Spatial autocorrelation: robustness of measures and tests

Marie Ernst and Gentiane Haesbroeck

University of Liege

London, December 14, 2015
Spatial Data

Spatial data:

- geographical positions
- non spatial attributes

Example

Mean price of land for sales (€/m²) in Belgian municipalities

Spatial Data

Spatial data:

- geographical positions
- non spatial attributes

Example
Mean price of land for sales (€/m²) in Walloon municipalities

Spatial autocorrelation

- Positive spatial autocorrelation
- Negative spatial autocorrelation
- No spatial autocorrelation

Weighting matrix $W$

- Binary weights,
- Row-standardized,
- Globally standardized, . . .

Convention: zero diagonal and $S_0 = \sum_i \sum_j W_{ij}$
Spatial autocorrelation

- Positive spatial autocorrelation
- Negative spatial autocorrelation
- No spatial autocorrelation

Weighting matrix $W$

- Binary weights,
- Row-standardized,
- Globally standardized, . . .

Convention: zero diagonal and $S_0 = \sum_i \sum_j W_{ij}$
Measures of spatial autocorrelation
The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$
Measures of spatial autocorrelation

The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$

Moran’s Index (1950)

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

The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$

Moran’s Index (1950)

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

The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$

Moran’s Index (1950)

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

$$\sum_{i=1}^{n} (z_i - \bar{z})^2$$

Geary’s ratio (1954)

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

$$\sum_{i=1}^{n} (z_i - \bar{z})^2$$
Measures of spatial autocorrelation

The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$

Moran’s Index (1950)

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

Geary’s ratio (1954)

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

The values \( z_1, \ldots, z_n \) are observed at locations \( s_1, \ldots, s_n \)

Moran’s Index (1950)

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

Geary’s ratio (1954)

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

Getis and Ord’s statistics (1992)

\[
G(z) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} \sum_{j=1,j \neq i}^{n} z_i z_j}
\]
Measures of spatial autocorrelation

The values $z_1, \ldots, z_n$ are observed at locations $s_1, \ldots, s_n$

**Moran’s Index (1950)**

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

**Geary’s ratio (1954)**

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

**Getis and Ord’s statistics (1992)**

$$G(z) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j}$$
Inference

Tests based on asymptotic normality

Without spatial autocorrelation, \( I, c \) and \( G \) are asymptotically **Gaussian** under normality (N) and/or randomisation (R) assumption.

Package spdep in R:

- `moran.test`: under N and R assumption
- `geary.test`: under N and R assumption
- `globalG.test`: under R assumption

Example

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Expected value</th>
<th>Variance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I = 0.53 )</td>
<td>-0.0038</td>
<td>0.0006</td>
<td>( &lt; 2 \times 10^{-16} )</td>
</tr>
<tr>
<td>( c = 0.42 )</td>
<td>1</td>
<td>0.0008</td>
<td>( &lt; 2 \times 10^{-16} )</td>
</tr>
<tr>
<td>( G = 0.046 )</td>
<td>0.038</td>
<td>( 2 \times 10^{-7} )</td>
<td>( &lt; 2 \times 10^{-16} )</td>
</tr>
</tbody>
</table>
Inference

Tests based on asymptotic normality

Without spatial autocorrelation, \( I \), \( c \) and \( G \) are asymptotically Gaussian under normality (N) and/or randomisation (R) assumption.

Package `spdep` in R:
- `moran.test`: under N and R assumption
- `geary.test`: under N and R assumption
- `globalG.test`: under R assumption

Example

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Expected value</th>
<th>Variance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I = 0.53 )</td>
<td>-0.0038</td>
<td>0.0006</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>( c = 0.42 )</td>
<td>1</td>
<td>0.0008</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>( G = 0.046 )</td>
<td>0.038</td>
<td>(2 \times 10^{-7})</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
</tbody>
</table>
Tests based on asymptotic normality

Without spatial autocorrelation, $I$, $c$ and $G$ are asymptotically Gaussian under normality (N) and/or randomisation (R) assumption.

