Spatial and Regional Economic Analysis Mini-Course:

Author: Vicente Rios

Lecture 2

- **Spatial Econometric Models**
- **Estimation**
- **Inference and Spillovers**
- **Spatial Bayesian Model Selection**

Spatial Econometric Models

- **Spatial econometric models deal with interaction effects among geographical units.**

- Examples are economic growth rates of OECD countries over T years, monthly unemployment rates of EU regions in the last decade, and annual tax rate changes of all jurisdictions in a country since the last election.

- **Spatial interaction effect models**

In modeling terms, three different types of interaction effects can be distinguished:

(i) Endogenous interaction effects among the dependent variable (WY)

(ii) Exogenous interaction effects among the independent variables (WX)

(iii) Interaction effects among the error terms (Wu).

- **Endogenous interaction effects**

Refer to the case where **the decision of a particular unit A** (or its economic decision makers) **to behave in some way depends on the decision taken by other units, among which, say, unit B**:

Dependent variable y of unit A ↔Dependent variable y of unit B

Endogenous interaction effects are typically considered as the **formal specification for the equilibrium outcome of a spatial or social interaction process**, in which the value of the dependent variable for one agent is jointly determined with that of the neighboring agents.

Literature on **strategic interaction among local governments**, for example, endogenous interaction effects are theoretically consistent with the situation where **taxation and expenditures on public services interact with taxation and expenditures on public services in nearby jurisdictions** (Brueckner 2003).

Spatial Econometric Models

- **Exogenous interaction effects**

Exogenous interaction effects, where the **decision of a particular unit to behave in some way depends on independent explanatory variables of the decision taken by other units**

- Independent variable x of unit B →Dependent variable y of unit A

- **Capital can flow across borders**; hence the amount an individual economy saves does not have to be the same as the amount it invests**. Per capita income in one economy also depends on the savings rates of neighboring economies**. Not only the savings rate but also other explanatory variables may affect per capita income in neighboring economies.
- In both the theoretical and the empirical literature on **economic growth and convergence among countries or regions** is not only taken to **depend on the initial income level and the rates of saving, population growth, technological change and depreciation in the own economy, but also those of neighboring economies** (Ertur and Koch 2007; Elhorst et al. 2010)

Spatial Econometric Models

Interaction effects among the error terms

- Error term u of unit A ↔Error term u of unit B

Interaction effects among the error terms **do not require a theoretical model for a spatial or social interaction process**, but, instead, is **consistent with a situation where determinants of the dependent variable omitted from the model are spatially autocorrelated,** and with a situation where **unobserved shock follow a spatial pattern.**

Spatial Econometric Models

Originally, **the central focus of spatial econometrics has been on one type of interaction effect** in a single equation cross-section setting.

Usually, **the point estimate of the coefficient of this interaction effect** was used to **test the hypothesis** as to whether **spatial spillover effects exist**.

Most of the work was **inspired by research questions arising in regional science and economic geography**, where the units of observations are geographically determined and **the structure of the dependence** among these units can somehow be **related to location and distance**.

However, more recently**, the focus has shifted to models with more than one type of interaction effects, to panel data, and to the marginal effects of the explanatory variables** in the model rather than the point estimates of the interaction effects

Spatial Lag Model

- Consider

$$y_i = \alpha + \rho \sum_{j=1}^{n} w_{ij}\, y_j + \boldsymbol{X}\beta + \varepsilon_i$$

where:

w_{ij} is the th element of W

$\sum_{j=1}^{n} w_{ij}\, y_j$ is the weighted average of the dependent variable

ε_i is the error term such that $\mathrm{E}(\varepsilon_i) = 0$

ρ is the spatial autoregressive parameter which measures the intensity of the spatial interdependence

$\rho > 0$ positive spatial interdependence

$\rho < 0$ negative spatial interdependence

$\rho = 0$ traditional OLS model

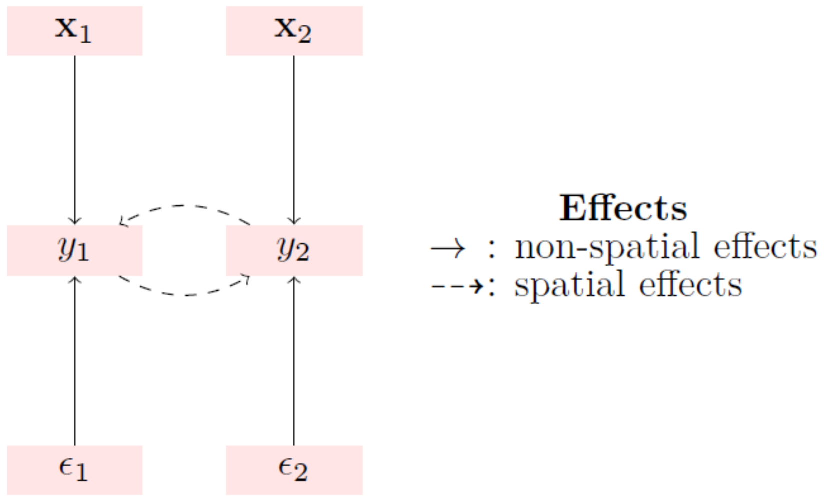
This model is known as Spatial Lag Model (SLM) or the Spatial Autoregressive Model (SAR)

