Spatial Econometric Models

Hanno Reuvers

Excellence Programme 2021–2022

My Conclusions

- The **blessings and curses** of economic growth
  - Carbon tax can resolve the "*tragedy of the commons*"
    - guide technological innovations
    - consumer and producer signalling
    - ... maybe also some peace of mind?
- Climate resilience dependence on **manageability** (and by extension on wealth?)
- Econometrics has a role in climate change modelling, especially on the economic side

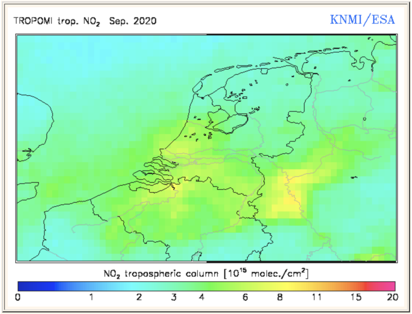
Question: Imagine NL with carbon tax, you are prime minister, what will you do with the revenues?

Climate Change has a huge spatial component

- Climate change always slips through border security
  - GHGs diffuse freely through space
  - Climate impacts are heterogeneous over space
- As a consequence, climate policy is spatial problem
  - Free-riding and/or climate regulation evasion are a valid concern
  - Should climate compensations vary over space (and over income)?
- Let us consider a "Dutch" example of climate spatial data

“De Stikstofcrisis”

Satellite image of NO_2 concentration in NL (source: Tropospheric Emission Monitoring Internet Service)



Question: What is the crisis about? Implications? Policy?

Arriving in Rotterdam, your Courses at Erasmus University...

Course	Typical Data Property
Statistics	i.i.d.
Intro to Multivariate Statistics	i.i.d.
Markov Processes	time series
Econometrics 1	cross-sectional regression
Econometrics 2	heteroskedasticity, serial correlation, endogeneity, limited dependence
Time Series Analysis	time series

=> ***Today*: Learn how to model spatial data** <=

Overview of the Lectures

1. General considerations on modelling/forecasting climate data
2. Three example cases:
   - Case 1: The Tip of the Iceberg
   - Case 2: Climate Change and Agriculture
   - Case 3: Uncomfortable Temperatures
3. Some general conclusions
4. Spatial econometrics: the Spatial Autoregressive (SAR) Model
5. Assignment

Specific Overview for this Lecture

4. Spatial econometrics: the Spatial Autoregressive (SAR) Model
   a. **Modeling Spatial Data**
   b. The spatial weight matrix
   c. Estimation
   d. On normalization

5. Assignment

Modeling Spatial Data

The Spatial Autoregressive (SAR) Model

The SAR model is

$$\boldsymbol{y}_n = \lambda \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n, \tag{1}$$

where $\boldsymbol{\varepsilon}_n = (\varepsilon_1, \ldots, \varepsilon_n)' \in \mathbb{R}^n$ is a random vector of innovations with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{V}ar(\varepsilon_i) = \sigma^2$.

=> By convention, $[\boldsymbol{W}_n]_{ii} = 0$ for all $i = 1, \ldots, n$ <=

Some questions:

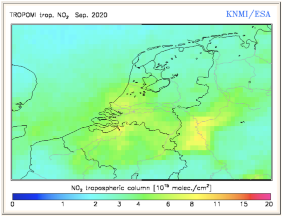
- What is the interpretation of model (1)?
- How to estimate the unknown parameters λ, $\boldsymbol{\beta}$, and σ^2?

SAR = Linear Regression Model + Spatial lag

- Let me (temporarily) delete a term: $\boldsymbol{y}_n =$ ~~$\lambda \boldsymbol{W}_n \boldsymbol{y}_n$~~ $+ \boldsymbol{X}_n \beta + \varepsilon_n$
- Your econometric Pavlov reaction might be:
  1. This is boring... it's just a multivariate linear regression
  2. Yes, OLS! I love $\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$
- The new feature is the so-called spatial lag $\boldsymbol{W}_n \boldsymbol{y}_n$
  - $\boldsymbol{W}_n \in \mathbb{R}^{n \times n}$ is the spatial weight matrix
  - Spatial autoregressive parameter λ
  - Given $\boldsymbol{W}_n$ and the observations $\boldsymbol{y}_n$, $\boldsymbol{W}_n \boldsymbol{y}_n$ is similar to any other regressor? Or maybe not?
  - After today, you can invent your own spatial model because:

known model + spatial lag $\Longrightarrow$ unknown new spatial model

"De Stikstofcrisis" (revisited)



$$\boldsymbol{y}_n = \lambda \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{X}_n \beta + \varepsilon_n$$

- $\boldsymbol{W}_n \boldsymbol{y}_n$: captures diffusion effects
- $\boldsymbol{X}_n$: temperature, traffic intensity, agricultural index, etc.

