and the kurtosis of Moran's $\mathcal{I}$ to that of the normal distribution, alternative approximation strategies, such as a saddlepoint approximation need to be employed (Tiefelsdorf, 2002).

The z-transformed Moran's $\mathcal{I}$ is for normally distributed regression residuals and well-behaved spatial link matrices under the assumption of spatial independence asymptotically normally distributed with expected value 0 and variance 1, and z($\mathcal{I}$) is defined as

$$z(\mathcal{I}) = \frac{\mathcal{I} - \mathrm{E}[\mathcal{I}]}{\sqrt{\mathrm{Var}[\mathcal{I}]}}, \tag{3.4}$$

see e.g. Tiefelsdorf (2000). Values of $z(\mathcal{I})$ are compared with the well known critical values of the $N(0,1)$ distribution and whenever $z(\mathcal{I})$ lies in the critical region, a hypothetical spatial Gaussian process (either AR or MA) can be assumed.

An analysis of the effect of objects, which are that far apart of all others that they lead to spatial weight matrices which do not have full rank, on the Moran's $\mathcal{I}$ test can be found in the next chapter.

3.1.3 Local Moran's $\mathcal{I}$

Moran's $\mathcal{I}$, given in (3.1), is a measure of global spatial autocorrelation. In the case of different spatial structures inherent in the dataset, i.e. if some regions have a positive spatial autocorrelation and some others have a negative spatial autocorrelation, they can compensate each other, and the global Moran's $\mathcal{I}$ might indicates spatial independence. Such local effects can be detected and tested via the local Moran's $\mathcal{I}$, which can be calculated for each region in the dataset. The local Moran's $\mathcal{I}$ is defined as a ratio of quadratic forms,

$$\mathcal{I}_i = \frac{\hat{\boldsymbol{\varepsilon}}' \mathbf{V}_i \hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}, \tag{3.5}$$

where $\mathbf{V}_i$ is a local spatial link matrix, see e.g. formula (2.2). The general distributional properties of global Moran's $\mathcal{I}$ also apply to local Moran's $\mathcal{I}_i$, but the specific distributional properties are different, local Moran's $\mathcal{I}$ is not asymptotically normally distributed, see Tiefelsdorf (2002). The average of all n local Moran's $\mathcal{I}_i$'s gives the global $\mathcal{I}$, i.e. $\mathcal{I} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}_i$.

3.2 Getis Statistic

Another measurement of the intensity of spatial connectivity is the so-called Getis statistic, $G_i(\delta)$, Getis & Ord (1992). It was originally developed as a diagnostic to reveal local spatial dependencies that are not properly captured by global measures as the Moran's $\mathcal{I}$ statistic. It is a distance-weighted and normalized average of observations $(x_1, ..., x_n)$ from a relevant variable x:

$$G_i(\delta) = \frac{\sum_j v_{ij}(\delta) x_j}{\sum_j x_j}, \; i \neq j, \tag{3.6}$$

where $v_{ij}(\delta)$ are the elements of a row-standardized spatial link matrix, δ is a locality parameter of the regional weighting scheme (typically δ is a distance parameter and observations which are further apart are down-weighted). Like the Moran's $\mathcal{I}$, the $G_i(\delta)$ statistic can be standardized to z_{Gi} which is approximately Normal (0,1) distributed and can therefore be directly compared with the well-known critical values. The expected value of $G_i(\delta)$ represents the realization at location i when no spatial autocorrelation occurs.

$$\mathrm{E}[G_i(\delta)] = \frac{\sum_{i \neq j} v_{ij}(\delta)}{(n-1)}. \tag{3.7}$$

The variance of $G_i(\delta)$ is given by:

$$\mathrm{Var}[G_i(\delta)] = \frac{\sum_{i \neq j} v_{ij}(n - 1 - \sum_{i \neq j} v_{ij})}{(n-1)^2(n-2)} \left(\frac{Y_{i2}}{Y_{i1}} \right),$$

where $Y_{i1} = \frac{\sum_j x_j}{n-1}$ and $Y_{i2} = \frac{\sum_j x_j^2}{n-1} - Y_{i1}^2$.

Chapter 4

Treatment of Far-Off Objects in Moran's $\mathcal{I}$ Test

4.1 Motivation

Observations that are far apart from all other objects in the sense that they have no spatial links to other design points will be called **far-off objects**. The treatment of far-off observations in spatial analysis is not really worked out in literature although it is an interesting question, because even when such objects are not spatially linked to others, they have influence on the spatial analysis. This can be seen e.g. when looking at the expected value of Moran's $\mathcal{I}$ in an intercept only model, here $\mathrm{E}[\mathcal{I}|H_0]$ should be equal to $-\frac{1}{n-1}$ but this relation holds only if $\mathbf{V}$ has full rank, i.e. all observations are derived from different locations, and all of them are somehow related to at least one other observation, and therefore no objects are completely separated from all the others. The problem of far-off objects can easily occur if neighbourhood-based spatial link matrices are used, and it might occur if the connectivity is based on distance matrices and a sill exists, i.e. from a certain distance onwards the connectivity is assumed to be negligible and therefore set to zero. The relevance of the topic of far-off objects can be seen in chapter 7.

One of the first ideas that comes into thought, of how to treat such observa-

[1]Chapter 4 is published as a working paper in the Research Report Series of the University of Economics and Business Administration, see Gumprecht (2007b).

tions, might be to simply ignore them insofar to run the standard procedures and do not care about the zero lines in $\mathbf{V}$ and accept that $\mathrm{E}[\mathcal{I}|H_0] \neq -\frac{1}{n-1}$. Another idea is to exclude them from the dataset and use only an $(n-r)$ points design, because no spatial connections to other objects should have no influence on the spatial autocorrelation of all other observations. But there is a difference whether far-off objects are excluded or not. Thus, the treatment of far-off observations has influence on measuring and testing spatial autocorrelation. Another possibility to avoid zero-lines and zero-columns in $\mathbf{V}$ is to add a very small value ν to all elements (except of the ones on the main diagonal) of the unstandardized spatial weight matrix $\mathbf{U}$, then no element is completely separated, and if ν is small enough it should has no influence on the general structure, but it prevents getting zero-lines. So, there are three kinds of specifications:

(s) Include the separated observations and work with a spatial weight matrix which does not have full rank.

(e) Exclude the separated observations, i.e. work with $(n-r)$ observations.

(ν) Include all n observations in the analysis, and use a modified unstandardized spatial weight matrix with elements $u_{ij} + \nu$ for all $i \neq j$ to avoid zero-lines and zero-columns.

Whichever specification is used, it influences $\mathcal{I}$, $\mathrm{E}(\mathcal{I})$ and $\mathrm{Var}(\mathcal{I})$ and therefore $z(\mathcal{I})$ and potentially the decision, whether to reject the null hypothesis or not.

4.2 Definitions

For reasons of simplicity it is assumed that only one object (the first one) is completely separated from all others. For the three different treatments of the far-off observation the spatial weight matrices are slightly different.

For treatment (s) the unstandardized spatial link matrix is a symmetric n by n matrix $\mathbf{U}^{(s)}$, the row-standardized weight matrix $\mathbf{V}^{(s)}$ is nonsymmetric:

$$\mathbf{U}^{(s)} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & u_{23} & u_{24} & \cdots & u_{2n} \\ 0 & u_{32} & 0 & u_{34} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & u_{n2} & u_{n3} & u_{n4} & \cdots & 0 \end{pmatrix},$$

$$\mathbf{V}^{(s)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{u_{23}}{\sum_{j=2}^{n} u_{2j}} & \cdots & \frac{u_{2n}}{\sum_{j=2}^{n} u_{2j}} \\ 0 & \frac{u_{32}}{\sum_{j=2}^{n} u_{3j}} & 0 & \cdots & \frac{u_{3n}}{\sum_{j=2}^{n} u_{3j}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{u_{n2}}{\sum_{j=2}^{n} u_{nj}} & \frac{u_{n3}}{\sum_{j=2}^{n} u_{nj}} & \cdots & 0 \end{pmatrix}.$$

If the first observation is excluded, i.e. treatment (e) is used, the unstandardized spatial weight matrix is a symmetric $(n-1)$ by $(n-1)$ matrix $\mathbf{U}^{(e)}$:

$$\mathbf{U}^{(e)} = \begin{pmatrix} 0 & u_{23} & u_{24} & \cdots & u_{2n} \\ u_{32} & 0 & u_{34} & \cdots & u_{3n} \\ u_{42} & u_{43} & 0 & \cdots & u_{4n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{n2} & u_{n3} & u_{n4} & \cdots & 0 \end{pmatrix},$$

$$\mathbf{V}^{(e)} = \begin{pmatrix} 0 & \frac{u_{23}}{\sum_{j=2}^{n} u_{2j}} & \cdots & \frac{u_{2n}}{\sum_{j=2}^{n} u_{2j}} \\ \frac{u_{32}}{\sum_{j=2}^{n} u_{3j}} & 0 & \cdots & \frac{u_{3n}}{\sum_{j=2}^{n} u_{3j}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{u_{n2}}{\sum_{j=2}^{n} u_{nj}} & \frac{u_{n3}}{\sum_{j=2}^{n} u_{nj}} & \cdots & 0 \end{pmatrix}.$$

For treatment (ν), i.e. a small value ν is added, $\mathbf{U}^{(\nu)}$ is again a symmetric n by n matrix:

$$\mathbf{U}^{(\nu)} = \begin{pmatrix} 0 & 0+\nu & 0+\nu & \cdots & 0+\nu \\ 0+\nu & 0 & u_{23}+\nu & \cdots & u_{2n}+\nu \\ 0+\nu & u_{32}+\nu & 0 & \cdots & u_{3n}+\nu \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0+\nu & u_{n2}+\nu & u_{n3}+\nu & \cdots & 0 \end{pmatrix},$$

the corresponding row-standardized weight matrix $\mathbf{V}^{(\nu)*}$ goes to $\mathbf{V}^{(\nu)}$ if $\nu \to 0$

$$\mathbf{V}^{(\nu)*} = \begin{pmatrix} 0 & \frac{\nu}{(n-1)\nu} & \cdots & \frac{\nu}{(n-1)\nu} \\ \frac{\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{2j}} & 0 & \cdots & \frac{u_{2n}+\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{2j}} \\ \frac{\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{2j}} & \frac{u_{32}+\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{3j}} & \cdots & \frac{u_{3n}+\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{3j}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{2j}} & \frac{u_{n2}+\nu}{(n-1)\nu+\sum_{j=2}^{n} u_{nj}} & \cdots & 0 \end{pmatrix},$$

$$\mathbf{V}^{(\nu)} = \begin{pmatrix} 0 & \frac{1}{(n-1)} & \frac{1}{(n-1)} & \cdots & \frac{1}{(n-1)} \\ 0 & 0 & \frac{u_{23}}{\sum_{j=2}^{n} u_{2j}} & \cdots & \frac{u_{2n}}{\sum_{j=2}^{n} u_{2j}} \\ 0 & \frac{u_{32}}{\sum_{j=2}^{n} u_{3j}} & 0 & \cdots & \frac{u_{3n}}{\sum_{j=2}^{n} u_{3j}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{u_{n2}}{\sum_{j=2}^{n} u_{nj}} & \frac{u_{n3}}{\sum_{j=2}^{n} u_{nj}} & \cdots & 0 \end{pmatrix}.$$

In the calculation of Moran's $\mathcal{I}$, $\mathrm{E}[\mathcal{I}|H_0]$ and $\mathrm{Var}[\mathcal{I}|H_0]$ the following terms are used: a symmetric spatial weight matrix $\mathbf{G} = \frac{1}{2}(\mathbf{V} + \mathbf{V}')$, the projection matrix for an intercept only model $\mathbf{M} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$ and $\mathbf{K} = \mathbf{M}'\mathbf{G}\mathbf{M}$. For the three different specification (s), (e) and (ν), these matrices have different forms, and all of them can be written in the structure of block matrices. Specification (s) gives:

$$\mathbf{G}^{(s)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & g_{23} & \cdots & g_{2n} \\ 0 & g_{32} & 0 & \cdots & g_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & g_{n2} & g_{n3} & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{1\times 1} & : & \mathbf{B}_{1\times(n-1)} \\ .. & : & .. \\ \mathbf{B}'_{(n-1)\times 1} & : & \mathbf{C}_{(n-1)\times(n-1)} \end{pmatrix},$$

$$\mathbf{M}^{(s)}_{n\times n} = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix} = \begin{pmatrix} \mathbf{M}^{(s1)}_{1\times 1} & : & \mathbf{M}^{(s2)}_{1\times(n-1)} \\ .. & : & .. \\ \mathbf{M}^{(s2)\prime}_{(n-1)\times 1} & : & \mathbf{M}^{(s3)}_{(n-1)\times(n-1)} \end{pmatrix}.$$

In case (e) the dimension reduces to $(n-1)$

$$\mathbf{G}^{(e)}_{(n-1)\times(n-1)} = \begin{pmatrix} 0 & g_{23} & \cdots & g_{2n} \\ g_{32} & 0 & \cdots & g_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n2} & g_{n3} & \cdots & 0 \end{pmatrix} = \mathbf{C}_{(n-1)\times(n-1)},$$

$$\mathbf{M}^{(e)}_{(n-1)\times(n-1)} = \begin{pmatrix} 1-\frac{1}{(n-1)} & -\frac{1}{(n-1)} & \cdots & -\frac{1}{(n-1)} \\ -\frac{1}{(n-1)} & 1-\frac{1}{(n-1)} & \cdots & -\frac{1}{(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{(n-1)} & -\frac{1}{(n-1)} & \cdots & 1-\frac{1}{(n-1)} \end{pmatrix}.$$

Finally for case (ν), the projection matrix $\mathbf{M}^{(\nu)}_{n\times n} = \mathbf{M}^{(s)}_{n\times n}$, and

$$\begin{aligned} \mathbf{G}^{(\nu)}_{n\times n} &= \begin{pmatrix} 0 & \frac{1}{2(n-1)} & \frac{1}{2(n-1)} & \cdots & \frac{1}{2(n-1)} \\ \frac{1}{2(n-1)} & 0 & g_{23} & \cdots & g_{2n} \\ \frac{1}{2(n-1)} & g_{32} & 0 & \cdots & g_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2(n-1)} & g_{n2} & g_{n3} & \cdots & 0 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{1\times 1} & : & \mathbf{B}^{(\nu)}_{1\times(n-1)} \\ .. & & .. \\ \mathbf{B}^{(\nu)\prime}_{(n-1)\times 1} & : & \mathbf{C}_{(n-1)\times(n-1)} \end{pmatrix}. \end{aligned}$$

These matrices are used in the next sections to quantify and compare the values of $\mathcal{I}$, $\mathrm{E}[\mathcal{I}|H_0]$, $\mathrm{Var}[\mathcal{I}|H_0]$ and $z(\mathcal{I})$ to find out if the choice of the treatment of a far-off object has influence on the Moran's test.

4.3 Moran's $\mathcal{I}$

Formulas for Moran's $\mathcal{I}$, $\mathrm{E}[\mathcal{I}|H_0]$ and $\mathrm{Var}[\mathcal{I}|H_0]$ are given in (3.1), (3.2) and (3.3). $\mathcal{I}$ contains $\mathbf{G}$ and the vector of the residuals $\hat{\boldsymbol{\varepsilon}}_{n\times 1} = [\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n]' = [\mathbf{a}' : \mathbf{b}']$, where $\mathbf{a}$ is simply the residual $\hat{\varepsilon}_1$ of the far-off object and $\mathbf{b}$ is an $(n-1)\times 1$ vector of the residuals $\hat{\varepsilon}_i$ $(i = 2, ..., n)$ of the 'well-behaved'

objects. Using the block structure helps to see the difference between the three specifications.

$$\begin{aligned}\mathcal{I} &= \frac{\hat{\boldsymbol{\varepsilon}}'\frac{1}{2}(\mathbf{V}+\mathbf{V}')\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{G}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{[a' : b']\begin{pmatrix} \mathbf{A} : \mathbf{B} \\ \ddot{\mathbf{B}}' : \ddot{\mathbf{C}} \end{pmatrix}\begin{bmatrix} a \\ \ddot{b} \end{bmatrix}}{[a' : b']\begin{bmatrix} a \\ \ddot{b} \end{bmatrix}} \\ &= \frac{\mathbf{a}'\mathbf{A}\mathbf{a} + \mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} \,. \end{aligned} \tag{4.1}$$

For the different treatments of the separated observation the blocks of the corresponding matrix $\mathbf{G}$ are inserted in (4.1). This gives for case (s), where the first element is separated and therefore $\mathbf{A} = 0$ and $\mathbf{B}$ is a vector of zeros

$$\mathcal{I}^{(s)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \tag{4.2}$$

In case (e) the separated element is excluded, $\mathcal{I}^{(e)}$ is given by

$$\mathcal{I}^{(e)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij}}{\sum_{i=2}^{n}\hat{\varepsilon}_i^2}. \tag{4.3}$$

For case (ν) $\mathbf{A} = 0$ but now $\mathbf{B} = \left[\frac{1}{2(n-1)}, \ldots, \frac{1}{2(n-1)}\right]$, therefore

$$\mathcal{I}^{(\nu)} = \frac{\mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij} - \frac{\hat{\varepsilon}_1^2}{n-1}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \tag{4.4}$$

This follows from the fact that $\mathbf{b}'\mathbf{B}'\mathbf{a} = \mathbf{a}'\mathbf{B}\mathbf{b} = \frac{1}{2(n-1)}\sum_{i=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_1$, because $\sum_{i=2}^{n}\hat{\varepsilon}_i = -\hat{\varepsilon}_1$, these terms can be written as $\frac{-\hat{\varepsilon}_1^2}{2(n-1)}$, thus $\mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} = \frac{-\hat{\varepsilon}_1^2}{(n-1)}$.

Comparing (4.2) and (4.3) shows that $\mathcal{I}^{(e)} \geq \mathcal{I}^{(s)}$ because the nominator is the same but the denominator of $\mathcal{I}^{(e)}$ is smaller or equal to the one of $\mathcal{I}^{(s)}$, equality holds only if $\hat{\varepsilon}_1 = 0$, i.e.

$$\mathcal{I}^{(e)} = \mathcal{I}^{(s)} + \frac{\mathbf{a}'\mathbf{a}\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{b}(\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b})} = \mathcal{I}^{(s)} + \frac{\hat{\varepsilon}_1^2\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij}}{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j\left(\hat{\varepsilon}_1^2 + \sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j\right)}.$$

Furthermore $\mathcal{I}^{(s)} \geq \mathcal{I}^{(\nu)}$ because here the denominators are the same but the nominator of $\mathcal{I}^{(\nu)}$ is smaller or equal than the one of $\mathcal{I}^{(s)}$, see (4.2) and (4.4). Again, equality holds only for $\hat{\varepsilon}_1 = 0$.

$$\mathcal{I}^{(s)} = \mathcal{I}^{(\nu)} + \frac{\hat{\varepsilon}_1^2}{(n-1)\sum_{i=1}^{n} \hat{\varepsilon}_i^2}.$$

Comparing all three treatments of a separated observation gives the following relationship of the Moran's $\mathcal{I}$ values:

$$\mathcal{I}^{(\nu)} \leq \mathcal{I}^{(s)} \leq \mathcal{I}^{(e)}. \tag{4.5}$$

This result is quite plausible. If the far-off observation is completely excluded it has no influence at all, the Moran's $\mathcal{I}$ measures only the connectivity between all other $(n-1)$ objects and is therefore greater than in case of including the far-off object, the observation that lies far apart is not taken into account. On the other hand if the far-off object is included but zero-weighted it effects the Moran's $\mathcal{I}$ only insofar as the corresponding residual is included in the denominator, whereas if the far-off object is weighted with $\frac{1}{2(n-1)}$ this object increases the denominator and at the same time decreases the overall connectivity (i.e. the nominator).

