

Eigenvector-Based Semiparametric Filtering of Spatial Autocorrelation in Regression Models

Sebastian Juhl

University of Mannheim



March 17, 2021

- spatial misspecification can lead to inefficient or even biased and inconsistent parameter estimates in regression models
- social scientists predominantly (exclusively?) apply parametric spatial regression models to *control* for spatial confounding
- semiparametric filtering techniques can offer an attractive alternative:
 - ease of estimation (standard OLS or ML estimators)
 - straightforward interpretation
 - accounts for SA at different scales/resolutions
 - easily adaptable to GLMs
- especially eigenvector-based spatial filtering (e.g., Griffith 1996 & 2003) is of practical utility

$$y = X\beta + e$$

$$y = X\beta + e$$

Spatial Error DGP

$$y = X\beta + \underbrace{(I - \rho W)^{-1}\epsilon}_{e_{SEM}}$$

Spatial-X DGP

$$y = X\beta + \underbrace{\rho W X + \epsilon}_{e_{SLX}}$$

Spatial Lag DGP

$$y = X\beta + \underbrace{\rho W y + \epsilon}_{e_{SAR}}$$

$$y = (I - \rho W)^{-1}(X\beta + \epsilon)$$

$$y = X\beta + e$$

Spatial Error DGP

$$y = X\beta + \underbrace{(I - \rho W)^{-1}\epsilon}_{e_{SEM}}$$

Spatial-X DGP

$$y = X\beta + \underbrace{\rho W X + \epsilon}_{e_{SLX}}$$

Spatial Lag DGP

$$y = X\beta + \underbrace{\rho W y + \epsilon}_{e_{SAR}}$$

$$y = (I - \rho W)^{-1}(X\beta + \epsilon)$$

$$(I - \rho W)^{-1} = (I + \rho W + \rho^2 W + \dots)$$

$$y = X\beta + e$$

Spatial Error DGP

$$y = X\beta + \underbrace{(I - \rho W)^{-1}\epsilon}_{e_{SEM}}$$

Spatial-X DGP

$$y = X\beta + \underbrace{\rho W X + \epsilon}_{e_{SLX}}$$

Spatial Lag DGP

$$y = X\beta + \underbrace{\rho W y + \epsilon}_{e_{SAR}}$$

$$y = (I - \rho W)^{-1}(X\beta + \epsilon)$$

$$(I - \rho W)^{-1} = (I + \rho W + \rho^2 W + \dots)$$

$$y = \left. \begin{array}{l} X\beta \\ + \rho W \epsilon \\ + \rho^2 W^2 \epsilon \\ + \dots + \epsilon \end{array} \right\} e_{SEM}$$

$$y = \left. \begin{array}{l} X\beta \\ + \rho W (X\beta + \epsilon) \\ + \rho^2 W^2 (X\beta + \epsilon) \\ + \dots + \epsilon \end{array} \right\} e_{SAR}$$

$$y = X\beta + e$$

Spatial Error DGP

$$y = X\beta + \underbrace{(I - \rho W)^{-1}}_{e_{SEM}} \epsilon$$

Spatial-X DGP

$$y = X\beta + \underbrace{\rho W X + \epsilon}_{e_{SLX}}$$

Spatial Lag DGP

$$y = X\beta + \underbrace{\rho W y + \epsilon}_{e_{SAR}}$$

$$y = (I - \rho W)^{-1} (X\beta + \epsilon)$$

$$(I - \rho W)^{-1} = (I + \rho W + \rho^2 W + \dots)$$

$$y = \left. \begin{array}{l} X\beta \\ + \rho W \epsilon \\ + \rho^2 W^2 \epsilon \\ + \dots + \epsilon \end{array} \right\} e_{SEM}$$

e_{SEM}

$$\underbrace{\sum_{r=1}^{\infty} \rho^r W^r \epsilon}_{\text{spatial part}} + \underbrace{\epsilon}_{\text{noise}}$$

$$y = \left. \begin{array}{l} X\beta \\ + \rho W (X\beta + \epsilon) \\ + \rho^2 W^2 (X\beta + \epsilon) \\ + \dots + \epsilon \end{array} \right\} e_{SAR}$$

