Spatial Eigenfunction Modeling of Geo-Referenced Data in the Social Sciences

Sebastian Juhl

University of Mannheim

June 24, 2021

Motivation

- many phenomena of interest to social scientists cluster in space
  - economic development, regime type, voting behavior, religious beliefs, etc.
- geographic distribution of variables provide valuable information about the underlying mechanism of interest
- spatial autocorrelation (SA) causes severe problems for common econometric methods

Adequately accounting for spatial structures is necessary to guard against false inferences and to utilize spatial information!

This Project

- so far, social scientists use a limited subset of techniques suitable to handle SA
  - exploratory spatial analysis:
    - different types of maps
    - local indicators of SA (LISA)
  - inferential models
    - parametric spatial regression models

This Project

UNIVERSITY OF MANNHEIM

- so far, social scientists use a limited subset of techniques suitable to handle SA
  - exploratory spatial analysis:
    - different types of maps
    - local indicators of SA (LISA)
  - inferential models
    - parametric spatial regression models
- spatial eigenfunction analysis – particularly Moran eigenvector maps (MEM) – is a simple yet powerful tool to analyze cross-sectional data structures
  1. identification & visualization of complex (multi-scale) spatial patterns
  2. specification, estimation, and interpretation of inferential models
  3. partitioning of the variation in $\boldsymbol{y}$ into individual components (space, covariates, joined)
  4. (structure-preserving) simulation of spatially autocorrelated data

SA & Spatial Eigenfunctions

Spatial eigenfunction alanysis is based on the spectral decomposition of a centered (and symmetric/ symmetrized) connectivity matrix $\boldsymbol{W}$:

$$\boldsymbol{MWM} = \boldsymbol{E\Lambda E}^{-1} = \boldsymbol{E\Lambda E}'$$

- demeaning projector $\boldsymbol{M} = (\boldsymbol{I} - \mathbf{11}'/n)$ – also ensures that all eigenvectors are orthogonal and uncorrelated
- $\boldsymbol{E}$ are all n eigenvectors
- $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues $\boldsymbol{\lambda}$

SA & Spatial Eigenfunctions

Spatial eigenfunction alanysis is based on the spectral decomposition of a centered (and symmetric/ symmetrized) connectivity matrix $\boldsymbol{W}$:

$$\boldsymbol{MWM} = \boldsymbol{E \Lambda E}^{-1} = \boldsymbol{E \Lambda E}'$$

- demeaning projector $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{11}'/n)$ – also ensures that all eigenvectors are orthogonal and uncorrelated
- $\boldsymbol{E}$ are all n eigenvectors
- $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues $\boldsymbol{\lambda}$

Direct relationship to the numerator of the global Moran coefficient:

$$MC(\boldsymbol{x}) = \frac{n}{\boldsymbol{1}'\boldsymbol{W}\boldsymbol{1}} \frac{\boldsymbol{x}'\boldsymbol{MWMx}}{\boldsymbol{x}'\boldsymbol{Mx}}$$

Eigenvectors as Synthetic Map Patterns

- $\boldsymbol{E}$ depict all possible distinct and mutually uncorrelated synthetic map patterns permitted by $\boldsymbol{W}$
- corresponding eigenvalues $\boldsymbol{\lambda}$ indicate the level of SA

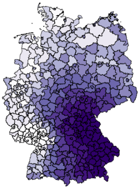
Example: 401 German NUTS-3 regions

Eigenvectors as Synthetic Map Patterns

- $\boldsymbol{E}$ depict all possible distinct and mutually uncorrelated synthetic map patterns permitted by $\boldsymbol{W}$
- corresponding eigenvalues $\boldsymbol{\lambda}$ indicate the level of SA

Example: 401 German NUTS-3 regions

Eigenvector 1

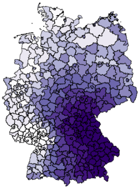


Eigenvectors as Synthetic Map Patterns

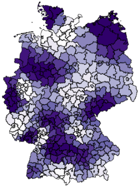
- $\boldsymbol{E}$ depict all possible distinct and mutually uncorrelated synthetic map patterns permitted by $\boldsymbol{W}$
- corresponding eigenvalues $\boldsymbol{\lambda}$ indicate the level of SA