4.4 Expected Value of $\mathcal{I}$

The expected value of $\mathcal{I}$ under the null hypothesis is $\mathrm{E}[\mathcal{I} \mid H_0] = \frac{\mathrm{tr}(\mathbf{K})}{n-k}$, see (3.2). Using the same block-structure for $\mathbf{M}$ and $\mathbf{G}$ as before, helps to find the difference between the three different treatments of the separated object. For $\mathrm{E}[\mathcal{I}|H_0]$ only the trace of $\mathbf{K}$ is relevant, $\mathrm{tr}(\mathbf{K}) = \mathrm{tr}(\mathbf{M}'\mathbf{G}\mathbf{M}) = \mathrm{tr}(\mathbf{M}\mathbf{M}'\mathbf{G}) = \mathrm{tr}(\mathbf{M}\mathbf{G})$. Block-structure notation gives:

$$\begin{aligned}
\mathrm{tr}(\mathbf{K}) &= \mathrm{tr}\left[\begin{pmatrix} \mathbf{M}^{(1)} & : & \mathbf{M}^{(2)} \\ .. & & .. \\ \mathbf{M}^{(2)\prime} & : & \mathbf{M}^{(3)} \end{pmatrix} \begin{pmatrix} \mathbf{A} & : & \mathbf{B} \\ .. & & .. \\ \mathbf{B}' & : & \mathbf{C} \end{pmatrix}\right] \\
&= \mathrm{tr}\left[\begin{pmatrix} \mathbf{M}^{(1)}\mathbf{A} + \mathbf{M}^{(2)}\mathbf{B}' & : & \mathbf{M}^{(1)}\mathbf{B} + \mathbf{M}^{(2)}\mathbf{C} \\ .. & & .. \\ \mathbf{M}^{(2)\prime}\mathbf{A} + \mathbf{M}^{(3)}\mathbf{B}' & : & \mathbf{M}^{(2)\prime}\mathbf{B} + \mathbf{M}^{(3)}\mathbf{C} \end{pmatrix}\right].
\end{aligned}$$

Since only the main diagonal elements enter the trace, it follows that

$$\mathrm{tr}(\mathbf{K}) = \mathbf{M}^{(1)}\mathbf{A} + \mathbf{M}^{(2)}\mathbf{B}' + \mathbf{M}^{(2)\prime}\mathbf{B} + \mathbf{M}^{(3)}\mathbf{C}.$$

Now the three different treatments can be evaluated.

Case (s), the far-off object is included and zero-weighted, $\mathbf{A} = 0$ and $\mathbf{B} = [0, \ldots, 0]$

$$\mathrm{tr}(\mathbf{K}^{(s)}) = \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}) = -\frac{1}{n}\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij} = -\frac{1}{n}(n-1). \tag{4.6}$$

$\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij} = n - 1$ follows directly from the construction of $\mathbf{G}$ and $\mathbf{C}$ respectively. The expected value for the intercept model ($k = 1$) is therefore

$$\mathrm{E}[\mathcal{I}^{(s)}|H_0] = -\frac{1}{n}. \tag{4.7}$$

Case (e), the far-off observation is deleted, the corresponding matrix $\mathbf{K}^{(e)}$ has to be used

$$\mathrm{tr}(\mathbf{K}^{(e)}) = \mathrm{tr}\left[\mathbf{M}^{(e)}\mathbf{C}\right] = -\frac{1}{n-1}\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij} = -\frac{1}{n-1}(n-1) = -1. \tag{4.8}$$

Here, the number of objects taken into account is $(n-1)$ and as $k = 1$ for the intercept model, the denominator of the expected value is $(n-1)-k = n-2$. Thus

$$\mathrm{E}[\mathcal{I}^{(e)}|H_0] = -\frac{1}{n-2}. \tag{4.9}$$

Case (ν), $\mathbf{K}^{(\nu)}$ is used, now $\mathbf{B}^{(\nu)}$ is not a vector of zeros like in case (s), and therefore

$$\begin{aligned} \mathrm{tr}(\mathbf{K}^{(\nu)}) &= \mathrm{tr}(\mathbf{M}^{(s1)}\mathbf{A}) + \mathrm{tr}(\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}) + \mathrm{tr}(\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}) + \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}) \\ &= 0 - \frac{1}{2n} - \frac{1}{2n} - \frac{n-1}{n} = -1. \end{aligned} \tag{4.10}$$

This leads to

$$\mathrm{E}[\mathcal{I}^{(\nu)}|H_0] = -\frac{1}{n-1}. \tag{4.11}$$

Case (ν) is the only one which leads for an intercept only model to the expected value given in Cliff & Ord (1981), see section 3.1.2.

Comparing the expected values of $\mathcal{I}^{(s)}$, $\mathcal{I}^{(e)}$ and $\mathcal{I}^{(\nu)}$ under the null, see (4.7), (4.9) and (4.11), gives the following order:

$$\mathrm{E}[\mathcal{I}^{(e)}] < \mathrm{E}[\mathcal{I}^{(\nu)}] < \mathrm{E}[\mathcal{I}^{(s)}]. \tag{4.12}$$

The absolute values of the expected values have the reverse order, that is: $|\mathrm{E}[\mathcal{I}^{(s)}]| < |\mathrm{E}[\mathcal{I}^{(\nu)})] < |\mathrm{E}[\mathcal{I}^{(e)}]|$. The expected value of $\mathcal{I}$ for an intercept model does neither depend on the locations nor on the attribute values of the design points, it just depends on the number of the objects and the kind of treatment of far-off objects.

4.5 Variance of $\mathcal{I}$

The variance of $\mathcal{I}$ under the null hypothesis is $\mathrm{Var}[\mathcal{I} \mid H_0] = \frac{2\{(n-k)\mathrm{tr}(\mathbf{K}^2)-\mathrm{tr}(\mathbf{K})^2\}}{(n-k)^2(n-k+2)}$, already given in (3.3), with $\mathrm{tr}(\mathbf{K}^2) = \mathrm{tr}(\mathbf{KK}) = \mathrm{tr}(\mathbf{MGMG})$. In block-notation this trace is:

$$\begin{aligned}\mathrm{tr}(\mathbf{KK}) &= \mathrm{tr}(\mathbf{MGMG}) \\ &= \mathrm{tr}\left[\begin{pmatrix} \mathbf{M}^{(1)} : \mathbf{M}^{(2)} \\ \ddot{\mathbf{M}}^{(2)\prime} : \ddot{\mathbf{M}}^{(3)} \end{pmatrix} \begin{pmatrix} \mathbf{A} : \mathbf{B} \\ \ddot{\mathbf{B}}' : \ddot{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \mathbf{M}^{(1)} : \mathbf{M}^{(2)} \\ \ddot{\mathbf{M}}^{(2)\prime} : \ddot{\mathbf{M}}^{(3)} \end{pmatrix} \begin{pmatrix} \mathbf{A} : \mathbf{B} \\ \ddot{\mathbf{B}}' : \ddot{\mathbf{C}} \end{pmatrix}\right].\end{aligned}$$

Matrix multiplication gives a matrix which can be written in the form of 4 blocks, for the trace only the elements of the main diagonal are crucial, they are denoted $\mathbf{D}_1$ which is an 1×1 matrix and $\mathbf{D}_2$ which is an $(n-1) \times (n-1)$ matrix. $\mathrm{tr}(\mathbf{KK}) = \mathrm{tr}(\mathbf{D}_1) + \mathrm{tr}(\mathbf{D}_2)$. The matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are composed of 8 other matrices $\mathbf{D}_1^{(1)}, \ldots, \mathbf{D}_1^{(8)}$ and $\mathbf{D}_2^{(1)}, \ldots, \mathbf{D}_2^{(8)}$ respectively, and $\mathrm{tr}(\mathbf{KK}) = \sum_{i=1}^{2}\sum_{j=1}^{8} \mathrm{tr}(\mathbf{D}_i^{(j)})$, as the trace of a sum of matrices is equal to the sum of the traces. The 16 different traces of matrix-blocks are:

$$\begin{array}{ll}
\mathrm{tr}(\mathbf{D}_1^{(1)}) = \mathrm{tr}(\mathbf{M}^{(1)}\mathbf{A}\mathbf{M}^{(1)}\mathbf{A}) & \mathrm{tr}(\mathbf{D}_2^{(1)}) = \mathrm{tr}(\mathbf{M}^{(2)\prime}\mathbf{A}\mathbf{M}^{(1)}\mathbf{B}) \\
\mathrm{tr}(\mathbf{D}_1^{(2)}) = \mathrm{tr}(\mathbf{M}^{(2)}\mathbf{B}'\mathbf{M}^{(1)}\mathbf{A}) & \mathrm{tr}(\mathbf{D}_2^{(2)}) = \mathrm{tr}(\mathbf{M}^{(3)}\mathbf{B}'\mathbf{M}^{(1)}\mathbf{B}) \\
\mathrm{tr}(\mathbf{D}_1^{(3)}) = \mathrm{tr}(\mathbf{M}^{(1)}\mathbf{B}\mathbf{M}^{(2)\prime}\mathbf{A}) & \mathrm{tr}(\mathbf{D}_2^{(3)}) = \mathrm{tr}(\mathbf{M}^{(2)\prime}\mathbf{B}\mathbf{M}^{(2)\prime}\mathbf{B}) \\
\mathrm{tr}(\mathbf{D}_1^{(4)}) = \mathrm{tr}(\mathbf{M}^{(2)}\mathbf{C}\mathbf{M}^{(2)\prime}\mathbf{A}) & \mathrm{tr}(\mathbf{D}_2^{(4)}) = \mathrm{tr}(\mathbf{M}^{(3)}\mathbf{C}\mathbf{M}^{(2)\prime}\mathbf{B}) \\
\mathrm{tr}(\mathbf{D}_1^{(5)}) = \mathrm{tr}(\mathbf{M}^{(1)}\mathbf{A}\mathbf{M}^{(2)}\mathbf{B}\prime) & \mathrm{tr}(\mathbf{D}_2^{(5)}) = \mathrm{tr}(\mathbf{M}^{(2)\prime}\mathbf{A}\mathbf{M}^{(2)}\mathbf{C}) \\
\mathrm{tr}(\mathbf{D}_1^{(6)}) = \mathrm{tr}(\mathbf{M}^{(2)}\mathbf{B}'\mathbf{M}^{(2)}\mathbf{B}') & \mathrm{tr}(\mathbf{D}_2^{(6)}) = \mathrm{tr}(\mathbf{M}^{(3)}\mathbf{B}'\mathbf{M}^{(2)}\mathbf{C}) \\
\mathrm{tr}(\mathbf{D}_1^{(7)}) = \mathrm{tr}(\mathbf{M}^{(1)}\mathbf{B}\mathbf{M}^{(3)}\mathbf{B}') & \mathrm{tr}(\mathbf{D}_2^{(7)}) = \mathrm{tr}(\mathbf{M}^{(2)\prime}\mathbf{B}\mathbf{M}^{(3)}\mathbf{C}) \\
\mathrm{tr}(\mathbf{D}_1^{(8)}) = \mathrm{tr}(\mathbf{M}^{(2)}\mathbf{C}\mathbf{M}^{(3)}\mathbf{B}') & \mathrm{tr}(\mathbf{D}_2^{(8)}) = \mathrm{tr}(\mathbf{M}^{(3)}\mathbf{C}\mathbf{M}^{(3)}\mathbf{C})
\end{array}$$

For the different treatments of the separated object the corresponding matrices $\mathbf{M}$ and $\mathbf{G}$ are used. Since $\mathbf{A}$ is always 0, the traces of all matrices that include $\mathbf{A}$ are zero, and traces which include matrix $\mathbf{B}$ are zero in case (s) and (e). The last term $\mathrm{tr}(\mathbf{D}_2^{(8)})$ is relevant for all three cases, it can be further simplified:

$$\begin{aligned}
\mathrm{tr}(\mathbf{D}_2^{(8)}) &= \mathrm{tr}(\mathbf{M}^{(3)}\mathbf{C}\mathbf{M}^{(3)}\mathbf{C}) \\
&= \mathrm{tr}([\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}']\mathbf{C}[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}']\mathbf{C}) \\
&= \mathrm{tr}(\mathbf{C}\mathbf{C} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{C} - \frac{1}{n}\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C} + \frac{1}{n^2}\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C}) \\
&= \mathrm{tr}(\mathbf{C}\mathbf{C}) - \frac{2}{n}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{C}) + \frac{1}{n^2}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C}).
\end{aligned}$$

From construction of $\mathbf{G}$ and $\mathbf{C}$ respectively, $\mathbf{G} = \mathbf{G}'$ and $\mathbf{C} = \mathbf{C}'$ follows.
$\mathrm{tr}(\mathbf{C}\mathbf{C}) = \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2$
$\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{C}) = \mathrm{tr}(\mathbf{1}'\mathbf{C}'\mathbf{C}\mathbf{1}) = \sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2$
$\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C}) = \mathrm{tr}(\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}) = \left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)$
$= (n-1)^2$.
Thus,

$$\mathrm{tr}(\mathbf{M}^{(3)}\mathbf{C}\mathbf{M}^{(3)}\mathbf{C}) = \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{1}{n^2}(n-1)^2. \quad (4.13)$$

Note that for case (e), where only the $(n-1)$ objects which are spatially connected are included, (4.13) is slightly different because the number of the design points is $(n-1)$ and $\mathbf{M}^{(e)}$ is used instead of $\mathbf{M}^{(3)}$, see section 4.2. Therefore $\mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) = \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{(n-1)}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{1}{(n-1)^2}(n-1)^2$. In general, the following relationship holds:

$$\mathrm{tr}(\mathbf{C}\mathbf{C}) \leq \mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{C}) \leq \mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C}),$$

equality holds only if all elements of $\mathbf{C}$ except for one are zero, this case is not relevant here, it would mean that all objects but two are far apart from each other.

For case (s) $\mathbf{B} = [0, \ldots, 0]$, therefore only $\text{tr}(\mathbf{D}_2^{(8)}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ plays a role. This leads to a variance

$$\text{Var}[\mathcal{I}^{(s)} \mid H_0] = \frac{2\{(n-k)\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) - \text{tr}(\mathbf{K}^{(s)})^2\}}{(n-k)^2(n-k+2)} \tag{4.14}$$

where $\text{tr}(\mathbf{K}^{(s)})^2 = \left(-\frac{(n-1)}{n}\right)^2$, $\text{tr}(\mathbf{K}^{(s)})$ is given in (4.6), and $\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ is given in (4.13).

For case (ν) $\mathbf{G}^{(\nu)}$ is used, $\mathbf{A} = [0]$, $\mathbf{B}^{(\nu)} = [\frac{1}{2(n-1)}, \ldots, \frac{1}{2(n-1)}]$, and considerably more terms are unequal zero and therefore relevant for the variance:
$\text{tr}(\mathbf{D}_1^{(6)}) = \text{tr}(\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}) = \frac{1}{4n^2}$,
$\text{tr}(\mathbf{D}_1^{(7)}) = \text{tr}(\mathbf{M}^{(s1)}\mathbf{B}^{(\nu)}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}) = \frac{1}{4n^2}$,
$\text{tr}(\mathbf{D}_1^{(8)}) = \text{tr}(\mathbf{M}^{(s2)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}) = -\frac{1}{2n^2}$,
thus, $\text{tr}(\mathbf{D}_1) = 0$.
$\text{tr}(\mathbf{D}_2^{(2)}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(s1)}\mathbf{B}^{(\nu)}) = \text{tr}(\mathbf{M}^{(s1)}\mathbf{B}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}) = \text{tr}(\mathbf{D}_1^{7}) = \frac{1}{4n^2}$,
$\text{tr}(\mathbf{D}_2^{(3)}) = \text{tr}(\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}) = \text{tr}(\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}) = \text{tr}(\mathbf{D}_1^{(6)}) = \frac{1}{4n^2}$,
$\text{tr}(\mathbf{D}_2^{(4)}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}) = -\frac{1}{2n^2}$,
$\text{tr}(\mathbf{D}_2^{(6)}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(s2)}\mathbf{C}) = \text{tr}(\mathbf{M}^{(s2)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}) = \text{tr}(\mathbf{D}_1^{(8)}) = -\frac{1}{2n^2}$,
$\text{tr}(\mathbf{D}_2^{(7)}) = \text{tr}(\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}\mathbf{M}^{(s3)}\mathbf{C}) = \text{tr}(\mathbf{D}_2^{(4)}) = -\frac{1}{2n^2}$,
$\text{tr}(\mathbf{D}_2^{(8)}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$,
thus, $\text{tr}(\mathbf{K}^2) = \sum_{j=1}^{n} = \text{tr}(\mathbf{D}_2^{(j)}) = -\frac{1}{n^2} + \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$. The variance is given by

$$\text{Var}[\mathcal{I}^{(\nu)} \mid H_0] = \frac{2\{(n-k)\left[-\frac{1}{n^2} + \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})\right] - \text{tr}(\mathbf{K}^{(\nu)})^2\}}{(n-k)^2(n-k+2)} \tag{4.15}$$

with $\text{tr}(\mathbf{K}^{(\nu)})^2 = (-1)^2$, see (4.10) and $\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ is again the one given in (4.13).

In case (e) where the separated observation is completely excluded from the dataset, $\text{tr}(\mathbf{K}^{(e)}\mathbf{K}^{(e)}) = \text{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C})$, and

$$\text{Var}[\mathcal{I}^{(e)} \mid H_0] = \frac{2\{(n-1-k)\text{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) - \text{tr}(\mathbf{K}^{(e)})^2\}}{(n-1-k)^2(n-1-k+2)} \tag{4.16}$$

with $\text{tr}(\mathbf{K}^{(e)})^2 = (-1)^2$, see (4.8).
$\text{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) = \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n-1}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{1}{(n-1)^2}(n-1)^2$,
the number of design points is $(n-1)$ instead of n.

Finding the relationship between the variances is not that simple as in case of the $\mathcal{I}$s and their expected values. The ordering of $\text{Var}[\mathcal{I}^{(\nu)}]$ and $\text{Var}[\mathcal{I}^{(s)}]$ is quite obvious, $\text{Var}[\mathcal{I}^{(\nu)}] < \text{Var}[\mathcal{I}^{(s)}]$, which follows directly from (4.14) and (4.15), the denominators are the same, and the nominator of $\text{Var}[\mathcal{I}^{(\nu)}]$ is smaller. The relationships between $\text{Var}[\mathcal{I}^{(e)}]$ and $\text{Var}[\mathcal{I}^{(s)}]$ is more complex, nominators as well as denominators are different, see (4.14) and (4.16). To find the difference $\text{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C})$ is expressed in terms of $\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$:
$\text{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) = \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2-1}{(n-1)n^2}$.
Hence, $\text{Var}[\mathcal{I}^{(e)}|H_0]$ is equal to:

$$\frac{2\left\{(n-1-k)\left[\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2-1}{(n-1)n^2}\right] - \text{tr}(\mathbf{K}^{(e)})^2\right\}}{(n-1-k)^2(n-1-k+2)}.$$

It can also be written in form of $\text{Var}[\mathcal{I}^{(s)}|H_0]$ plus additional terms:

$$\begin{aligned}\text{Var}[\mathcal{I}^{(e)}|H_0] &= \text{Var}[\mathcal{I}^{(s)}|H_0] \\ &+ \frac{2\left\{(n-1-k)\left[\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n(\text{tr}(\mathbf{1'CC1}))+1}{(n-1)n^2}\right] - 1\right\}}{(n-1-k)^2(n-k)} \\ &- \frac{2(n-k)\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) - \frac{(n-1)^2}{n^2}}{(n-k)^2(n-k+2)}.\end{aligned}$$

The magnitude of the difference depends on the concrete values of the spatial link matrix via $\text{tr}(\mathbf{MCMC})$ and $\text{tr}(\mathbf{11'CC})$ which appear in the formula. To show the difference, two extreme cases are examined:

(1) All objects (except for the first, which is far-off) are neighbours of all others, this can happen if e.g. a critical distance d_c is defined and within this distance each object is regarded as neighbour, if d_c is large enough and the observations are near each other, each object is a neighbour of every other object - of course except for

the far-off one; unstandardized- and row-standardized spatial link matrices are given by:

$$\mathbf{U}^{(1)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \cdots & 0 \end{pmatrix}, \mathbf{V}^{(1)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{n-2} & \cdots & \frac{1}{n-2} \\ 0 & \frac{1}{n-2} & 0 & \cdots & \frac{1}{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{n-2} & \frac{1}{n-2} & \cdots & 0 \end{pmatrix},$$

and matrix $\mathbf{G} = \frac{1}{2}(\mathbf{V} + \mathbf{V}')$ is:

$$\mathbf{G}^{(1)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{n-2} & \cdots & \frac{1}{n-2} \\ 0 & \frac{1}{n-2} & 0 & \cdots & \frac{1}{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{n-2} & \frac{1}{n-2} & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{1\times 1} & : & \mathbf{B}_{1\times(n-1)} \\ .. & & .. \\ \mathbf{B}'_{(n-1)\times 1} & : & \mathbf{C}^{(1)}_{(n-1)\times(n-1)} \end{pmatrix}.$$

(2) Each object has only one neighbour (except of the first far-off one), i.e. there are only separated pairs of neighbourships (further assumption needed: $n-1$ is even). The corresponding spatial link matrices are:

$$\begin{aligned} \mathbf{U}^{(2)}_{n\times n} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} = \mathbf{V}^{(2)}_{n\times n} = \mathbf{G}^{(2)}_{n\times n} \\ &= \begin{pmatrix} \mathbf{A}_{1\times 1} & : & \mathbf{B}_{1\times(n-1)} \\ .. & & .. \\ \mathbf{B}'_{(n-1)\times 1} & : & \mathbf{C}^{(2)}_{(n-1)\times(n-1)} \end{pmatrix}. \end{aligned}$$

These matrices correspond to case (s) where the far-off object is included in the analysis. If the far-off observation is excluded, case (e), the corresponding matrices are slightly different. The dimension reduces to $(n-1)\times(n-1)$ by

simply deleting the first line and the first column. For these extreme cases the variances of $\mathcal{I}^{(s)}$ and $\mathcal{I}^{(e)}$ can be explicitly specified and in the following the standardized Moran's $\mathcal{I}$ values can be compared. For the variances (3.3), of case (s) and (e), the corresponding matrices $\mathbf{K}^{(s)} = \mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}$ and $\mathbf{K}^{(e)} = \mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}$ are needed, the projection matrix $\mathbf{M}$ depends on the treatment of the far-off object (separation or exclusion), spatial weight matrix $\mathbf{C}$ depends on the extreme case (all objects are neighbours, or only pairs of neighbours).