Spatial Lag Model

- In this model, a change in a regressor k in spatial unit i ΔX_{ik} is directly transmitted to the dependent variable of spatial unit i producing a change Δy_i
- However, the effect on i is also transmitted to j: $\Delta y_i \rightarrow \Delta y_j$

and the effect in j is transmitted back to i: $\Delta y_j \rightarrow \Delta y_i$

and back to j….



Spatial Lag Model

- The SLM specification with covariates in matrix form can be written as:

$$\boldsymbol{y} = \alpha \boldsymbol{\iota_n} + \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\beta + \varepsilon$$

where $\boldsymbol{y}$ is a n x 1 vector of observations of the dependent variable, $\boldsymbol{X}$ is an n x K matrix of observations on the explanatory variables, β is a k x 1 vector of parameters and $\boldsymbol{\iota_n}$ is an n x 1 vector of ones.

It is also important to find the **reduced form of the process**.

The reduced form of a system of equations is the result of solving the system for the endogenous variables.

This gives the **endogenous variables as functions of exogenous variables**.

For example, the general expression of a structural form is $\mathrm{f}(\boldsymbol{y}, \boldsymbol{X}, \varepsilon) = 0$ whereas the reduced form is given by $\boldsymbol{y} = \mathrm{g}(X, \varepsilon)$, with g as function.

Spatial Lag Model

Reduced form of the SLM

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\beta + \varepsilon$$

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

We need $(I - \rho \boldsymbol{W})^{-1}$ to be invertible. From standard algebra theory any matrix A is invertible if det(A) is non zero.

If $\boldsymbol{W}$ is not row-normalized $(I - \rho \boldsymbol{W})^{-1}$is invertible if:

$$\omega_{min}^{-1} < \rho < \omega_{max}^{-1}$$

$\omega_{min}, \omega_{max}$ are the minimum and maximum eigenvalues of $\boldsymbol{W}$

If W is row-normalized, then $(I - \rho \boldsymbol{W})^{-1}$ is invertible if:

$$|\rho| < 1$$

Therefore, the spatial structure embodied in W is closely connected to ρ.
Spatial stationariy is garanteed if $\omega_{min}^{-1} < \rho < 1$ (different from time-series)

Spatial Lag Model

The expectation is given by:

$$\boldsymbol{E}(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{W}) = E[(I-\rho\boldsymbol{W})^{-1}\{\alpha\boldsymbol{\iota_n} + \boldsymbol{X}\beta\} + (I-\rho\boldsymbol{W})^{-1}\varepsilon|X,W\ ]$$

To understand this expresion, we need to know the Leontief expansion:

If $|\rho| < 1$, then $(I-\rho\boldsymbol{W})^{-1} = \sum_{i=0}^{\infty}(\rho\boldsymbol{W})^{i}$

Then, we can rewrite the model as.

$$\boldsymbol{y} = \left(I + \rho\boldsymbol{W} + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}+..\right)\{\alpha\boldsymbol{\iota_n} + \boldsymbol{X}\beta\} + \left(I + \rho\boldsymbol{W} + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}+..\right)\varepsilon$$
$$= \alpha\boldsymbol{\iota_n} + \rho\boldsymbol{W}\boldsymbol{\iota_n}\alpha + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\boldsymbol{\iota_n}\alpha +..+\boldsymbol{X}\beta + \rho\boldsymbol{W}\boldsymbol{X}\beta + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\boldsymbol{X}\beta+..$$
$$+\ \varepsilon + \rho\boldsymbol{W}\varepsilon + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\varepsilon$$

This expresion can be simplified since the infinite series:

$$\alpha\boldsymbol{\iota_n} + \rho\boldsymbol{W}\boldsymbol{\iota_n}\alpha + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\boldsymbol{\iota_n}\alpha \rightarrow \frac{\alpha\boldsymbol{\iota_n}}{(1-\rho)}$$