Package spdep in R:

- `moran.test`: under N and R assumption
- `geary.test`: under N and R assumption
- `globalG.test`: under R assumption

Example

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Expected value</th>
<th>Variance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I = 0.53$</td>
<td>-0.0038</td>
<td>0.0006</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>$c = 0.42$</td>
<td>1</td>
<td>0.0008</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>$G = 0.046$</td>
<td>0.038</td>
<td>$2 \times 10^{-7}$</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
</tbody>
</table>
Inference

Permutation tests
Computation of the index using random permutations of the variable on the spatial domain.
Permutation tests
Computation of the index using random permutations of the variable on the spatial domain.

\[
\begin{array}{c|c|c|c}
1 & 2 & 3 \\
4 & 5 & 6 \\
\end{array}
\Rightarrow
\begin{array}{c|c|c|c}
2 & 5 & 6 \\
1 & 3 & 4 \\
\end{array}
\]
Inference

Permutation tests
Computation of the index using random permutations of the variable on the spatial domain.

\[
p-value = \frac{\#\{I_{permuted} \geq I_{obs}\}}{\#\{simulations\} + 1}
\]
Permutation tests
Computation of the index using random permutations of the variable on the spatial domain.

\[
p-value = \frac{\#\{I_{permuted} \geq I_{obs}\}}{\#\{simulations\} + 1}
\]

- **Benefits**: No assumption on the distribution
- **Drawbacks**: Simulations (randomness, computational cost,...)
Inference

Permutation tests
Computation of the index using random permutations of the variable on the spatial domain.

\[
p-value = \frac{\# \{ I_{\text{permuted}} \geq I_{\text{obs}} \}}{\#\{\text{simulations}\} + 1}
\]

- **Benefits**: No assumption on the distribution
- **Drawbacks**: Simulations (randomness, computational cost,...)

Example (nsim=5000)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed rank</th>
<th>p-value</th>
<th>Package spdep in R:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I = 0.53$</td>
<td>5001</td>
<td>$2 \times 10^{-4}$</td>
<td>moran.mc</td>
</tr>
<tr>
<td>$c = 0.42$</td>
<td>1</td>
<td>$2 \times 10^{-4}$</td>
<td>geary.mc</td>
</tr>
<tr>
<td>$G = 0.046$</td>
<td>5001</td>
<td>$2 \times 10^{-4}$</td>
<td>/</td>
</tr>
</tbody>
</table>
Inference

Dray’s test based on $I$

Decomposition of positive / negative influence on autocorrelation

$$I(z) = \sum_{I(u_k) < E[I]} I(u_k) \text{cor}^2(u_k, z) + \sum_{I(u_k) > E[I]} I(u_k) \text{cor}^2(u_k, z)$$

$$S_I^-(z) \quad S_I^+(z)$$

where $u_k$ are eigenvector of $HWH$, $H = I_n - \frac{1}{n}11'$.

- Simultaneous positive and negative spatial autocorrelation
- Permutation test

Example (nsim=5000)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed rank</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I = 0.53$</td>
<td>5001</td>
<td>$2 \times 10^{-4}$</td>
</tr>
<tr>
<td>$S_I^+ = 0.57$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_I^- = -0.04$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inference

Dray’s test based on $I$

Decomposition of positive / negative influence on autocorrelation

$$I(z) = \sum_{I(u_k) < E[I]} I(u_k) \text{cor}^2(u_k, z) + \sum_{I(u_k) > E[I]} I(u_k) \text{cor}^2(u_k, z)$$

$$S_I^-(z) + S_I^+(z)$$

where $u_k$ are eigenvector of $HWH$, $H = I_n - \frac{1}{n} 11'$.

- Simultaneous positive and negative spatial autocorrelation
- Permutation test

Example ($\text{nsim}=5000$)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed rank</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I = 0.53$</td>
<td>5001</td>
<td>$2 \times 10^{-4}$</td>
</tr>
<tr>
<td>$S_I^+ = 0.57$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_I^- = -0.04$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Inference

<table>
<thead>
<tr>
<th>Moran’s $I$</th>
<th>Geary’s $c$</th>
<th>Getis and Ord’s $G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test under $R$</td>
<td>Test under $R$</td>
<td>Test under $R$</td>
</tr>
<tr>
<td>Test under $N$</td>
<td>Test under $N$</td>
<td></td>
</tr>
<tr>
<td>Permutation test</td>
<td>Permutation test</td>
<td>Permutation test</td>
</tr>
<tr>
<td>Dray’s test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Robustness of the tests

How contamination influences the result of the tests?