Under treatment (s) and extreme case (1)

$$\begin{aligned}
\mathrm{tr}(\mathbf{K}^{(s1)}\mathbf{K}^{(s1)}) &= \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}^{(1)}\mathbf{M}^{(s3)}\mathbf{C}^{(1)}) \\
&= \mathrm{tr}(\mathbf{C}^{(1)}\mathbf{C}^{(1)}) - \frac{2}{n}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}\mathbf{C}^{(1)}) + \frac{2}{n^2}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}) \\
&= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{n^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\
&= \frac{(n-1)}{(n-2)} - \frac{2}{n}(n-1) + \frac{2}{n^2}(n-1)^2
\end{aligned}$$

for $\mathrm{tr}(\mathbf{K}^{(s1)})^2$, see (4.6), it does not depend on the concrete values of the spatial link matrix. The variance of $\mathcal{I}^{(s1)}$ under the null can be written as:

$$\begin{aligned}
\mathrm{Var}[\mathcal{I}^{(s1)}|H_0] &= \frac{2\left\{(n-k)\left[\frac{n-1}{n-2} - \frac{2}{n}(n-1) + \frac{2}{n^2}(n-1)^2\right] - \frac{(n-1)^2}{n^2}\right\}}{(n-k)^2(n-k+2)} \\
&= \frac{2n^2-3n+6}{n^2(n^2-n-2)}. \qquad (4.17)
\end{aligned}$$

By contrast, excluding the far-off object needs

$$\begin{aligned}
&\mathrm{tr}(\mathbf{K}^{(e1)}\mathbf{K}^{(e1)}) = \mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}^{(1)}\mathbf{M}^{(e)}\mathbf{C}^{(1)}) \\
&= \mathrm{tr}(\mathbf{C}^{(1)}\mathbf{C}^{(1)}) - \frac{2}{n-1}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}\mathbf{C}^{(1)}) + \frac{2}{(n-1)^2}\mathrm{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}\mathbf{1}\mathbf{1}'\mathbf{C}^{(1)}) \\
&= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n-1}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{(n-1)^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\
&= \frac{(n-1)}{(n-2)} - \frac{2}{n-1}(n-1) + \frac{2}{(n-1)^2}(n-1)^2 = \frac{n-1}{n-2}.
\end{aligned}$$

Again, $\text{tr}(\mathbf{K}^{(e1)})^2$ does not depend on the design points, it is the one given in (4.8). The variance of $\mathcal{I}^{(e1)}$ under the null is given by:

$$\begin{aligned}\text{Var}[\mathcal{I}^{(e1)}|H_0] &= \frac{2\left\{(n-1-k)\frac{n-1}{n-2}-1\right\}}{(n-1-k)^2(n-1-k+2)} \\ &= \frac{2\left\{(n-2)\frac{n-1}{n-2}-1\right\}}{(n-2)^2 n} = \frac{2}{n(n-2)}. \end{aligned} \tag{4.18}$$

Comparing the variances of $\mathcal{I}$ for extreme case (1) shows that $\text{Var}[\mathcal{I}^{(e1)}|H_0] \geq \text{Var}[\mathcal{I}^{(s1)}|H_0]$, see (4.18) and (4.17). The difference is $\frac{5n-6}{n^2(n^2-n-2)}$, equality holds for $n = \frac{6}{5}$, for $n = -1, n = 0, n = 2$ the denominator is zero and the difference is not defined, for $n \to \infty$ the difference between the variances goes to zero. Thus, the bigger the design, the less important is the treatment of the far-off object.

For treatment (s) and extreme case (2), $\text{tr}(\mathbf{K}^{(s2)}\mathbf{K}^{(s2)})$ and $\text{tr}(\mathbf{K}^{(s2)})^2$ are used.

$$\begin{aligned}\text{tr}(\mathbf{K}^{(s2)}\mathbf{K}^{(s2)}) &= \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}^{(2)}\mathbf{M}^{(s3)}\mathbf{C}^{(2)}) \\ &= \text{tr}(\mathbf{C}^{(2)}\mathbf{C}^{(2)}) - \frac{2}{n}\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)}\mathbf{C}^{(2)}) + \frac{2}{n^2}\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)}\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)}) \\ &= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{n^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\ &= (n-1) - \frac{2}{n}(n-1) + \frac{2}{n^2}(n-1)^2\end{aligned}$$

$\text{tr}(\mathbf{K}^{(s2)})^2$ is given in (4.6).

$$\text{Var}[\mathcal{I}^{(s2)}|H_0] = \frac{2\{(n-k)[n-1+\frac{2}{n^2}-\frac{2}{n}]-\frac{(n-1)^2}{n2}\}}{(n-k)^2(n-k+2)} = \frac{2(n^2-3)}{n^2(n+1)}. \tag{4.19}$$

Excluding the first observation, case (e), gives for extreme case (2)

$$\begin{aligned}\text{tr}(\mathbf{K}^{(e2)}\mathbf{K}^{(e2)}) &= \text{tr}(\mathbf{M}^{(e)}\mathbf{C}^{(2)}\mathbf{M}^{(e)}\mathbf{C}^{(2)}) \\ &= \text{tr}(\mathbf{C}^{(2)}\mathbf{C}^{(2)}) - \frac{2}{n}\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)}\mathbf{C}^{(2)}) + \frac{2}{n^2}\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)}\mathbf{1}\mathbf{1}'\mathbf{C}^{(2)})\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{n^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\
&= (n-1) - \frac{2}{(n-1)}(n-1) + \frac{2}{(n-1)^2}(n-1)^2 = (n-1)
\end{aligned}
$$

and $\text{tr}(\mathbf{K}^{(e1)})^2 = (-1)^2$, see (4.8), this leads to

$$
\text{Var}[\mathcal{I}^{(e2)}|H_0] = \frac{2n(n-2)+2k(1-n)}{(n-1-k)^2(n-k+1)} = \frac{2(n^2-3n+1)}{n(n-2)^2}. \tag{4.20}
$$

Comparing the variances of $\mathcal{I}$ for the second extreme case shows the same relationship as for extreme case (1): $\text{Var}[\mathcal{I}^{(e2)}|H_0] \geq \text{Var}[\mathcal{I}^{(s2)}|H_0]$, see (4.20) and (4.19), taking the far-off object into account leads to a smaller variance. The difference is $\frac{6n^2-26n+24}{(n-2)^2(n-1)n^2}$, equality holds for $n = \frac{4}{3}$ and $n = 3$, for $n = 0, n = 1, n = 2$ the denominator is zero and the difference is not defined, and as in extreme case (1): the greater n, the smaller the difference, for $n \to \infty$ the difference between the variances goes to zero.

For both extreme cases of the spatial link matrix the variance of $\mathcal{I}$ is smaller if the far-off observation is included and zero weighted. Under the assumptions given above, the order of the variances for the three different treatments of a far-off object is:

$$
\text{Var}[\mathcal{I}^{(\nu)}|H_0] < \text{Var}[\mathcal{I}^{(s)}|H_0] < \text{Var}[\mathcal{I}^{(e)}|H_0]. \tag{4.21}
$$

4.6 Standardized Moran's $\mathcal{I}$

The test statistic for the Moran's test $z(\mathcal{I})$ is given in (3.4). To find the difference between the three treatments of a far-off observation, the corresponding values of $\mathcal{I}$, $\text{E}[\mathcal{I}|H_0]$ and $\text{Var}[\mathcal{I}|H_0]$ are used.

The difference between $z(\mathcal{I}^{(\nu)})$ and $z(\mathcal{I}^{(s)})$ depends on the value of $\hat{\varepsilon}_1$ as $\mathcal{I}^{(\nu)}$ includes this residual whereas it does not appear in $\mathcal{I}^{(s)}$. For reasons of simplicity $\hat{\varepsilon}_1 = 0$ is assumed. Then $\mathcal{I}^{(\nu)} = \mathcal{I}^{(s)}$, and the nominator $\mathcal{I}^{(\nu)} - \text{E}[\mathcal{I}^{(\nu)}]$ of $z[\mathcal{I}^{(\nu)}]$ is greater than the one of $z[\mathcal{I}^{(s)}]$, as $\text{Var}[\mathcal{I}^{(\nu)}] < \text{Var}[\mathcal{I}^{(s)}]$ it follows that

$$
z[\mathcal{I}^{(\nu)}] > z[\mathcal{I}^{(s)}].
$$

Thus, adding a small value ν to all elements of the spatial link matrix has an influence on the Moran's test. If the residual of the far-off object is zero, the Moran's test is more likely to reject the null if ν is added to $\mathbf{U}$ than in case of putting zero weight to the observation which is far apart. Whereas if $\hat{\varepsilon}_1 \neq 0$ the difference of the nominators depends on the magnitude of this residual and a general ordering can not be given.

To clarify the difference between excluding and including but zero-weighting the far-off observation the two extreme cases from the previous section have to be considered. The difference of $\mathcal{I}^{(e)}$ and $\mathcal{I}^{(s)}$ is again dependent on the value of $\hat{\varepsilon}_1$. To get rid of this difference $\hat{\varepsilon}_1$ is again assumed to be zero, now $\mathcal{I}^{(e)} = \mathcal{I}^{(s)}$ holds. For extreme case (1) the two different treatments (s) and (e) give:

$$z[\mathcal{I}^{(e1)}] = \frac{\mathcal{I}^{(1)} + \frac{1}{n-1}}{\sqrt{\frac{2}{n(n-2)}}} \text{ and } z[\mathcal{I}^{(s1)}] = \frac{\mathcal{I}^{(1)} + \frac{1}{n}}{\sqrt{\frac{2n^2-3n+6}{n^2(n^2-n-2)}}}.$$

For extreme case (2), where each object, except of the far-off one, has only one neighbour, standardized Moran's $\mathcal{I}$s for case (s) and (e) are:

$$z[\mathcal{I}^{(e2)}] = \frac{\mathcal{I}^{(2)} + \frac{1}{n-1}}{\sqrt{\frac{n(2n^2-6n+3)}{(n-2)^2}}} \text{ and } z[\mathcal{I}^{(s2)}] = \frac{\mathcal{I}^{(2)} + \frac{1}{n}}{\sqrt{\frac{2(n^2-3)}{n^2(n+1)}}}.$$

The relationships between the variances depend on the number of design objects, see Figure 4.1. For both extreme cases the difference between treatment (s) and (e) is positive, but becomes negligible if n increases. For extreme case (1) the difference between the variances goes faster to zero than for extreme case (2). From n approximately greater than 30 the difference is nearly zero, i.e. it does not play a role which treatment is used. Note that $z(\mathcal{I})$ is only approximately $N(0,1)$-distributed, therefore it should not be used for smaller n anyway. For small designs an exact test should be used. The exact distribution of Moran's $\mathcal{I}$ can be found in (Tiefelsdorf & Boots, 1995).

The relationships of the standardized Moran's $\mathcal{I}$ values depend on the number of objects and additionally on the value of $\mathcal{I}$ - this dependence can not be factored out by the zero-assumption on $\hat{\varepsilon}_1$, because even if $\mathcal{I}$ is the same for both treatments, the size of $\mathcal{I}$ still plays a role for the nominator of $z[\mathcal{I}]$.

The relationships for the two extreme cases and the two different treatments of the far-off observation for some different values of $\mathcal{I}$ can be seen in Figures 4.2, 4.3, 4.4, and 4.5. The first plot always shows the standardized Moran's $\mathcal{I}$ depending on n, the second one shows the difference of the z-values between treatment (s) and (e) for the two extreme cases under the assumption that $\hat{\varepsilon}_1 = 0$. In the plots on the left hand side it can be seen that for all examined values of $\mathcal{I}$, $z(\mathcal{I})$ of extreme case (1) is always bigger than of case (2) - at least for designs with more than 5 points, for small designs the approximation of the $N(0,1)$ distribution should not be used anyway. The differences between treatment (s) and (e) do not converge to zero for all values of $\mathcal{I}$. It depends much on the assumed level of autocorrelation. A positive difference means that treatment (e) leads to a more conservative test, i.e. the test based on treatment (s) rejects the null hypothesis earlier. If the difference is negative, the test based on treatment (e) rejects the null earlier. For both extreme cases the differences have the same sign for designs which are large enough to use the normal approximation. For very small values of $\mathcal{I}$, e.g. $\mathcal{I} = 0.01$ treatment (s) rejects the null earlier in small designs, see Figure 4.2, the difference becomes negative if n increases. For medium and high values of $\mathcal{I}$ treatment (e) rejects the null earlier, see Figure 4.4 and 4.5.

For large designs the following ordering of the standardized values of $\mathcal{I}$, under assumption $\hat{\varepsilon}_1 = 0$, can be given:

$$z[\mathcal{I}^{(e)}] \leq z[\mathcal{I}^{(s)}] \leq z[\mathcal{I}^{(\nu)}]. \tag{4.22}$$

4.7 Findings

The standardized Moran's $\mathcal{I}$ depends on the value of $\mathcal{I}$, the residuals of the far-off objects and the concrete form of the spatial link matrix. Therefore, the relationships between the values of $\mathcal{I}$, $\mathrm{Var}[\mathcal{I}|H_0]$ and $z[\mathcal{I}]$ from the different methods of treating far-off objects, given in (4.5), (4.21) and (4.22), do not hold in general as the assumptions that were made are very restrictive. Only the behavior of the expected values $\mathrm{E}[\mathcal{I}|H_0]$ holds without such assumptions, see (4.12). Nevertheless, the influence of far-off observations on the behavior of Moran's $\mathcal{I}$ and the corresponding spatial autocorrelation test is better understood now. The problem of far-off objects arose during the work on a paper about optimal designs for spatial data (Gumprecht, Müller,

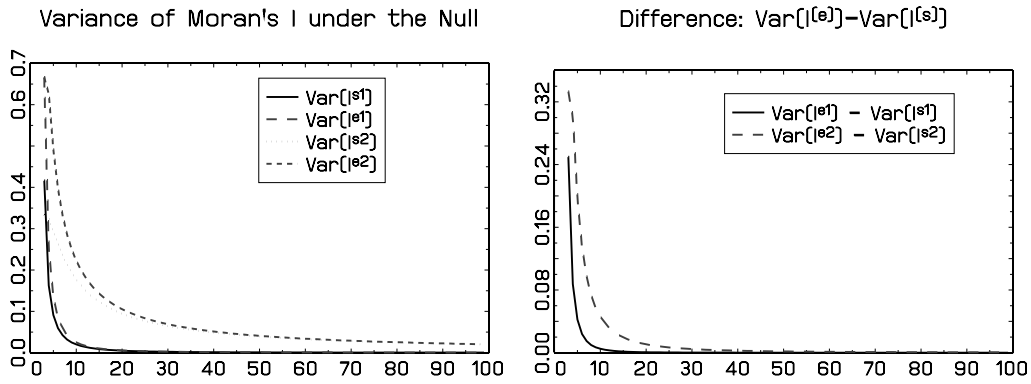


Figure 4.1: Variances of $\mathcal{I}$ and differences between the variances within the extreme cases

& Rodríguez-Díaz, 2007), see chapter 7. For a given set of spatial objects the task is to find the optimal design to detect spatial dependence that might be in the data. The best design is found by an algorithm which evaluates a lot of different possible designs concerning a special design criterion. If these designs are subsets of all n objects, it can easily happen that they include far-off objects. If the number of the design points is large, it does not make a difference which treatment is used for the far-off object. Due to practical reasons, treatment (e) is not recommended, because even if an observation is not connected to others it might be important in the design.

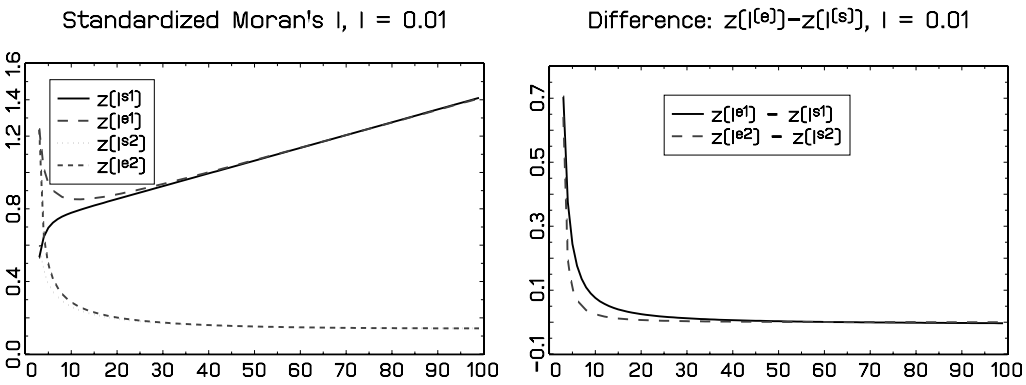


Figure 4.2: $z(\mathcal{I})$ for $\mathcal{I} = 0.01$, and differences within the extreme cases

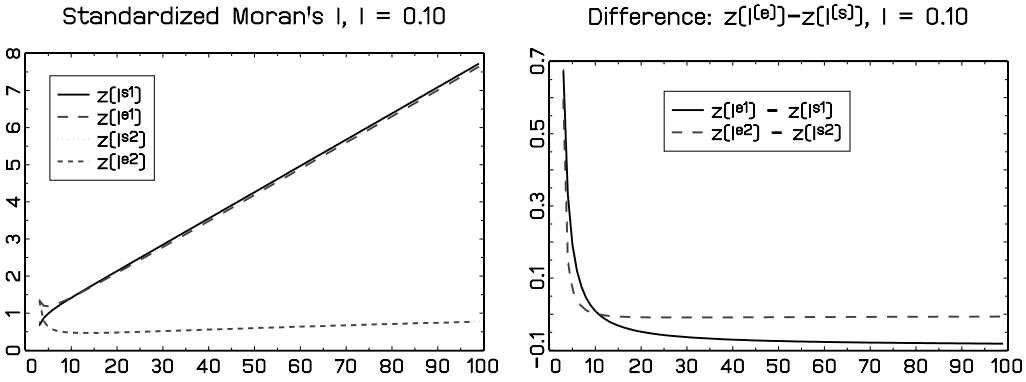


Figure 4.3: $z(\mathcal{I})$ for $\mathcal{I} = 0.10$, and differences within the extreme cases

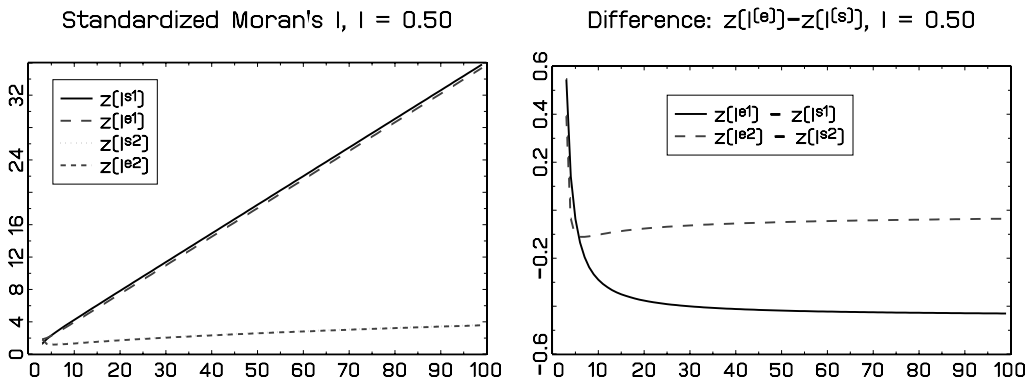


Figure 4.4: $z(\mathcal{I})$ for $\mathcal{I} = 0.50$, and differences within the extreme cases

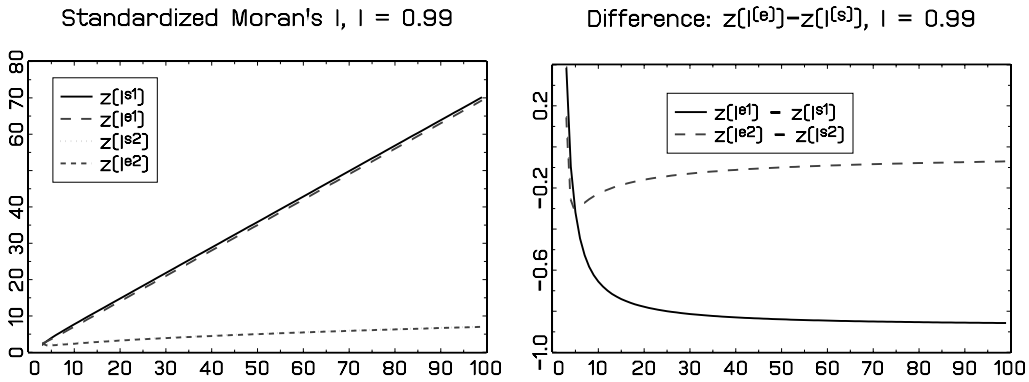


Figure 4.5: $z(\mathcal{I})$ for $\mathcal{I} = 0.99$, and differences within the extreme cases

Chapter 5

Handling of Spatial Data

If there is some spatial dependence existent in the data, there are mainly two possibilities to deal with it. The first alternative is to filter out the spatial effect and use standard statistical methods for the analysis, e.g. use OLS for a regression model. The second one is to use some special spatial estimation techniques, e.g. Spatial Two Stage Least Squares technique or Maximum Likelihood technique in spatial econometrics and variogram estimation and kriging in spatial statistics.