Multiplier effect
Diffusion effect

$$\boldsymbol{y} = \frac{\alpha\boldsymbol{\iota_n}}{(1-\rho)} + \boldsymbol{X}\beta + \rho\boldsymbol{W}\boldsymbol{X}\beta + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\boldsymbol{X}\beta+..+\varepsilon + \rho\boldsymbol{W}\varepsilon + \rho^{\mathbf{2}}\boldsymbol{W}^{\mathbf{2}}\varepsilon+..$$

Spatial Durbin Model

- The DGP is:

$$y = \alpha\iota_n + \rho Wy + X\beta + WX\theta + \varepsilon$$

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

The SDM results in a spatial aturoregressive model of a special form, **including** not only **the spatially lagged dependent variable** and the explanatory variables, but also **the spatially lagged explanatory variables, WX.**

y depends **on own-regional factors** from matrix *(X)*, **plus the same factors averaged over the n neighboring regions** *(WX)*

This model can be defined written as a SAR model by defining:

$Z = [\iota_n \; X \;\; WX]$ and $\delta = [\alpha \;\; \beta \;\; \theta]$

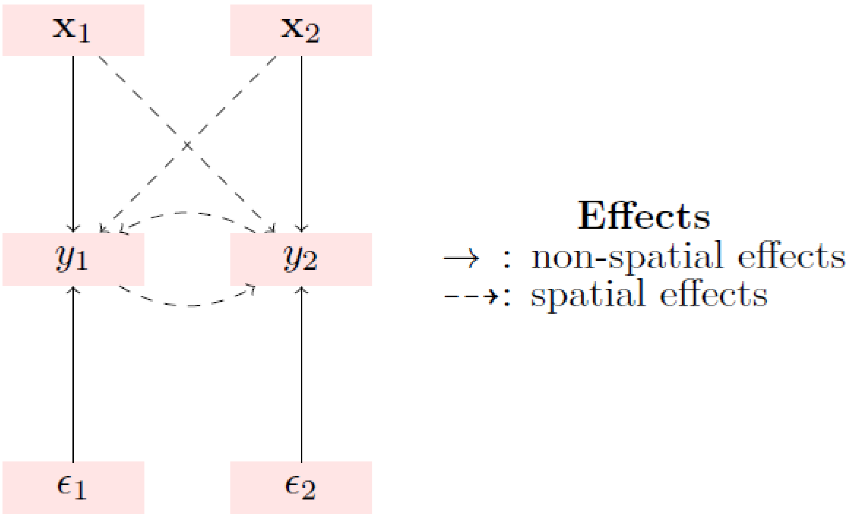
$$y = \rho Wy + Z\delta + \epsilon$$

Spatial Durbin Model

- In the SDM, a change in a regressor k in spatial unit j ΔX_{jk} is directly transmitted to both the dependent variable of spatial unit i producing a change Δy_i and to the dependent variable of j Δy_j
- However, the effect on i is also transmitted to j: $\Delta y_i \rightarrow \Delta y_j$

and the effect in j is transmitted back to i: $\Delta y_j \rightarrow \Delta y_i$

and back to j….



Spatial Error Model

- We can also use spatial lags to reflect dependence in the disturbance process, which lead to the SEM:

$$y = X\beta + u$$
$$u = \lambda W u + \varepsilon$$

The reduced form is given by:

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

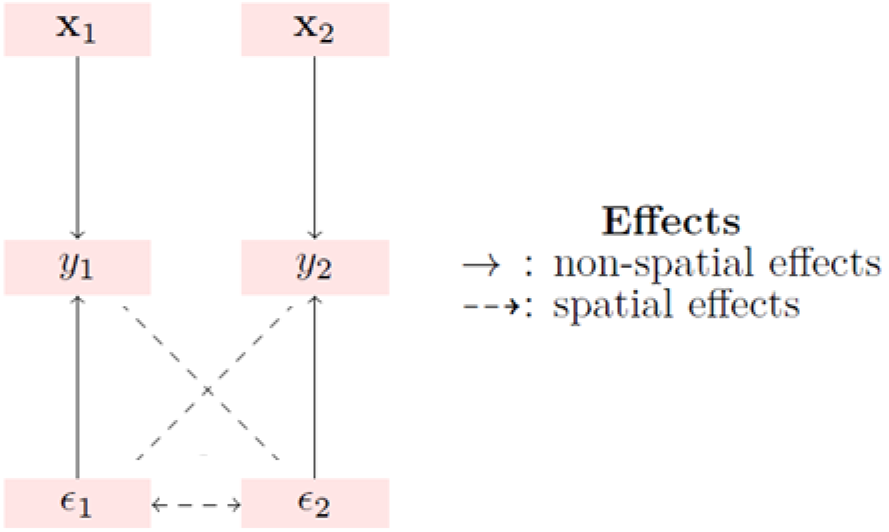
where λ is the autoregressive parameter for the error lag (to distinguish the notation from the spatial autoregressive coefficient in a spatial lag model) and ε is a generally i.i.d noise

Spatial Error Model

- In the SEM a random innovation in spatial unit i ε_i affects the residuals of i, which are spatially correlated, thus affecting the dependent variable.
- In this model, a shock $\varepsilon_i$is directly transmitted to the error term u_i variable of spatial unit i producing a change Δy_i
- But also, the effect of the shock $\varepsilon_i \rightarrow u_j$

and the effect in j is transmitted back to i: $u_j \rightarrow u_i$

and back to j….