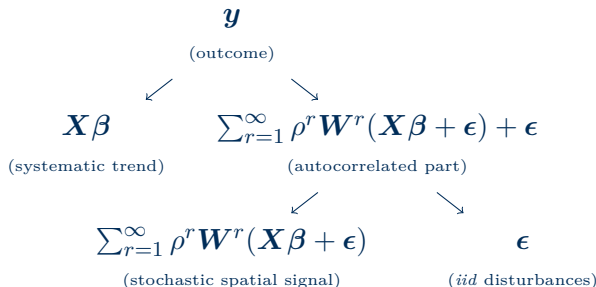
e_{SAR}

$$\underbrace{\sum_{r=1}^{\infty} \rho^r W^r (X\beta + \epsilon)}_{\text{spatial part}} + \underbrace{\epsilon}_{\text{noise}}$$

Intuition:

- include a proxy variable for omitted spatial effects in the regression's systematic part
- ESF partitions the response variable into i) a systematic trend, ii) a stochastic spatial signal, and iii) *iid* disturbances

Example: SAR DGP



ESF is based on the spectral decomposition of a centered (and symmetric/ symmetrized) connectivity matrix \mathbf{W} :

$$\mathbf{M}\mathbf{W}\mathbf{M} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{-1} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}'$$

- demeaning projector $\mathbf{M} = (\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ – also ensures that all eigenvectors are orthogonal and uncorrelated
- \mathbf{E} are all n eigenvectors
- $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues λ

ESF is based on the spectral decomposition of a centered (and symmetric/ symmetrized) connectivity matrix W :

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

- demeaning projector $M = (I - \mathbf{1}\mathbf{1}'/n)$ – also ensures that all eigenvectors are orthogonal and uncorrelated
- E are all n eigenvectors
- Λ is a diagonal matrix of the eigenvalues λ

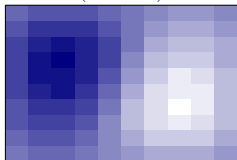
Direct relationship to the numerator of the global Moran coefficient:

$$MC(x) = \frac{n}{\mathbf{1}'W\mathbf{1}} \frac{x'\textcolor{red}{MWM}x}{x'Mx}$$

- eigenvectors \mathbf{E} depict distinct – and uncorrelated – synthetic map patterns
- \mathbf{E} contains all possible spatial patterns permitted by \mathbf{W}
- corresponding eigenvalues indicate the level of SA

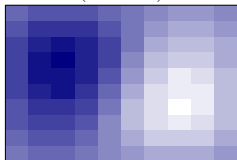
- eigenvectors E depict distinct – and uncorrelated – synthetic map patterns
- E contains all possible spatial patterns permitted by W
- corresponding eigenvalues indicate the level of SA

Eigenvector 1
(MC = 1)

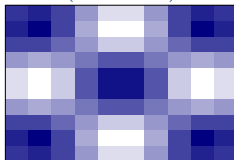


- eigenvectors E depict distinct – and uncorrelated – synthetic map patterns
- E contains all possible spatial patterns permitted by W
- corresponding eigenvalues indicate the level of SA

Eigenvector 1
(MC = 1)

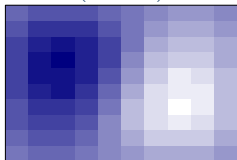


Eigenvector 10
(MC = 0.73)

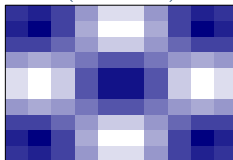


- eigenvectors E depict distinct – and uncorrelated – synthetic map patterns
- E contains all possible spatial patterns permitted by W
- corresponding eigenvalues indicate the level of SA

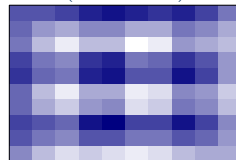
Eigenvector 1
(MC = 1)



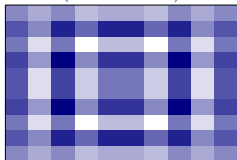
Eigenvector 10
(MC = 0.73)



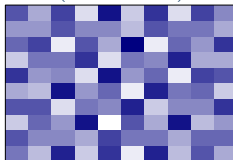
Eigenvector 20
(MC = 0.454)