5.1 Spatial Filtering

The basic idea of spatial filtering is to separate the regional interdependencies by partitioning the original variable into two parts: a filtered non-spatial (so called 'spaceless') variable, and a residual spatial variable, and use conventional statistic techniques that are based on the assumption of spatially uncorrelated errors for the filtered ('spaceless') variables. There are different spatial filtering techniques available, one of these methods is based on the local spatial autocorrelation statistic $G_i(\delta)$ from Getis and Ord (1992), see (3.6). Other techniques are based on an eigenfunction decomposition related to the global spatial autocorrelation statistic Moran's $\mathcal{I}$ (Getis and Griffith, 2002). The first method is equally effective but computationally simpler and therefore described in more detail.

The ratio of the expected value of $G_i(\delta)$, given in (3.7), and the original variable indicates the local magnitude of spatial dependence. The filtered

observations are given by:

$$\tilde{x_i} = \frac{x_i \mathrm{E}[G_i(\delta)]}{G_i(\delta)} = \frac{x_i \sum_i v_{ij}(\delta)/(n-1)}{G_i(\delta)}.$$

The purely spatial component of the variable is then given by $(x_i - \tilde{x_i})$.
If δ is chosen properly, the standardized value of $G_i(\delta)$ corresponding to $\tilde{x_i}$ is insignificant (demonstrated by Getis and Griffith, 2002). This means: filtering all variables (dependent and independent ones) in a regression model removes the spatial dependence and allows one to use a conventional regression model in which the parameters are estimated by ordinary least squares. A practical problem, when using this filtering technique is the choice of the structure of the spatial link matrix $\mathbf{V}$ and the choice of the locality parameter δ of the regional weighting scheme. One possibility to model the distance decay is to use a negative exponential function (2.1) for the unstandardized matrix $\mathbf{U}$

$$u_{ij} = \exp(-\delta d_{ij}),\ 0 \leq \delta \leq \infty,$$

where d_{ij} denotes the (e.g. geographic) distance between the locations i and j. The choice of the structure does not have decisive impact on the outcomes, whereas the choice of δ is more problematic. Several methods to determine δ are discussed in Getis (1995), one of these methods to choose δ properly is: $\tilde{\delta} = \arg\max_\delta \sum_i |z_{Gi}(\delta)|$.

5.2 Estimation in Spatial Econometrics

Another possibility to deal with spatially dependent data is to use spatial estimation techniques. In this case the spatial effect is not excluded from the data, like in the spatial filtering approach, but adequately included in the estimation. There are different estimation methods for spatial data, one can e.g. use the Maximum Likelihood (ML) technique (first outlined by Ord, 1975), or a Spatial Two Stage Least Squares (S2SLS) method based on Instrumental Variable (IV) estimation (see e.g. Kelejian and Robinson, 1993; or Kelejian and Prucha, 1998), or based on a Generalized Method of Moments (GMM) (Kelejian and Prucha, 1999).

5.2.1 Spatial Two Stage Least Square Estimation

Kelejian and Prucha (1999) suggest to use the following procedure to estimate a SAR model: For a spatial autoregressive model, given in equation (2.5): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ and $\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$, the covariance matrix is given by equation (2.7): $\boldsymbol{\Omega}(\rho, \sigma^2) = \sigma^2[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$. The auxiliary parameters ρ and σ^2 are estimated via the Generalized Method of Moments technique. The Generalized Moments estimator of ρ and σ^2 is a non-linear least squares estimator:

$$(\tilde{\rho}, \tilde{\sigma}^2) = \arg\min_{\rho,\sigma^2} \{[\boldsymbol{\Gamma}(\rho, \rho^2, \sigma^2)' - \boldsymbol{\gamma}]'[\boldsymbol{\Gamma}(\rho, \rho^2, \sigma^2)' - \boldsymbol{\gamma}]\}$$

where $\rho \in [-a, a]$ with $a \geq 1$ and $\sigma^2 \in [0, b]$, they are elements of the vector (ρ, ρ^2, σ^2). Matrix $\boldsymbol{\Gamma}$ and vector $\boldsymbol{\gamma}$ are functions of the OLS residuals derived from the moment conditions, and $(\boldsymbol{\Gamma}(\rho, \rho^2, \sigma^2)' - \boldsymbol{\gamma})$ can be viewed as a vector of residuals. Detailed specifications can be found in Kelejian and Prucha (1999, p.8). The parameter $\boldsymbol{\beta}$ of the regression model is then a Feasible Generalized Least Squares (FGLS) estimator:

$$\tilde{\boldsymbol{\beta}} = [\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\tilde{\boldsymbol{\Omega}}^{-1}\mathbf{y}, \tag{5.1}$$

where $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\tilde{\rho}, \tilde{\sigma}^2)$.

For more complex spatial models, i.e. models containing spatial lags in the dependent variable, the exogenous variables, and the disturbances and additionally allowing for unknown heteroskdasticity in the innovations, a modified Generalized Moments (GM) estimator for the spatial autoregressive parameter in the disturbances and an Instrumental Variable (IV) estimator for the regression parameters of the model are given in Kelejian and Prucha (2006).

5.2.2 Maximum Likelihood Estimation

An alternative to the S2SLS technique from Kelejian and Prucha (1999) for estimating spatial regression models is the Maximum Likelihood method, see e.g. Anselin (1988) or Anselin (1999). It is based on the assumption of normally distributed error terms of the spatial regression model. The joint likelihood follows from the multivariate normal distribution of the dependent variable. Considering e.g. a SAR error model (2.5) with $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\rho))$ and

$\mathbf{\Omega}(\rho) = \sigma^2[(\mathbf{I} - \rho\mathbf{V})'(\mathbf{I} - \rho\mathbf{V})]^{-1}$, see (2.7), the log likelihood is given by:

$$\text{lnL} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|\mathbf{I} - \rho\mathbf{V}| - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \rho\mathbf{V})'(\mathbf{I} - \rho\mathbf{V})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The ML estimates of $\boldsymbol{\beta}$ and σ^2 are:

$$\hat{\boldsymbol{\beta}}_{ML} = \left[(\mathbf{X} - \rho\mathbf{V}\mathbf{X})'(\mathbf{X} - \rho\mathbf{V}\mathbf{X})\right]^{-1}(\mathbf{X} - \rho\mathbf{V}\mathbf{X})'(\mathbf{X} - \rho\mathbf{V}\mathbf{X}),$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{n}(\hat{\mathbf{u}} - \rho\mathbf{V}\hat{\mathbf{u}})'(\hat{\mathbf{u}} - \rho\mathbf{V}\hat{\mathbf{u}}),$$

with $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML}$, respectively. The estimate for the spatial parameter ρ follows a maximization of a concentrated likelihood function.

Both approaches, the ML estimation as well as the GMM based estimation, can be used for spatial models. In the ML procedure the computation of the logarithm of the determinant of the Jacobian is problematic, especially in samples with a large number of observations. Larch et al. (2007) compare the performance of the GMM based estimation technique from Kelejian and Prucha (1999) and the ML technique with different approaches of computing the determinant of the Jacobian in huge samples. They conclude that GMM as well as ML with a Monte Carlo estimator of the determinant of the Jacobian are the most suitable methods for large sample sizes.

5.3 Measuring and Estimating Spatial Relations in Geostatistics

In the field of geostatistics the most widespread methods to measure spatial relations are the covariance, the correlogram, and the variogram. Covariance and correlogram show how the correlation behaves with the distance between the points, and the variogram shows the dissimilarity between two different located points $Z(\mathbf{x})$ and $Z(\mathbf{x} + \mathbf{h})$ for a certain class of stochastic processes, namely the intrinsic stochastic processes (definitions are given in chapter 2).

There is a close relationship between the Moran's plot respectively the spatial autocorrelation level and the variogram plot respectively the variogram. A variogram plot or variogram cloud is a scatterplot with the distances on the

horizontal- and the variogram values on the vertical axis, see section 5.3.3, it is an important tool in geostatistics. A shallow slope of the variogram plot indicates a strong positive spatial autocorrelation, whereas a steep slope of the variogram plot denotes very small or zero spatial autocorrelation (Griffith & Layne, 1999). In other words, the higher the positive spatial autocorrelation the smaller the value of the variogram and vice versa.

5.3.1 Covariogram and Correlogram

Stationary random functions $Z(\mathbf{x})$, see section 2.4, are characterized by their mean

$$\mathrm{E}[Z(\mathbf{x})] \tag{5.2}$$

and their covariance, often referred as covariogram:

$$\mathrm{C}(\mathbf{h}) = \mathrm{E}[\{Z(\mathbf{x}) - \mathrm{E}[Z(\mathbf{x})]\}\{Z(\mathbf{x}+\mathbf{h}) - \mathrm{E}[Z(\mathbf{x}+\mathbf{h})]\}], \tag{5.3}$$

where $\mathbf{h}$ is a vector of the spatial lag or spatial separation. It holds that

$$\mathrm{C}(\mathbf{h}) = \mathrm{C}(-\mathbf{h}) \text{ and } |\mathrm{C}(\mathbf{h})| \leq |\mathrm{C}(\mathbf{0})|,$$

$\mathrm{C}(\mathbf{0})$ represents the covariance at lag $\mathbf{0}$ - this is simply the variance. Linked to the covariance function is the so called correlogram function, which is defined as

$$\rho(\mathbf{h}) = \frac{\mathrm{C}(\mathbf{h})}{\mathrm{C}(\mathbf{0})}. \tag{5.4}$$

The correlogram is a correlation coefficient between $Z(\mathbf{x})$ and $Z(\mathbf{x}+\mathbf{h})$.

Covariance and correlogram depend on the length and the direction of vector $\mathbf{h}$, for isotropic processes (see section 2.4) they depend only on the length of $\mathbf{h}$ (Chilès & Delfiner, 1999).

5.3.2 Theoretical Variogram

The theoretical variogram $\gamma(\mathbf{h})$ describes the variation in space of an intrinsic random function, see section 2.4. It shows how the dissimilarity between $Z(\mathbf{x})$ and $Z(\mathbf{x}+\mathbf{h})$ evolves with separation $\mathbf{h}$ (Chilès & Delfiner, 1999). For intrinsic stochastic processes the mean of the increments (= difference between values at pairs of points $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$: $Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$) is invariant

for any translation of a given vector $\mathbf{h}$ and it is assumed to be zero: $\mathrm{E}[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})] = 0$ - this mean is often called drift, and the variance of the increments have a finite value $2\gamma(\mathbf{h}) = \mathrm{Var}[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]$ which depends only on the length and the orientation of $\mathbf{h}$ but not on the placement. One half of this variance $2\gamma(\mathbf{h})$ is called theoretical variogram (or sometimes also semi-variogram)

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}))^2\right]. \tag{5.5}$$

A theoretical variogram has the following characteristics:

(i) At the origin a variogram is zero by definition: $\gamma(\mathbf{h}) = 0$

(ii) All values of a variogram are positive: $\gamma(\mathbf{h}) \geq 0$

(iii) Variograms are symmetric in space: $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$

(iv) Variograms grow slower than $|\mathbf{h}|^2$: $\lim_{|\mathbf{h}|\to\infty} \frac{\gamma(\mathbf{h})}{|\mathbf{h}|^2} = 0$

(v) Variograms can be derived from a covariance: $\gamma(\mathbf{h}) = \mathrm{C}(\mathbf{0}) - \mathrm{C}(\mathbf{h})$

Definitions can e.g. be found in Wackernagel (1998). This theoretical variogram has to be estimated in practice.

5.3.3 Variogram Cloud and Sample Variogram

In a variogram cloud the dissimilarities between pairs of sample values of a regionalized variable $z(\mathbf{x})$ are plotted against the (geographic) distance of these pairs. For a separation $\mathbf{h}$ (which is a specified distance and direction) of sample points, the dissimilarity between pairs of data values $z(\mathbf{x})$ and $z(\mathbf{x}+\mathbf{h})$ depends on the spacing and the orientation of the data pair. It is given by

$$\gamma^*(\mathbf{h}) = \frac{1}{2}\left(z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})\right)^2. \tag{5.6}$$

Thus, the variogram shows the semi-variance on the vertical axis and the absolute distance between the values on the horizontal axis. In the variogram

cloud the spread of values at different lags is shown, this might help to identify outliers or anomalies (Webster & Oliver, 2001).

Instead of the original dissimilarities given in (5.6) averages of these semivariances are used for the experimental variogram. Therefore classes of vectors H are built. In a certain class H_k all n_k point pairs that can be linked by a vector $\mathbf{h}$ (which belongs to H_k) are included. The average semivariance or experimental variogram is given by:

$$\hat{\gamma}_S(H_k) = \frac{1}{2N_k} \sum_{i=1}^{N_k} \left[z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h}) \right]^2 \text{ with } \mathbf{h} \in H_k. \tag{5.7}$$

As a rule of thumb, vectors $\mathbf{h}$ used for the experimental variogram should have a length lower half the diameter of the whole region (Wackernagel, 1998).

5.3.4 Regional Variogram

The best sample variogram of a region can be calculated if the whole domain D is perfectly known. It is called regional variogram.

$$\gamma_R(\mathbf{h}) = \frac{1}{2|D(\mathbf{h})|} \int_{D(\mathbf{h})} \left[z(\mathbf{x} + \mathbf{h}) - z(\mathbf{x}) \right]^2, \tag{5.8}$$

where $D(\mathbf{h})$ is the intersection of the domain D with a translation $D_{-\mathbf{h}}$, i.e. $\mathbf{x}$ and $\mathbf{x}(\mathbf{h})$ belong to $D \cap D_{\mathbf{h}}$, and $|D(\mathbf{h})|$ is the measure of $D \cap D_{\mathbf{h}}$. In practice the value of the variable $z(\mathbf{x})$ is only known for sparsely locations, therefore $z(\mathbf{x})$ is considered as realizations of random variables $Z(\mathbf{x})$. Using the random variables for formula (5.8) gives the randomized version of the regional variogram. The expectation of it defines the theoretical variogram $\gamma(\mathbf{h})$, given in (5.5), see e.g. Wackernagel (1998).

5.3.5 Variogram Models

The experimental variogram gives a summary of the spatial relationships in the data. For each lag $\mathbf{h}$ the calculated semivariance is an estimate of a mean semivariance for that lag. As an estimate it is influenced by an error. For each lag the sample variogram gives one estimate - this gives an discrete

function, whereas the true variogram is a continuous function. One is not interested in the point to point fluctuation but in the general trend. So, a simple and sensible function is fit to the sample variogram (Webster & Oliver, 2001).

There are many different variogram models in use. The linear model e.g. is the simplest one, the values on the vertical axis are $\gamma(\mathbf{0}) = 0$ and $\gamma(\mathbf{h}) = C_0 + p\mathbf{h}$ (for $\mathbf{h} > 0$) respectively, parameter p gives the slope of the line and C_0 is the intercept, often called nugget effect, values on the horizontal axis are the distances $\mathbf{h}$ between the pairs. Other commonly used models are the generalized linear model, the spherical model, the exponential model, the Gaussian model, the hole effect model, etc. Descriptions of these models can e.g. be found in Clark and Harper (2000).

5.3.6 Upper-Austria SO_2 Monitoring Network

This example stems from the field of environment pollution, namely air pollution, it is one of the illustrating examples in Müller (2001). Data are taken from the Upper-Austria sulphur dioxide (SO_2) monitoring network. The network includes 17 gauging stations, most of them are located in or near the capital city Linz, which is also the main industrial region, see Figure 5.1. The observation sites are: Lenzing, Linz-Hauserhof, Linz-Urfahr, Traun, Asten, Wels, Vöcklabruck, Perg, Steyr, Braunau, Linz-Kleinmünchen, Linz-Ursulinenhof, Linz-ORF-Zentrum, Linz-24er-Turm, Linz-Berufsschulzentrum, Steyregg-Weih, Schöeneben. The dataset contains 288 daily averages of SO_2 concentrations in mg/m^3 during the period from January 1994 until December 1995. Müller (2001) estimates a spherical variogram from data of a typical day (1^{st} of March 1994). The formula of a spherical variogram is given by:

$$\gamma_S(h,\theta) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_1 + \theta_2 \left[\frac{3}{2}\left(\frac{h}{\theta_3}\right) - \frac{1}{2}\left(\frac{h}{\theta_3}\right)^3 \right] & \text{if } 0 < h \leq \theta_3 \\ \theta_1 + \theta_2 & \text{if } h > \theta_3 \end{cases} \tag{5.9}$$

The parameters are estimated from the variogram cloud (see section 5.3.3). The estimated parameters are $\hat{\theta}_1 = 0.164$, $\hat{\theta}_2 = 0.323$ and $\hat{\theta}_3 = 24.25$, i.e. $\hat{\theta}_1$ is the estimated nugget effect, $\hat{\theta}_1 + \hat{\theta}_2$ is the estimated sill value, which gives the height of the variogram, and $\hat{\theta}_3$ is an estimate for the range of influence.

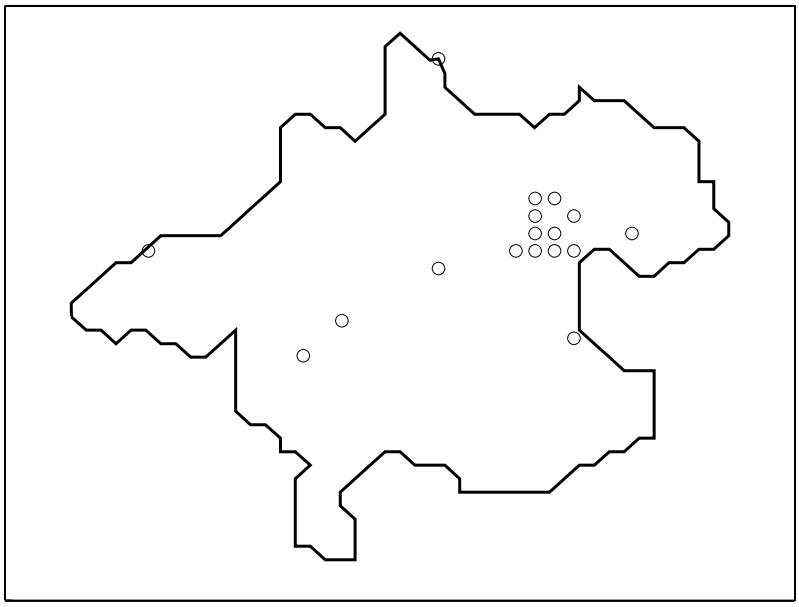


Figure 5.1: Upper-Austria SO_2 monitoring network

5.3.7 Kriging

Kriging is an estimation method to find adequate estimates of values of regionalized variables at locations that are not observed. The idea behind is, to use weighted averages of the known values of the observed points and additionally include the knowledge of the covariances or the variogram between the values at the points involved. Kriging gives the best unbiased linear estimates of values of regionalized variables, best in the sense of minimum variance. Depending on the level of knowledge and the variable of interest there are different kriging procedures on-hand. If the (true) mean of the regionalized variable is known, one uses the so-called 'Simple Kriging' to find the best predictors of unobserved sample points. Whereas if the mean is not known, one can estimate it via the procedure 'Kriging the mean', and if unknown values should be estimated, one runs 'Ordinary Kriging' method. Details about Kriging can be found in (nearly) all textbooks about geostatistics, see e.g. Armstrong (1998) or Wackernagel (1998).