Some Economic Examples

- Cigarette demand in the US
- Modeling house prices
- Crime analysis
- SAR model is closely related to social interaction models (see slide 18)

Reasons to include a spatial lag are:

1. There are spatial spillover effects
2. "Borrowing" omitted variables from neighboring locations

Specific Overview for this Lecture

4. **Spatial econometrics: the Spatial Autoregressive (SAR) Model**
   a. Modeling Spatial Data
   b. **The spatial weight matrix**
   c. Estimation
   d. On normalization
5. **Assignment**

The Spatial Weight Matrix

Specifications of the Spatial Weight Matrix

- The choice of $\boldsymbol{W}_n$ is crucial because it determines how neighboring information is incorporated
- Two main approached to specify $\boldsymbol{W}_n$:
  - Contiguity
  - Distance-based
- In practice, try several $\boldsymbol{W}_n$ and compare findings and fit

Contiguity (1)

Contiguity: "*the state of bordering or being in contact with something*"

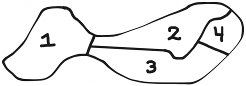
- Binary contiguity

$$[\boldsymbol{W}_n]_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{if } i \text{ and } j \text{ are not contiguous (and } i = j) \end{cases}$$

- We still need to give a geographic meaning to "contiguous"

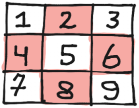
Contiguity (2)

- Sharing a common border



$$W_n = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

- Rook contiguity for lattice



$$W_n = \begin{bmatrix} 0&1&0&1&0&0&0&0&0 \\ 1&0&1&0&1&0&0&0&0 \\ 0&1&0&0&0&1&0&0&0 \\ 1&0&0&0&1&0&1&0&0 \\ 0&\color{red}{1}&0&\color{red}{1}&0&\color{red}{1}&0&\color{red}{1}&0 \\ 0&0&1&0&1&0&0&0&1 \\ 0&0&0&1&0&0&0&1&0 \\ 0&0&0&0&1&0&1&0&1 \\ 0&0&0&0&0&1&0&1&0 \end{bmatrix}$$

Question (for chess players): What are bishop and queen contiguity?

Contiguity (3)

Question: What does the following $\boldsymbol{W}_n$ reflect?

$$\boldsymbol{W}_n = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Weight Matrices based on Distances

- Let d_{ij} denote the distance between spatial units i and j
  - Geophysical distance based on, say longitude+latitude
  - Distance between centroids of regions
  - "Economic" distance based on economic similarity

- Examples:
  - Inverse distance: $[\boldsymbol{W}_n]_{ij} = 1/d_{ij}^{\alpha}$ (typically: $\alpha = 1$ or $\alpha = 2$)
  - Exponential decay: $[\boldsymbol{W}_n]_{ij} = \exp(-d_{ij}/\alpha)$
  - Threshold distance: $[\boldsymbol{W}_n]_{ij} = \mathbf{1}_{\{d_{ij} \leq d^*\}}$

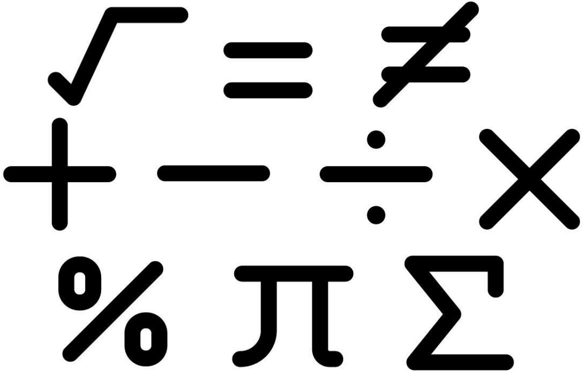
=> **Do not forget:** $[\boldsymbol{W}_n]_{ii} = 0$ **for all** $i = 1, \ldots, n$ <=

Specific Overview for this Lecture

4. **Spatial econometrics: the Spatial Autoregressive (SAR) Model**
   - a. Modeling Spatial Data
   - b. The spatial weight matrix
   - c. **Estimation**
   - d. On normalization

5. **Assignment**

Estimating the SAR Parameters



Estimation by OLS? Yes or no?