Impact on the decision of the test

Example
Robustness of the tests

How contamination influences the result of the tests?

Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I under R</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>I under N</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>I Permutation</td>
<td>$&lt; 2 \times 10^{-4}$</td>
</tr>
<tr>
<td>Dray test</td>
<td>$&lt; 2 \times 10^{-4}$</td>
</tr>
<tr>
<td>c under R</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>c under N</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>c Permutation</td>
<td>$&lt; 2 \times 10^{-4}$</td>
</tr>
<tr>
<td>G under R</td>
<td>$&lt; 2 \times 10^{-16}$</td>
</tr>
<tr>
<td>G Permutation</td>
<td>$&lt; 2 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Robustness of the tests

How contamination influences the result of the tests?

Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>I under R</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>I under N</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>I Permutation</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>Dray test</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>c under R</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>c under N</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>c Permutation</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>G under R</td>
<td>$RH_0$</td>
</tr>
<tr>
<td>G Permutation</td>
<td>$RH_0$</td>
</tr>
</tbody>
</table>
Robustness of the tests

How contamination influences the result of the tests?

Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>I under R</td>
<td>RH₀</td>
</tr>
<tr>
<td>I under N</td>
<td>RH₀</td>
</tr>
<tr>
<td>I Permutation</td>
<td>RH₀</td>
</tr>
<tr>
<td>Dray test</td>
<td>RH₀</td>
</tr>
<tr>
<td>c under R</td>
<td>RH₀</td>
</tr>
<tr>
<td>c under N</td>
<td>RH₀</td>
</tr>
<tr>
<td>c Permutation</td>
<td>FTR</td>
</tr>
<tr>
<td>G under R</td>
<td>RH₀</td>
</tr>
<tr>
<td>G Permutation</td>
<td>RH₀</td>
</tr>
</tbody>
</table>

Paliseul: 690€/m²
Robustness of the tests

How contamination influences the result of the tests?

Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>l under R</td>
<td>RH₀</td>
</tr>
<tr>
<td>l under N</td>
<td>RH₀</td>
</tr>
<tr>
<td>l Permutation</td>
<td>RH₀</td>
</tr>
<tr>
<td>Dray test</td>
<td>RH₀</td>
</tr>
<tr>
<td>c under R</td>
<td>FTR</td>
</tr>
<tr>
<td>c under N</td>
<td>RH₀</td>
</tr>
<tr>
<td>c Permutation</td>
<td>FTR</td>
</tr>
<tr>
<td>G under R</td>
<td>RH₀</td>
</tr>
<tr>
<td>G Permutation</td>
<td>RH₀</td>
</tr>
</tbody>
</table>

Paliseul: 730€/m²
Robustness of the tests

How contamination influences the result of the tests?
Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{I} ) under R</td>
<td>FTR</td>
</tr>
<tr>
<td>( \bar{I} ) under N</td>
<td>( RH_0 )</td>
</tr>
<tr>
<td>( \bar{I} ) Permutation</td>
<td>( RH_0 )</td>
</tr>
<tr>
<td>Dray test</td>
<td>( RH_0 )</td>
</tr>
<tr>
<td>( c ) under R</td>
<td>FTR</td>
</tr>
<tr>
<td>( c ) under N</td>
<td>FTR</td>
</tr>
<tr>
<td>( c ) Permutation</td>
<td>FTR</td>
</tr>
<tr>
<td>( G ) under R</td>
<td>( RH_0 )</td>
</tr>
<tr>
<td>( G ) Permutation</td>
<td>( RH_0 )</td>
</tr>
</tbody>
</table>

Paliseul \( \approx \) max(Belgium) (1200€)
Robustness of the tests

How contamination influences the result of the tests?
Impact on the decision of the test

