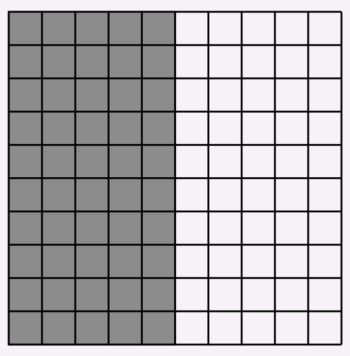
# Robustness of spatial autocorrelation tests


ERNST Marie — Department of Mathematics, University of Liege, Belgium — m.ernst@ulg.ac.be

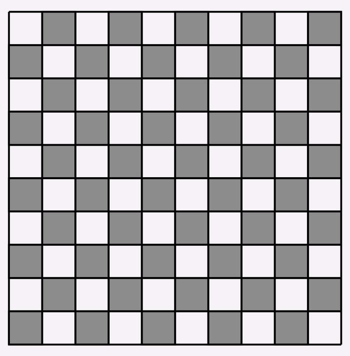
Joint work with HAESBROECK Gentiane


## Spatial autocorrelation

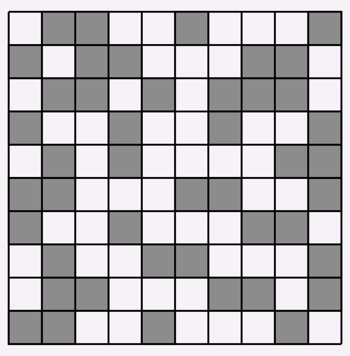
Measure of the dependence between values at neighbouring locations.



Positive spatial autocorrelation



Negative spatial autocorrelation



No spatial autocorrelation

### Notation

- Spatial process $\{Z(s_i) : s_i \in D\}$ over a fixed and discrete domain $D$.
- Sample data points $z = \{z_1, \dots, z_n\}$ for the spatial locations $\{s_1, \dots, s_n\}$.

**Weighting matrix** $W = (w_{ij})_{1 \le i,j \le n}$ describes spatial neighbours (not necessarily symmetric; with zero diagonal).

Notations: $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$ ; $\quad w_{i\bullet} = \sum_{j=1}^{n} w_{ij}$ ; $\quad w_{\bullet i} = \sum_{k=1}^{n} w_{ki}$

## Spatial autocorrelation indexes

Spatial autocorrelation measures usually used by geographers in the literature:

- Moran's index [9] is a global indicator of the spatial autocorrelation:

$$I(z) = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$$

- Geary's ratio [6] is based on comparisons between pairs of observations:

$$c(z) = \frac{n-1}{2S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - z_j)^2}{\sum_i (z_i - \bar{z})^2}$$

- Other index: general Getis and Ord's statistics [7]

## Spatial autocorrelation tests

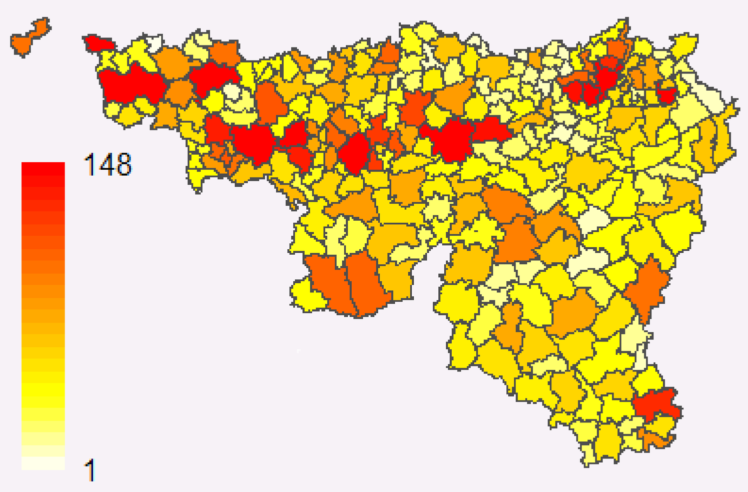
Let focus on tests based on **asymptotic normality** of $I$ and $c$ (see [4, 11]).

- Under N assumption: observations are the results of $n$ independent drawings from a normal population.
- Under R assumption: the set of values $\{z_1, \dots, z_n\}$ is fixed and observations are randomly permuted on the locations $\{s_1, \dots, s_n\}$.

**Other tests**: permutation tests and Dray's test [5]

## Illustration: school establishments

Number of school establishments in each Walloon municipality in 2008. Two municipalities are neighbors if they share a boundary.




| Name | Index | p-value |
|---|---|---|
| Moran | $I = 0.13$ | $7.4 \times 10^{-5}$ |
| Geary | $c = 1.48$ | 0.0007 |

## Robustness of the tests

As the set of locations is finite, one need to work with empirical devices instead of usual robust tools which are based on functional, i.e., correspond to asymptotic values.

