Daniela Gumprecht

Treatment of Far-Off Objects in Moran's I Test

Working Paper

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

Daniela Gumprecht

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 46
January 2007

http://statmath.wu-wien.ac.at/

Treatment of far-off objects in Moran's $\mathcal{I}$ test

Daniela Gumprecht

January 2007

Abstract

Spatial dependency is commonly measured and tested with Moran's $\mathcal{I}$ statistic. The question to be answered is, whether far-off objects affect this statistic and influence the test. Far-off objects are observations that are far apart from all other objects in the dataset, i.e. they do not have spatial links to other design points. In the paper different possibilities of treating such objects are discussed, and their influence on Moran's $\mathcal{I}$ and the corresponding spatial autocorrelation test is analysed.

Keywords: Moran's I, far-off objects

1 Introduction

One of the most famous methods for measuring and testing spatial dependency is based on Moran's $\mathcal{I}$ statistic Moran (1950), which measures the intensity of spatial autocorrelation in a spatial stochastic process. For a linear regression model $\mathcal{I}$ is given by

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

where $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n)'$ are the normally distributed regression residuals and $\mathbf{V}$ is a standardized spatial weight matrix, see e.g., Tiefelsdorf (2000). For testing the null hypothesis of no spatial autocorrelation against the alternative of positive or negative spatial autocorrelation, the expected value and the variance of $\mathcal{I}$ under the null are needed. Both can be written in form of the projection matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{K} = \mathbf{M}'\frac{1}{2}(\mathbf{V} + \mathbf{V}')\mathbf{M}$. The expected value under the null hypothesis is:

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

If the spatial dependency of a single variable is of interest, one simply uses a regression on only an intercept and takes the corresponding residuals for calculating $\mathcal{I}$, i.e. $k = 1$ and $\mathbf{M} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. Under the assumption that $\mathbf{V}$ has full rank, the expected value under the null hypothesis reduces to $\mathrm{E}[\mathcal{I}|H_0] = -\frac{1}{n-1}$, (cf. Cliff and Ord, 1981), which is independent of the design points, i.e. the location of the observations is not relevant. The variance of $\mathcal{I}$ under the null is equally given by

$$\begin{aligned}\mathrm{Var}[\mathcal{I} \mid H_0] &= \frac{\mathrm{tr}(\mathbf{MVMV}') + \mathrm{tr}(\mathbf{MV})^2 + \{\mathrm{tr}(\mathbf{MV})\}^2}{(n-k)(n-k+2)} - \{\mathrm{E}[\mathcal{I} \mid H_0]\}^2 \\ &= \frac{2\{(n-k)\mathrm{tr}(\mathbf{K}^2) - \mathrm{tr}(\mathbf{K})^2\}}{(n-k)^2(n-k+2)}. \end{aligned} \tag{3}$$

For $k = 1$ and $\mathrm{rank}(\mathbf{V}) = n$, this variance can be given in terms of the eigenvalues ν_i of matrix $\mathbf{K} = \mathbf{M}'\frac{1}{2}(\mathbf{V} + \mathbf{V}')\mathbf{M}$ as $\mathrm{Var}[\mathcal{I}|H_0] = \frac{2n}{n^2-1}\sum_{i=1}^{n}(\nu_i - \bar{\nu})^2 = \frac{2n}{n^2-1}\sigma_\nu^2$, (cf. Cliff and Ord, 1981).
The standardized Moran's $\mathcal{I}$ is for normal distributed regression residuals and well-behaved spatial link matrices under the assumption of spatial independence asymptotically $N(0, 1)$- distributed, and is given by

$$z(\mathcal{I}) = \frac{\mathcal{I} - \mathrm{E}(\mathcal{I})}{\sqrt{\mathrm{Var}(\mathcal{I})}}. \tag{4}$$

Values $z(\mathcal{I})$ are simply compared with the critical values of the $N(0,1)$ distribution to test the null hypothesis of no spatial autocorrelation against the alternative of positive or negative spatial autocorrelation, (cf. Tiefelsdorf, 2000).

2 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. Far-off objects are observations that are far apart from all other objects but still of interest in a certain point of view, i.e. they belong to the design but have no spatial links to other design points. And even such objects are not spatially linked to others, they have influence on spatial analysis. This can be seen e.g., when looking at the expected value of $\mathcal{I}$ 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 occure if neighbourhood-based spatial link matrices are used, and it might occur if the connectivity is based on distance matrices and a sill, i.e. from a certain distance onwards the connectivity is assumed to be negligible and therefore set to zero.

One of the first ideas that comes into thought, of how to treat such observations, 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, then no element is completely separated, and if ν is small enough it should have no influence on the general structure, but it prevents getting zero-lines. So, there are three kinds of specifications:

(i) Include the separated observation and work with a spatial weight matrix which does not have full rank, this case is denoted by (s) in the following.

(ii) Exclude the separated observations, i.e. work with $(n-r)$ observations, notation for this treatment will be (e).

(iii) 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, denoted by (ν).

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.