Example: 401 German NUTS-3 regions

Eigenvector 1



Eigenvector 25

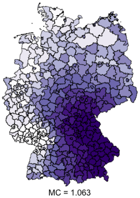


Eigenvectors as Synthetic Map Patterns

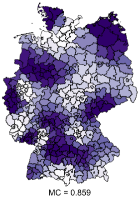
- $\boldsymbol{E}$ depict all possible distinct and mutually uncorrelated synthetic map patterns permitted by $\boldsymbol{W}$
- corresponding eigenvalues $\boldsymbol{\lambda}$ indicate the level of SA

Example: 401 German NUTS-3 regions

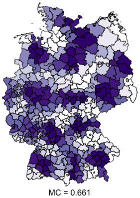
Eigenvector 1



Eigenvector 25



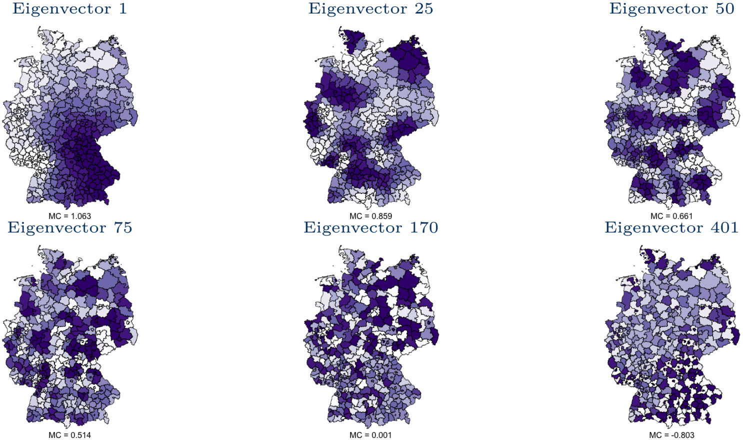
Eigenvector 50



Eigenvectors as Synthetic Map Patterns

- $\boldsymbol{E}$ depict all possible distinct and mutually uncorrelated synthetic map patterns permitted by $\boldsymbol{W}$
- corresponding eigenvalues $\boldsymbol{\lambda}$ indicate the level of SA

Example: 401 German NUTS-3 regions

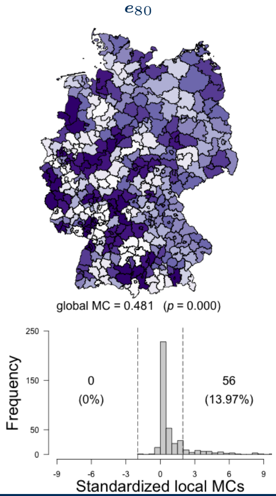


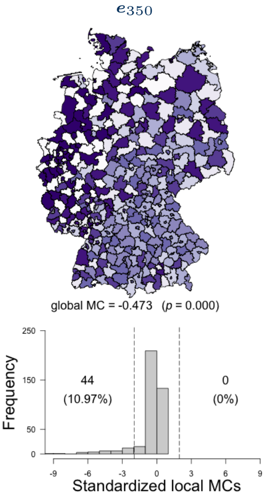
1. Identification of Complex Spatial Structures

Identification of Complex SA Structures

- global and local MC statistics might fail to identify mixtures of positive and negative SA
- Example:

- global and local MC statistics might fail to identify mixtures of positive and negative SA
- Example:

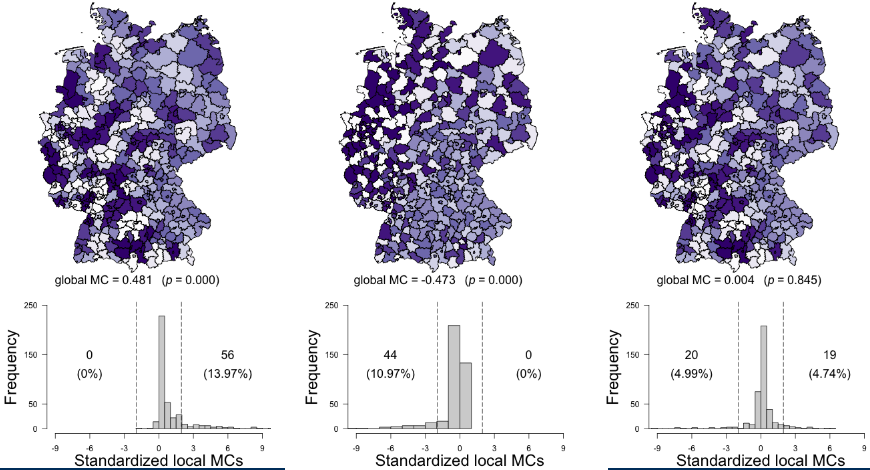




Identification of Complex SA Structures

- global and local MC statistics might fail to identify mixtures of positive and negative SA
- Example:

e_{80} | e_{350} | $e_{80} + e_{350}$



Identification of Complex SA Structures

- eigenfunction analysis allows researchers to decompose the global MC into positively and negatively autocorrelated parts
  - global MC $= \mathrm{MC}^{+} + \mathrm{MC}^{-}$
- decomposing the global MC helps identifying complex non-random spatial patterns

Identification of Complex SA Structures

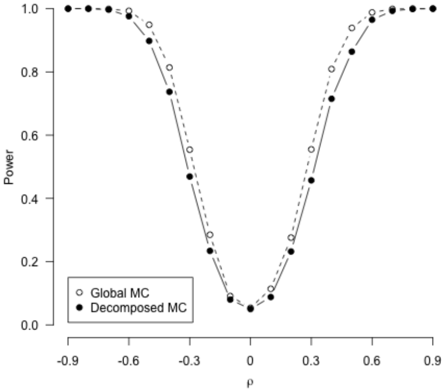
- eigenfunction analysis allows researchers to decompose the global MC into positively and negatively autocorrelated parts
  - global MC = $\mathrm{MC}^{+} + \mathrm{MC}^{-}$
- decomposing the global MC helps identifying complex non-random spatial patterns

Monte Carlo setup:

- $N = 100$ units ordered on a regular 10×10 grid
- $\boldsymbol{W}$ is a symmetric contiguity matrix (rook scheme)
- ρ ranges from $-.9$ to $.9$ in steps of $.1$
- Scenario 1: simple spatial structure
  - $\boldsymbol{x} = (\boldsymbol{I} - \rho \boldsymbol{W}_g)^{-1} \boldsymbol{u}$
- Scenario 2: mixture of positive and negative SA
  - $\boldsymbol{x} = (\boldsymbol{I} - \rho \boldsymbol{W}_g)^{-1} \boldsymbol{u} + (\boldsymbol{I} - (-\rho) \boldsymbol{W}_g)^{-1} \boldsymbol{v}$

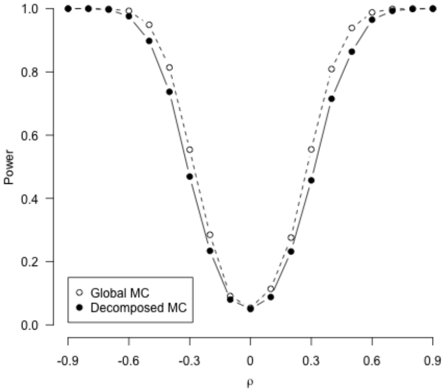
How does the decomposed MC based on spatial eigenfunctions performs compared to the global MC in terms of power?