Other spatial models

- The Spatial Durbin Error is given by:

$$y = X\beta + WX\theta + u$$
$$u = \lambda W u + \varepsilon$$

The reduced form is given by:

$$y = X\beta + WX\theta + (I - \lambda W)^{-1}\varepsilon$$

- The SLX is given by:

$$y = X\beta + WX\theta + u$$

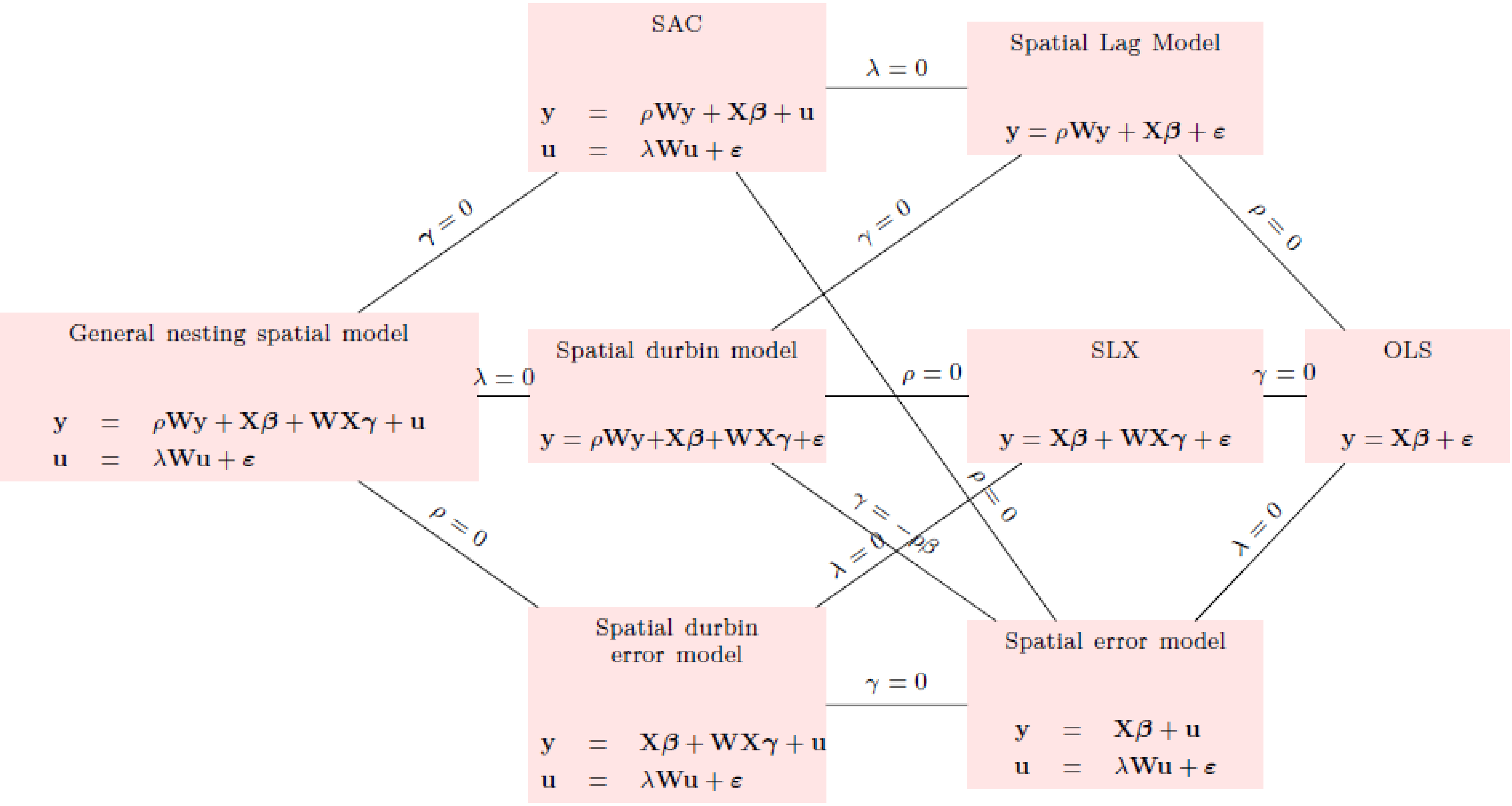
- The SARAR (or SAC) is given by:

$$y = \rho W y + X\beta + u$$
$$u = \lambda W u + \varepsilon$$

- The General Nesting Model (GNS) is given by:

$$y = \rho W y + X\beta + WX\theta + u$$
$$u = \lambda W u + \varepsilon$$

Taxonomy of spatial models



Taxonomy of spatial models

- Traditionally, the **parameter restrictions** that allow the researcher to find the specification for the empirical analysis were analyzed by means of **Likelihood Ratio Tests / Lagrange Multiplier tests**

- **Approaches**
  - Specific to general approach (LM tests, Anselin 1988; RLM Anselin, 1996)

  Start with the OLS, check if SAR/SEM fit better the data

  If they do, then check if the SDM/SDEM fit better the data tan SAR/SEM….

  - General to specific approach (Florax et al., 2003)

  Start with the SDM/SDEM and check if the SDM/SDEM can be simplified to SAR/SEM models.

  If they can, then check if SAR/SEM can be simplified to OLS

  Problem: Depending on your starting point you may get a different specification. Not very reliable.

  Nowadays: Bayesian Model Selection

Motivating Spatial Models

- **Long-run equilibrium motivation**

Consider the following:

$$y_t = \rho W y_{t-1} + X\beta + \varepsilon_t$$

y_t: dependent variable at time t (i.e, house selling price)

Wy_{t-1}: space-time lag (average value of neighbors past year)

X_t: characteristics of regions remain relatively fixed over time

$$X_t = X$$

Note that we can replace $y_{t-1} = \rho W y_{t-2} + X\beta + \varepsilon_{t-1}$ producing:

$$y_t = \rho W(\rho W y_{t-2} + X\beta + \varepsilon_{t-1}) + X\beta + \varepsilon_t$$

$$y_t = X\beta + \rho W X\beta + \rho^2 W^2 y_{t-2} + \varepsilon_t + \rho W \varepsilon_{t-1}$$

Motivating Spatial Models

- Recursive substitution for past values of the vector $\boldsymbol{y}_{t-r}$ on the right hand side of previous expresion over q periods leads to:

$$y_t = \rho W(\rho W y_{t-2} + X\beta + \varepsilon_{t-1}) + X\beta + \varepsilon_t$$

$$y_t = (I + \rho W + \rho^2 W^2 + .. + \rho^{q-1} W^{q-1})X\beta + \rho^q W^q y_{t-q} + u_t$$
$$u_t = \varepsilon_t + \rho W \varepsilon_{t-1} + \rho^2 W^2 \varepsilon_{t-2} + .. + \rho^{q-1} W^{q-1} \varepsilon_{t-(q-1)}$$

Noting that:

$\mathrm{E}(y_t) = (I + \rho W + \rho^2 W^2 + .. + \rho^{q-1} W^{q-1})X\beta + \rho^q W^q y_{t-q}$

where we used the fact that $\mathrm{E}(\varepsilon_{t-r}) = 0$, *r=0,…,q-1* which also implies that $E(u_t) = 0$

Taking the limit:

$$\lim_{q\to\infty} \mathrm{E}(y_t) = (I - \rho W)^{-1} X\beta$$

Note that we use the fact that the magnitude of $\rho^q W^q y_{t-q}$ tends to zero for large q, under the assumption that $|\rho|<1$ and being W row-normalized

Point: A SLM/SAR can emerge as a consequence of a dynamic process where past neighboring decisions are taken into account

Motivating Spatial Models

- **Omitted variable motivation**

Consider the process

$$y = X\beta + \mathrm{z}\delta$$

where *x* and *z* are uncorrelated random vectors of dimension n x 1, and the vector *z* follows the following spatial autoregressive process:

$$z = \lambda W z + r$$

$$z = (I - \lambda W)^{-1} r$$

where $r \sim N(0, \sigma^2 I)$. Examples of *z* are culture, social capital, neighborhood prestige. If z is not observed directly, then:

$$y = X\beta + \mathrm{u}$$
$$u = (I - \lambda W)^{-1} \varepsilon$$

where $\varepsilon = \delta r \;\; \rightarrow \boldsymbol{y} = \boldsymbol{X\beta} + (\boldsymbol{I} - \boldsymbol{\lambda W})^{-\mathbf{1}} \boldsymbol{\delta r} \leftrightarrow \boldsymbol{y} = \boldsymbol{X\beta} + (\boldsymbol{I} - \boldsymbol{\lambda W})^{-\mathbf{1}} \boldsymbol{\varepsilon}$

Then we have the DGP for the spatial error model.