2.1 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 (i) 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:

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

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

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

For treatment (iii), if a small value ν is added, $\mathbf{U}^{(\nu)}$ is again a symmetric $n \times n$ matrix:

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

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

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

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

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 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 (i), (ii) and (iii), these matrices have different forms, and all of them can be written in the structure of block matrices. Specification (i) gives:

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

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

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

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

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

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

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

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.

2.2 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 (1), (2) and (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 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. Using the block structure helps to see the difference between the three specifications.

$$\mathcal{I} = \frac{\hat{\boldsymbol{\varepsilon}}'\frac{1}{2}(\mathbf{V}+\mathbf{V}')\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{G}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = \frac{[a' : b']\begin{pmatrix} \mathbf{A} : \mathbf{B} \\ \ddot{\mathbf{B}}' : \ddot{\mathbf{C}} \end{pmatrix}\begin{bmatrix} a \\ \ddot{b} \end{bmatrix}}{[a' : b']\begin{bmatrix} a \\ \ddot{b} \end{bmatrix}} = \frac{\mathbf{a}'\mathbf{A}\mathbf{a} + \mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} \quad (5)$$

For the different treatments of the separated observation the blocks of the corresponding matrix $\mathbf{G}$ are inserted in (5). This gives for case (i), where the first element is separated and therefore $\mathbf{A} = 0$ and $\mathbf{B}$ is a vector of zeros

$$\mathcal{I}^{(s)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \quad (6)$$

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

$$\mathcal{I}^{(e)} = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij}}{\sum_{i=2}^{n}\hat{\varepsilon}_i^2}. \tag{7}$$

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

$$\mathcal{I}^{(\nu)} = \frac{\mathbf{b}'\mathbf{B}'\mathbf{a} + \mathbf{a}'\mathbf{B}\mathbf{b} + \mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b}} = \frac{\sum_{i=2}^{n}\sum_{j=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_j g_{ij} - \frac{\hat{\varepsilon}_1^2}{n-1}}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \tag{8}$$

This follows from the fact that $\mathbf{b}'\mathbf{B}'\mathbf{a} = \mathbf{a}'\mathbf{B}\mathbf{b} = \frac{1}{2(n-1)}\sum_{i=2}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_1$, as $\sum_{i=2}^{n}\hat{\varepsilon}_i = -\hat{\varepsilon}_1$, 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 (6) and (7) 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 (6) and (8). 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:

$$\mathcal{I}^{(\nu)} \leq \mathcal{I}^{(s)} \leq \mathcal{I}^{(e)}. \tag{9}$$

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 residuum 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).

2.3 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 (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:

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

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

$$\mathrm{tr}(\mathbf{K}^{(s)}) = \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}) = -\frac{1}{n}\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij} = -\frac{1}{n}(n-1). \tag{10}$$

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

$$\mathrm{E}[\mathcal{I}^{(s)}|H_0] = -\frac{1}{n}. \tag{11}$$

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

$$\mathrm{tr}(\mathbf{K}^{(e)}) = \mathrm{tr}\left[\mathbf{M}^{(e)}\mathbf{C}\right] = -\frac{1}{n-1}\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij} = -\frac{1}{n-1}(n-1) = -1. \tag{12}$$

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

$$\mathrm{E}[\mathcal{I}^{(e)}|H_0] = -\frac{1}{n-2}. \tag{13}$$

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

$$\begin{aligned}\mathrm{tr}(\mathbf{K}^{(\nu)}) &= \mathrm{tr}(\mathbf{M}^{(s1)}\mathbf{A}) + \mathrm{tr}(\mathbf{M}^{(s2)}\mathbf{B}^{(\nu)\prime}) + \mathrm{tr}(\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}) + \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C})) \\ &= 0 - \frac{1}{2n} - \frac{1}{2n} - \frac{n-1}{n} = -1. \end{aligned} \tag{14}$$

This leads to

$$\mathrm{E}[\mathcal{I}^{(\nu)}|H_0] = -\frac{1}{n-1}. \tag{15}$$

Comparing the expected values of $\mathcal{I}^{(s)}$, $\mathcal{I}^{(e)}$ and $\mathcal{I}^{(\nu)}$ under the null, see (11), (13) and (15), gives the following order:

$$\mathrm{E}[\mathcal{I}^{(e)}] < \mathrm{E}[\mathcal{I}^{(\nu)}] < \mathrm{E}[\mathcal{I}^{(s)}]. \tag{16}$$

and the absolute values of the expected values have the reverse order, i.e. $|\mathrm{E}[\mathcal{I}^{(s)}]| < |\mathrm{E}[\mathcal{I}^{(\nu)})] < |\mathrm{E}[\mathcal{I}^{(e)}]|$. 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.

