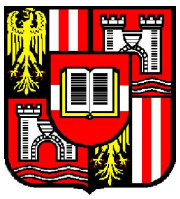


Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2007-20

Optimal Design for Detecting Spatial Dependence

Daniela Gumprecht[a], Werner G. Müller[b], Juan M. Rodríguez-Díaz[c]

January 2007

[a]Vienna University of Economics and Business Administration
[b]Johannes Kepler University Linz
[c]University of Salamanca

Abstract

The aim of this paper 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 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). 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 data set 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.

1 Introduction

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 our 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).

Let us exemplify the ideas of the paper on a well-known case. In Anselin (1988) the author presents us a study which became a classic testground for spatial analysis: the Columbus, Ohio crime dataset. The data stem from 49 contiguous planning neighbourhoods in Columbus, Ohio, USA, see Figure 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. Although the example does not fit well for our practical purposes, as the data are freely available for all the neighbourhoods, we have chosen it for its familiarity amongst the readership. 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 our considerations will continue to hold.

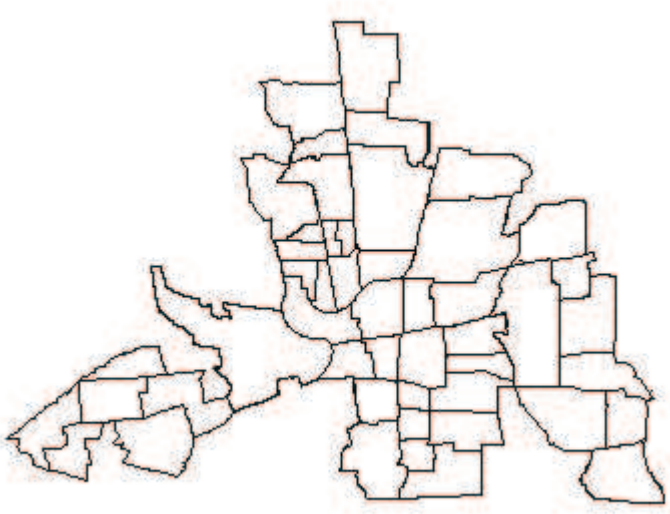


Figure 1: Neighbourhoods in Columbus, Ohio

As a measure for the intensity of the spatial dependence and consequently a valid test statistic for detecting its potential existence we confine ourselves in this paper to the probably most popular statistic, the Moran's $\mathcal{I}$, cf. Moran (1950), although a considerable number of alternatives are available. Note however, that our principal considerations are not affected by this choice.

1.1 Spatial Link Matrices

In spatial modeling the sampling design primarily affects the so called spatial link matrices (or spatial weighting matrices), which represent the spatial relationships between the observations. In general, spatial link matrices measure similarities, e.g. connectivity, neighbourhoods or inverse distances. A spatial link matrix $\mathbf{G}$ is an n by n matrix (n is the number of observations) with the following properties:

(i) $g_{ij} = 0$ for $i = j$;

(ii) $g_{ij} > 0$ if i and j are spatially connected.

There are in principle two ways to define a a spatial weight matrix. First as a function of the Euclidean distances between the locations of the observations,

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

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 $i = j$. By $\boldsymbol{\xi} = \{s_1, \ldots, s_n\}$ we denote the collection of coordinates, the sampling design, defined on a design space given by the set $\mathcal{S}$.

Another approach for building a spatial weight matrix is to use neighbourhood (contiguity) matrices. These are symmetric, binary $n \times n$ matrices with entries $g_{ij} = 1$, if two observations i and j are neighbours, and $g_{ij} = 0$, if i and j are not neighbours or if $i = j$. There are different definitions of neighbourhood used in practice. Here, we used the so-called Queen's criterion, where adjacent areas are neighbours if they share zero-length or nonzero-length boundaries.

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 object. A widely used coding scheme is the row-sum standardized coding scheme where the sum of each row of the standardized link matrix is equal to one. Other coding-schemes are the globally standardized-, and the variance stabilizing coding scheme (Tiefelsdorf, 2000). In the following, the spatial link matrix $\mathbf{V}(\boldsymbol{\xi})$ is the row-standardized version of matrix $\mathbf{G}$ with entries g_{ij}.

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), we rather try to find the optimal locations of the observations, the aim is not, to find a more robust test, but to better detect a spatial effect that is potentially present in the data.

1.2 Models

We intend to estimate an ordinary linear model and use the residuals for the test of spatial dependency, 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)

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, we either want to accept or reject the null hypothesis of spatial independence of Moran's $\mathcal{I}$ test (see section 2) to make a correct decision. The aim of this paper 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.

In this paper we restrict ourselves to Gaussian spatial processes, which are based on normally distributed regression disturbances. 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).

2 Some General Issues about Moran's $\mathcal{I}$

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 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). For regression residuals Moran's $\mathcal{I}$ is defined as scale invariant ratio of quadratic forms in the normally distributed regression residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n)'$, i.e.