Chapter 6

R&D Spillovers - An Application in Spatial Econometrics

The methods presented in the previous chapters are commonly used for 'classical' spatial problems, in the sense that geo-referenced data are of interest, i.e. the locational information is usually given in coordinates. The distances or weights used in the analysis typically represent geographic distances or a function of them. However, spatial methods are not restricted to this kind of geo-referenced data, they can be used in much more cases, even ones that are not obviously spatial. Some of the methods from the field of spatial econometrics and the application to a 'non-spatial' problem, namely the classical 'International R&D Spillover' dataset from Coe and Helpman (1995), are exemplified in the following.

6.1 R&D Spillovers Data

In 1995 Coe and Helpman published an article titled "International R&D Spillovers" (Coe & Helpman, 1995), which became quite famous in the economic society. It was discussed, re-analysed and also criticised many times. The main intention of this study was to answer the question whether coun-

[2]This chapter is part of a working paper published in the Research Report Series of the Vienna University of Economics and Business Administration (Gumprecht, 2005), furthermore it is published in the Austrian Journal of Statistics (Gumprecht, 2007a).

tries can profit from the 'knowledge' of their trade partner countries in the sense that a higher level of 'knowledge' has a positive effect on their own productivity, i.e. whether there exists a spillover effect of the knowledge. In general the 'knowledge' of a whole country is delicate to define and even more complicate to measure. Under the term 'knowledge' of a country, Coe and Helpman catch innovations and technical progress and measured it via the R&D capital stock of a country. The problem of R&D spillover effects is well known nowadays, none the less it is still not really understood and there are different opinions whether such an R&D spillover phenomenon exists or not.

This chapter is structured as follows: The first part includes the non-spatial analysis of Coe and Helpmans (1995) dataset, and some of the criticisms and suggestions for improvement are discussed and the corresponding outcomes are presented, the second part contains a spatial analysis of this problem.

6.1.1 'Non-Spatial' Analysis and Results

In the economic background of Coe and Helpmans (1995) analysis are the theories of economic growth, that treat commercially oriented innovation efforts as a major engine of technological progress and productivity growth (Romer, 1990), (Grossman & Helpman, 1991). This means that on one hand innovation profits from knowledge that results from R&D spending and on the other hand innovation contributes to this stock of knowledge. This means, the higher the R&D expenditures of a country, the higher is the productivity growth. Coe and Helpman (1995) go one step further, and claim that the productivity of a global economy depends on its own stock of knowledge as well as the knowledge of its trade partners. They study the extent to which a country's productivity level depends on the domestic and the foreign stock of knowledge. The domestic stock of knowledge is quantified by the cumulative expenditures for R&D of a country, the foreign stock of knowledge is quantified via an import-weighted sum of cumulated R&D spending of the trade partners of a country. The importance of the R&D capital stock is measured by the elasticity of total factor productivity with respect to the R&D capital stock. Their empirical analysis is based on a panel dataset that contains 22 countries (21 OECD countries plus Israel) and 20 years (during the period from 1971 to 1990). The dependent variable of the model is the total factor productivity (TFP), and the independent variables are domestic R&D capital stock (DRD) and foreign R&D capital

stock (FRD), all are constructed as indices with basis 1985 (1985=1). The dataset is available on the homepage of Elhanan Helpman (Helpman, 2003), which is accessible via the internet address:

`http://post.economics.harvard.edu/faculty/helpman/data.html`

In their paper Coe and Helpman (1995) used a variety of quite similar specifications to model the effects of DRD and FRD on TFP. Only the first one is selected to demonstrate the methods and results, the following conclusions, however, are not limited to this particular case but rather apply to all of the suggested models (for a more complete analysis see Gumprecht, 2003). This first model contains three variables: TFP as the regressand, and DRD and FRD as the regressors. The impact of domestic and foreign R&D expenditures is supposed to be the same for all countries. The equation - with regional index i and temporal index t - has the following form:

$$logF_{it} = \alpha_{it}^0 + \alpha_{it}^d logS_{it}^d + \alpha_{it}^f logS_{it}^f, \tag{6.1}$$

where F_{it} denotes TFP, S_{it}^d denotes DRD and S_{it}^f stands for FRD, which is defined as a bilateral import-share weighted average of the DRD of the trade partners:

$$S_{it}^f = \sum_{i \neq j} b_{ijt} S_{jt}^d \tag{6.2}$$

where b_{ijt} denotes the bilateral import-shares of country i from country j in period t, $b_{ijt} \neq b_{jit}$ and $\sum_j b_{ijt} = 1$. α_{it}^0 stands for the country-specific intercepts. This means the intercepts are allowed to vary across countries. This assumption was made for mainly two reasons: first, there may exist country specific effects on productivity that are not included in the variables of this model, and second, all variables are transformed into index numbers, TFP is measured in the country specific currency whereas DRD and FRD are measured in U.S. dollars. α_{it}^d and α_{it}^f denote the regression coefficients, α_{it}^d corresponds to the elasticity of TFP with respect to DRD, and α_{it}^f determines the elasticity of TFP with respect to FRD. According to standard practice in time series literature Coe and Helpman (1995) used a panel data model with fixed effects for their estimations. The aim of Coe and Helpman was to estimate the long-run relationship between TFP and the domestic and foreign R&D capital stocks. Therefore and because almost all of the data exhibit a clear trend, they estimate cointegrated equations. The OLS estimate of such a cointegrated equation is said to be 'super consistent', that is, the estimate

converges on the true parameter value much faster than in the case where the variables are stationary (Stock, 1987). Coe and Helpman (1995) give the following OLS estimates of the fixed effects model (6.1)

$$logF_{it} = \alpha_{it}^0 + 0.097 logS_{it}^d + 0.0924 logS_{it}^f. \tag{6.3}$$

Coe and Helpman (1995) took these estimation results, both with a positive regression coefficient, as a confirmation of their hypothesis that TFP of a country depends on both domestic and foreign R&D capital stock, although they did not calculate t- or p-values for the parameter estimators. Therefore this model was estimated once again, now using the Least Squares Dummy Variable (LSDV) method for the estimation and including the tests for the parameter estimators. The coefficients are the same as the ones from Coe and Helpman, the t-value for α^d is 10.6834, the one for α^f is 5.8673. Both coefficients are positive and significant on a 1% level; the fit of the model is moderate with a *pseudo* $R^2 = 0.5584$, which is calculated as the squared correlation between $\hat{y}_{it}$ and y_{it}. These results are given in column 'Model 2' in the left part of Table 6.2.

6.1.2 Other Non-Spatial Approaches

Coe and Helpmans article was discussed many times and many suggestions for improvement of the model and estimation were made. Kao, Chiang and Chen (1999) e.g. criticized (among other points) that in spite of the super consistency of the time-series estimator, the bias of the estimation can be quite substantial for small samples and there is no reason to assume that this bias becomes negligible by the inclusion of a cross section dimension in panel data. Therefore they used different estimation methods for the international R&D spillovers regression, namely the OLS, the Fully-Modified (FM) and the Dynamic OLS (DOLS) estimation, and compared the empirical consequences from these methods. They claim that the DOLS estimation is the best solution for this problem because in the given setting the DOLS estimator exhibits no bias and is asymptotically normal. The DOLS estimation of Coe and Helpmans fixed effects model, given in (6.1), can be found in column 'Model 2' in the left part of Table 6.3.

Another major issue in the panel data estimation literature is the choice of the model, more precisely whether one should regard the region specific

or other effects as random. This poses a valuable alternative to the fixed coefficient model. In the context of R&D spillovers Müller and Nettekoven (1999) suggest to use a random coefficient model for the R&D dataset. In a random coefficient model the parameters are assumed to vary randomly around a common mean. This model is well compatible with the data, nevertheless the conclusions drawn from this estimation are contradictory to the ones from Coe and Helpman (1995) in the sense that the estimate of the FRD changes sign, although this is not statistically significant. Contrary to Coe and Helpman's conclusions from the fixed effects model, a random coefficients model indicates that there does not exist a significant effect of the foreign R&D expenditures. See column 'Model 2' in the left part of Table 6.4.

After a detailed examination of Coe and Helpmans (1995) work and the various critics of it, where the focus laid rather on the econometric model and estimation technique than on the economic model and specification, the following changes and modifications were suggested by Gumprecht et al. (2004): use of a random coefficient model and use of the DOLS technique for its estimation. The DOLS random coefficient estimation yields

$$logF_{it} = \alpha_{it}^0 + 0.3529 logS_{it}^d - 0.085 logS_{it}^f. \tag{6.4}$$

The t-value for α^d is 7.7946 and is significant on a 5% level, the one for α^f is -1.1866 and is not significant, and $pseudo\ R^2 = 0.9736$. The results of the panel cointegration model with random coefficient and dynamic regressors do not support Coe and Helpman's hypothesis, that the TFP of a country depends on domestic and foreign R&D knowledge (measured by the R&D expenditures). The effect of the knowledge of the trade partners of a country is not significant. It seems from (6.4) that foreign R&D do rather not affect the TFP of a country.

These different results do not help to find an answer to the original question, whether the foreign R&D expenditures effect the productivity of a country. Depending on the econometric model and estimation method one prefers, the conclusions diverge.

6.2 A Spatial Approach for the Analysis of R&D Spillovers

A spatial analysis of the R&D spillover data can be examined if the countries are regarded as regions - this can be done - and if an appropriate spatial link matrix is available. A spatial link matrix measures connectivity or distance between the regions, see chapter 2. The first thought of a distance matrix is one of geographic distances. But, in a global economy not the geographic distances but rather the trade intensity between two countries is relevant for R&D spillover effects. Given the economical and theoretical background of Coe and Helpmans (1995) analysis, and to be consistent with them, it is self-evident to use the bilateral import shares (in year 1990) as a row-standardized spatial link matrix. To define a kind of economic distance a symmetric connectivity measure is needed, therefore a new quantity for the trade intensity between two countries is defined, simply being the average of the bilateral import-shares of these countries. The elements of the symmetric spatial connectivity matrix are therefore calculated by:

$$v_{ij} = \frac{b_{ij} + b_{ji}}{2} \text{ for } i \neq j,$$

where b_{ij} are the bilateral import-shares of country i from country j in period 1990, and by definition $v_{ij} = 0$ for $i = j$. It was assumed that the trade intensity is the same for all periods, this means the same spatial link matrix is used for all years. The economic distances between two countries are simply the inverse connectivity:

$$d_{ij} = \frac{1}{v_{ij}}$$

and by definition $d_{ii} = 0$. Now these distances can be used to produce a 'trade-intensity' landscape by projecting all distances from the 21-dimensional space to the 2-dimensional space. For this projection a Multidimensional Scaling method is used. The squared sums of the distances between the original and the projected points (the points represent the countries) are minimized. This gives an approximation of all 231 distances between the 22 countries in the 2-dimensional space, and provides a quite good survey of the relationships in the data set, Figure 6.1. Here the countries are quite evenly scattered, nevertheless some clusters can be identified, e.g. Australia, New

Zealand and Israel are quite far apart from the rest of the countries, this means they have a small trade intensity with other countries and a relative high trade intensity within their group. The U.S. are settled in the center, it can be interpreted in the way that the U.S. are an important trade partner for all countries.

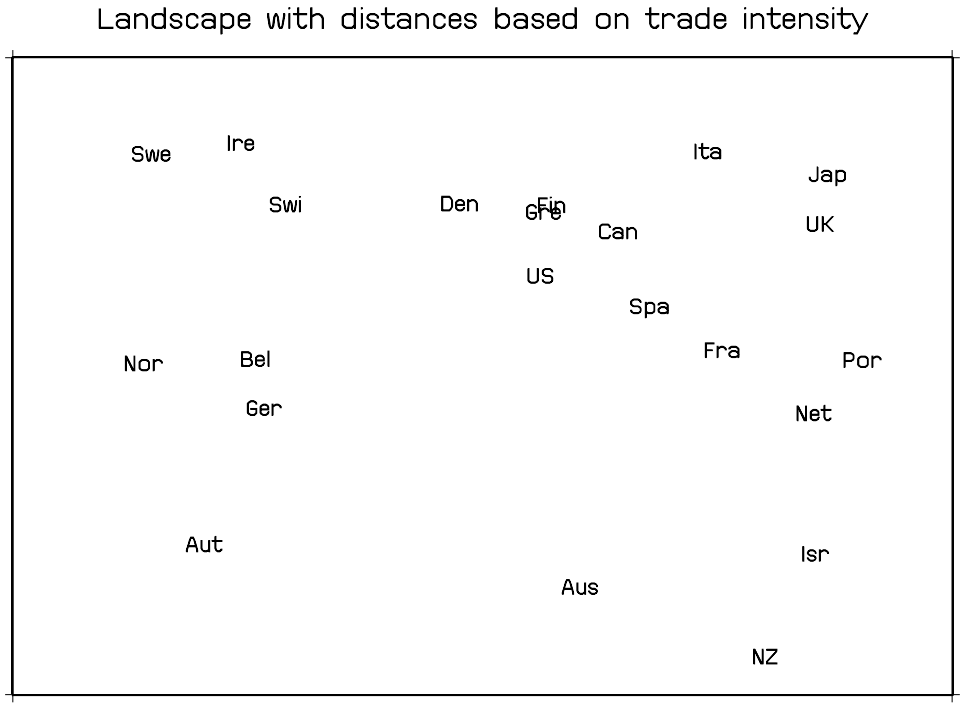


Figure 6.1: Landscape based on trade-intensities between the countries

From the inverse trade intensity matrix a neighbourhood matrix can be easily constructed, e.g. via specifying a certain radius, and all countries which lie within this area are regarded as neighbours. Doing this can help to see how different values of the spatial parameter ρ influence the dependencies between the countries. For a radius equal to 0.25 and $\rho = 1$ the U.S. have 17 neighbouring countries (not neighbours are: Austria, Belgium, Greece and Portugal), if $\rho = 0.5$ the U.S. have just 8 neighbouring countries (Japan, Germany, UK, Canada, Australia, Ireland, Israel and New Zealand), and if $\rho = 0.1$ the only neighbour of the U.S. is Canada.

6.2.1 Results of a Spatial Analysis

The spatial link matrix for the spatial regression model is the row-standardized bilateral import-shares matrix $\mathbf{V}$ from Coe and Helpmans dataset. The first step in the spatial analysis is to estimate a fixed effect model without any foreign R&D spending and without any spatial structure:

$$logF_{it} = \alpha_{it}^0 + \alpha_{it}^d logS_{it}^d$$

which gives an $\hat{\alpha_{it}^d} = 0.1362$ (significant on 1% level) and to calculate and test Moran's $\mathcal{I}$ for the residuals of this model for each period separately (see section 3.1). As spatial link matrix the bilateral import shares (matrix $\mathbf{V}$) are used. Nearly all values are not significant (see Table 6.1), this means there is no global spatial effect in the error term. Even if there is no global spatial effect, local spatial effects can be included, and e.g. if there are positive and negative local spatial effects in the data these effects can compensate each other and the global Moran's $\mathcal{I}$ test shows no significant global spatial effect. If there is a spatial effect in the error term, one should use an adequate estimation technique for the SAR regression model, given in (2.5), e.g. the FGLS estimation from Kelejian and Prucha (1999), see section 5.2.1. This leads to similar results as the non-spatial analysis, the results of the non-spatial model are given in column 'Model 1' in the left part of Table 6.2, the results of the spatial model are given in column 'Model 1' in the right part of Table 6.2. A fixed effect SAR model including the foreign R&D spending (original definition from Coe and Helpman) is estimated to compare the results with the ones from Coe and Helpman (1995), given in (6.3). For the results see Table 6.2, column 'Model 2' in the right part. DRD as well as FRD have a positive and significant effect on TFP.

The foreign R&D spending can be regarded as spatially lagged domestic R&D spending. To avoid the logarithms of the independent variables and as all of the values of S_{it}^d are around one, a Taylor series approximation is used for the logarithm. In general the Taylor series approximation of a function f around a value $x = a$ is given by:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

For the R&D data, $logS_{it}^d$ is substituted by the first two terms of the Taylor series approximation:

$$logS_{it}^d \simeq S_{it}^d - 1$$

and $log(\sum_{i \neq j} b_{ijt} S_{jt}^d)$ is substituted by

$$log(\sum_{i \neq j} b_{ijt} S_{jt}^d) \simeq \sum_{i \neq j} b_{ijt} S_{jt}^d - 1.$$

This leads to the following fixed effects model with one spatially lagged exogenous variable:

$$logF_{it} = \tilde{\alpha_{it}}^0 + \alpha_{it}^d S_{it}^d + \alpha_{it}^f b_{ijt} S_{jt}^d \ . \tag{6.5}$$

The fixed effects change from α_{it}^0 to $\tilde{\alpha_{it}}^0 = \alpha_{it}^0 - \alpha_{it}^d - \alpha_{it}^f$. Ignoring an additional spatial effect that might be included in the error term, the fixed effect panel regression, specified in (6.5), can be estimated by LSDV. This simple approach gives positive and significant parameter estimators for DRD as well as FRD.

$$logF_{it} = \tilde{\alpha_{it}}^0 + 0.0673 S_{it}^d + 0.1787 b_{ijt} S_{jt}^d \ ,$$

see Table 6.3 column 'Model 3' in the left part. Under the assumption of a SAR error model, where a spatial effect is included in the error term, see formula 2.5, a FGLS estimation based on GM estimators of the autoregressive parameter and the noise variance leads to unbiased estimators. Such a spatial estimation gives for the fixed effects model a positive effect of DRD and a negative effect of FRD, and both of them are significant (column 'Model 3' in the right part of Table 6.2). This result is neither in line with any other result so far, nor with the common beliefs and theories in the economic community. This result is dubious anyway, as all of the critics of the original, non-spatial analysis of Coe and Helpman are also legitimate in the spatial context. Therefore a dynamic random coefficients panel regression should be executed.

The dynamic random coefficients model with a spatially lagged exogenous variable and a spatial effect in the error term, estimated by FGLS yields

$$logF_{it} = \tilde{\alpha_{it}}^0 + 0,0809 S_{it}^d + 0.0161 b_{ijt} S_{jt}^d.$$

The parameter estimator for DRD is positive and significant on 10% level, whereas the estimator for FRD is not significant (see column 'Model 3' in the right part of Table 6.5). This is in consensus with the non-spatial results

and again in contrast with the original conclusions from Coe and Helpman (1995). Nevertheless, the unusual high value of $\hat{\rho}$ (=0.72) indicates overcompensation. This is caused by the fact, that the spatial effect is already included as a spatially lagged independent variable and an additional spatial effect in the error term leads to an overcompensation.

Thus, the preferred method is the DOLS estimation of the random coefficients model with approximated variables, a spatially lagged independent variable but no additional spatial effect in the error term, which yields

$$logF_{it} = \tilde{\alpha_{it}}^{0} + 0,1252S_{it}^{d} + 0.1663b_{ijt}S_{jt}^{d}.$$

with *pseudo* $R^2 = 0.976$. This model has the best fit of all examined models and the result is in consensus with the original conclusions from Coe and Helpman (1995).

6.3 Findings

In the R&D dataset an adequate spatial contiguity matrix is already given by the bilateral import shares, even if it is not used in this way in the original analysis. Anyway, it is quite simple to use these relationships for correcting an additional spatial dependence that is not properly captured by the given regressors. The aim of the analysis of the R&D spillovers dataset was to answer the question, whether domestic and foreign R&D spending have an effect on the total factor productivity of a country. Concerning domestic R&D spending the answer is quite obvious, all different estimation techniques (fixed effects- and random coefficients model) and both non-spatial and spatial approach lead to the conclusion that domestic R&D spending have a positive effect on the total factor productivity of a country. Concerning the foreign R&D spending the answer is not that clear, because different estimation techniques lead to different conclusions. Results for all different models can be found in Tables 6.2, 6.3, 6.4 and 6.5. Some results support Coe and Helpman's (1995) conclusion that the foreign R&D expenditures have a positive effect on the total factor productivity, some do not. Nevertheless if one takes the dynamic random coefficient model with a spatially lagged independent variable as the superior specification, the effect of foreign R&D expenditures seems to be existent.