Recall the SAR model: $\boldsymbol{y}_n = \lambda \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n$

1. Define $\boldsymbol{X}_n^* = \begin{bmatrix} \boldsymbol{W}_n \boldsymbol{y}_n & \boldsymbol{X}_n \end{bmatrix}$ and $\boldsymbol{\beta}^* = \begin{bmatrix} \lambda \\ \boldsymbol{\beta} \end{bmatrix}$
2. In the new notation, $\boldsymbol{y}_n = \boldsymbol{X}_n^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_n$

Question: Can we estimate the unknown parameters by OLS, i.e.

$$\begin{bmatrix} \widehat{\lambda} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix} = \widehat{\boldsymbol{\beta}}^* = (\boldsymbol{X}_n^{*\prime} \boldsymbol{X}_n^*)^{-1} \boldsymbol{X}_n^{*\prime} \boldsymbol{y}_n ?$$

Maximum Likelihood of the SAR Model (1)

$$\boxed{\boldsymbol{y}_n = \lambda \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{X}_n \beta + \varepsilon_n}$$

- Many results originate from the seminal paper is Lee (2004):
  - $>$1000 citations
  - Quasi-maximum likelihood assuming $\varepsilon_n \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$
  - Detailed analysis of the asymptotic properties of the QMLE
- We will explore a subset of Lee's (2004) results

Deriving the Log-likelihood (Assignment 1)

1. Define $\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n$ and assume that $\boldsymbol{S}_n^{-1}(\lambda)$ exists. Prove that the log-likelihood under $\boldsymbol{\varepsilon}_n \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ is equal to

$$\log L_n(\boldsymbol{\theta}) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) + \log\big(\det(\boldsymbol{S}_n(\lambda))\big) - \tfrac{1}{2\sigma^2}\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big)'\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big),$$

where $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}', \sigma^2)'$.

Steps:

1. Solve for $\boldsymbol{y}_n$
2. Use $\boldsymbol{\varepsilon}_n \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ to determine distribution of $\boldsymbol{y}_n$
3. The distribution provides the log-likelihood

A Closer Look at the Log-Likelihood (1)

$$\log L_n(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \log\big(\det(\boldsymbol{S}_n(\lambda))\big) - \frac{1}{2\sigma^2}\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\beta\Big)'\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\beta\Big)$$

Question: How would you maximize this log-likelihood?

A Closer Look at the Log-Likelihood (2)

$$\log L_n(\boldsymbol{\theta}) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) + \log\big(\det(\boldsymbol{S}_n(\lambda))\big) - \tfrac{1}{2\sigma^2}\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big)'\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big)$$

Observations:

- $\log L_n(\boldsymbol{\theta})$ is quadratic in $\boldsymbol{\beta}$
- The dependence on σ^2 mimics the "traditional" Gaussian MLE
- The spatial autoregressive parameter enters nonlinearly

=> Concentrate the log-likelihood w.r.t. both $\boldsymbol{\beta}$ and σ^2 <=

Vector Differentiation (recap)

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ and $f : \mathbb{R}^k \to \mathbb{R}$, then

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_k} \end{bmatrix} \in \mathbb{R}^k$$

- A vector derivative is often called the gradient
- The notation $\nabla f(\boldsymbol{\beta})$ is also used
- Proofs work "element-wise" (examples on next slide)

Some Vector Derivatives (recap)[1]

- For $\boldsymbol{x} \in \mathbb{R}^k$, $\boxed{\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\beta}' \boldsymbol{x} = \boldsymbol{x}}$, because

$$\left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\beta}' \boldsymbol{x}\right]_\ell = \frac{\partial}{\partial \beta_\ell} \sum_{i=1}^{k} \beta_i x_i = \sum_{i=1}^{k} \frac{\partial \beta_i}{\partial \beta_\ell} x_i = \sum_{i=1}^{k} \delta_{i\ell} x_i = x_\ell = [\boldsymbol{x}]_\ell$$