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

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 classical Moran's $\mathcal{I}$ as the two-dimensional analog of a test for univariate time series correlation is given in Cliff and Ord (1981). For a random variable Y, measured in each of the n non-overlapping subareas of the whole study area, Moran's $\mathcal{I}$ is defined from the residuals of an intercept only regression, i.e. $\hat{\boldsymbol{\varepsilon}} = \mathbf{My}$ where $\mathbf{y}' = (y_1, \ldots, y_n)$, $\mathbf{M} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$, $\mathbf{I}_n$ is an $n \times n$ identity matrix and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. In this case, and if the spatial link matrix $\mathbf{V}$ has full rank, i.e. there is no observation completely separated from all others, the expected value of the test statistic $\mathcal{I}$ under independence is given by $\mathrm{E}[\mathcal{I}|H_0] = -\frac{1}{n-1}$, and the variance of $\mathcal{I}$ can be given in terms of the eigenvalues γ_i of matrix $\mathbf{K} = \mathbf{M}'\frac{1}{2}(\mathbf{V} + \mathbf{V}')\mathbf{M}$ as $\mathrm{Var}[\mathcal{I}|H_0] = \frac{2n}{n^2-1}\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2 = \frac{2n}{n^2-1}\sigma_\gamma^2$, and $\mathcal{I}$ is asymptotically normally distributed.

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. Thus, from now on $\rho \geq 0$ will be assumed. The z-transformed Moran's $\mathcal{I}$ is for normal distributed regression residuals and well-behaved spatial link matrices under certain regularity conditions (see e.g. Tiefelsdorf, 2000) asymptotically standard normally distributed, i.e. $z(\mathcal{I})$ is defined as

$$z(\mathcal{I}) = \frac{\mathcal{I} - \mathrm{E}[\mathcal{I}|H_0]}{\sqrt{\mathrm{Var}[\mathcal{I}|H_0]}} \sim N(0, 1). \tag{3}$$

The exact small sample distribution of Moran's $\mathcal{I}$ was obtained by Tiefelsdorf and Boots (1995), but we refrain from using it here as it would be a restrictive computational burden on our algorithm.

We shall later see that a special class of objects is relevant especially for design purposes. These are observations that belong to a design but are far apart from all other objects, in the sense that they have no spatial links to other observations; they will be called far-off objects in the following. If e.g. one observation is completely separated from all other, $\mathbf{V}$ has one line and one column with zero elements, nevertheless this does not lead to the same results as in case of excluding this separated observation completely from the analysis, i.e. taking only an $(n-1)$ point design. This problem is relevant in case of neighbourhood-based spatial link matrices. On the contrary if the connectivity is based on distances, all elements - except for the ones on the main diagonal which are zero by definition - are typically unequal zero and $\mathbf{V}$ has full

rank (with the notable exception of the case when a sill is defined, i.e. from a certain distance onwards the connectivity is assumed to be negligible and therefore set to zero). Thus, far-off objects influence a spatial analysis, nevertheless the treatment of far-off objects is not really worked out in literature. A discussion of far-off objects in Moran's $\mathcal{I}$ test is given in Gumprecht (2007). There are three obvious possibilities how to handle such observations:

(s) Include the separated observation and work with a spatial weight matrix which does not have full rank. Thus, simply ignore the fact that an object is far apart.

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

(ν) Include all n observations in the analysis, and use a modified unstandardized spatial weight matrix $\mathbf{U}$ with elements $u_{ij}+\nu$, with $\nu > 0$ and $\nu \to 0$, for all $i \neq j$ to avoid zero-lines and zero-columns.

Whichever specification is used, it influences $\mathcal{I}$, $\mathrm{E}[\mathcal{I}|H_0]$ and $\mathrm{Var}[\mathcal{I}|H_0]$ and therefore $z(\mathcal{I})$ and potentially the decision whether to reject the null hypothesis or not. Each treatment leads to a specific spatial link matrix. For the case of only one far-off object (namely the first one) the Moran's $\mathcal{I}$ values have the following relationships:

$$\mathcal{I}^{(\nu)} \leq \mathcal{I}^{(s)} \leq \mathcal{I}^{(e)}, \tag{4}$$

where equality holds only for $\hat{\varepsilon}_1 = 0$, i.e. the residuum of the far-off object is zero. A proof can be found in the Appendix A.