Eigenvector 44
(MC = 0.066)

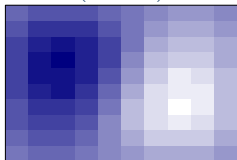


Eigenvector 80
(MC = -0.454)

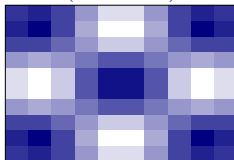


- eigenvectors E depict distinct – and uncorrelated – synthetic map patterns
- E contains all possible spatial patterns permitted by W
- corresponding eigenvalues indicate the level of SA

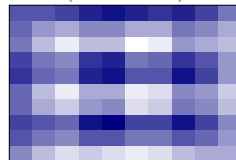
Eigenvector 1
(MC = 1)



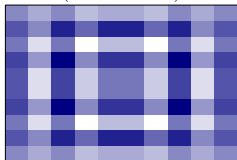
Eigenvector 10
(MC = 0.73)



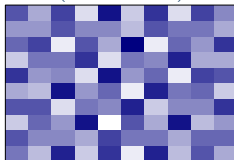
Eigenvector 20
(MC = 0.454)



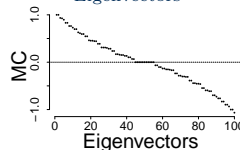
Eigenvector 44
(MC = 0.066)



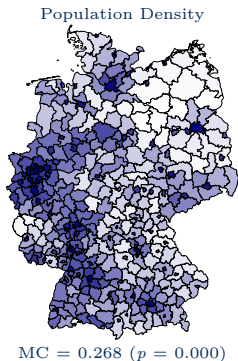
Eigenvector 80
(MC = -0.454)



MCs of all
Eigenvectors

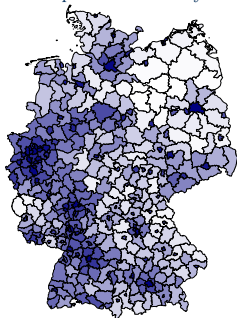


More realistic (and complex) patterns can be obtained by a linear combination of eigenvectors:



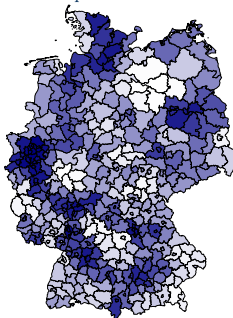
More realistic (and complex) patterns can be obtained by a linear combination of eigenvectors:

Population Density



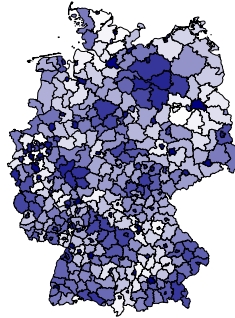
MC = 0.268 ($p = 0.000$)

Spatial Filter



MC = 0.268 ($p = 0.000$)

Filtered Residuals



MC = -0.078 ($p = 0.486$)

However, it is impossible to include all n eigenvectors as regressors

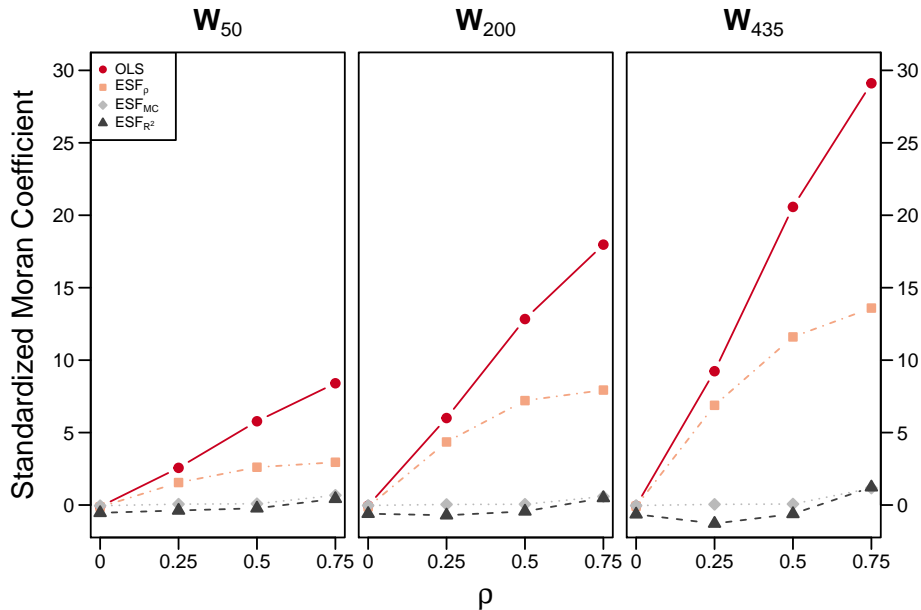
- ① identification of a candidate set $\mathbf{E}^C \subset \mathbf{E}$ based on
 - sign of SA
 - strength of SA
- ② select relevant map patterns $\mathbf{E}^* \subset \mathbf{E}^C$ using supervised or unsupervised stepwise regression