2.4 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), with $\mathrm{tr}(\mathbf{K}^2) = \mathrm{tr}(\mathbf{KK}) = \mathrm{tr}(\mathbf{MGMG})$. In block-notation this trace is:

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

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 $\mathbf{D}$ are composed of 8 other matrices $\mathbf{D}_1^{(1)}, \ldots, \mathbf{D}_8^{(1)}$ and $\mathbf{D}_1^{(2)}, \ldots, \mathbf{D}_8^{(2)}$ respectively, and $\mathrm{tr}(\mathbf{KK}) = \sum_{i=1}^{2}\sum_{j=1}^{8}\mathrm{tr}(\mathbf{D}_j^{(i)})$ (the trace of a sum of matrices is equal to the sum of the traces). The 16 different traces of matrix-blocks are:

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

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 (i) and (ii). The last term $\text{tr}(\mathbf{D}_2^{(8)})$ is relevant for all three cases, it can be further simplified.

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

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

$$\text{tr}(\mathbf{MCMC})=\sum_{i=2}^n\sum_{j=2}^n g_{ij}^2-\frac{2}{n}\sum_{i=2}^n\left(\sum_{j=2}^n g_{ij}\right)^2+\frac{1}{n^2}(n-1)^2 \tag{17}$$

Note that for case (ii), where only the $(n-1)$ objects which are spatially connected are included, (17) is slightly different because the number the design points is $(n-1)$, it is $\text{tr}(\mathbf{MCMC})=\sum_{i=2}^n\sum_{j=2}^n g_{ij}^2-\frac{2}{(n-1)}\sum_{i=2}^n\left(\sum_{j=2}^n g_{ij}\right)^2+\frac{1}{(n-1)^2}(n-1)^2$. In general the following relationship holds: $\text{tr}(\mathbf{C}\mathbf{C})\le\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{C})\le\text{tr}(\mathbf{1}\mathbf{1}'\mathbf{C}\mathbf{1}\mathbf{1}'\mathbf{C})$, 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 (i) $\mathbf{B}=[0,\ldots,0]$, therefore only $\text{tr}(\mathbf{D}_8^2)=\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ plays a role. This leads to a variance of

$$\text{Var}[\mathcal{I}^{(s)}\mid H_0]=\frac{2\{(n-k)\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})-\text{tr}(\mathbf{K}^s)^2\}}{(n-k)^2(n-k+2)} \tag{18}$$

where $\text{tr}(\mathbf{K}^{(s)})^2=\left(-\frac{(n-1)}{n}\right)^2$, $\text{tr}(\mathbf{K}^{(s)})$ is given in (10), and $\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ is given in (17).

For case (iii) $\mathbf{G}^{(\nu)}$ 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 $\text{tr}(\mathbf{D}_1)=0$.
$\text{tr}(\mathbf{D}_2^{(2)})=\text{tr}(\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(s1)}\mathbf{B}^{(\nu)})=\text{tr}(\mathbf{M}^{(s1)}\mathbf{B}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime})=\text{tr}(\mathbf{D}_1^7)=\frac{1}{4n^2}$
$\text{tr}(\mathbf{D}_2^{(3)})=\text{tr}(\mathbf{M}^{(2)\prime}\mathbf{B}^{(\nu)}\mathbf{M}^{(2)\prime}\mathbf{B}^{(\nu)})=\text{tr}(\mathbf{M}^{(2)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(2)}\mathbf{B}^{(\nu)\prime})=\text{tr}(\mathbf{D}_1^{(6)})=\frac{1}{4n^2}$
$\text{tr}(\mathbf{D}_2^{(4)})=\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)})=-\frac{1}{2n^2}$
$\text{tr}(\mathbf{D}_2^{(6)})=\text{tr}(\mathbf{M}^{(3)}\mathbf{B}^{(\nu)\prime}\mathbf{M}^{(2)}\mathbf{C})=\text{tr}(\mathbf{M}^{(s2)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{B}^{(\nu)\prime})=\text{tr}(\mathbf{D}_1^{(8)})=-\frac{1}{2n^2}$
$\text{tr}(\mathbf{D}_2^{(7)})=\text{tr}(\mathbf{M}^{(s2)\prime}\mathbf{B}^{(\nu)}\mathbf{M}^{(s3)}\mathbf{C})=\text{tr}(\mathbf{D}_2^{(4)})=-\frac{1}{2n^2}$
$\text{tr}(\mathbf{D}_2^{(8)})=\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$
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

$$\text{Var}[\mathcal{I}^{(\nu)}\mid H_0]=\frac{2\{(n-k)\left[-\frac{1}{n^2}+\text{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})\right]-\text{tr}(\mathbf{K}^{(\nu)})^2\}}{(n-k)^2(n-k+2)} \tag{19}$$

with $\mathrm{tr}(\mathbf{K}^{(\nu)})^2 = (-1)^2$, see (14) and $\mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$ is again the one given in (17).