For practical reasons the relationships of the z-transformed Moran's $\mathcal{I}$ values are more interesting, as theses values might influence the test decision. The relations of the z-values are more complex as they include $\mathcal{I}$, $\mathrm{E}[\mathcal{I}|H_0]$ and $\mathrm{Var}[\mathcal{I}|H_0]$, see (3). Under certain assumptions the following relationships between the standardized Moran's $\mathcal{I}$ values of case (s), (e) and (ν) hold:

$$z[\mathcal{I}^{(e)}] \leq z[\mathcal{I}^{(s)}] \leq z[\mathcal{I}^{(\nu)}]. \tag{5}$$

It can be seen that treatment (e) leads to the most conservative test statistic, treatment (s) rejects the null earlier, and treatment (ν) is the one which rejects the null hypothesis first. All assumptions and a more detailed derivation of (5) can be found in Gumprecht (2007). Unfortunately, a general statement on the behaviour of $z(\mathcal{I})$ can not be made, nevertheless it turns out that when the number of design points is large, it does not make a difference which treatment is used for the far-off objects. Due to practical reasons, treatment (e) is not recommended, because even if an observations is not connected to others it might be important for the design.

2.1 Status of the World: Spaceless

Let us consider the first case of section 1.2, where we estimate a model under the assumption of spatial independence, and the true model is of the same form. The aim is then to accept the null hypothesis (=spatial independence). For the approximate test we require the moments of Moran's $\mathcal{I}$, which can be expressed in terms of the eigenvalues of the matrix $\mathbf{K}$ (Tiefelsdorf, 2000), with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ denoting the general projection matrix. Since only the moments are of interest, the evaluation of eigenvalues can be by-passed by making use of the

trace operator tr(.). In this case under the assumption of spatial independence, expected value and variance of $\mathcal{I}$ are then given by

$$\mathrm{E}[\mathcal{I} \mid H_0] = \frac{\mathrm{tr}(\mathbf{K})}{n-k} = \frac{\mathrm{tr}\{\mathbf{M}\frac{1}{2}(\mathbf{V}+\mathbf{V}')\mathbf{M}\}}{n-k} = \frac{\mathrm{tr}(\mathbf{MV})}{n-k} \tag{6}$$

and

$$\begin{aligned} \mathrm{Var}[\mathcal{I} \mid H_0] &= \frac{\mathrm{tr}(\mathbf{MVMV}') + \mathrm{tr}(\mathbf{MV})^2 + \{\mathrm{tr}(\mathbf{MV})\}^2}{(n-k)(n-k+2)} - \{\mathrm{E}[\mathcal{I} \mid H_0]\}^2 \\ &= \frac{2\{(n-k)\mathrm{tr}(\mathbf{K}^2) - \mathrm{tr}(\mathbf{K})^2\}}{(n-k)^2(n-k+2)} \end{aligned} \tag{7}$$

respectively, see Henshaw (1966).

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 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.

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:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbf{u} = \rho\mathbf{V}\mathbf{u} + \boldsymbol{\varepsilon} \tag{8}$$

with $\boldsymbol{\varepsilon}$ i.i.d. Here the spatial dependence appears in the form of a spatially lagged error term $\mathbf{u}$. This model is called 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 is

$$\boldsymbol{\Omega}(\rho) = \mathrm{E}[\mathbf{u}\mathbf{u}'] = \sigma^2[(\mathbf{I}-\rho\mathbf{V})'(\mathbf{I}-\rho\mathbf{V})]^{-1} \tag{9}$$

To ensure that $\boldsymbol{\Omega}(\rho)$ is positive definite, ρ is restricted to the interval $]\frac{1}{\lambda_{min}}; \frac{1}{\lambda_{max}}[$, where λ_{min} and λ_{max} denote the smallest and largest eigenvalue of $\mathbf{V}$.

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

$$\mathrm{E}[\mathcal{I}|H_A] = \int_0^\infty \prod_{i=1}^{n-k}(1+2\lambda_i t)^{-\frac{1}{2}} \cdot \sum_{i=1}^{n-k} \frac{h_{ii}^*}{1+2\lambda_i t}\, dt \tag{10}$$

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

$$\mathrm{Var}[\mathcal{I}|H_A] = \mathrm{E}[\mathcal{I}^2|H_A] - \mathrm{E}[\mathcal{I}|H_A]^2 \tag{11}$$

where

$$\mathrm{E}[\mathcal{I}^2|H_A] = \int_0^\infty \left[\prod_{i=1}^{n-k}(1+2\lambda_i t)^{-\frac{1}{2}}\right] \cdot \left[\sum_{i=1}^{n-k}\sum_{j=1}^{n-k} \frac{h_{ii}^* h_{jj}^* + 2(h_{ij}^*)^2}{(1+2\lambda_i t)(1+2\lambda_j t)}\right] t\, dt \tag{12}$$

and $\mathrm{E}[\mathcal{I}|H_A]$ is given in equation (10). 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 (10) of

$$\left[\frac{(n-k)h_{max}}{2\lambda_1^{\frac{n-k}{2}}}\left(\frac{n-k}{2}-1\right)\frac{1}{\epsilon}\right]^{\frac{1}{\frac{n-k}{2}-1}} = \tau_1 \tag{13}$$

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 $\mathrm{E}[\mathcal{I}^2|H_A]$, (12), is

$$\left[\frac{3(n-k)^2 h_{max}^{(2)}}{(2\lambda_1)^{\frac{n-k}{2}}}\left(\frac{n-k}{2}-2\right)\frac{1}{\epsilon}\right]^{\frac{1}{\frac{n-k}{2}-2}} = \tau_2 \tag{14}$$

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 (10) and (12) are based on a GAUSS-code implemented by M. Tiefelsdorf and the results were checked with a code in Mathematica programm implemented by the authors.