Table 6.1: Moran's $\mathcal{I}$ for residuals of the Fixed Effects model with independent variable $logS_{it}$; for all periods $\mathrm{E}[\mathcal{I}] = -0.0476$ and $\mathrm{Var}[\mathcal{I}] = 0.0053$.

Period	Moran's $\mathcal{I}$	z($\mathcal{I}$)	Period	Moran's $\mathcal{I}$	z($\mathcal{I}$)
1990	0.0532	1.3818	1980	-0.0323	0.2101
1989	-0.1777	-1.7822	1979	-0.0860	-0.5263
1988	-0.1184	-0.9693	1978	-0.0426	-0.0694
1987	-0.0607	-0.1789	1977	0.0623	1.5064
1986	-0.0330	-0.1789	1976	-0.1337	-1.1788
1985	-0.0258	0.2993	1975	-0.0584	-0.1480
1984	-0.0702	-0.3088	1974	-0.0473	0.0042
1983	-0.0101	0.5144	1973	0.0198	0.9241
1982	0.0506	1.3451	1972	-0.0053	0.5795
1981	0.0910	1.8989	1971	-0.0328	0.2026

Table 6.2: Results for R&D Spillovers: Static Fixed Effects Model.

			Static Fixed Effects, LSDV			Static Fixed Effects, FGLS		
			Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Original Variables	ln(drd)	$\hat{\alpha}$	0,1362	0,0970		0,1383	0,0961	
		t-ratio	21,3317	10,6834		22,2154	10,5393	
		p-value	0,0000	0,0000		0,0000	0,0000	
	ln(frd)	$\hat{\alpha}$		0,0924			0,0956	
		t-ratio		5,8673			6,1200	
		p-value		0,0000			0,0000	
Taylor Series Approximation of ln	1+drd	$\hat{\alpha}$			0,0673			0,1410
		t-ratio			4,1483			6,1766
		p-value			0,0000			0,0000
	1+drd*V'	$\hat{\alpha}$			0,1787			-0,0498
		t-ratio			8,2235			-1,8678
		p-value			0,0000			0,0312
	Moran's $\mathcal{I}$	z($\mathcal{I}$)	0,2022	0,3613	0,2551	-0,0430	0,1409	-0,5056
	spatial p.	ρ				0,1369	0,1636	0,2279
	variance	σ^2				0,0025	0,0023	0,0019
	Model Fit	pseudo R^2	0,5218	0,5584	0,6240	0,5420	0,5799	0,2685
		pseudo adj.R^2	0,4966	0,5339	0,6032	0,5179	0,5566	0,2280

Table 6.3: Results for R&D Spillovers: Dynamic Fixed Effects Model.

			Dyn. Fixed Effects, LSDV			Dyn. Fixed Effects, FGLS		
			Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Original Variables	ln(drd)	$\hat{\alpha}$	0,1461	0,1078		0,2124	0,0667	
		t-ratio	17,1916	13,6515		8,6564	2,3232	
		p-value	0,0000	0,0000		0,0000	0,0104	
	ln(frd)	$\hat{\alpha}$		0,0464			0,3831	
		t-ratio		3,7133			8,6208	
		p-value		0,0000			0,0000	
Taylor Series Approximation of ln	1+drd	$\hat{\alpha}$			0,1887			0,0227
		t-ratio			27,2654			1,9333
		p-value			0,0000			0,0270
	1+drd*V'	$\hat{\alpha}$			0,0187			0,0800
		t-ratio			1,9464			5,9329
		p-value			0,0262			0,0000
	Moran's $\mathcal{I}$	z($\mathcal{I}$)	-0,5460	-0,3964	-0,6359	0,2580	0,4142	0,8376
	spatial p.	ρ				-0,2180	-0,5336	-0,1424
	variance	σ^2				0,0008	0,0005	0,0002
	Model Fit	pseudo R^2	0,8050	0,8758	0,9468	0,6533	0,8151	0,9001
		pseudo adj.R^2	0,7919	0,8671	0,9431	0,6301	0,8021	0,8931

Table 6.4: Results for R&D Spillovers: Static Random Coefficients Model

			Static Random Coeff., LSDV			Static Random Coeff., FGLS		
			Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Original Variables	ln(drd)	$\hat{\alpha}$	0,2443	0,2874		0,1826	0,2061	
		t-ratio	9,0446	7,3441		9,3238	6,9246	
		p-value	0,0000	0,0000		0,0000	0,0000	
	ln(frd)	$\hat{\alpha}$		-0,0603			-0,0046	
		t-ratio		-0,9155			-0,0949	
		p-value		0,1802			0,4622	
Taylor Series Approximation of ln	1+drd	$\hat{\alpha}$			-0,0205			0,1871
		t-ratio			-0,4279			3,3408
		p-value			0,3345			0,0005
	1+drd*V'	$\hat{\alpha}$			0,3787			-0,2104
		t-ratio			5,6590			-3,4582
		p-value			0,0000			0,0003
	Moran's $\mathcal{I}$	z($\mathcal{I}$)	0,4301	0,3630	0,4095	-0,2893	-0,2752	-0,3779
	spatial p.	ρ				0,4977	0,5042	0,3669
	variance	σ^2				0,0061	0,0075	0,0034
	Model Fit	pseudo R^2	0,9061	0,9135	0,9164	0,8792	0,8923	0,7054
		pseudo adj.R^2	0,9012	0,9087	0,9118	0,8728	0,8863	0,6891

Table 6.5: Results for R&D Spillovers: Dynamic Random Coefficients Model

			Dyn. Random Coeff., LSDV			Dyn. Random Coeff., FGLS		
			Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Original Variables	ln(drd)	$\hat{\alpha}$	0,2431	0,3529		0,1631	0,2522	
		t-ratio	9,1011	7,7946		7,4995	6,5971	
		p-value	0,0000	0,0000		0,0000	0,0000	
	ln(frd)	$\hat{\alpha}$		-0,0850			-0,0160	
		t-ratio		-1,1866			-0,2727	
		p-value		0,1181			0,3926	
Taylor Series Approximation of ln	1+drd	$\hat{\alpha}$			0,1252			0,0809
		t-ratio			2,2895			1,4394
		p-value			0,0113			0,0755
	1+drd*V'	$\hat{\alpha}$			0,1663			0,0161
		t-ratio			2,1853			0,2508
		p-value			0,0148			0,4011
	Moran's $\mathcal{I}$	z($\mathcal{I}$)	-0,1107	-0,1043	-0,1908	0,0566	-0,0600	0,1842
	spatial p.	ρ				0,3208	0,3754	0,7199
	variance	σ^2				0,0041	0,0071	0,0030
	Model Fit	pseudo R^2	0,9378	0,9736	0,9760	0,8963	0,9564	0,9603
		pseudo adj.R^2	0,9337	0,9717	0,9743	0,8894	0,9534	0,9575

Chapter 7

Optimal Design for Spatial Data

The field of optimal design is quite large and for standard cases (especially non-spatial problems) the theory of finding an optimal design is well established. In general the problem is the following: A random variable y is observed, and the distribution of y depends on some variables x controlled by the researcher, some parameters θ which are fixed and unknown, and some nuisance parameters ν also fixed and unknown. And the main question to be answered is: Which N observations should be included in the design, (Silvey, 1980). In principle the procedure is to search for the design that optimizes an adequate criterion function. In the theory of optimal design some paradox cases appear, e.g. it might happen that more observations worsen the criterion. Problems occur when deviations from the standard assumptions happen. A special case is - as always - the treatment of spatially correlated data.

The topic of this chapter is the application and adaption of optimal design theory to spatial datasets. The aim is to find optimal or nearly optimal designs for experiments to detect spatial dependence that might be in the data. The question to be answered is, how to optimally select predictor values to detect the spatial structure - if it is existent, and how to avoid to spuriously detect spatial dependence if there is no such structure. The starting point of

[3]The main parts of this chapter stem from an article from Gumprecht et al. (2007) which was submitted in January 2007.

this analysis are two different linear regression models: (1) an ordinary linear regression model with i.i.d. error terms - the non-spatial case, and (2) a regression model with a spatially autocorrelated error term, a so called spatial autoregressive error model (SAR error model), see (2.5). The procedure can be divided into two main parts: firstly, use of an exchange algorithm to find the optimal design for the respective data collection process; for its evaluation an artificial dataset was generated and used. Secondly, estimation of the parameters of the regression model and calculation of Moran's $\mathcal{I}$ which is used as an indicator for spatial dependence in the data set. The method is illustrated by applying it to a well-known case study in spatial analysis. Furthermore it is applied to the SO_2 monitoring network data from chapter 5 and to the R&D spillover data from chapter 6.

7.1 Motivation

When one is concerned with the analysis of spatial data, before all there is the desire to detect whether there is any spatial dependence in them or not. Should they be spatially independent, the respective statistical analysis usually reduces to the application of a classical and well established toolbox. Thus, the decision of whether one can confine oneself to this well understood body of knowledge or whether one has to resort to the rather freshly developed methodologies of spatial statistics (cf. Anselin, 1988 or Cressie, 1993) is a crucial element of any serious spatial investigation.

Besides the nature of the investigated process, what has the most influence on the ability to isolate spatial effects are the locations in space, where the data are collected, the so called spatial sampling design. There have been made considerable efforts to make this design as efficient as possible for the purpose of confirmatory spatial analysis, see e.g. Müller (2001). However, it seems a little like negligence that this has never been considered for the very first phase of a spatial study.

One explanation for this is that very frequently the sampling design is fixed beforehand. The spatial data comes from a predefined lattice of locations or a given number of contiguous areas. Usually this data comes at no or little cost at all design points/regions and thus there is no need for posing the question: where (to measure)? However, this is not always the case, since one can easily imagine that the data may come only at considerable costs and the decision of which data to collect can be of great relevance. Moreover,

it is well known that in spatial analysis it can be sometimes an advantage not to employ the full potential data set (cf. the well known Smit's paradox in Smit, 1961).

The idea of this chapter is exemplified on a well-known case, namely the Columbus, Ohio crime study from Anselin (1988), which became a classic testground for spatial analysis. Data of the Columbus, Ohio crime dataset stem from 49 contiguous planning neighbourhoods in Columbus, Ohio, USA, see Figure 7.1. The dependent variable is an index of criminal activity, it includes residential burglaries and vehicle thefts per thousand households in a region, the explanatory variables are household income and housing values in thousand dollars. The Moran's plot for the dependent variable is given in chapter 3, Figure 3.1. Although the example does not fit well for the practical purposes of this topic, as the data are freely available for all the neighbourhoods, it was chosen for its familiarity amongst the 'spatial community'. It is evident that one can easily replace the crime index by another characteristic that may only be measured at a high cost and all the considerations will continue to hold.

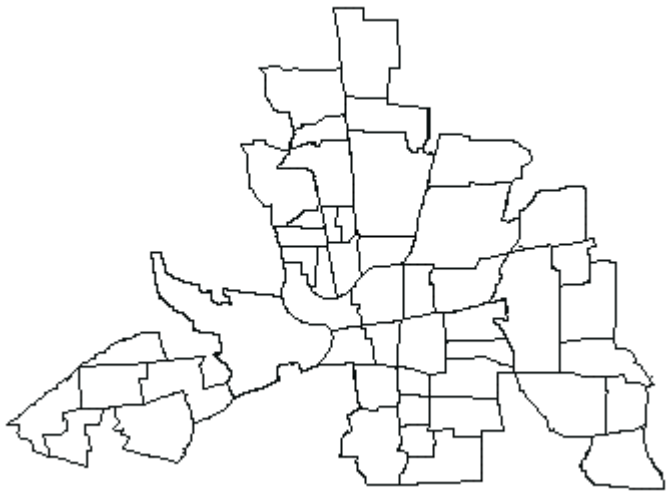


Figure 7.1: Neighbourhoods in Columbus, Ohio

As a measure for the intensity of the spatial dependence and consequently a valid test statistic for detecting its potential existence the Moran's $\mathcal{I}$ statistic (3.1) was chosen, although there is a considerable number of alternatives available. However, the principal considerations are not affected by this choice.

7.1.1 Spatial Link Matrices

In spatial modeling the sampling design primarily affects the spatial link matrices (or spatial weighting matrices), which represent the spatial relationships between the observations, see section 2.1 on page 5 for more details.
Two different types of spatial link matrices are used in the following. First, use a function of the Euclidean distances between the locations of the observations, cf. formula (2.1),

$$g_{ij} = e^{-\delta d_{ij}} - \mathbf{1}_{\{i=j\}} \tag{7.1}$$

with $d_{ij} = \|s_i - s_j\|$, where s_i and s_j $(i, j = 1, ..., n)$ are the coordinates of the locations, δ is some decay parameter, and $\mathbf{1}_{\{i=j\}}$ is an indicator function for $i = j$, it is included to obtain zeros on the main diagonal. By $\boldsymbol{\xi} = \{s_1, \ldots, s_n\}$ the collection of coordinates, the sampling design, defined on a design space given by the set $\mathcal{S}$, is denoted.
The other approach is the use of neighbourhood (contiguity) matrices, based on the Queen's criterion, cf. page 6.
For all computations the row-standardized versions of the spatial link matrices, described in section 2.1.3, are used.

The idea of choosing the spatial weight matrix was differently used by Kooijman (1976). He maximized Moran's $\mathcal{I}$ by choosing an appropriate spatial link matrix $\mathbf{V}$ (under certain constraints), to increase the robustness of the test. In contrast to Kooijman (1976), the aim is to find the optimal locations of the observations, it is not, to find a more robust test, but to better detect a spatial effect that is potentially present in the data.

7.1.2 Models

The intention is to estimate an ordinary linear model and use the residuals for the test of spatial dependence, i.e. estimation of the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

under the assumption $\boldsymbol{\varepsilon}$ i.i.d. The real data generating process, the true but unknown status of the world, is one of the following:

H_0, spaceless: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}$ i.i.d. (ordinary linear model)

H_A, spatial: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ and $\mathbf{u} = \rho\mathbf{V}\mathbf{u} + \boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}$ i.i.d. (SAR error model, cf. 2.5)

where $\mathbf{y}$ is an $n \times 1$ vector of the depending variable, $\mathbf{X}$ is an $n \times k$ matrix of the regressors (which may also depend upon $\boldsymbol{\xi}$), $\boldsymbol{\beta}$ is the $k \times 1$ parameter vector, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of i.i.d. errors, $\mathbf{u}$ is an $n \times 1$ vector of spatially correlated errors, ρ is the spatial autocorrelation parameter, and $\mathbf{V}$ which depends upon $\boldsymbol{\xi}$ is the $n \times n$ spatial weight matrix.

Depending on the two examined cases, one either wants to accept or reject the null hypothesis of spatial independence of Moran's $\mathcal{I}$ test (see section 3.1.2) to make a correct decision. The aim is to find an optimal or nearly optimal design for a test strategy to receive either acceptation or rejection of the null hypothesis for derivation of a model that matches the real status of the world.

The following discussion is restricted to Gaussian spatial processes, which are based on normally distributed regression disturbances, see page 12. A Gaussian spatial process is parameterized by the expected values of the observations $\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ and their mutual covariance matrix, which denotes the spatial interaction between the objects $\mathrm{Cov}(\mathbf{y}\mathbf{y}') = \mathrm{E}(\mathbf{y}\mathbf{y}') = \boldsymbol{\Omega}(\rho)$, it depends on the spatial autocorrelation parameter ρ. The disturbances are $N(\mathbf{0}, \boldsymbol{\Omega}(\rho))$ distributed, see e.g. Tiefelsdorf (2000).

7.2 Moran's $\mathcal{I}$

General issues about Moran's $\mathcal{I}$ are given in section 3.1. As already mentioned, for a standard regression model it is crucial to know whether the residuals are spatially dependent or not. If there is no spatial dependence in the residuals, one can use standard estimation methods, like OLS, but if the residuals show spatial dependence, one has to use special methods like the S2SLS or the ML estimation technique (cf. section 5.2), because spatial autocorrelation in the error term leads to biased estimates of the residual variance and inefficient estimates of the regression coefficients when the OLS

estimation method is applied, see e.g. Cliff and Ord (1981). Moran's $\mathcal{I}$ (3.1) of the OLS regression residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n)'$ is given by

$$\mathcal{I} = \frac{\hat{\boldsymbol{\varepsilon}}' \frac{1}{2}(\mathbf{V} + \mathbf{V}')\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} \ ,$$

where $\mathbf{V}$ is a standardized spatial weight matrix ($\sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij} = n$), see e.g. Tiefelsdorf (2000).

The Moran's $\mathcal{I}$ test is here used to test positive spatial autocorrelation which is the much more relevant case in real life ($H_0 : \rho = 0$ against $H_A : \rho > 0$). Thus, from now on $\rho \geq 0$ will be assumed. The z-transformed Moran's $\mathcal{I}$, given in (3.4), is for normally distributed regression residuals and well-behaved spatial link matrices under certain regularity conditions (see e.g. Tiefelsdorf, 2000) asymptotically standard normally distributed,

$$z(\mathcal{I}) = \frac{\mathcal{I} - \mathrm{E}[\mathcal{I}|H_0]}{\sqrt{\mathrm{Var}[\mathcal{I}|H_0]}} \sim N(0, 1).$$

The exact small sample distribution of Moran's $\mathcal{I}$ was obtained by Tiefelsdorf and Boots (1995), but is not used here as it would be a restrictive computational burden on the algorithm.

7.2.1 Status of the World: Spaceless

In the first case of section 7.1.2 a model under the assumption of spatial independence is estimated, and the true model is of the same form. The aim is then to accept the null hypothesis (=spatial independence). For the approximate test the moments of Moran's $\mathcal{I}$ under the null are required. The moments, are given in section 3.1. Under the assumption of spatial independence, expected value (3.2) and variance (3.3) of $\mathcal{I}$ under the null are:

$$\mathrm{E}[\mathcal{I} \mid H_0] = \frac{\mathrm{tr}(\mathbf{K})}{n-k}$$

and

$$\mathrm{Var}[\mathcal{I} \mid H_0] = \frac{2\{(n-k)\mathrm{tr}(\mathbf{K}^2) - \mathrm{tr}(\mathbf{K})^2\}}{(n-k)^2(n-k+2)}$$

respectively, see Henshaw (1966).
The null case is the simpler one, there is no spatial effect in the data, data

follow an ordinary linear model, the correct model is estimated and the null hypothesis of no spatial dependence should be accepted. The intention is to find an optimal design which gives the best locations for the observations in the sense that the rejection of the null hypothesis is minimized.