- For $\boldsymbol{A} \in \mathbb{R}^{k \times k}$, $\boxed{\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\beta}' \boldsymbol{A} \boldsymbol{\beta} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\beta}}$, because

$$\begin{aligned}
\left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\beta}' \boldsymbol{A} \boldsymbol{\beta}\right]_\ell &= \frac{\partial}{\partial \beta_\ell} \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_i A_{ij} \beta_j = \sum_{i=1}^{k} \sum_{j=1}^{k} \left(\frac{\partial \beta_i}{\partial \beta_\ell} A_{ij} \beta_j + \beta_i A_{ij} \frac{\partial \beta_j}{\partial \beta_\ell} \right) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} (\delta_{i\ell} A_{ij} \beta_j + \beta_i A_{ij} \delta_{j\ell}) = \sum_{j=1}^{k} A_{\ell j} \beta_j + \sum_{i=1}^{k} A_{i\ell} \beta_i \\
&= [(\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\beta}]_\ell
\end{aligned}$$

[1]Note: The Kronecker delta δ_{ij} equals 1 if $i = j$, and 0 if $i \neq j$.

Concentrating the Log-likelihood w.r.t. $\boldsymbol{\beta}$

Using the results from the previous slide, we find

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}\log L_n(\boldsymbol{\theta}) &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial\boldsymbol{\beta}}\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big)'\Big(\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\Big)\\ &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial\boldsymbol{\beta}}\Big(\boldsymbol{\beta}'\boldsymbol{X}_n'\boldsymbol{X}_n\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{X}_n'\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n\Big)\\ &= -\frac{1}{\sigma^2}\Big(\boldsymbol{X}_n'\boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{X}_n'\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n\Big)\end{aligned}$$

Given λ, the optimal choice for $\boldsymbol{\beta}$ is

$$\boxed{\widehat{\boldsymbol{\beta}}(\lambda) = \left(\boldsymbol{X}_n'\boldsymbol{X}_n\right)^{-1}\boldsymbol{X}_n'\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n}$$

Concentrating the Log-likelihood w.r.t. σ^2 (Assignment 2)

2. Concentrate the log-likelihood with respect to σ^2 and show that

$$\widehat{\sigma}^2(\lambda) = \frac{1}{n}\boldsymbol{y}_n'\boldsymbol{S}_n'(\lambda)\boldsymbol{M}_X\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n,$$

where $\boldsymbol{M}_X = \boldsymbol{I}_n - \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'$.

Steps:

1. Set $\frac{\partial}{\partial\sigma^2}\log L_n(\boldsymbol{\theta})$ equal to zero
2. Insert the expression for $\widehat{\boldsymbol{\beta}}(\lambda)$

It's all about the λ... (Assignment 3)

- We have optimal choices for $\boldsymbol{\beta}$ and σ^2 given the value for λ
- It remains to compute the concentrated log-likelihood

3. Derive the concentrated log-likelihood

$$\log L_n(\lambda) = -\tfrac{n}{2}\Big(\log(2\pi) + 1\Big) - \tfrac{n}{2}\log\left(\widehat{\sigma}^2(\lambda)\right) + \log\left(\det(\boldsymbol{S}_n(\lambda))\right).$$

The Estimation Approach (1)

A list of all the results so far:

- $\log L_n(\lambda) = -\frac{n}{2}\Big(\log(2\pi)+1\Big)-\frac{n}{2}\log\left(\widehat{\sigma}^2(\lambda)\right)+\log\left(\det(\boldsymbol{S}_n(\lambda))\right)$
- $\widehat{\boldsymbol{\beta}}(\lambda) = \left(\boldsymbol{X}_n'\boldsymbol{X}_n\right)^{-1}\boldsymbol{X}_n'\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n$
- $\widehat{\sigma}^2(\lambda) = \frac{1}{n}\boldsymbol{y}_n'\boldsymbol{S}_n'(\lambda)\boldsymbol{M}_X\boldsymbol{S}_n(\lambda)\boldsymbol{y}_n$

Question 1: How to interpret the formulae for $\widehat{\boldsymbol{\beta}}(\lambda)$ and $\widehat{\sigma}^2(\lambda)$?