3 Optimal Design Considerations

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

$$\Psi(\boldsymbol{\xi}) = 1 - \Phi\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{\mathrm{Var}[\mathcal{I}|H_0]} + \mathrm{E}[\mathcal{I}|H_0] - \mathrm{E}[\mathcal{I}|H_A]}{\sqrt{\mathrm{Var}[\mathcal{I}|H_A]}}\right) \tag{15}$$

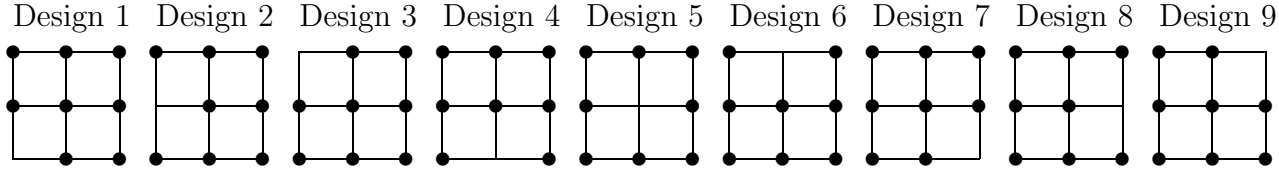
where Φ denotes the cdf of the standard normal distribution. The maximization of Ψ over $\boldsymbol{\xi} \in \mathcal{S}$ gives the final optimal locations of the observations and thus maximizes the power of the Moran's $\mathcal{I}$ test. To calculate Ψ, the expected value (6) and the variance (7) of $\mathcal{I}$ under the null hypothesis, and the expected value (10) and the variance (11) 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 we can not employ the well developed optimum design theory (cf. Silvey, 1980) but must resort to algorithmic approaches.

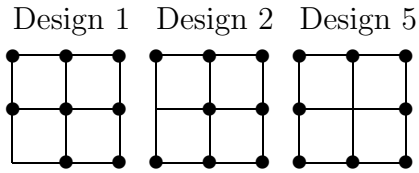
3.2 Design Algorithms

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 (13) and (14). 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. A detailed discussion of this properties can be found in Appendix B. We call the respective designs 'symmetric'. 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 (10) and (12) 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 illustration of this problem assume a regular 9-point grid, 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 other, really different 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 (10) and an additional one for the variance under the alternative (12). 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 (10) and (11) 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 a initial design $\boldsymbol{\xi}_0 = \mathcal{S}$, called 'base' design. Thus in the first iteration the number of points r in $\boldsymbol{\xi}_0$ is m.

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 a 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 alternatively methods of stochastic optimization such as in Haines (1987).

Fedorov exchange algorithm
As an alternative and sort of compromise, we suggest an exchange type algorithm based on Fedorov (1972). 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 $\boldsymbol{\xi}_0$ (we call it '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 diag($\mathbf{H}$) 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 our 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 section 2). In case of designs with far-off observations, excluding the far-off 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. In the following treatment (s), which leads to a more conservative test, is used.

4 Examples

4.1 Artificial Data set

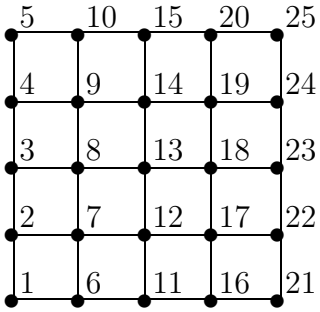
We estimate an OLS regression model with i.i.d. error terms. The OLS residuals are used to calculate Moran's $\mathcal{I}$ and its expected values and variances under the null and under the alternative, see (2), (6), (7), (10) and (11), which are needed for evaluating the design criterion Ψ (15). The best design is the one with the largest Ψ:

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi} \in \mathcal{S}} \Psi(\boldsymbol{\xi}) \tag{16}$$

Since Φ is a monotonous function we need to minimize its argument only, for computational simplicity. The observations are taken on a regular 25-point grid $[-1; 1]^2$. All three algorithms described in the previous section were used. The null hypothesis is spatial independence, the alternative hypothesis is spatial dependence with a spatial autoregressive parameter $\rho = 0.5$.