Scenario 1: Simple Spatial Structure

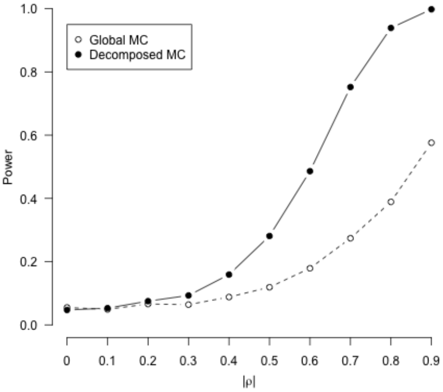


Identification of Complex SA Structures

Scenario 1: Simple Spatial Structure



Scenario 2: Mixture of SA



2. Model Specification, Estimation, and Interpretation

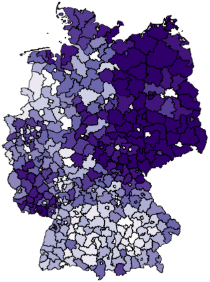
Specification, Estimation, and Interpretation

- SA causes severe problems for common econometric inferential techniques
- depending on the spatial DGP, SA can lead to
  1. incorrect standard errors
  2. biased and inconsistent parameter estimates
- spatial regression models address these problems but require many more or less rigid assumptions
  - knowledge of the true DGP
  - functional form assumptions
  - exact specification of SA in each regressor
  - difficult to estimate in a GLM framework

- semiparametric spatial filtering methods use Moran eigenvectors to construct a synthetic proxy variable that controls for SA

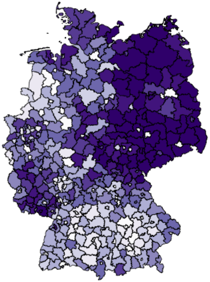
- a subset of eigenvectors can be combined to reproduce real-world map patterns
- Example: Median age in German NUTS-3 regions

Observed Median Age

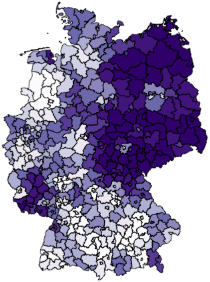


- a subset of eigenvectors can be combined to reproduce real-world map patterns
- Example: Median age in German NUTS-3 regions

Observed Median Age

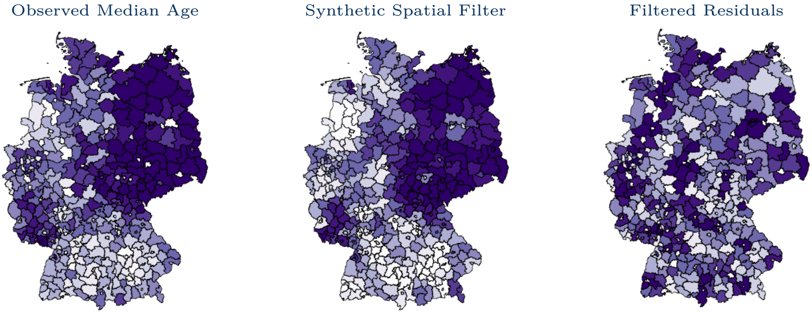


Synthetic Spatial Filter



Specification, Estimation, and Interpretation

- a subset of eigenvectors can be combined to reproduce real-world map patterns
- Example: Median age in German NUTS-3 regions



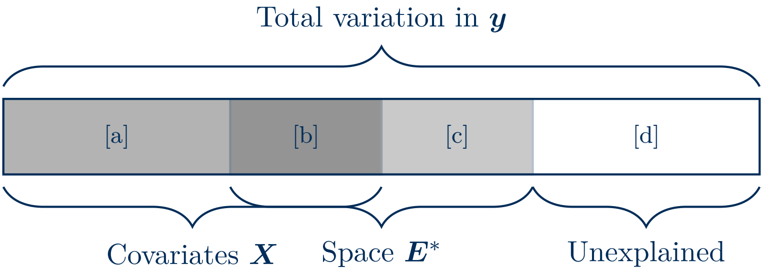
- a judiciously selected subset of eigenvectors controls for the underlying spatial pattern
- straightforward parameter estimation & interpretation

3. Variation Partitioning

Variation Partitioning

How much variation in $\boldsymbol{y}$ is caused by $\boldsymbol{X}$?

- a common spatial structure spuriously inflates the share of variation explained by the predictors
- disentangle the individual contribution of the covariates and the spatial structure



4. (Structure-Preserving) Simulation of Spatially Autocorrelated Data