Question 2: Given this interpretation, what is an intuitive implementation?

The Estimation Approach (2)

1. Maximize the concentrated log-likelihood

$$\log L_n(\lambda) = -\frac{n}{2}\Big(\log(2\pi)+1\Big) - \frac{n}{2}\log\left(\widehat{\sigma}^2(\lambda)\right) + \log\left(\det(\boldsymbol{S}_n(\lambda))\right)$$

   to determine the MLE $\widehat{\lambda}$

   - "Easy" univariate optimization problem (grid search, Newton-Rhapson, etc.) due to likelihood concentration

2. Consider $\boldsymbol{y}_n - \widehat{\lambda}\boldsymbol{W}_n\boldsymbol{y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n$
   - Newly computed dependent variable $\boldsymbol{y}_n - \widehat{\lambda}\boldsymbol{W}_n\boldsymbol{y}_n$
   - OLS estimator equal $\widehat{\boldsymbol{\beta}}(\widehat{\lambda})$
   - Average SSR is $\widehat{\sigma}^2(\widehat{\lambda})$

Specific Overview for this Lecture

4. **Spatial econometrics: the Spatial Autoregressive (SAR) Model**
   - a. Modeling Spatial Data
   - b. The spatial weight matrix
   - c. Estimation
   - d. **On normalization**

5. **Assignment**

On normalization

$$\lambda W_n = \left(\frac{\lambda}{c}\right)(cW_n)$$

Which λ-values?

- Two remaining (and related) issues
  $\longrightarrow$ $\lambda \boldsymbol{W}_n = \left(\frac{\lambda}{C}\right)(C\boldsymbol{W}_n)$ for any constant C
  $\longrightarrow$ We assumed $\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n$ to be invertible

- These issues are important because they determine the parameter space over which to optimize the concentrated log-likelihood

- Normalize the spatial weight matrix $\boldsymbol{W}_n$. Two normalization schemes are popular:
  1. Spectral-normalization
  2. Row-normalization

Focusing on *Symmetric* Weight Matrices

- For simplicity, assume $\boldsymbol{W}_n = \boldsymbol{W}_n'$ (restrictive in theory/practice?)

- Recall eigenvalue decomposition $\boldsymbol{W}_n = \boldsymbol{V}_n \boldsymbol{\Lambda}_n \boldsymbol{V}_n'$
  - $\boldsymbol{V}_n \in \mathbb{R}^{n \times n}$ stacks eigenvectors in columns
  - $\boldsymbol{\Lambda}_n = \text{diag}(\kappa_1(\boldsymbol{W}_n), \ldots, \kappa_n(\boldsymbol{W}_n)) \in \mathbb{R}^{n \times n}$ where
    $\longrightarrow$ $\kappa_i(\boldsymbol{W}_n)$ denotes the i^{th} eigenvalue of $\boldsymbol{W}_n$
    $\longrightarrow$ $\kappa_i(\boldsymbol{W}_n)$ is a real
    $\longrightarrow$ $\kappa_1(\boldsymbol{W}_n) \geq \kappa_2(\boldsymbol{W}_n) \geq \cdots \geq \kappa_n(\boldsymbol{W}_n)$
  - $\boldsymbol{V}_n' \boldsymbol{V}_n = \boldsymbol{V}_n \boldsymbol{V}_n' = \boldsymbol{I}_n$

- The key insight is

$$\boxed{\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n = \boldsymbol{V}_n \boldsymbol{V}_n' - \lambda \boldsymbol{V}_n \boldsymbol{\Lambda}_n \boldsymbol{V}_n' = \boldsymbol{V}_n \left(\boldsymbol{I}_n - \lambda \boldsymbol{\Lambda}_n\right) \boldsymbol{V}_n'}$$

=> $\boldsymbol{S}_n(\lambda)$ has eigenvalues $1 - \lambda \kappa_i(\boldsymbol{W}_n)$ ($i = 1, \ldots, n$) <=

Spectral-normalization

- $\boldsymbol{S}_n(\lambda)$ is invertible if all its eigenvalues are unequal to zero

$$1 - \lambda\kappa_i(\boldsymbol{W}_n) \neq 0 \qquad \text{for all } i = 1, \ldots, n. \tag{2}$$