Example

<table>
<thead>
<tr>
<th>Test</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ under R</td>
<td>FTR</td>
</tr>
<tr>
<td>/ under N</td>
<td>FTR</td>
</tr>
<tr>
<td>/ Permutation</td>
<td>FTR</td>
</tr>
<tr>
<td>Dray test</td>
<td>FTR</td>
</tr>
<tr>
<td>c under R</td>
<td>FTR</td>
</tr>
<tr>
<td>c under N</td>
<td>FTR</td>
</tr>
<tr>
<td>c Permutation</td>
<td>FTR</td>
</tr>
<tr>
<td>G under R</td>
<td>FTR</td>
</tr>
<tr>
<td>G Permutation</td>
<td>FTR</td>
</tr>
</tbody>
</table>

Paliseul: 1610€/m²
Robustness of the tests

Resistance of a test (Ylvisaker, 1977)

- **Resistance to acceptance**: smallest proportion of fixed observations which will always implies the *non rejection* of $H_0$.
- **Resistance to rejection**: smallest proportion $m_0$ of fixed observations which will always implies the *rejection* of $H_0$.

Result: Resistance to rejection of considered tests: $\frac{1}{n}$
Robustness of the tests

Impact on the $p$-value

Hairplot: $\xi \mapsto p\text{-value}(z + \xi e_i)$ for $i = 1, \ldots, n$
Robustness of the tests

Impact on the p-value

Hairplot: $\xi \mapsto \text{p-value}(z + \xi e_i)$ for $i = 1, \ldots, n$
Robustness of the tests

Impact on the p-value

Hairplot: $\xi \mapsto \text{p-value}(z + \xi e_i)$ for $i = 1, \ldots, n$
Robustness of the tests

Impact on the p-value

Hairplot: $\xi \mapsto \text{p-value}(z + \xi e_i)$ for $i = 1, \ldots, n$

$\Rightarrow$ robust alternative must be considered
Robust Moran

Moran scatterplot

Moran can be interpreted as the slope in a LS regression of $Wz$ over $z$
Robust Moran

Moran scatterplot

Moran can be interpreted as the slope in a LS regression of $Wz$ over $z$
Robust Moran

Moran scatterplot

Moran can be interpreted as the slope in a LS regression of $Wz$ over $z$

Idea: Robustly estimate this slope

Robust techniques

- Least Trimmed Squares (Rousseeuw, 1984)
- M-estimator (Huber, 1973)
Robust regression

Regression: $y = ax + b$ and residuals: $r_i = y_i - \hat{y}_i$

Least squares regression
Minimisation of $\sum_{i=1}^{n} r_i^2$

Least Trimmed Squares (Rousseeuw, 1984)
Minimisation of the trimmed sum: $\sum_{i=1}^{n-h} r_{(i)}^2$

M-estimator (Huber, 1973)
Minimisation of another function: $\sum_{i=1}^{n} \rho \left( \frac{r_i}{\hat{\sigma}} \right)$
e.g., $\rho(r) = r^2$, $\rho(r) = |r|$
Robust regression

Regression: $y = ax + b$ and residuals: $r_i = y_i - \hat{y}_i$

Least squares regression
Minimisation of $\sum_{i=1}^{n} r_i^2$

Least Trimmed Squares (Rousseeuw, 1984)
Minimisation of the trimmed sum: $\sum_{i=1}^{n-h} r_{(i)}^2$

M-estimator (Huber, 1973)
Minimisation of another function: $\sum_{i=1}^{n} \rho \left( \frac{r_i}{\hat{\sigma}} \right)$
e.g., $\rho(r) = r^2$, $\rho(r) = |r|$
Robust regression

Regression: $y = ax + b$ and residuals: $r_i = y_i - \hat{y}_i$

Least squares regression
Minimisation of $\sum_{i=1}^{n} r_i^2$

Least Trimmed Squares (Rousseeuw, 1984)
Minimisation of the trimmed sum: $\sum_{i=1}^{n-h} r_{(i)}^2$