This makes the **SEM a very useful specification when you have omitted variables z uncorrelated with X** that exhibit a spatially correlated pattern (we do not always have a measurement for everything)

SDM and Omitted Variables Motivation

Now suppose that X and z are correlated and given by the following process:

$$y = X\beta + z$$
$$z = \lambda W z + u$$
$$u = X\gamma + v$$
$$v \sim N(0, \sigma^2 I)$$

where the scalar parameters γ and σ^2 govern the strength of the relationship between X and z $= (I - \lambda W)^{-1} u$. Inserting $u = X\gamma + v$ into the SEM we obtain:

$$\begin{aligned} y_t &= X\beta + (I - \lambda W)^{-1} u \\ &= X\beta + (I - \lambda W)^{-1}(X\gamma + v) \\ &= X\beta + (I - \lambda W)^{-1} X\gamma + (I - \lambda W)^{-1} v \end{aligned}$$

$$y(I - \rho W) \quad = (I - \rho W) X\beta + X\gamma + v$$

$$\boldsymbol{y = \rho W y + X(\beta + \gamma) + WX(-\rho\beta) + v}$$

This is the **SDM,** which included a spatial lag of the dependent variable $\boldsymbol{y}$, as well as the explanatory variables $\boldsymbol{X}$.

SDM: useful when we have omitted variables z that follow a spatial patern and that are correlated with the regressors

Estimation: OLS bias

In a **time series** context, **the OLS estimator remains consistent even when a lagged dependent variable is present**, as long as the error term does not show serial correlation

While the OLS estimator may be biased in small simples it can still be used for asymptotic inference.

In spatial context, this rule does not hold, irrespective of the properties of the error term.

Consider the most basic SAR model (with covariates omitted):

$$y = \rho W y + \varepsilon$$

The OLS estimator of ρ would be:

$$\hat{\rho} = \left((Wy)'(Wy)\right)^{-1}(Wy)'y \rightarrow (X'X)^{-1}X'y$$

$$\hat{\rho} = \left((Wy)'(Wy)\right)^{-1}(Wy)'[\rho W y + \varepsilon]$$

$$\hat{\rho} = \left((Wy)'(Wy)\right)^{-1}(Wy)'\rho W y + \left[\left((Wy)'(Wy)\right)^{-1}(Wy)'\varepsilon\right]$$

$$\hat{\rho} = \rho + \left[\left((Wy)'(Wy)\right)^{-1}(Wy)'\varepsilon\right]$$

The second term does not equal to zero and the estimator will be biased

Estimation: OLS bias

If we open the term $(Wy)'\varepsilon$ in the following expression

$$\hat{\rho} = \rho + \left[\left((Wy)'(Wy)\right)^{-1} (Wy)'\varepsilon \right]$$

$(Wy)' = \varepsilon' W' (I - \rho W)^{-1} \varepsilon$ and take its expected value for a W matrix of the form:

$W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ or its row-normalized version $W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

we get

$$E[\varepsilon' W' (I - \rho W)^{-1} \varepsilon] = \frac{2\sigma^2 \rho (3 - \rho^2)}{D} \text{ where } D = 1 - 3\rho^2 + \rho^4$$