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

$$\mathrm{Var}[\mathcal{I}^{(e)} \mid H_0] = \frac{2\{(n-1-k)\mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) - \mathrm{tr}(\mathbf{K}^{(e)})^2\}}{(n-1-k)^2(n-1-k+2)} \tag{20}$$

with $\mathrm{tr}(\mathbf{K}^{(e)})^2 = (-1)^2$, see (12).
$\mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) = \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n-1}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{1}{(n-1)^2}(n-1)^2$, 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 $\mathrm{Var}[\mathcal{I}^{(\nu)}]$ and $\mathrm{Var}[\mathcal{I}^{(s)}]$ is quite obvious, $\mathrm{Var}[\mathcal{I}^{(\nu)}] < \mathrm{Var}[\mathcal{I}^{(s)}]$, which follows directly from (18) and (19), the denominators are the same, and the nominator of $\mathrm{Var}[\mathcal{I}^{(\nu)}]$ is smaller. The relationships between $\mathrm{Var}[\mathcal{I}^{(e)}]$ and $\mathrm{Var}[\mathcal{I}^{(s)}]$ is more complex, nominators as well as denominators are different, see (18) and (20). To find the difference $\mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C})$ is expressed in terms of $\mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C})$:
$\mathrm{tr}(\mathbf{M}^{(e)}\mathbf{C}\mathbf{M}^{(e)}\mathbf{C}) = \mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2-1}{(n-1)n^2}$.
Hence,

$$\begin{aligned}
\mathrm{Var}[\mathcal{I}^{(e)}|H_0] &= \frac{2\left\{(n-1-k)\left[\mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2-1}{(n-1)n^2}\right] - \mathrm{tr}(\mathbf{K}^{(e)})^2\right\}}{(n-1-k)^2(n-1-k+2)} \\
&= \mathrm{Var}[\mathcal{I}^{(s)}|H_0] \\
&\quad + \frac{2\left\{(n-1-k)\left[\mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) + \frac{2n^2-3n-2n(\mathrm{tr}(\mathbf{1}'\mathbf{C}\mathbf{C}\mathbf{1}))+1}{(n-1)n^2}\right] - 1\right\}}{(n-1-k)^2(n-k)} \\
&\quad - \frac{2(n-k)\mathrm{tr}(\mathbf{M}^{(s3)}\mathbf{C}\mathbf{M}^{(s3)}\mathbf{C}) - \frac{(n-1)^2}{n^2}}{(n-k)^2(n-k+2)}.
\end{aligned}$$

The magnitude of the difference depends on the concrete values of the spatial link matrix via $\mathrm{tr}(\mathbf{MCMC})$ and $\mathrm{tr}(\mathbf{11'CC})$ 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; unstandardized- and row-standardized spatial link matrices are given by:

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

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

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

(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:

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

These matrices correspond to case (i) where the far-off object is included in the analysis. If the far-off observation is excluded, case (ii), 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), of case (i) and (iii), 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, 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 (i) 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 (10), 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:

$$\mathrm{Var}[\mathcal{I}^{(s1)}|H_0] = \frac{2\left\{(n-k)\left[\frac{n-1}{n-2} - \frac{2}{n}(n-1) + \frac{2}{n^2}(n-1)^2\right] - \frac{(n-1)^2}{n^2}\right\}}{(n-k)^2(n-k+2)} = \frac{2n^2-3n+6}{n^2(n^2-n-2)}. \quad (21)$$

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, $\mathrm{tr}(\mathbf{K}^{(e1)})^2$ does not depend on the design points, it is the one given in (12). The variance of $\mathcal{I}^{(e1)}$ under the null is given by:

$$\mathrm{Var}[\mathcal{I}^{(e1)}|H_0] = \frac{2\left\{(n-1-k)\frac{n-1}{n-2} - 1\right\}}{(n-1-k)^2(n-1-k+2)} = \frac{2\left\{(n-2)\frac{n-1}{n-2} - 1\right\}}{(n-2)^2 n} = \frac{2}{n(n-2)}. \quad (22)$$

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 (22) and (21). 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 (i) and extreme case (2), $\text{tr}(\mathbf{K}^{(s2)}\mathbf{K}^{(s2)})$ and $\text{tr}(\mathbf{K}^{(s2)})^2$ are used.

$$\begin{aligned}
\text{tr}(\mathbf{K}^{(s2)}\mathbf{K}^{(s2)}) &= \text{tr}(\mathbf{M}^{(s3)}\mathbf{C}^{(2)}\mathbf{M}^{(s3)}\mathbf{C}^{(2)}) \\
&= \text{tr}(\mathbf{C}^{(2)}\mathbf{C}^{(2)}) - \frac{2}{n}\text{tr}(\mathbf{11}'\mathbf{C}^{(2)}\mathbf{C}^{(2)}) + \frac{2}{n^2}\text{tr}(\mathbf{11}'\mathbf{C}^{(2)}\mathbf{11}'\mathbf{C}^{(2)}) \\
&= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{n^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\
&= (n-1) - \frac{2}{n}(n-1) + \frac{2}{n^2}(n-1)^2
\end{aligned}$$

$\text{tr}(\mathbf{K}^{(s2)})^2$ is given in (10).