M-estimator (Huber, 1973)
Minimisation of another function: $\sum_{i=1}^{n} \rho \left( \frac{r_i}{\hat{\sigma}} \right)$
e.g., $\rho(r) = r^2$, $\rho(r) = |r|$
Robust regression

Regression: $y = ax + b$ and residuals: $r_i = y_i - \hat{y}_i$

Least squares regression
Minimisation of $\sum_{i=1}^{n} r_i^2$

Least Trimmed Squares (Rousseeuw, 1984)
Minimisation of the trimmed sum: $\sum_{i=1}^{n-h} r_{(i)}^2$

M-estimator (Huber, 1973)
Minimisation of another function: $\sum_{i=1}^{n} \rho \left( \frac{r_i}{\hat{\sigma}} \right)$
e.g., $\rho(r) = r^2$, $\rho(r) = |r|$
Robust inference based on Moran

Impact on the decision of the test: resistance

Linked to the Breakdown point of the robust regression

Least Trimmed Squares  M-estimator

\[ \frac{h}{n} \leq 0.5 \]

Impact on the p-value: hairplot
Robust inference based on Moran

Impact on the decision of the test: resistance
Linked to the Breakdown point of the robust regression
Least Trimmed Squares M-estimator
\[ h/n \leq 0.5 \]

Impact on the p-value: hairplot
Robust inference based on Moran

Impact on the decision of the test: resistance
Linked to the Breakdown point of the robust regression
Least Trimmed Squares M-estimator
\[ h/n \leq 0.5 \]

Impact on the p-value: hairplot

What about level and power?
Robust inference based on Moran

Impact on the decision of the test: resistance

Linked to the Breakdown point of the robust regression

Least Trimmed Squares M-estimator

\[ \frac{h}{n} \leq 0.5 \]

Impact on the p-value: hairplot

What about level and power? \( \Rightarrow \) simulations
Simulation

Model
The Spatial autoregressive (SAR) model: \( Z = \rho WZ + \varepsilon \)

Simulation settings (Holmberg and Lundevaller, 2015)

- Spatial domain: grid \((n = 100, 400, 900, 1600)\)
- Weighting matrix: queen and rook contiguity
  
  ![Contiguity Diagram]

- Spatial correlation coefficient: \( \rho = 0, 0.1, 0.2, 0.3 \)
- Distribution of \( \varepsilon \): \( N(0, 1), \mathcal{P}(1), \text{Bern}(0.5) \)

Contamination process
A number \( \delta \) of observations are contaminated
Simulations

Example

\( n = 100, \) queen contiguity, \( \rho = 0.2 \) and \( \varepsilon \sim N(0, 1) \)

\[ \delta = 0 \]

\[ \delta = 1 \]
Comparison

1. Without contamination \((\delta = 0)\)
   Similar results are expected for classical and robust tests

2. With contamination \((\delta > 0)\)
   Robust tests should resist to contamination and be powerful
Results ($\delta = 0$)

Robust tests vs initial tests
Good correlation between the p-value of robust and classical Moran's tests ($\geq 90\%$)

Level ($n = 100$)

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>W</th>
<th>Moran</th>
<th>M-estimator</th>
<th>LTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAD</td>
<td>Huber</td>
<td>75%</td>
</tr>
<tr>
<td>N(0,1) Queen</td>
<td>0.055</td>
<td>0.061</td>
<td>0.062</td>
<td>0.072</td>
</tr>
<tr>
<td>N(0,1) Rook</td>
<td>0.051</td>
<td>0.051</td>
<td>0.050</td>
<td>0.053</td>
</tr>
<tr>
<td>$\mathcal{P}(1)$ Queen</td>
<td>0.044</td>
<td>0.054</td>
<td>0.053</td>
<td>0.080</td>
</tr>
<tr>
<td>$\mathcal{P}(1)$ Rook</td>
<td>0.052</td>
<td>0.056</td>
<td>0.055</td>
<td>0.057</td>
</tr>
<tr>
<td>Bern(0.5) Queen</td>
<td>0.067</td>
<td>0.086</td>
<td>0.087</td>
<td>0.091</td>
</tr>
<tr>
<td>Bern(0.5) Rook</td>
<td>0.041</td>
<td>0.061</td>
<td>0.050</td>
<td>0.076</td>
</tr>
</tbody>
</table>
Results ($\delta = 0$)