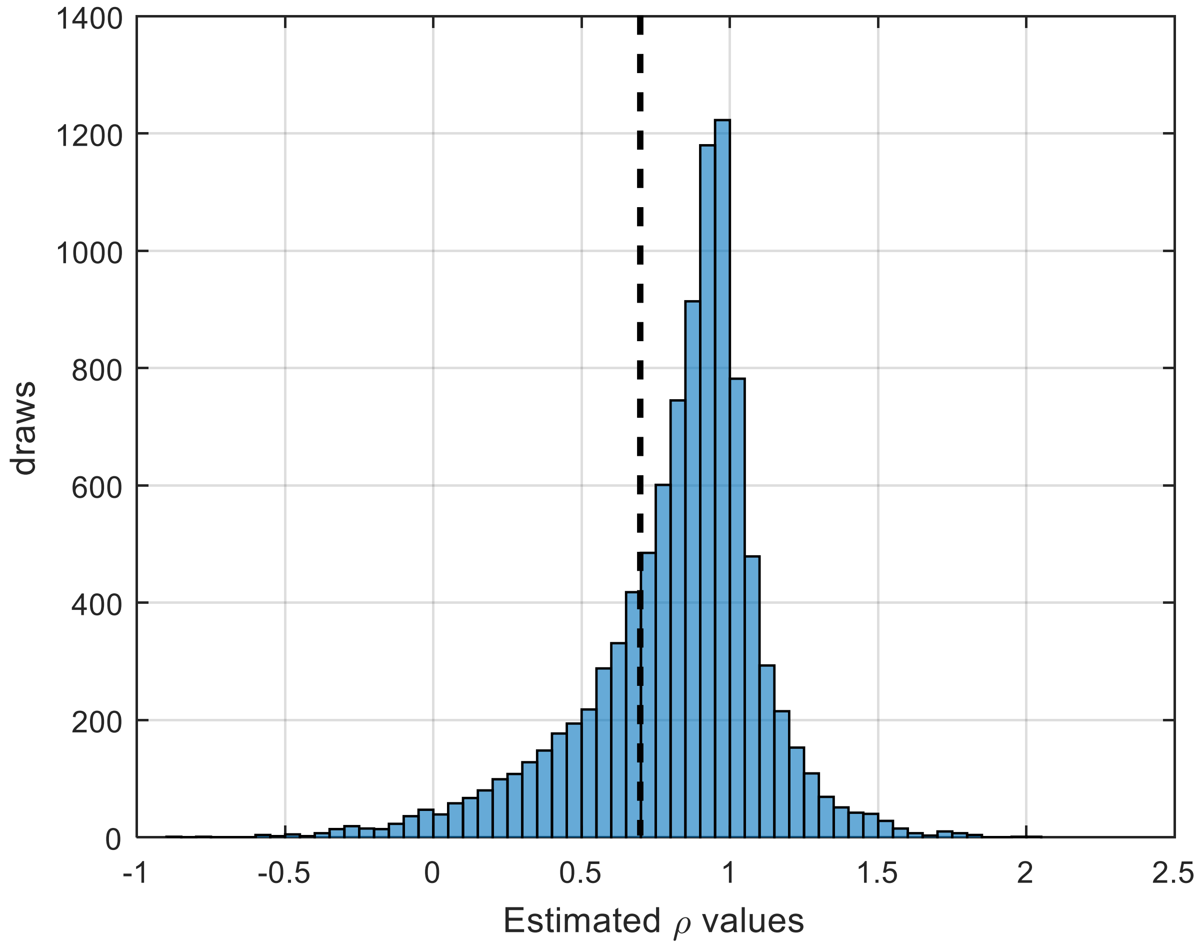
$$E[\varepsilon' W' (I - \rho W)^{-1} \varepsilon] = \frac{2\sigma^2 \rho \left(\frac{5}{4} - \rho^2\right)}{D} \text{ where } D = 1 - \left(\frac{5}{4}\right)\rho^2 + \left(\frac{1}{4}\right)\rho^4$$

which can only be 0 if ρ=0 (the other solutions for ρ violate the condition $|\rho|$<1

In the first case $\rho = +/- \sqrt{3/2}$ while in the second one $\rho = +/- \sqrt{5/4}$