7.2.2 Status of the World: Spatial Dependence

Under the alternative the (wrongly) estimated model is still: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}$ i.i.d. but now the true assumed (but unknown) data generating process is a SAR error process (2.5):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbf{u} = \rho\mathbf{V}\mathbf{u} + \boldsymbol{\varepsilon}$$

with $\boldsymbol{\varepsilon}$ i.i.d. Here the spatial dependence appears in the form of a spatially lagged error term $\mathbf{u}$. This is the SAR error model, the parameter ρ is a spatial autoregressive coefficient. This model can be transformed into a form with i.i.d. error terms, $\mathbf{y} = \rho\mathbf{V}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \rho\mathbf{V}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, being an exposition with a spatially lagged dependent variable $\mathbf{V}\mathbf{y}$ and a set of spatially lagged exogenous variables $\mathbf{V}\mathbf{X}$. The variance-covariance matrix $\boldsymbol{\Omega}(\rho)$ of the error terms, given in (2.7) on page 13, is

$$\boldsymbol{\Omega}(\rho) = \mathrm{E}[\mathbf{u}\mathbf{u}'] = \sigma^2[(\mathbf{I} - \rho\mathbf{V})'(\mathbf{I} - \rho\mathbf{V})]^{-1}.$$

The model is estimated via OLS and the residuals $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ are used for the calculation of Moran's $\mathcal{I}$. If the real data generating process follows a SAR error process, the aim is to reject the null hypothesis of no spatial dependence. The task is to maximize the power of the test, i.e. the probability to reject the null hypothesis given the alternative (spatial dependence). For the normal approximation again only the conditional moments are needed. The conditional expectation of Moran's $\mathcal{I}$ can be evaluated by the improper integral

$$\mathrm{E}[\mathcal{I}|H_A] = \int_0^\infty \prod_{i=1}^{n-k}(1 + 2\lambda_i t)^{-\frac{1}{2}} \cdot \sum_{i=1}^{n-k} \frac{h_{ii}^*}{1 + 2\lambda_i t}\, dt\ , \tag{7.2}$$

where h_{ii}^* are the diagonal elements of matrix $\mathbf{H} = \mathbf{P}'\mathbf{A}\mathbf{P}$ with $\mathbf{A} = \boldsymbol{\Omega}'^{\frac{1}{2}}\mathbf{M}\frac{1}{2}(\mathbf{V}+\mathbf{V}')\mathbf{M}\boldsymbol{\Omega}^{\frac{1}{2}}$, and $\mathbf{P}$ is the matrix of the normalized eigenvectors of matrix $\mathbf{B} = \boldsymbol{\Omega}'^{\frac{1}{2}}\mathbf{M}\boldsymbol{\Omega}^{\frac{1}{2}}$, the eigenvalues and their associated eigenvectors are resequenced so that $0 < \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{n-k}$. The variance of $\mathcal{I}$ under

the alternative is given by

$$\mathrm{Var}[\mathcal{I}|H_A] = \mathrm{E}[\mathcal{I}^2|H_A] - \mathrm{E}[\mathcal{I}|H_A]^2, \tag{7.3}$$

where

$$\mathrm{E}[\mathcal{I}^2|H_A] = \int_0^\infty \left[\prod_{i=1}^{n-k}(1+2\lambda_i t)^{-\frac{1}{2}}\right] \cdot \left[\sum_{i=1}^{n-k}\sum_{j=1}^{n-k} \frac{h_{ii}^* h_{jj}^* + 2(h_{ij}^*)^2}{(1+2\lambda_i t)(1+2\lambda_j t)}\right] t \; dt \tag{7.4}$$

and $\mathrm{E}[\mathcal{I}|H_A]$ is given in equation (7.2). The upper truncation points for the integrals can be approximated by a formula given by De Gooijer (1980). Following him leads to an approximation of the upper bound for the expected value (7.2) of

$$\left[\frac{(n-k)h_{max}}{2\lambda_1^{\frac{n-k}{2}}}\left(\frac{n-k}{2}-1\right)\frac{1}{\epsilon}\right]^{\frac{1}{\frac{n-k}{2}-1}} = \tau_1, \tag{7.5}$$

where h_{max} is the biggest absolute value of the elements of the diagonal of matrix $\mathbf{H}$. An approximation of the upper bound for $\mathrm{E}[\mathcal{I}^2|H_A]$, (7.4), is

$$\left[\frac{3(n-k)^2 h_{max}^{(2)}}{(2\lambda_1)^{\frac{n-k}{2}}}\left(\frac{n-k}{2}-2\right)\frac{1}{\epsilon}\right]^{\frac{1}{\frac{n-k}{2}-2}} = \tau_2, \tag{7.6}$$

with $h_{max}^{(2)}$ denoting the biggest absolute value of the elements of matrix $\mathbf{H}$. Tiefelsdorf (2000) suggests to use $\frac{1}{n-k}\sum_{i=1}^{n-k}\lambda_i$ instead of λ_1.

The calculations of (7.2) and (7.4) are based on a GAUSS-code implemented by M. Tiefelsdorf and the results were checked with a code in Mathematica programm implemented by J. Rodríguez-Díaz.

7.3 Optimal Design Considerations

7.3.1 A Criterion

In both cases, where a linear regression model is estimated and the corresponding residuals are used to calculate Moran's $\mathcal{I}$ test, the aim, whether to accept or reject the null hypothesis of no spatial autocorrelation in the

error term, depends on the true data generating process. As the true process is unknown, a general design criterion Ψ (which does not depend on the knowledge of the true data generating process), is needed. The aim is to minimize the probability that, given the alternative, the Moran's $\mathcal{I}$ test accepts the null hypothesis of no spatial autocorrelation. The test statistic $Z = \frac{\mathcal{I}-\mathrm{E}(\mathcal{I}|H_0)}{\sqrt{\mathrm{Var}(\mathcal{I}|H_0)}}$ is asymptotically normally distributed, and therefore the aim is:

$$\min_{H_A} \mathrm{P}\left(\frac{\mathcal{I}-\mathrm{E}(\mathcal{I}|H_0)}{\sqrt{\mathrm{Var}(\mathcal{I}|H_0)}} \leq \Phi^{-1}(1-\alpha)\right).$$

This leads to

$$\min_{H_A} \mathrm{P}\left(\mathcal{I} \leq \Phi^{-1}(1-\alpha)\sqrt{\mathrm{Var}(\mathcal{I}|H_0)} + \mathrm{E}(\mathcal{I}|H_0)\right).$$

Using the z-transformation for $\mathcal{I}$ under the alternative gives $\frac{\mathcal{I}-\mathrm{E}[\mathcal{I}|H_A]}{\sqrt{\mathrm{Var}[\mathcal{I}|H_A]}}$ which is also asymptotically standard normal distributed. The final criterion to be maximized is therefore given by

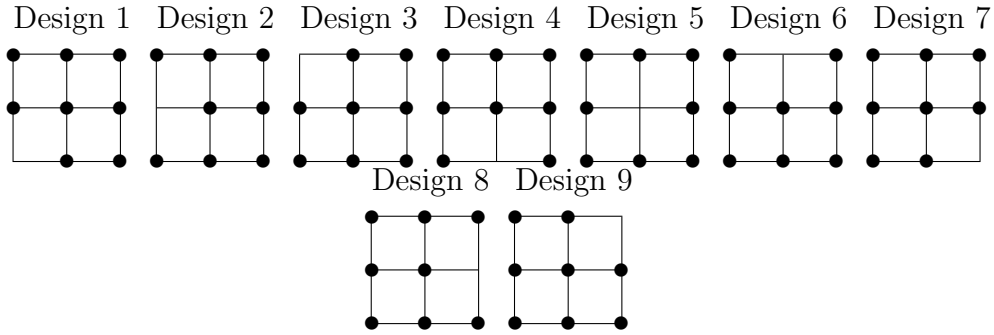
$$\Psi(\boldsymbol{\xi}) = 1 - \Phi\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{\mathrm{Var}[\mathcal{I}|H_0]} + \mathrm{E}[\mathcal{I}|H_0] - \mathrm{E}[\mathcal{I}|H_A]}{\sqrt{\mathrm{Var}[\mathcal{I}|H_A]}}\right), \qquad (7.7)$$

where Φ denotes the cdf of the standard normal distribution. The maximization of Ψ over $\boldsymbol{\xi} \in \mathcal{S}$ gives the final optimal locations of the observations and thus maximizes the power of the Moran's $\mathcal{I}$ test. To calculate Ψ, the expected value (3.2) and the variance (3.3) of $\mathcal{I}$ under the null hypothesis, and the expected value (7.2) and the variance (7.3) of $\mathcal{I}$ under the alternative hypothesis are needed. For the calculation of $\mathrm{E}[\mathcal{I}|H_A]$ and $\mathrm{Var}[\mathcal{I}|H_A]$ one has to assume a particular spatial process.
Unfortunately the given criterion is not convex and thus the well developed optimum design theory (cf. Silvey, 1980) can not be employed and algorithmic approaches are needed.

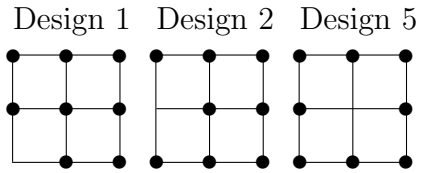
7.3.2 Design Algorithms

Full enumeration
Evidently, the global optimal design can be found by evaluating all possible designs, i.e. in an m-point grid there are $\binom{m}{r}$ possible r-point designs, r goes from $4 + k + 1$ to m, where k is the number of the regressors in the model.

This minimum number of points in a design follows from the approximation of the upper truncation points for the integrals (7.5) and (7.6). The number of possible designs increases very fast with the size of the grid. This leads to a high runtime, as the numerical integration needs some time. From this point of view it is worth to notice that not all possible designs are different in the sense that they have different criterion values. Some of the r-point designs are only rotations, reflections or translations of other r-point designs, and therefore give the same value of the criterion Ψ. The respective designs will be called 'symmetric' in the following. To avoid calculating Ψ for those designs which are known to be symmetric to others, an appropriate symmetry check can be done before the computation of Ψ. From formula (7.2) and (7.4) it can be seen, that designs give the same Ψ if the absolute value of the elements of the lower triangular matrix of $\mathbf{H}$, and vector $\boldsymbol{\lambda}$ are the same. For illustrating this problem a regular 9-point grid is assumed, and the model is a regression on the intercept. The number of all possible 8-point designs is $\binom{9}{8} = 9$, they are illustrated here:



Some of them are rotations or reflections of others, really different, in the sense that they lead to different criterion values Ψ, are only the following three designs. All others have the same values $|h_{ij}|$ and $\boldsymbol{\lambda}$ and therefore the same Ψ like one of these three:



For evaluating Ψ two integrals are needed, one for the expected value (7.2) and an additional one for the variance (7.4) under the alternative. Ignoring

symmetric designs means: there is only need to compute $2 \cdot 3 = 6$ numerical integrals instead of $2 \cdot 9 = 18$.
The implementation of this symmetry check improves the runtime of the algorithm as the calculations of the numeric integrals (7.2) and (7.3) take quite a long time. A further advantage is, that the number of the 'really' different designs, different in the sense of non-symmetric, can be counted. A disadvantage is the high memory capacity needed for the symmetry check. Nevertheless, the number of non-symmetric designs, that have to be evaluated, becomes large if the number of points in the grid increases, e.g. in an intercept regression model on a 25-point grid there are 1081575 different 17-point designs and still 108963 are non-symmetric. The complete evaluation of all 'really' different designs can only be done for very small grids and therefore is not relevant for practical use.

Simple search algorithm
A possibility for finding a 'nearly' optimal design is the use of a simple search algorithm. This algorithm is much faster than the full enumeration algorithm as for the r-point design the number of evaluated $(r-1)$-point designs is r. This algorithm can also be done in an acceptable time for quite large grids. The procedure is quite simple:

1. Start with an initial design $\boldsymbol{\xi}_0 = \mathcal{S}$, called 'base' design, and compute Ψ_0. Thus in the first iteration the number of points r in $\boldsymbol{\xi}_0$ is m (= full design).

2. Delete each point, one at a time, to get $(r-1)$ designs $\boldsymbol{\xi}_e$, and compute Ψ_e. The symmetries can be checked before the criterion is calculated.

3. Take the best $(r-1)$ design $\boldsymbol{\xi}_e$, i.e. the design with the largest Ψ_e, and put it as new base design.

   Go to step 2.

The algorithm stops if $r = (4 + k + 1)$. The r-point design that gives the largest Ψ is the 'nearly' optimal one. The disadvantage of this algorithm is, that once an r-point design is chosen, all smaller $r - i$ point designs are restricted to this set of points, it can happen quite easily that one is trapped in a local maximum. To avoid this one could employ alternative methods of stochastic optimization such as in Haines (1987).

Fedorov exchange algorithm

As an alternative and sort of compromise, an exchange type algorithm based on Fedorov (1972) can be used. The 'nearly' optimal r-point design, when equal points in the design are not allowed, is found via an exchange type procedure. The aim is to improve the design by exchanging points from it, one at a time, as follows:

1. Start with an initial r-point design, $\boldsymbol{\xi}_0 = \{s_1, \ldots s_r\}$, the points are chosen at random and should be different. Compute the design criterion Ψ_0 for the initial design.

2. Take one point s_i from $\boldsymbol{\xi}_0$ (it is again called 'base' design) and exchange it with a point not in $\boldsymbol{\xi}_0$ - these points are called candidate points, the set of all candidate points is $\boldsymbol{\xi}_c = \{\mathcal{S}|\boldsymbol{\xi}_0\} = \{s_{r+1}, \ldots s_m\}$. Do this for all candidate points in $\boldsymbol{\xi}_c$ and all points in the base design $\boldsymbol{\xi}_0$ and compute Ψ_e for each different combination (design). Before the criterion is computed, the symmetry check based on $\mathbf{H}$ and $\boldsymbol{\lambda}$ can be done.

3. Get the best r-point design ($\boldsymbol{\xi}_e$), i.e. the design with the largest Ψ_e, from the previous exchange step and put it as new base design $\boldsymbol{\xi}_0$.

   Go to step 2.

The algorithm stops if there is no further improvement in the criterion, i.e. if Ψ_e is worse than Ψ of the base design. In this way 'nearly' optimal r-point designs are computed for $r = 4 + k + 1, ..., m$, the overall best design is the best one of all r-point designs found by the algorithm. A refinement of this algorithm, which could be useful also in this context is the so-called coordinate exchange algorithm by Meyer and Nachtsheim (1995).

Algorithms that evaluate many different designs, like the ones given here, which are chosen by random and/or via exchanging points, will most probably also lead to designs with far-off objects (see the discussion in chapter 4). In case of designs with far-off observations, excluding the far-off object, i.e. use treatment (e), can make an algorithm much faster, designs which include far-off objects do not have to be evaluated because they give the same criterion value as the ones without this far-off point. As this treatment is

not recommended, treatment (s), which leads to a more conservative test, is used in the following.

7.4 Examples

7.4.1 Artificial Dataset

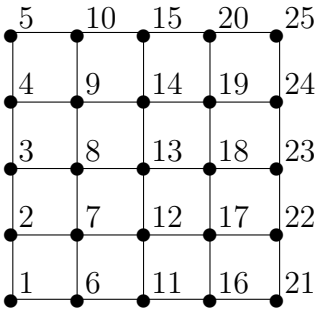
An OLS regression model with i.i.d. error terms is estimated. The OLS residuals are used to calculate Moran's $\mathcal{I}$ and its expected values and variances under the null and under the alternative, see (3.1), (3.2), (3.3), (7.2) and (7.3), which are needed for evaluating the design criterion Ψ (7.7). The best design is the one with the largest Ψ:

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}\in\mathcal{S}} \Psi(\boldsymbol{\xi}). \tag{7.8}$$

Since Φ is a monotonous function one needs to minimize its argument only, for computational simplicity. The observations are taken on a regular 25-point grid $[-1;1]^2$. The simple search and the Fedorov exchange algorithm, described in the previous section, are used. The null hypothesis is spatial independence, the alternative hypothesis is spatial dependence with a spatial autoregressive parameter $\rho = 0.5$.

Regression on Intercept

The considered model is a regression of $\mathbf{y}$ on an intercept: $\mathbf{y} = \mathbf{1}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. It is assumed that all observations derive from different locations. The full 25-point design (with numbering of points) is simply:



The spatial link matrix $\mathbf{V_c}$ is a row-standardized contiguity matrix (based on Queen's criterion, see chapter 2), neighbours of point number i are given

in row i, the spatial link matrix $\mathbf{V}_d$ is a row-standardized distance matrix based on (7.1) with parameter $\delta = 5.76$. This setting gives the same criterion value for the full design as in case of using $\mathbf{V}_c$. The correlation structure with corresponding exponential function is displayed in Figure 7.2.

$$\mathbf{V}_c = \begin{pmatrix} 0 & 0.33 & 0 & \cdots & 0 \\ 0.20 & 0 & 0.20 & \cdots & 0 \\ 0 & 0.20 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\mathbf{V}_d = \begin{pmatrix} 0 & 0.402 & 0.023 & 0.001 & \cdots & 0.000 \\ 0.262 & 0 & 0.262 & 0.015 & \cdots & 0.000 \\ 0.014 & 0.256 & 0 & 0.256 & \cdots & 0.000 \\ 0.001 & 0.015 & 0.262 & 0 & \cdots & 0.000 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.000 & 0.000 & 0.000 & 0.000 & \cdots & 0 \end{pmatrix}$$

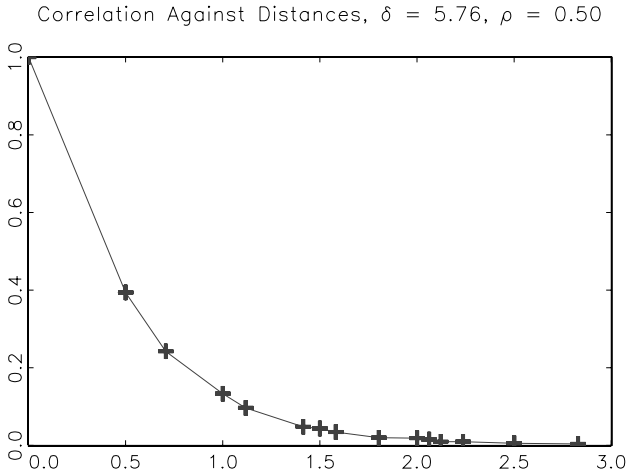
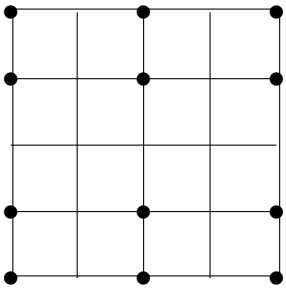


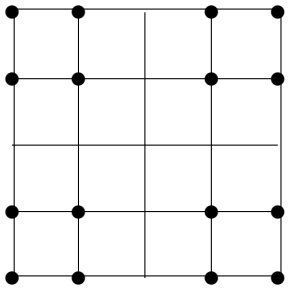
Figure 7.2: Correlation for a 25 point grid

Simple Search Algorithm: Executing the Simple Search algorithm gives, for a distance based spatial link matrix, a 16-point design with four points in each corner as the best one ($\Psi = 0.603$). Using a neighbourhood matrix leads to the same 'optimal' design (with $\Psi = 0.659$).

Fedorov exchange algorithm: Running the Fedorov exchange algorithm for this example finds the same best design when the spatial link matrix is based on a distance matrix ($\Psi = 0.603$). For the neighbourhood-based spatial link matrix a 12-point design is the best one ($\Psi = 0.720$):



Contiguity
12 points
$E_0 = -0.091$
$Var_0 = 0.153$
$E_A = 0.692$
$Var_A = 0.058$
$\Psi = 0.720$



Distance
16 points
$E_0 = -0.067$
$Var_0 = 0.037$
$E_A = 0.318$
$Var_A = 0.067$
$\Psi = 0.603$

The development of the 'nearly' optimal designs found by the search- and the exchange algorithm can be seen in Figure 7.3.

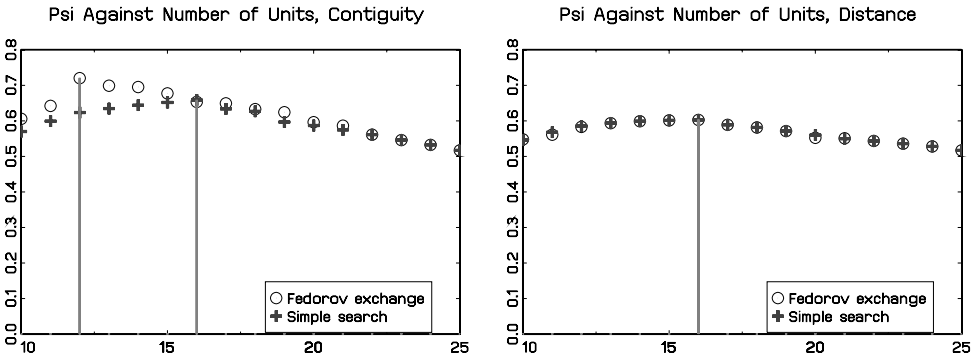


Figure 7.3: 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices

Linear Trend Model

Now the considered model is a regression of $\mathbf{y}$ on an intercept and on the horizontal s_1- and vertical s_2 coordinates of the observations: $\mathbf{y} = \mathbf{1}_n\boldsymbol{\beta}_0 + \mathbf{s}_1\boldsymbol{\beta}_1 + \mathbf{s}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. For this example again the artificial dataset is used and the simple search and the Fedorov exchange algorithm are executed for both a distance based and a neighbourhood based spatial link matrix with parameters $\rho = 0.5$ and for the distance based link matrix parameter $\delta = 0.543$

respectively. Here the best designs (12 points) coincide, the criteria are different, for the contiguity matrix $\Psi = 0.625$ whereas for the distance matrix $\Psi = 0.462$.