**Power**

proportion of rejected null hypothesis with $\rho = 0.1$, 0.2 or 0.3
Results ($\delta = 0$)

Power

proportion of rejected null hypothesis with $\rho = 0.1$, 0.2 or 0.3
Results ($\delta > 0$)

500 simulations for $n = 100$, $\varepsilon \sim N(0, 1)$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>W</th>
<th>Moran</th>
<th>M-estimator</th>
<th>LTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Queen</td>
<td>0.064</td>
<td>0.072</td>
<td>0.140</td>
</tr>
<tr>
<td>1</td>
<td>Rook</td>
<td>0.026</td>
<td>0.030</td>
<td>0.094</td>
</tr>
<tr>
<td>2</td>
<td>Queen</td>
<td>0.050</td>
<td>0.068</td>
<td>0.134</td>
</tr>
<tr>
<td>2</td>
<td>Rook</td>
<td>0.040</td>
<td>0.052</td>
<td>0.096</td>
</tr>
<tr>
<td>3</td>
<td>Queen</td>
<td>0.054</td>
<td>0.062</td>
<td>0.104</td>
</tr>
<tr>
<td>3</td>
<td>Rook</td>
<td>0.054</td>
<td>0.076</td>
<td>0.086</td>
</tr>
<tr>
<td>4</td>
<td>Queen</td>
<td>0.026</td>
<td>0.048</td>
<td>0.092</td>
</tr>
<tr>
<td>4</td>
<td>Rook</td>
<td>0.050</td>
<td>0.064</td>
<td>0.090</td>
</tr>
<tr>
<td>5</td>
<td>Queen</td>
<td>0.058</td>
<td>0.072</td>
<td>0.124</td>
</tr>
<tr>
<td>5</td>
<td>Rook</td>
<td>0.040</td>
<td>0.058</td>
<td>0.062</td>
</tr>
<tr>
<td>10</td>
<td>Queen</td>
<td>0.052</td>
<td>0.060</td>
<td>0.072</td>
</tr>
<tr>
<td>10</td>
<td>Rook</td>
<td>0.068</td>
<td>0.074</td>
<td>0.078</td>
</tr>
</tbody>
</table>
Results ($\delta > 0$)

500 simulations for $n = 100$, $\varepsilon \sim N(0, 1)$

Power

Graphs showing the power of different estimators for $n=100$, $\rho=0.1$, with 'rook' and 'queen' Gaussian distributions.
Results ($\delta > 0$)

500 simulations for $n = 100$, $\varepsilon \sim N(0, 1)$

Power

![Graphs showing power for different delta values with different estimators: M-est (MAD), M-est (Huber), LTS (75%), LTS (95%).](image)
Results ($\delta > 0$)

500 simulations for $n = 100$, $\varepsilon \sim N(0, 1)$

**Power**

![Graphs showing power for different estimators with $n=100$, $\rho=0.3$, and different distance measures (rook, queen).](image-url)
Results ($\delta > 0$)

500 simulations for $n = 100$, $\varepsilon \sim N(0, 1)$

Preliminary observations

- Robust test using Least Trimmed Square gives good results
- Robust test using M-estimator is more sensitive to leverage points
Perspective

In progress

- Simulations and comparison need to pursue
- Proposition of a robust test based on the others statistics

Perspective

- Study robustness of local measures and local tests
In progress

- Simulations and comparison need to pursue
- Proposition of a robust test based on the others statistics
  For Geary’s $c$:
  - $\hat{\gamma}(h) = S_z^2 c$ for $\hat{\gamma}(h)$ the empirical variogram
  - Robust estimation of the variogram using $Q_n$
  - $\hat{c} \propto \frac{1}{2} \left( \frac{Q_n(V_n)}{Q_n(z)} \right)^2$

Perspective

- Study robustness of local measures and local tests
Spatial autocorrelation:

Robustness:

References