Eigenvector-Based Semiparametric Filtering of Spatial Autocorrelation in Regression Models

Sebastian Juhl

University of Mannheim

March 17, 2021

Motivation

- spatial autocorrelation can lead to inefficient or even biased and inconsistent parameter estimates in regression models
- social scientists primarily address this issue in one of two ways:
    1. ignore it altogether and assume *iid* observations
    2. apply (parametric) spatial regression models

- semiparametric filtering techniques can offer an attractive alternative
    - ease of estimation (standard OLS or ML estimators)
    - straightforward interpretation
    - accounts for spatial autocorrelation at different scales/resolutions
    - easily adaptable to GLMs

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

UNIVERSITY
OF MANNHEIM

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$$

Spatial Error Model

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

Spatial-X Model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \underbrace{\rho\boldsymbol{WX} + \boldsymbol{\epsilon}}_{\boldsymbol{e}_{SLX}}$$

Spatial Lag Model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \underbrace{\rho\boldsymbol{Wy} + \boldsymbol{\epsilon}}_{\boldsymbol{e}_{SAR}}$$

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

$$y = X\beta + e$$

Spatial Error Model

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

Spatial-X Model

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

Spatial Lag Model

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

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

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

$$y = \begin{array}{l} X\beta \\ +\rho W\epsilon \\ +\rho^2 W^2\epsilon \\ +... + \epsilon \end{array} \left.\rule{0pt}{2.5em}\right\} \epsilon_{SEM}$$

$$y = \begin{array}{l} X\beta \\ +\rho W(X\beta + \epsilon) \\ +\rho^2 W^2(X\beta + \epsilon) \\ +... + \epsilon \end{array} \left.\rule{0pt}{2.5em}\right\} \epsilon_{SAR}$$

$$y = X\beta + e$$

Spatial Error Model

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

Spatial-X Model

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

Spatial Lag Model

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

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

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

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

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

$$\underbrace{e_{SEM}}$$

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

$$\underbrace{e_{SLX}}$$

$$\underbrace{\rho WX}_{\text{spatial part}} + \underbrace{\epsilon}_{\text{noise}}$$

$$\underbrace{e_{SAR}}$$

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

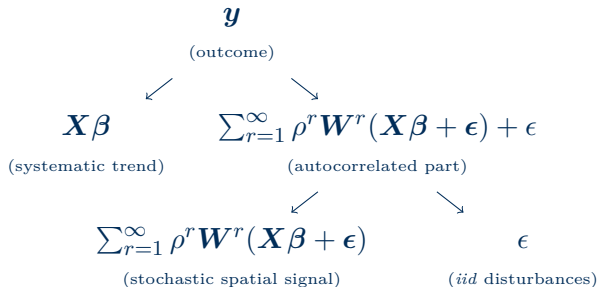# Eigenvector-Based Spatial Filtering

Intuition:

ESF partitions the response variable into i) a systematic trend, ii) a stochastic spatial signal, and iii) *iid* disturbances

Example: SAR DGP

$$\boldsymbol{y}$$
(outcome)

$$\boldsymbol{X\beta} \qquad \sum_{r=1}^{\infty} \rho^r \boldsymbol{W}^r (\boldsymbol{X\beta} + \boldsymbol{\epsilon}) + \epsilon$$
(systematic trend)   (autocorrelated part)

$$\sum_{r=1}^{\infty} \rho^r \boldsymbol{W}^r (\boldsymbol{X\beta} + \boldsymbol{\epsilon}) \qquad \epsilon$$
(stochastic spatial signal)   (*iid* disturbances)

ESF technique utilizes the spectral decomposition of a centered and symmetrized connectivity matrix: $\boldsymbol{V} = \frac{1}{2}(\boldsymbol{W} + \boldsymbol{W}')$:

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

- demeaning projector $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}'/n)$ ensures that all eigenvectors are orthogonal und uncorrelated
- asymptotically, the eigenfunctions obtained from $\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}$ converge to those of matrix $\boldsymbol{V}$
- $\boldsymbol{E}$ are all n eigenvectors
- $\boldsymbol{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues $\boldsymbol{\lambda}$

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

$$\Rightarrow \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \underbrace{\rho\boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}'\boldsymbol{y} + \boldsymbol{\epsilon}}_{\boldsymbol{e}_{SAR}}$$

$$y = X\beta + \rho E \Lambda \underbrace{E'y}_{g} + \epsilon$$

- g is the OLS estimator of y regressed on E

  $(E'E)^{-1}Ey \Rightarrow Ey$ since $E'E = I$

$$y = X\beta + \rho E \Lambda \underbrace{E'y}_{g} + \epsilon$$

- g is the OLS estimator of y regressed on E

  $(E'E)^{-1}Ey \Rightarrow Ey$ since $E'E = I$

$$y = X\beta + \rho E \Lambda g + \epsilon$$

- $\rho E \Lambda$: $\rho \Lambda$ essentially rescales E

  $\Rightarrow \gamma = \rho \Lambda g$

$$y = X\beta + \rho E \Lambda \underbrace{E'y}_{g} + \epsilon$$

- g is the OLS estimator of y regressed on E

  $(E'E)^{-1}Ey \Rightarrow Ey$ since $E'E = I$

$$y = X\beta + \rho E \Lambda g + \epsilon$$

- $\rho E \Lambda$: $\rho \Lambda$ essentially rescales E

  $\Rightarrow \gamma = \rho \Lambda g$

$$y = X\beta + \rho E \Lambda E'y + \epsilon$$
$$= X\beta + E\gamma + \epsilon$$

However, we cannot include all n eigenvectors as regressors!

Constructing the Spatial Filter

- the set of $\boldsymbol{E}$ depict <u>all</u> possible patterns of autocorrelation permitted by the connectivity matrix $\boldsymbol{W}$
- stepwise regression techniques help to select the most relevant patterns $\boldsymbol{E}^*$
- thereby, we can craft a synthetic variable – the spatial filter

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{E\gamma} + \boldsymbol{\epsilon}$$
$$\approx \boldsymbol{X\beta} + \underbrace{\boldsymbol{E}^*\boldsymbol{\gamma}}_{filter} + \underbrace{\boldsymbol{\epsilon}}_{noise}$$

- different criteria can be used to select eigenvectors, e.g.:
  - model fit
  - significance
  - spatial autocorrelation

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
  some modifications)

Cons

- "removes" indirect spillovers
- computationally demanding for
  large N
- over- or undercorrection of
  spatial autocorrelation possible

(Preliminary)

Monte Carlo Evidence

Reduction in Residual SA