4.1.1 Regression on Intercept

The considered model is a regression of $\mathbf{y}$ on an intercept: $\mathbf{y} = \mathbf{1}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We assume 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), 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 (1) with parameter $\delta = 5.76$, this setting gives the same criterion value for the full design. The correlation structure with corresponding exponential function is displayed in Figure 2.

$$\mathbf{V}_c = \begin{pmatrix} 0 & 0.33 & 0 & \cdots & 0 \\ 0.20 & 0 & 0.20 & \cdots & 0 \\ 0 & 0.20 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \mathbf{V}_d = \begin{pmatrix} 0 & 0.402 & 0.023 & 0.001 & \cdots & 0.000 \\ 0.262 & 0 & 0.262 & 0.015 & \cdots & 0.000 \\ 0.014 & 0.256 & 0 & 0.256 & \cdots & 0.000 \\ 0.001 & 0.015 & 0.262 & 0 & \cdots & 0.000 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.000 & 0.000 & 0.000 & 0.000 & \cdots & 0 \end{pmatrix}$$

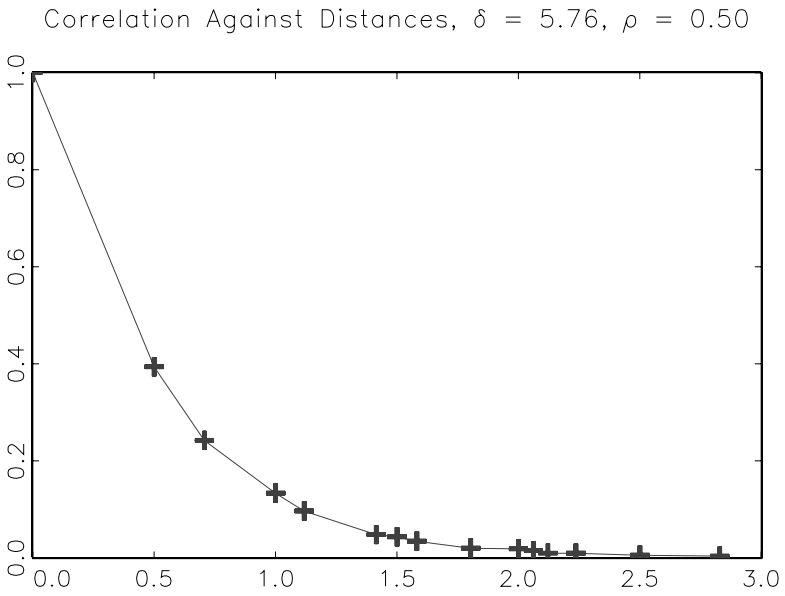
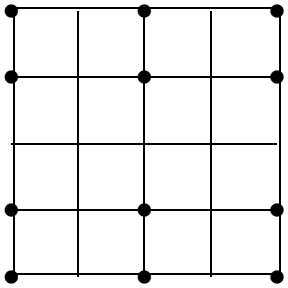


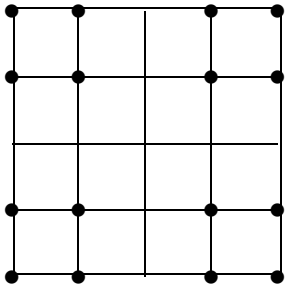
Figure 2: Correlation for 25 point grid

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-point design
$\mathrm{E}[\mathcal{I}|H_0] = -0.091$
$\mathrm{Var}[\mathcal{I}|H_0] = 0.153$
$\mathrm{E}[\mathcal{I}|H_A] = 0.692$
$\mathrm{Var}[\mathcal{I}|H_A] = 0.058$
$\Psi = 0.720$



Distance
16-point design
$\mathrm{E}[\mathcal{I}|H_0] = -0.067$
$\mathrm{Var}[\mathcal{I}|H_0] = 0.037$
$\mathrm{E}[\mathcal{I}|H_A] = 0.318$
$\mathrm{Var}[\mathcal{I}|H_A] = 0.067$
$\Psi = 0.603$

The development of the 'nearly' optimal designs found by the search- and the exchange algorithm can be seen in Figure 3.