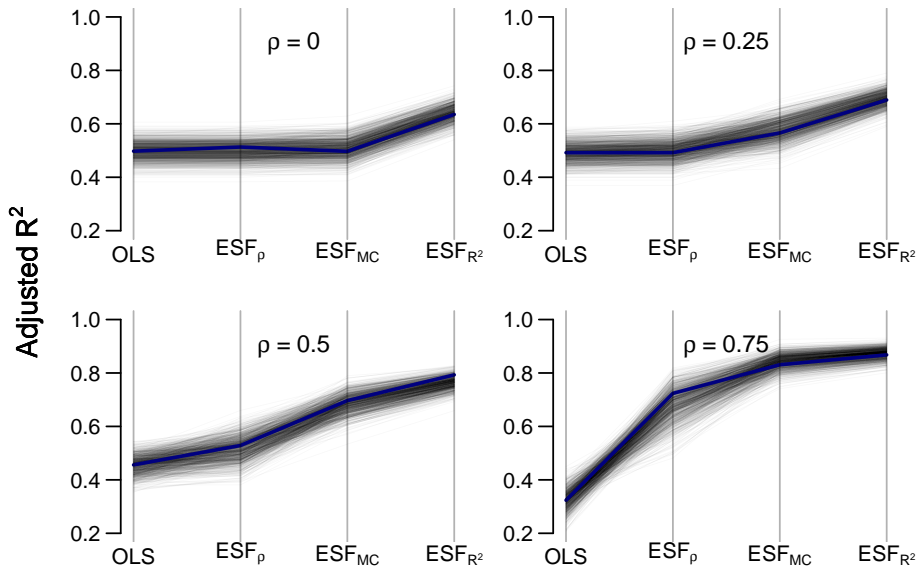
$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \overbrace{\mathbf{E}\boldsymbol{\gamma}}^{e_{OLS}} + \boldsymbol{\epsilon} \\ &\approx \mathbf{X}\boldsymbol{\beta} + \underbrace{\mathbf{E}^*\boldsymbol{\gamma}}_{\text{filter}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}} \end{aligned}$$

- different selection criteria can be used:
 - maximization of model fit (e.g., AIC, BIC, R^2 , ...)
 - significance level of eigenvectors or residual SA
 - minimization of residual SA
 - LASSO
 - ...

(Preliminary)

Monte Carlo Evidence





- ESF offers numerous advantages for political scientists
- things to consider:

Pros

- + ease of model estimation
- + straightforward interpretation of parameters
- + flexibility
(no need to specify the spatial pattern in each variable)
- + generalizability
(also applicable to GLMs – with slight modifications)

Cons

- “space” as misspecification problem
- computationally demanding for large N
- over- or undercorrection of SA possible

- applicability depends on research context (as always)
- ESF is not a replacement for parametric spatial regression models!

- **spfilterR** package:
CRAN: <https://CRAN.R-project.org/package=spfilterR>
GitHub: <https://github.com/sjuhl/spfilterR>
- further projects & working papers
 - ❶ introductory paper on the **spfilterR** package (under review)
 - ❷ project on Moran eigenvector maps and spatial eigenfunction analysis

Feedback & suggestions are highly appreciated!

`sebastian.juhl@gess.uni-mannheim.de`

`www.sebastianjuhl.com`