Spatial analysis for better policies

Introduction: Spatial data analysis and its implications

R Bergs

ROBUST workshop Essen, 21 February 2019

ROBUST receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727988.

Purpose and reference to ROBUST

- Overcoming the information deficit in coarse regional data from official sources - need for feasible micro-spatial data processing to ensure prudent local policies

- Spatial functionality, spatial regularities, spatial heterogeneity and spatial dependence can be made visible for different fields of analysis, such as
  - functional segmentation (urban … rural);
  - Change over time by descriptive analysis;
  - spatial autocorrelation; and
  - and different procedures of spatial regression.

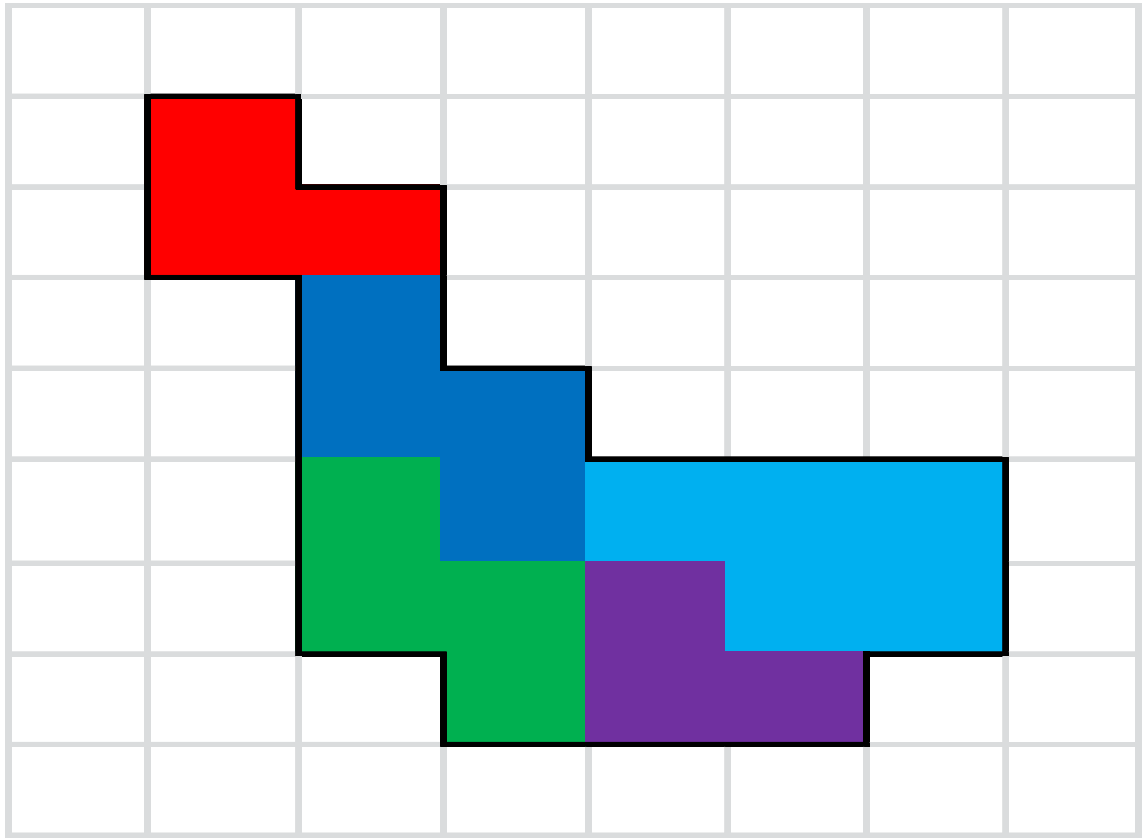
- **Stakeholders and policy representatives should recognize the enormous added value of using and processing micro-spatial data**

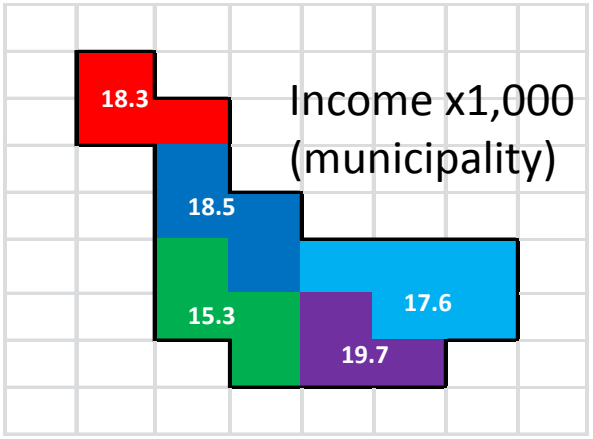
A brief (over-simplified) example to illustrate the purpose

- Just open Excel and build a database with two moderately or stronger correlated variables and let's say 20-30 or, if time allows, >500 observations
- Run a simple linear OLS regression by the Excel-RGP-command and look at the resulting coefficient, the constant, the standard errors and the R^2
- Then combine the observations into five groups by calculating their means and then run the same procedure
- You might find very different coefficients, standard errors and goodness of fit