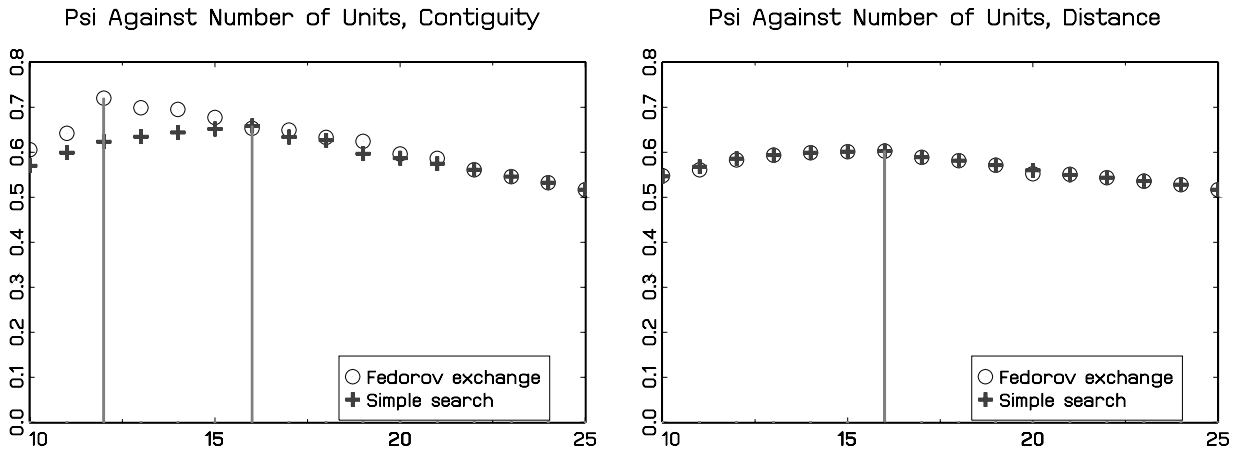


Figure 3: 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices

4.1.2 Linear Trend Model

Let the considered model be 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 was used and the simple search and the Fedorov exchange algorithm were 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 $\Psi = 0.462$.

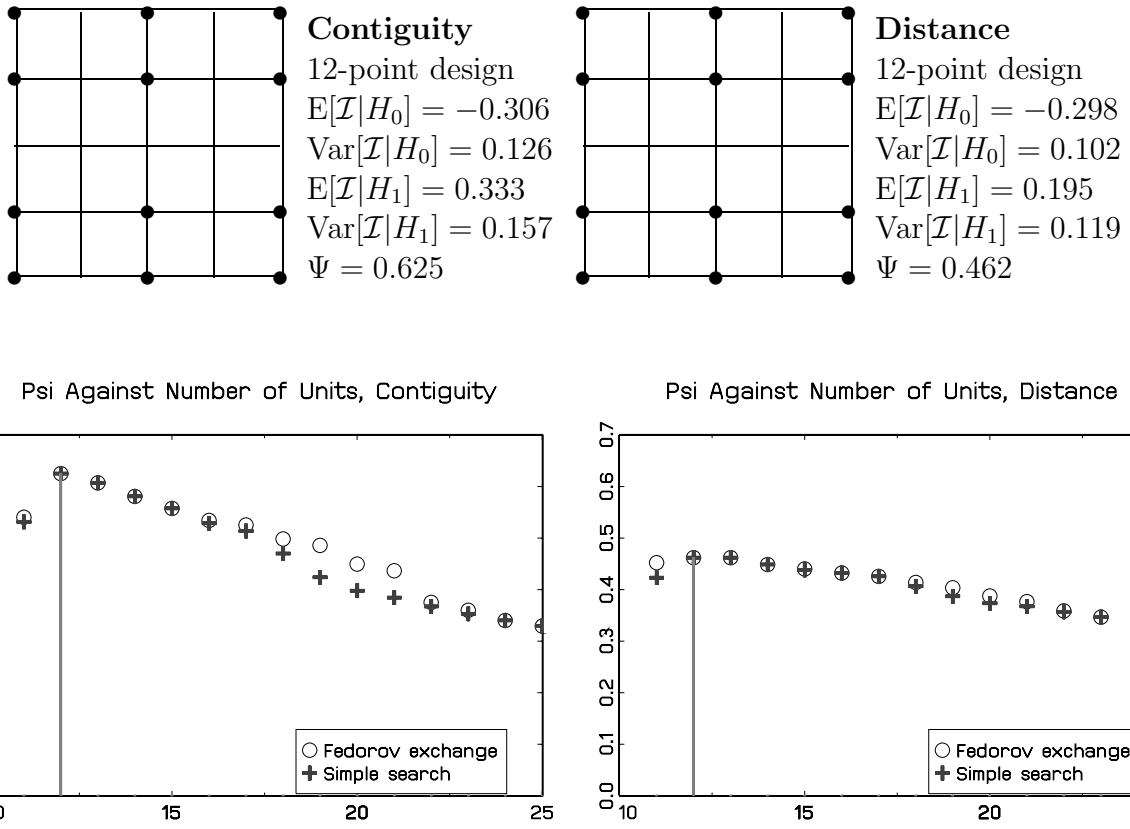


Figure 4: 25-Point grid, simple search & exchange algorithm, contiguity and distance matrices

4.2 Columbus Crime Data

Data for this example stem from the classical Columbus Crime dataset from Anselin (1988), see section 1 and Figure 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 a 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 $z(\mathcal{I}) = 5.675$.

When we try to find the optimal design, we use 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 $\Psi = 0.983$, see Figure 5, the dark grey locations are the ones which were selected by both algorithms, and a Moran's $\mathcal{I} = 0.417$ and $z(\mathcal{I}) = 1.914$ which is significant on the 5%-level. 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. 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 6.

Figure 5: Left: Simple search algorithm. Right: Fedorov exchange algorithm.

5 Conclusions

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. The great generality of the approach evidently allows improvements in many directions.

Thus, our suggestions for further work include the use of the exact distribution of Moran's $\mathcal{I}$ rather than the normal approximation; the development of more efficient numerical integration routines and search- or exchange algorithms (including heuristic techniques such as genetic algorithms); more elaborate checks of symmetries of the designs for decreasing runtime. Finally, the implementation of other traditional criteria for checking spatial dependence might be fruitful.