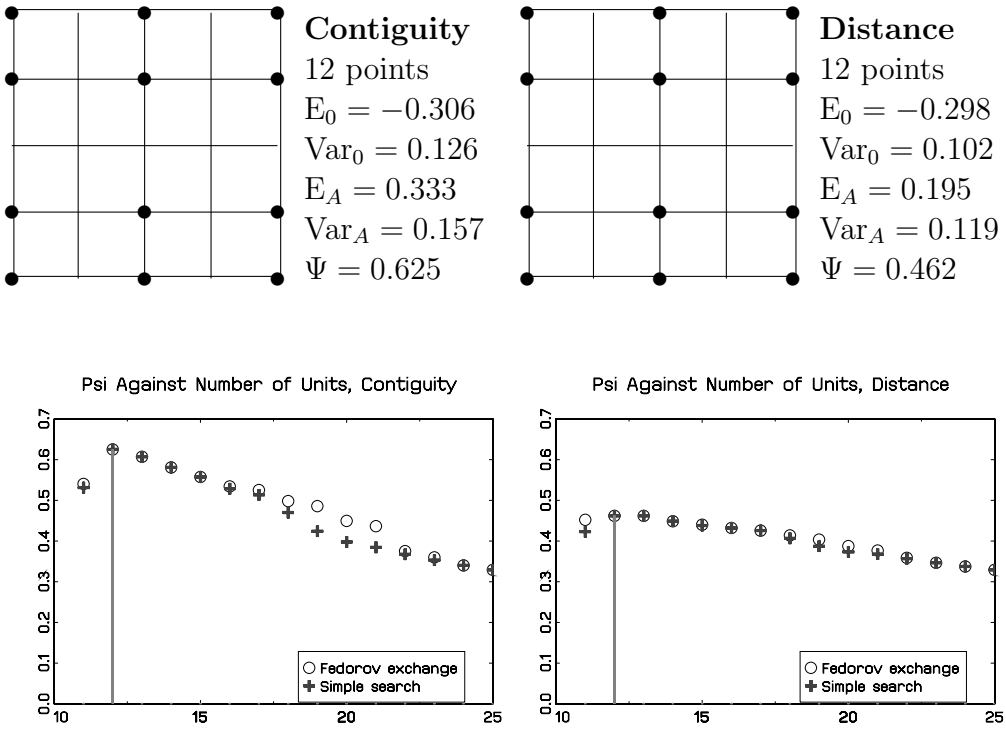


Figure 7.4: 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices

7.4.2 Columbus Crime Data

Data for this example stem from the classical Columbus Crime dataset from Anselin (1988), see section 7.1 and Figure 7.1. The spatial weight matrix $\mathbf{V}$ is the row-standardized neighbourhood matrix, and the spatial autoregressive parameter $\rho = 0.562$, this value is the Maximum Likelihood estimator of a linear regression model with an intercept, the two regressors and a spatially dependent error term, for the estimation the contiguity matrix was used, see Anselin (1988). The dependent variable 'crime' is spatially autocorrelated with Moran's $\mathcal{I}$ of 0.5109 which is significant with $z(\mathcal{I}) = 5.675$.

The optimal design search is based on the regression model with only an intercept. The idea behind this approach is, that of course normally one looks for the design first, and then data is collected on the corresponding locations, i.e. one does not know the values of the regressors in the design generating process. Running the Fedorov exchange algorithm gives a 'best' design with 29 locations with a criterion value $\Psi = 0.983$, and a Moran's $\mathcal{I} = 0.417$ and $z(\mathcal{I}) = 1.914$ which is significant on the 5%-level, see Figure 7.5. The 'best' design found by the simple search algorithm is one with 31 locations with $\Psi = 0.973$, and $\mathcal{I} = 0.519$ with $z(\mathcal{I}) = 2.705$, which is significant on the 1%-level, see Figure 7.5, the dark grey locations are the ones which were selected by both algorithms. It is remarkable that the border regions are included in both cases. The improvement over the full design is for both the Fedorov- and the simple search algorithm 22%. The values for design criteria Ψ for all different numbers of locations can be found in Figure 7.6.

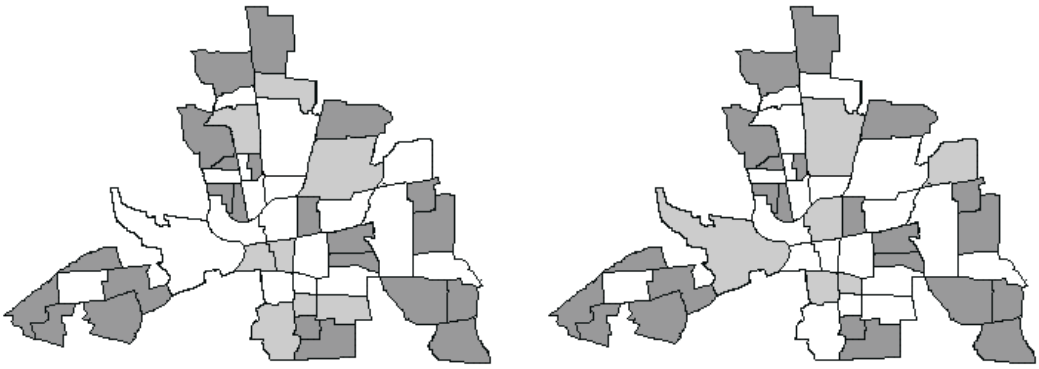


Figure 7.5: Left: Simple search. Right: Fedorov exchange algorithm.

7.4.3 Upper-Austria SO_2 Monitoring Network

This example deals with the Upper-Austria SO_2 monitoring network from section 5.3. The full design are the 17 observation sites presented in Figure 5.1. The estimated parameters of the spherical variogram (5.9), $\hat{\theta} = (0.164, 0.323, 24.25)'$, are used to build a spatial weight matrix $\mathbf{U}$, constructed via the inverse function. Using this spatial dependence structure leads to

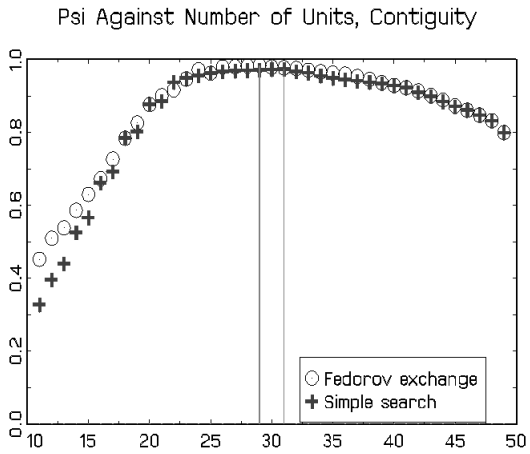


Figure 7.6: Columbus crime data, simple search & exchange algorithm, contiguity matrix

three far-off locations: Steyr, Braunau and Schöneben, these observations are zero-weighted because their distances to all other objects are greater than $\hat{\theta}_3$. For these objects treatment (s) as well as treatment (ν) with different values of ν (0.01, 0.001, 0.0001 and 0.00001) are used, see chapter 4. In the search- and exchange algorithms the corresponding row-standardized spatial link matrices $\mathbf{V}$ are used. The regression model of the optimal design procedure is an intercept only model. As an estimate of the spatial autoregressive parameter ρ, which is needed for the alternative, GM estimates are used (see section 5.2). Depending on the treatment of the far-off objects, estimates $\hat{\rho}$ and the criterion values Ψ differ, treatment (s) gives $\hat{\rho} = 0.2437$ and $\Psi = 0.2510$, treatment (ν) gives for $\nu = 0.01, 0.001, 0.0001, 0.00001$ values of $\hat{\rho}$ equal to $0.6625, 0.4409, 0.3825$ and 0.3753 and criterion values Ψ equal to $0.4451, 0.2634, 0.2225$ and 0.2177. Nevertheless, the optimal design is always the same 10-point design, in Figure 7.7 these points are the filled black ones. The best locations are: Linz-Hauserhof, Wels, Vöcklabruck, Steyr*, Braunau*, Linz-Kleinmünchen, Linz-ORF-Zentrum, Linz-Berufsschulzentrum, Lenzing and Schöneben*, the *-observations are far-off ones. Here it can be seen that treatment (e) should not be used as it would exclude these objects, even they are part of the optimal design.

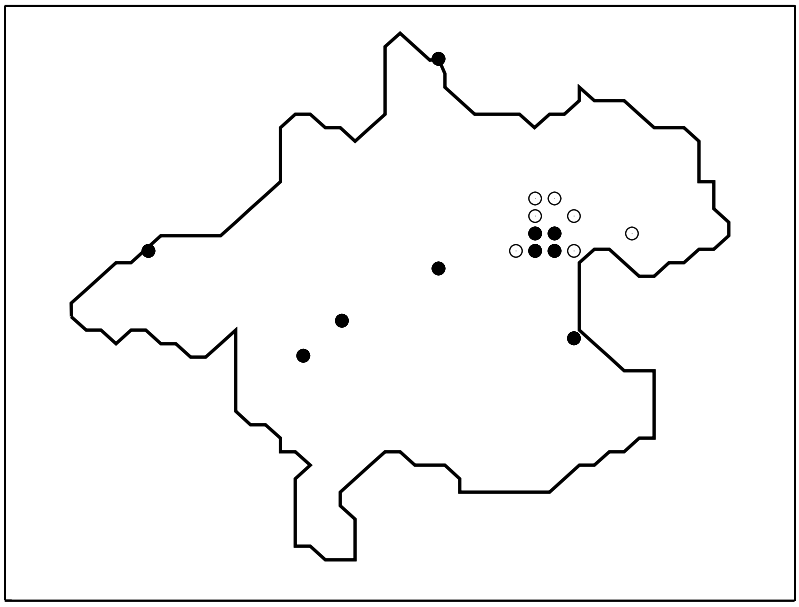


Figure 7.7: Upper-Austria SO_2 monitoring network

7.4.4 R&D Spillovers

The last example concerns again the R&D Spillover dataset from Coe and Helpman (1995), see also chapter 6. For the optimal design procedure not the whole panel dataset but only year 1990 was regarded. The spatial weight matrix is the row-standardized bilateral import-shares matrix in year 1990. The final design depends on the assumed value of ρ, i.e. the concrete form of the alternative, see section 7.3.1. To see the influence of the choice of ρ, two different values ($\rho = 0.1$ and $\rho = 0.5$) are assumed, and for both of them the optimal design is calculated with the Fedorov exchange algorithm as well as the simple search algorithm. For $\rho = 0.5$ Fedorov exchange- and simple search algorithm give the same optimal design with a criterion value $\Psi = 0.334$. It includes the following 13 countries: USA, Canada, Australia, Austria, Denmark, Finland, Israel, New Zealand, Norway, Portugal, Spain, Sweden and Switzerland. Figure 7.8 shows again the 'economic' landscape (Figure 6.1) and the optimal design countries are singled out.

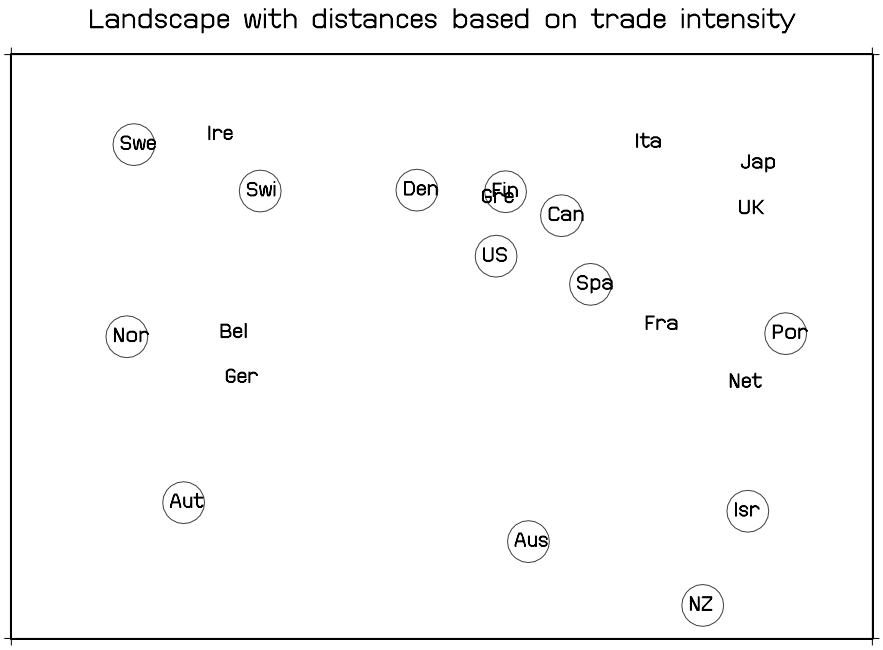


Figure 7.8: Economic landscape with optimal design points

For assumption $\rho = 0.1$, Fedorov exchange- and simple search algorithm lead to different designs, although the number of points is in both design equal to nine. The exchange type algorithm suggests to use USA, Canada, Australia, Denmark, New Zealand, Norway, Portugal, Spain and Sweden, the criterion value Ψ is equal to 0.082. The simple search algorithm leads to the countries USA, Canada, Australia, Belgium, Denmark, The Netherlands, New Zealand, Norway and Sweden, here the criterion is a bit smaller, $\Psi = 0.081$.

7.5 Findings

This new course of action combines the fields of optimal design theory and spatial analysis (via the design criterion). It helps to select the best locations for an empirical analysis of spatial data, especially if the data collecting process is expensive and/or time-demanding, and there is no or little knowledge about a potential spatial dependence. Using this procedure can lead not only to more economic but also more efficient networks.

Chapter 8

Conclusions

In this thesis different aspects and methods for analysing spatial data are shown. The great generality of spatial methods is illustrated by applying some of the techniques to various examples that stem from the field of geostatistics as well as spatial econometric, and also to one dataset which is not obviously spatial in the first view. This 'non-spatial' problem is about R&D spillovers, which is a well known and often analysed and re-analysed topic in economics. The spatial approach supports the theory that R&D spillover effects exist. Nevertheless there are still questions and aspects left open, e.g. whether one should use other models or better definitions of the variables.

This work also included a theoretical discussion of a special class of spatial datasets, namely ones with observations that are far apart from all others ('far-off' objects). Three possible methods of treating such observations are suggested and the influence on Moran's $\mathcal{I}$ test is worked out. A straightforward extension of this discussion would be to analyse the effect of more than one far-off object, and the influence of such objects on other spatial autocorrelation measures e.g. the Getis statistic.

The last part of the thesis deals with optimal design theory. Here a new criterion, especially useful for spatial datasets is presented. It is an instrument which helps to collect data at the 'best' locations, best in the sense that the true and unknown data generating process can be detected in the sample design. Due to the generality of this approach there are many directions for improvements available. Evidently one can try other algorithmic approaches to decrease the runtime or the memory capacity needed. An other fruitful extension might be to use other spatial autocorrelation test statistics in the design criterion.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models* (1^{st} ed.). Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1999). Spatial Econometrics. *Bruton Center, School of Social Sciences, University of Texas at Dallas, Richardson, TX 75083-0688.*

Armstrong, M. (1998). *Basic Linear Geostatistics* (1^{st} ed.). Berlin: Springer.

Badinger, H., Müller, W., & Tondl, G. (2002). Regional Convergence in the European Union, 1985-1999: A Spatial Dynamic Panel Analysis. *Regional Studies*, *38(3)*, 241-253.

Chilès, J.-P., & Delfiner, P. (1999). *Geostatistics, Modeling Spatial Uncertainty* (1^{st} ed.). New York: John Wiley & Sons.

Clark, I., & Harper, W. (2000). *Practical Geostatistics* (1^{st} ed.). Ohio: Greyden Press.

Cliff, A., & Ord, J. (1981). *Spatial Processes. Models & Applications.* London: Pion Limited.

Coe, D., & Helpman, E. (1995). International R&D Spillovers. *European Economic Review*, *39*, 859-887.

Cressie, N. (1993). *Statistics for Spatial Data* (rev. ed.). New York: Wiley.

Davis, J. (2002). *Statistics and Data Analysis in Geology* (3^{rd} ed.). New York: Wiley & Sons.

De Gooijer, J. (1980). Exact Moments of the Sample Autocorrelations from Series Generated by General ARIMA Processes of Order (p,d,q), d=0 or 1. *Journal of Econometrics*, *14*, 365-379.

Fedorov, V. (1972). *Theory of Optimal Experiments.* New York: Academic Press.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression.* New York: Wiley.

Getis, A., & Ord, J. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*, 189-206.

Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering.* Berlin: Springer.

Griffith, D., & Layne, L. (1999). *A Casebook for Spatial Statistical Data Analysis.* New York: Oxford University Press.

Grossman, G., & Helpman, E. (1991). *Innovation and Growth in the Global Economy.* Cambridge, MA: MIT Press.

Gumprecht, D. (2003). Ein Panel Kointegrationsmodell für internationale Forschungs- und Entwicklungs Spillovers. *unpublished Master Thesis, University of Vienna.*

Gumprecht, D. (2005). Spatial Methods in Econometrics: An Application to R&D Spillovers. *Research Report Series, Vienna University of Economics and Business Administration, 26.*

Gumprecht, D. (2007a). R&D Spillovers: A Non-Spatial and a Spatial Examination. *Austrian Journal of Statistics, 36,* 3-15.

Gumprecht, D. (2007b). Treatment of Far-Off Objects in Moran's $\mathcal{I}$ Test. *Research Report Series, Vienna University of Economics and Business Administration, 46.*

Gumprecht, D., Gumprecht, N., & Müller, W. (2004). Some Current Issues in the Statistical Analysis of Spillovers. In G. Meier & S. Sedlacek (Eds.), *Spillovers and Innovations: City, Environment, and the Economy, Interdisciplinary Studies in Economics and Management, Vol. 4* (p. 51-70). Wien: Springer.

Gumprecht, D., Müller, W., & Rodríguez-Díaz, J. (2007). Optimal Design for Detecting Spatial Dependence. *submitted.*

Haines, L. (1987). The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models. *Technometrics, 29,* 439-447.

Helpman, E. (2003). *Professor Elhanan Helpman's data on the web.* Retrieved January 24, 2003, from Harvard University, Economics Department Web site: http://post.economics.harvard.edu/faculty/helpman/data.html.

Henshaw, R. (1966). Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances. *Econometrica, 34,* 646-660.

Kao, C., Chiang, M., & Chen, B. (1999). International R&D Spillovers: An Application of Estimation and Inference in Panel Cointegration. *Oxford Bulletin of Economics and Statistics, 61,* 693-711.

Kelejian, H., & Prucha, I. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic*

Review, *40*, 509-533.

Kelejian, H., & Prucha, I. (2006). Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. *Working Paper, University of Maryland*.

Kooijman, S. (1976). Some Remarks on the Statistical Analysis of Grids Especially with Respect to Ecology. *Annals of Systems Research*, *5*, 113-132.

Larch, M., Tappeiner, G., & Walde, J. (2007). Performance Contest between MLE and GMM for Huge Spatial Autoregressive Models. *Journal of Statistical Computation and Simulation*.

Meyer, R., & Nachtsheim, C. (1995). The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs. *Technometrics*, *37*, 60-69.

Moran, P. (1948). The Interpretation of Statistical Maps. *Biometrika*, *35*, 255-260.

Moran, P. (1950a). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*, 17-23.

Moran, P. (1950b). A Test for the Serial Dependence of Residuals. *Biometrika*, *37*, 178-181.

Müller, W. (2001). *Collecting Spatial Data* (2^{nd} ed.). Heidelberg: Physica.

Müller, W., & Nettekoven, M. (1999). A Panel Data Analysis: Research and Development Spillover. *Economics Letters*, *64*, 37-41.

Rao, M. (1979). *Stochastic Processes and Integration* (1^{st} ed.). The Netherlands: Sijthoff & Noordhoff.

Ripley, B. (1981). *Spatial Statistics* (1^{st} ed.). New York: Wiley & Sons.

Romer, P. (1990). Endogenous Technical Change. *Journal of Political Economy*, *98*, 71-102.

Silvey, S. (1980). *Optimal Design*. London: Chapman and Hall.

Smit, J. (1961). Estimation of the Mean of a Stationary Stochastic Process by Equidistant Observations. *Trabajos de Estadistica*, *12*, 35-45.

Tiefelsdorf, M. (2000). *Modelling Spatial Processes*. Berlin: Springer.

Tiefelsdorf, M. (2002). The Saddlepoint Approximation of Moran's I's and Local Moran's I's Reference Distributions and Their Numerical Evaluation. *Geographical Analysis*, *34*, 187-206.

Tiefelsdorf, M., & Boots, B. (1995). The Exact Distribution of Moran's I. *Environment and Planning A*, *27*, 985-999.

Wackernagel, H. (1998). *Multivariate Geostatistics* (2^{nd} ed.). Berlin Heidelberg: Springer.

Webster, R., & Oliver, M. (2001). *Geostatistics for Environmental Scientists.* England: Wiley.