- For spectral-normalization, condition (2) is satisfied using
  1. Renormalize $\boldsymbol{W}_n$ to $\boldsymbol{W}_n^* = \boldsymbol{W}_n / \max_{i=1,\ldots,n} |\kappa_i(\boldsymbol{W}_n)|$
  2. Ensure that the maximum eigenvalue of $\boldsymbol{W}_n^*$ is equal to one[2]
  3. Optimize concentrated log-likelihood over $\lambda < 1$, then

$$\lambda k_1(\boldsymbol{W}_n^*) < 1 \qquad \Longrightarrow \qquad \boldsymbol{S}_n(\lambda) \text{ invertible}$$

[2] This is either trivially satisfied or requires multiplying $\boldsymbol{W}_n^*$ by -1

Row-normalization

- $\boldsymbol{W}_n$ often has non-negative elements
- For row-normalization, condition (2) is satisfied using
  1. Renormalize $\boldsymbol{W}_n$ using $[\boldsymbol{W}_n^*]_{ij} = [\boldsymbol{W}_n]_{ij} / \sum_{j=1}^{n} [\boldsymbol{W}_n]_{ij}$. In words: all row-sums of $\boldsymbol{W}_n^*$ are made equal to one
  2. Optimize concentrated log-likelihood over $\lambda < 1$, then

$$\lambda k_1(\boldsymbol{W}_n^*) < 1 \qquad \Longrightarrow \qquad \boldsymbol{S}_n(\lambda) \text{ invertible}$$

- This works because $\boldsymbol{W}_n^*$ is now a row-stochastic matrix

$\longrightarrow$ $\boldsymbol{W}_n^*$ has an eigenvalue equal to 1 (why?)

$\longrightarrow$ No eigenvalues of $\boldsymbol{W}_n^*$ can exceed 1 because all elements of $\boldsymbol{W}_n^* \boldsymbol{x}$ are convex combinations of $x_1, \ldots, x_n$

The Estimation Approach (3)

1. Select $\boldsymbol{W}_n$ and normalize

2. Maximize the concentrated log-likelihood

$$\log L_n(\lambda) = -\frac{n}{2}\Big(\log(2\pi)+1\Big) - \frac{n}{2}\log\left(\widehat{\sigma}^2(\lambda)\right) + \log\left(\det(\boldsymbol{S}_n(\lambda))\right)$$

   to determine the MLE $\widehat{\lambda}$
   - "Easy" univariate optimization problem (grid search, Newton-Rhapson, etc.) due to likelihood concentration
   - Normalization provides admissible values for λ

3. Consider $\boldsymbol{y}_n - \widehat{\lambda}\boldsymbol{W}_n\boldsymbol{y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n$
   - Newly computed dependent variable $\boldsymbol{y}_n - \widehat{\lambda}\boldsymbol{W}_n\boldsymbol{y}_n$
   - OLS estimator equal $\widehat{\boldsymbol{\beta}}(\widehat{\lambda})$
   - Average SSR is $\widehat{\sigma}^2(\widehat{\lambda})$

Extra 1: Asymptotic Results in Lee (2004)

- The objective function is nonlinear $\implies$ no finite sample results
- Lee (2004) provides explicit conditions guaranteeing **consistency** of the MLE
  - Normally distributed errors are not needed... assuming $\boldsymbol{\varepsilon}_n \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ is just a convenient way of finding an estimator (compare OLS)
- Asymptotic normality also holds (albeit under more restrictive conditions) and allows for inference

Extra 2: Other Estimation Approaches

- IV estimation in Kelejian and Prucha (1998)
- Lin and Lee (2010) investigate GMM

Remarks:

- IV and GMM allow for heteroskedasticity
  - Instruments are readily available (why?)
- ML is most efficient ("smallest" asymptotic covariance matrix) under homoskedasticity

Specific Overview for this Lecture

4. **Spatial econometrics: the Spatial Autoregressive (SAR) Model**
   a. Modeling Spatial Data
   b. The spatial weight matrix
   c. Estimation
   d. On normalization

5. **Assignment** $\Longrightarrow$ You have seen the three exercises!

References

Kelejian and Prucha (1998), *A Generalized Spatial Two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances*, The Journal of Real Estate Finance and Economics

Lee (2004), *Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models*, Econometrica

Lin and Lee (2010), *GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity*, Journal of Econometrics