Acknowledgments

This paper was initiated during a research visit of Daniela Gumprecht and Werner G. Müller in Salamanca financed by the ÖAD (project number 11/2004). The main work by Daniela Gumprecht was funded by the Jubiläumsfonds of the OeNB (Austrian central bank), project "Common Structures in Spatial Econometrics", project number 11052. We would like to thank Michael Tiefelsdorf for providing the GAUSS-code implementation of the expected value and the variance of Moran's $\mathcal{I}$ under the alternative hypothesis.

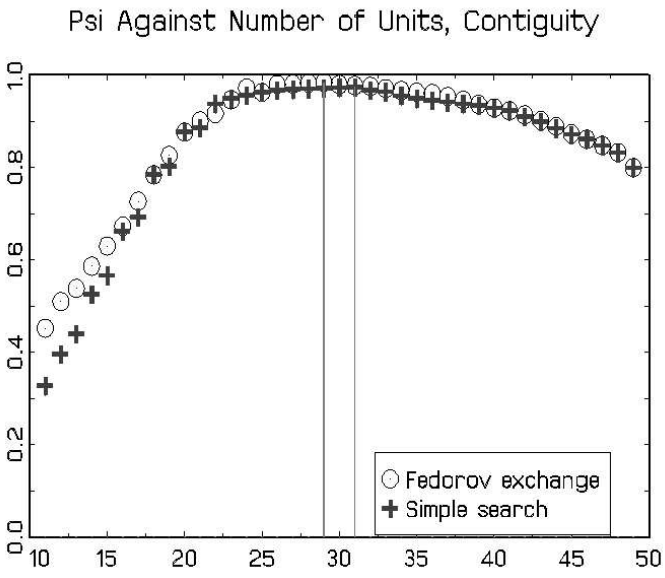


Figure 6: Columbus crime data, simple search- & exchange algorithm, contiguity matrix

6 Appendix

6.1 Appendix A

Under the assumption that only one object (the first one) is far apart from all others, the influence on Moran's $\mathcal{I}$ can be easily derived. The far-off object affects the spatial link matrix, according to the three treatments given above, three different weight matrices are relevant for the calculation of Moran's $\mathcal{I}$. The unstandardized and symmetric spatial link matrix is denoted by $\mathbf{U}$, the row-standardized (nonsymmetric) weight matrix is denoted by $\mathbf{V}$ where $v_{ij} = \frac{u_{ij}}{\sum_j u_{ij}}$, and $\mathbf{Q} = \frac{1}{2}(\mathbf{V} + \mathbf{V}')$, $\mathbf{Q}$ is used in Moran's $\mathcal{I}$, see (2).
In case (s), all n objects are taken into account:

$$\mathbf{U}^{(s)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & u_{23} & u_{24} & \cdots & u_{2n} \\ 0 & u_{32} & 0 & u_{34} & \cdots & u_{3n} \\ \vdots & & & & & \vdots \\ 0 & u_{n2} & u_{n3} & u_{n4} & \cdots & 0 \end{pmatrix}, \mathbf{V}^{(s)}_{n\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{u_{23}}{\sum_{j=2}^n u_{2j}} & \cdots & \frac{u_{2n}}{\sum_{j=2}^n u_{2j}} \\ 0 & \frac{u_{32}}{\sum_{j=2}^n u_{3j}} & 0 & \cdots & \frac{u_{3n}}{\sum_{j=2}^n u_{3j}} \\ \vdots & & & & \vdots \\ 0 & \frac{u_{n2}}{\sum_{j=2}^n u_{nj}} & \frac{u_{n3}}{\sum_{j=2}^n u_{nj}} & \cdots & 0 \end{pmatrix}.$$

In case (e), the far-off object is excluded, i.e. there are only $(n-1)$ design points left:

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

In case (ν) again all n objects are included in the analysis, the row-standardized spatial weight $\mathbf{V}^{(\nu)*}$ goes to $\mathbf{V}^{(\nu)}$ if $\nu \to 0$.

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

The corresponding matrices $\mathbf{Q}$ are written in the structure of block matrices to simplify the comparison of the Moran's $\mathcal{I}$ values for the treatments (s), (e) and (ν):
Case (s):

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

case(e):

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

case (ν):

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

Now Moran's $\mathcal{I}$ given in formula (2) can be written in block structure notation. Therefore the vector of the residuals is partitioned into two blocks: $\hat{\boldsymbol{\varepsilon}}_{n\times 1} = [\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n]' = [\mathbf{a}' : \mathbf{b}']$, where $\mathbf{a}$ is simply the residuum $\hat{\varepsilon}_1$ of the far-off object and $\mathbf{b}$ is a $(n-1)\times 1$ vector of the residuals $\hat{\varepsilon}_i$ $(i = 2, ..., n)$ of the 'well-behaved' objects.