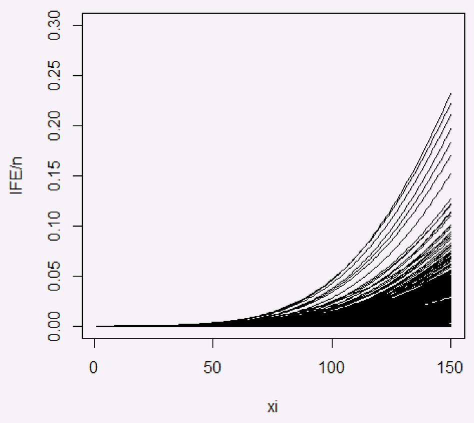
- **Resistance of a test** (analogous to BDP of an estimator)

  The resistance to acceptance (rejection)[12] of a test is the size of smallest subset of fixed values which always implies the acceptance (rejection) of $H_0$, no matter what the other values are.

  **Result:** Moran's and Geary's tests: $1/n$ and Getis and Ord's test: $2/n$.

- **Empirical influence function on p-value**

  Measure of the strength of the evidence against the decision to reject $H_0$:



$$IFE(\xi, i; I) = \frac{\text{p-value}(z + \xi e_i) - \text{p-value}(z)}{1/n}$$

**Result:** asymptotic test (under N) based on Moran

$$IFE(\xi, i; I) = 2n \left[ \Phi\left( - \left| \frac{n}{S_0 \sigma(I)} \frac{P(\xi)}{Q(\xi)} + \frac{1}{n\sigma(I)} \right| \right) - \Phi\left( - \left| \frac{I(z) - E(I)}{\sigma(I)} \right| \right) \right]$$

$$\rightarrow 2n \left[ \Phi\left( - \frac{2\,|S_0 - n w_{i\bullet}|}{(n-1) S_0 \sigma(I)} \right) - cste \right] \quad \text{if } \xi \rightarrow \infty$$

where $\sigma(I)$ is constant, $P(\xi)$ and $Q(\xi)$ are 2-degree polynomials in $\xi$.

**Other results**: under R assumption and using the other indexes.

## Robust alternatives

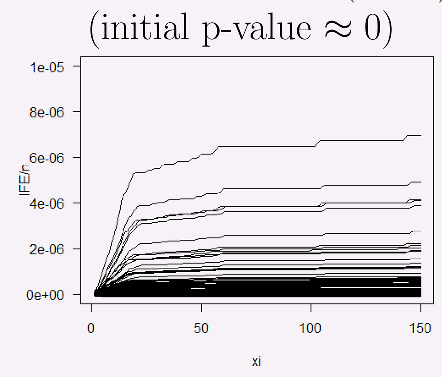
- **Test based on rank**

  Observations are replaced by their rank to compute the indexes. The asymptotic and permutation tests can be adapted.

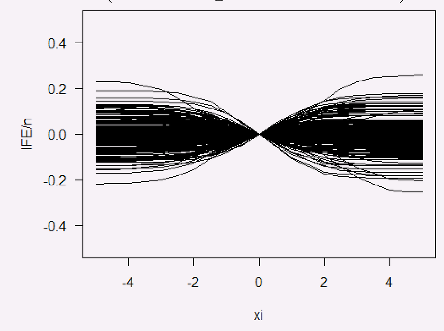
  **Robustness:**

  – Non significant impact of a unique contamination on empirical IF.

IFE on School data (zoom) (initial p-value $\approx 0$)



IFE on independent variables (initial p-value 0.53)



  – Resistance similar to BDP of rank correlation [2, 3].

- **Robust regression**

  Moran's index can be interpreted as the scope of a bivariate linear regression of the spatially lagged variable on the original variable [1].

  Robust regression can be used: Least Trimmed Squares [10] and M-estimator [8].

  **Robustness:**

  – No impact of a unique contamination on empirical IF.

  – Resistance linked to the BDP of the robust regression (resp. $h/n$ and at most 0.5) and to the sensibility of leverage points.

## On going research

- Comparison between the efficiency of classical tests and robust tests with or without contamination (using simulations).
- Testing the presence of a main direction, for instance Sambre-Meuse line in Wallonia.

### References :


[1] Anselin, L. (1996), The Moran Scatterplot [...], *Spatial Analytical Perspectives on GIS*: 111–125, London: Taylor and Francis.

[2] Boudt, K. et al. (2012). The Gaussian rank correlation estimator: robustness properties. *Statistics and Computing* 22(2): 471–483.

[3] Capéraà, P. and Guillem, A.I.G. (1997). Taux de résistance des tests de rang d'indépendance. *Canadian Journal of Stat.* 25(1): 113–124.

[4] Cliff, A.D. and Ord, J.K. (1973). Spatial autocorrelation, Pion London.

[5] Dray, S. (2011). A new perspective about Moran's coefficient [...]. *Geogr. analysis* 43(2): 127–141.

[6] Geary, R.C. (1954). The contiguity ratio and statistical mapping, *The incorporated statistician*: 115–146.

[7] Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geogr. analysis* 24(3): 189–206.

[8] Huber, P. (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. *Annals of Statistics* 1(5): 799–821.

[9] Moran, P.A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika* 37: 17–23.

[10] Rousseeuw, P.J., (1984). Least Median of Squares Regression. *Journal of the American statistical association* 79(388): 871–880.

[11] Sen, A. (1976). Large Sample-Size Distribution of Statistics Used In Testing for Spatial Correlation. *Geogr. analysis* 8(2): 175–184.

[12] Ylvisaker, D. (1977). Test resistance. *Journal of the American Statistical Association* 72(359): 551–556.