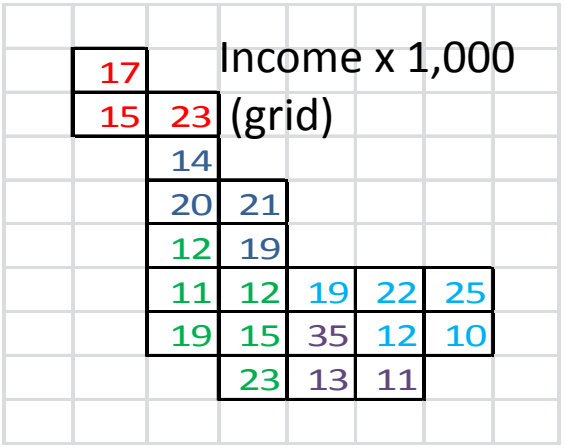
Just look and test it on your own PC …

A district with five municipalities

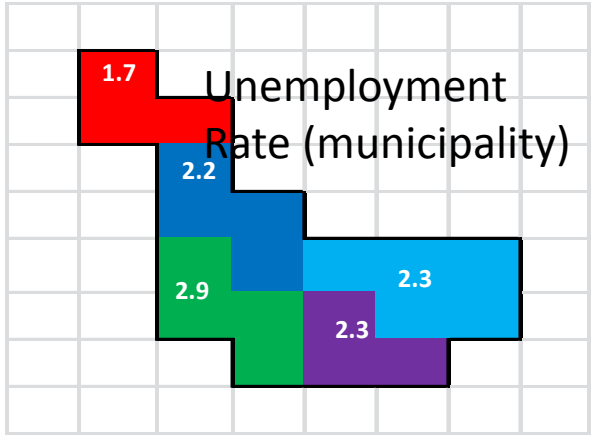




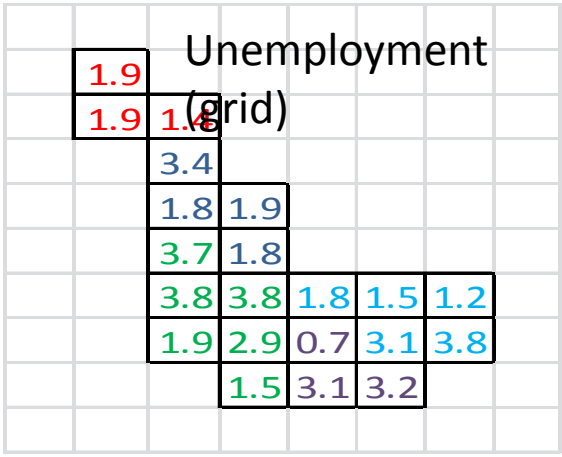
*Income=-2.5*Unemployment rate + 23.73*
(1.61) (3.74)



*Income=-5.6*Unemployment rate + 30.88*
(0.61) (1.55)



R^2=0.45



R^2=0.81

Findings *

- The left side shows an (insignificant**) estimate based on spatial mean values, where income is regressed on unemployment.
- The right side shows the same relationship based on the „true“ spatial variance of the data (the standard errors and R^2 reveal a stronger significance and determination)
- The negative coefficient on the right side is stronger, i.e. unemployment appears to be a significantly more serious problem than suggested by the mean value estimate.

* For the purpose of simplification the regression does not contain spatial distance/contiguity effects.
** E.g a higher number of observations with same means and proportion would show the same estimate, but then a significant one.

What does this example tell us?

- Mean values of variables conceal the real relationships and influences among variables because the spread underneath the average is unobserved
- Sufficient significance and a larger R^2 can appear for both types of models (with means and raw data), but coefficients may still differ.
- A policy conclusion determined from the model with means might be inadequate
- In former times, when micro-spatial data were not available, the risk of poor local policy conclusions was thus stronger.

What else does this example tell us?

- Small-scale spatial data are of utmost relevance for local policy representatives.
- Local policy representatives and other related stakeholders should recognize the information potential of the data and should understand the underlying logic of statistical procedures in spatial analysis, especially the rural-urban relationship.
- Local spatial data analysis and econometrics are worth to become a participatory science

But …

- Spatial analysis, especially spatial econometrics, has peculiar properties that need to be carefully considered when trying to understand the logic under
  - contiguity or distance effects of the dependent variable (Spatial Autoregressive or Spatial Error Model),
  - …such spatial effects of the independent variables (SLX model), or
  - …such spatial effects for both, dependent and independent variables (Spatial Durbin Model or the General Nesting Spatial Model)
- The choice and specification is usually demanding and to be advised by an econometrician
- So, results need to be well illustrated and imparted, so that sound policy decisions are possible.

Some more important features of micro-spatial data

- Population grid data (such as microm) allow a better classification of functional space (e.g. urban – peri-urban – rural)
- In combination with factor flow analyses (e.g. commuters among municipalities) it is possible to translate factor flows at the level of functional classes of space
- Night satellite images (e.g. VIIRS) are closely correlated with economic activity and spatial function and allow a segmentation of natural cities and non-urban space and its change by GIS and image analysis
- Further to that, change of pixel values (or radiance at a 1 km grid) allows observation or future simulation of spatial dependence at smallest spatial scale and thus informs about possible risks of urban sprawl and environmental hazard implied by investment into the built environment

Participatory data analysis

"… Sequential and iterative approaches—which we call more specifically *participatory econometrics*—seek varying degrees of dialogue between the qualitative and quantitative traditions at all phases of the research cycle […]. Respondents should be actively involved in the analysis and interpretation of findings. …"

V. Rao & M. Woolcock (2003) Integrating qualitative and quantitative approaches in program evaluation. In: Bourguignon FJ, Pereira da Silva L (eds.) The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools. New York: Oxford University Press, 165–190

Participatory data analysis

- Participatory spatial econometrics should generate policy relevant information by inferential analysis through the use of local data at small spatial scale. Data and methods are subject to a participatory reflection.
- It is not that much an approach of statistical sophistication in public authorities but rather awareness raising and better understanding.
- The objective is that stakeholders understand the estimates, their reliability, and the different steps to obtain such results.
- At least one regional representative needs to be qualified in spatial econometrics to facilitate the discussion and to do the analyses. Since every partner region is also represented by a research partner this should be ensured.