$$\mathcal{I} = \frac{\hat{\boldsymbol{\varepsilon}}'\frac{1}{2}(\mathbf{V}+\mathbf{V}')\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{Q}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{[a' : b'] \begin{pmatrix} \mathbf{A} : \mathbf{B} \\ \ddot{\mathbf{B}}' : \ddot{\mathbf{C}} \end{pmatrix} \begin{bmatrix} a \\ \ddot{b} \end{bmatrix}}{[a' : b'] \begin{bmatrix} a \\ \ddot{b} \end{bmatrix}} = \frac{\mathbf{a}'\mathbf{A}\mathbf{a} + \mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} \tag{17}$$

For the different treatments of the far-off object, the corresponding matrix $\mathbf{Q}$ is used in (17). Case (s), the first element is separated, $\mathbf{A} = 0$ and $\mathbf{B}$ is a vector of zeros, and Moran's $\mathcal{I}$ is

$$\mathcal{I}^{(s)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j q_{ij}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \tag{18}$$

In case (e) the far-off object is excluded from the analysis and Moran's $\mathcal{I}$ is given by

$$\mathcal{I}^{(e)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j q_{ij}}{\sum_{i=2}^{n}\hat{\varepsilon}_i^2}. \tag{19}$$

Finally case (ν), $\mathbf{A} = 0$ but $\mathbf{B} = \left[\frac{1}{2(n-1)}, \ldots, \frac{1}{2(n-1)}\right]$, Moran's $\mathcal{I}$ is given by

$$\mathcal{I}^{(\nu)} = \frac{\mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j q_{ij} - \frac{\hat{\varepsilon}_1^2}{n-1}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}, \tag{20}$$

because given that $\mathbf{b}'\mathbf{B}'\mathbf{a} = \mathbf{a}'\mathbf{B}\mathbf{b} = \frac{1}{2(n-1)}\sum_{i=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_1$, and $\sum_{i=2}^{n}\hat{\varepsilon}_i = -\hat{\varepsilon}_1$.
Now the Moran's $\mathcal{I}$ values of the various treatments given in (18), (19) and (20) can be compared. This gives relationship (4):

$$\mathcal{I}^{(\nu)} \leq \mathcal{I}^{(s)} \leq \mathcal{I}^{(e)},$$

where equality holds only for $\hat{\varepsilon}_1 = 0$.

6.2 Appendix B

A rotation of an angle θ (counter-clockwise) is given by the matrix

$$\mathbf{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

For other hand, the *basic reflection* (respect to the y-axis) is given by

$$\mathbf{S}_0 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad ,$$

while a general reflection, in which the reflection axis goes through the origin and form an angle θ with the y-axis, is represented by the matrix

$$\mathbf{S}_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$

Rotations and reflections are very related each other:

- the θ-angle rotation, $\mathbf{R}_\theta$, can be represented as the composition of two reflections with reflection axis keeping an angle $\theta/2$ between them
- the reflection $\mathbf{S}_\theta$ is equivalent to first rotate an angle θ to convert the reflection axis in the y-axis, then perform the *basic reflection* and finally undo the previous rotation; i.e.: $\mathbf{S}_\theta = \mathbf{R}_\theta \mathbf{S_0} \mathbf{R}_{-\theta}$

and they verify many nice properties. Especially interesting for our work are the following:

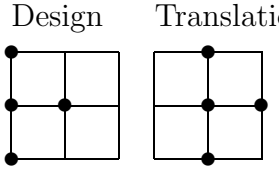
1. Both rotation and reflection matrices are orthogonal. Furthermore, the reflection matrix is trivially idempotent. Orthogonality is essential for this paper since it means that the transformation does not modify the matrix $\mathbf{M} = \mathbf{M_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$: if $\mathbf{R}$ is a rotation or reflection then $\mathbf{R}' = \mathbf{R^{-1}}$ and for $\mathbf{Z}' = \mathbf{R}\mathbf{X}'$ we have

$$\mathbf{M_Z} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{I} - \mathbf{X}\mathbf{R}(\mathbf{R}\mathbf{X}'\mathbf{X}\mathbf{R})^{-1}\mathbf{R}\mathbf{X}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M_X}$$

2. Both transformations maintain the distances between points as well as the contiguity, thus they have no influence on the matrix $\mathbf{V}$.

From the above properties it is clear that they do not modify the matrices $\mathbf{\Omega}$, $\mathbf{B}$, $\mathbf{P}$, $\mathbf{A}$ and specially $\mathbf{H}$, what means that they have no influence in the criterion function $\mathbf{\Psi}$.

In fact not only rotations or reflections, but any movements of the design space verifying properties (1) and (2) do not modify the criterion function. There are even more transformations giving equivalent designs (respect to the criterion function), for instance a transformation that in general can move points out of the design space but that when applied to specific designs keep all the points's image in the design space, like some translations:



References

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

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

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

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.

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

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

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

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

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

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

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

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.