$$\text{Var}[\mathcal{I}^{(s2)}|H_0] = \frac{2\{(n-k)[n-1+\frac{2}{n^2}-\frac{2}{n}] - \frac{(n-1)^2}{n2}\}}{(n-k)^2(n-k+2)} = \frac{2(n^2-3)}{n^2(n+1)}. \tag{23}$$

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

$$\begin{aligned}
\text{tr}(\mathbf{K}^{(e2)}\mathbf{K}^{(e2)}) &= \text{tr}(\mathbf{M}^{(e)}\mathbf{C}^{(2)}\mathbf{M}^{(e)}\mathbf{C}^{(2)}) \\
&= \text{tr}(\mathbf{C}^{(2)}\mathbf{C}^{(2)}) - \frac{2}{n}\text{tr}(\mathbf{11}'\mathbf{C}^{(2)}\mathbf{C}^{(2)}) + \frac{2}{n^2}\text{tr}(\mathbf{11}'\mathbf{C}^{(2)}\mathbf{11}'\mathbf{C}^{(2)}) \\
&= \sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}^2 - \frac{2}{n}\sum_{i=2}^{n}\left(\sum_{j=2}^{n} g_{ij}\right)^2 + \frac{2}{n^2}\left(\sum_{i=2}^{n}\sum_{j=2}^{n} g_{ij}\right)^2 \\
&= (n-1) - \frac{2}{(n-1)}(n-1) + \frac{2}{(n-1)^2}(n-1)^2 = (n-1)
\end{aligned}$$

and $\text{tr}(\mathbf{K}^{(e1)})^2 = (-1)^2$, see (12), this leads to

$$\text{Var}[\mathcal{I}^{(e2)}|H_0] = \frac{2n(n-2)+2k(1-n)}{(n-1-k)^2(n-k+1)} = \frac{2(n^2-3n+1)}{n(n-2)^2}. \tag{24}$$

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 (24) and (23), 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:

$$\text{Var}[\mathcal{I}^{(\nu)}|H_0] < \text{Var}[\mathcal{I}^{(s)}|H_0] < \text{Var}[\mathcal{I}^{(e)}|H_0]. \tag{25}$$

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

The test statistic for the Moran's test $z(\mathcal{I})$ is given in (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 residuum whereas is 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 $\mathcal{I}^{(\nu)} - \mathrm{E}[\mathcal{I}^{(\nu)}]$ of $z[\mathcal{I}^{(\nu)}]$ is greater than the one of $z[\mathcal{I}^{(s)}]$, as $\mathrm{Var}[\mathcal{I}^{(\nu)}] < \mathrm{Var}[\mathcal{I}^{(\nu)}]$ it follows that

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

The same relation holds if $\hat{\varepsilon}_1$ is positive, if $\hat{\varepsilon}_1$ is negative the order of the z-values might change. Thus, adding a small value ν to all elements of the spatial link matrix has an influence on the Moran's test. If the residuum 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 residuum 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 (i) and (ii) 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 (i) and (ii) 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 1. For both extreme cases the difference between treatment (i) and (ii) 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 and 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 dependency 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 2, 3, 4, 5, 6, 7 and 8. 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 (i) and (ii) 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 difference between treatment (i) and (ii) 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 (ii) leads to a more conservative test, i.e. the test based on treatment (i) rejects the null hypothesis earlier. If the difference is negative, the test based on treatment (ii) 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 (i) rejects the null earlier in small designs, see Figure 2, the difference becomes negative if n increases. For medium and high values of $\mathcal{I}$ treatment (ii) rejects the null earlier, see Figure 4, 5, 6, 7, 8.

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

$$z[\mathcal{I}^{(e)}] \leq z[\mathcal{I}^{(s)}] \leq z[\mathcal{I}^{(\nu)}]. \tag{26}$$

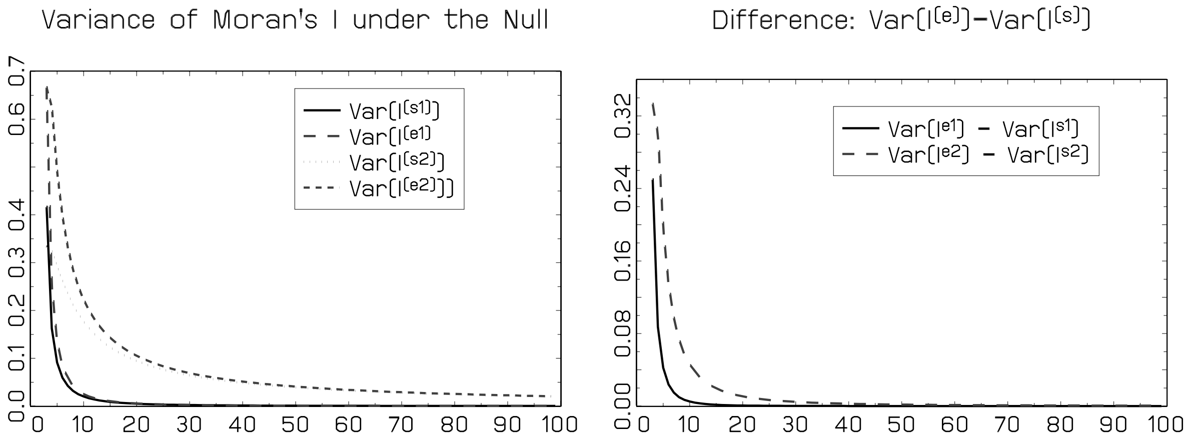

Figure 1: Variances of $\mathcal{I}$ and differences between the variances within the extreme cases

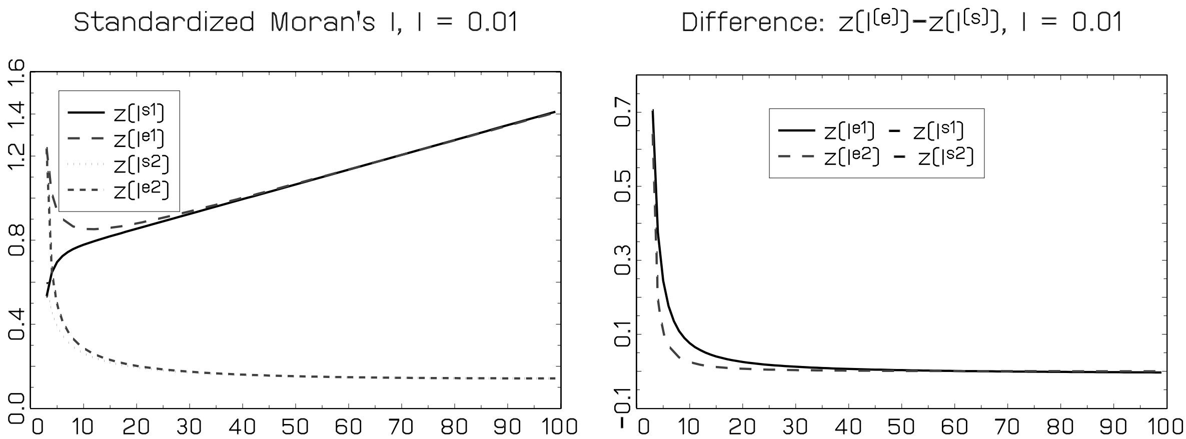

Figure 2: $z(\mathcal{I})$ with $\mathcal{I} = 0.01$ and differences between the $z(\mathcal{I})$ within the extreme cases

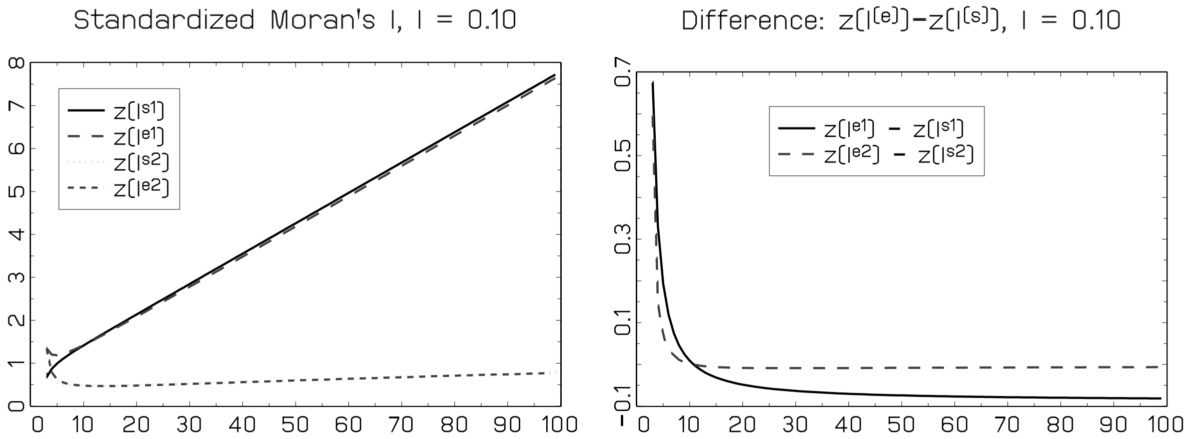

Figure 3: $z(\mathcal{I})$ with $\mathcal{I} = 0.10$ and differences between the $z(\mathcal{I})$ within the extreme cases

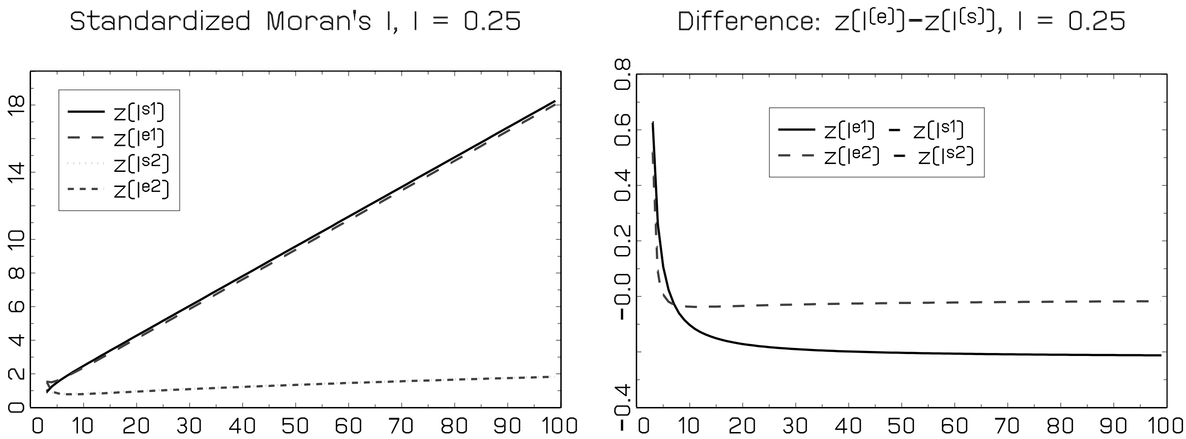

Figure 4: $z(\mathcal{I})$ with $\mathcal{I} = 0.25$ and differences between the $z(\mathcal{I})$ within the extreme cases

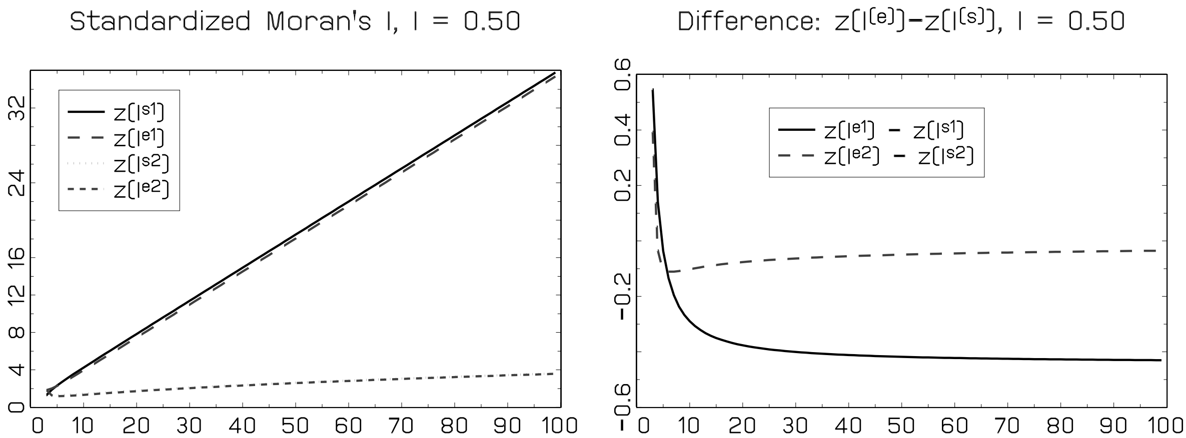

Figure 5: $z(\mathcal{I})$ with $\mathcal{I} = 0.50$ and differences between the $z(\mathcal{I})$ within the extreme cases

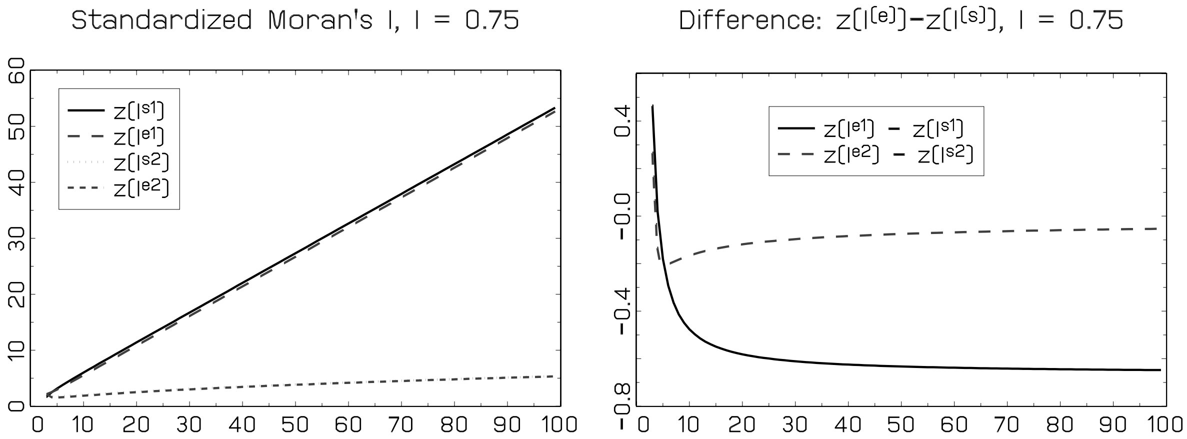

Figure 6: $z(\mathcal{I})$ with $\mathcal{I} = 0.75$ and differences between the $z(\mathcal{I})$ within the extreme cases

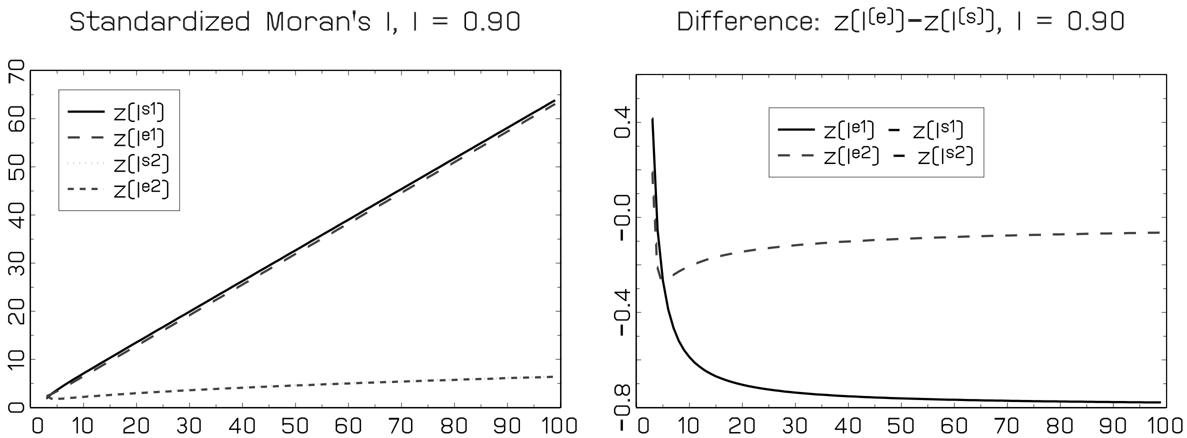

Figure 7: $z(\mathcal{I})$ with $\mathcal{I} = 0.90$ and differences between the $z(\mathcal{I})$ within the extreme cases

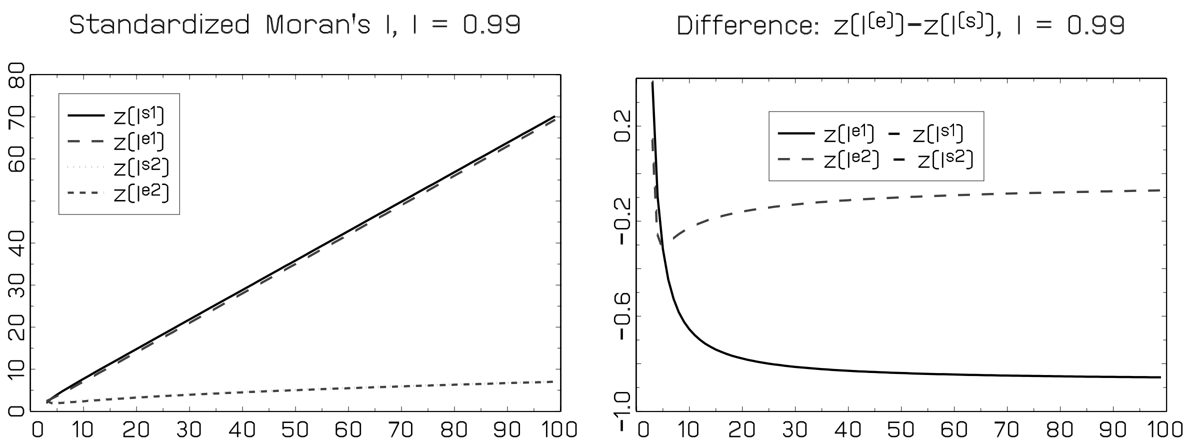

Figure 8: $z(\mathcal{I})$ with $\mathcal{I} = 0.99$ and differences between the $z(\mathcal{I})$ within the extreme cases

3 Conclusion

The standardized Moran's $\mathcal{I}$ depends on the values 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 $\mathcal{I}$, $\text{Var}[\mathcal{I}|H_0]$ and $z[\mathcal{I}]$ from the different methods of treating far-off objects given in (9), (25) and (26) do not hold in general as the assumptions that were made are very restrictive. Only the behavior the expected values $\text{E}[\mathcal{I}|H_0]$ holds without such assumptions, see (16). 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 et al. (2007). 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 (ii) is not recommended, because even if an observation is not connected to others it might be important in the design.

Acknowledgments

I would like to thank Werner G. Müller from the Johannes Kepler University Linz for his help within all steps during the preparation of the manuscript. I would also like to thank the Jubiläumsfonds of the OeNB (Austrian central bank) for funding the project "Common Structures in Spatial Econometrics" (project number 11052) in the course of which this paper arose.

References

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

D. Gumprecht, W. Müller, and J. Rodríguez-Díaz. Optimal Design for Detecting Spatial Dependence. *Submitted for publication*, 2007.

P. Moran. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37:17–23, 1950.

M. Tiefelsdorf. *Modelling Spatial Processes*. Springer, Berlin, 2000.

M. Tiefelsdorf and B. Boots. The Exact Distribution of Moran's I. *Environment and Planning A*, 27:985–999, 1995.