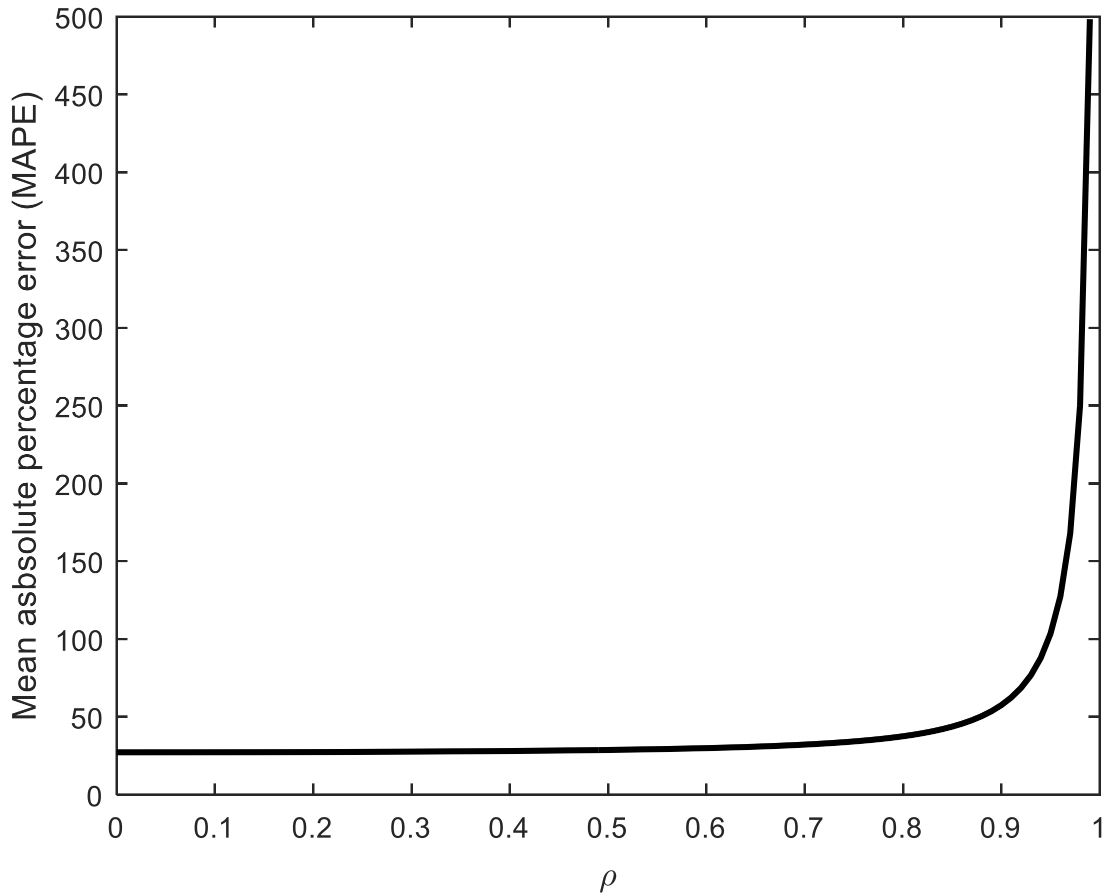
Estimation: OLS bias

$$y = 0.5Wy + \varepsilon, \varepsilon \sim N(0,1)$$



Estimation: OLS bias

- we can see how increased spatial dependence overshoots the bias



Estimation

- Estimation of Spatial Models
  - IV/2SLS regression
  - Maximum Likelihood
  - Bayesian

  (we will focus on SAR/SDM models but similar procedures apply for SEM/SDEM models)

  These estimation techniques both have advantages and disadvantages.

IV-2SLS Estimation

IV/2SLS regression in spatial models attempts to instrument the endogenous spatial term WY in SAR/SDM:

$$y = \rho W y + X\beta + \epsilon$$

The problem when making inferences about the effect of a change in X on y through β is that the things here are not constant because:

$$\Delta X \rightarrow \Delta y \rightarrow W\Delta y \rightarrow \Delta y ..$$

We need some other variables Q (instruments) that correlate with Wy and that at the same time are not caused by y such that we can break

$$\Delta X \rightarrow \Delta y \cancel{\rightarrow} Wy \rightarrow \Delta y$$

Because of $\Delta Q / \Delta y$ = $\Delta Q / \Delta \epsilon$= 0

such that when we look at β, we are effectively picking up:

$$\Delta X \rightarrow \Delta y$$

IV-2SLS Estimation

This **endogeneity** issue can be **addressed with two-stage methods** based on the existence of a **set of instruments Q** which are **correlated with the original variables** *Z=[Wy X]* but uncorrelated with the error term.

If Q is the same column dimension as Z the IV estimate of the parameters of the model $\eta = (\rho, \beta)$ is:

$$\eta = [Q'Z]^{-1}Q'y$$

In the general case where Q is larger than Z, the problem is a minimization of f:

$$\min f(\eta) = (y - Z\eta)'Q(Q'Q)^{-1}Q'(y - Z\eta)$$

with solution:

$$\eta_{IV} = [Z'PZ]^{-1}Z'Py$$

where $P = Q[Q'Q]^{-1}Q'$ is an indempotent projection matrix (i.e, PP=P)

P can be seen as a matrix of predicted values from regressions of Z on the instruments in Q (i.e $\hat{Z}$= f (Q)) in a first stage.

IV-2SLS Estimation

To see this more clearly, the idea would be that of running in a first stage, the following regression:

$$Wy = Q\pi + v$$

and then used predicted values of $\widehat{Wy}$ instead of Wy in:

$$y = \rho Wy + X\beta + \epsilon \rightarrow y = \rho\widehat{Wy} + X\beta + \epsilon$$

To obtain an unbiased/consistent estimator of ρ and β

The full thing/complication on IV/2SLS regressions is therefore "what to pick" as Q.

Main proposals so far:

Kelejian and Prucha: $Q = [X, WX, W^2X]$

Spatial ML Estimation

$$y = \rho W y + \alpha \iota_n + X\beta + WX\theta + \epsilon$$

$$y = (I_n - \rho W)^{-1}(\alpha \iota_n + X\beta + WX\theta + \epsilon\,)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

This model can be defined written as a SAR model by defining:

$Z = [\iota_n \; X \;\; WX]$ and $\delta = [\alpha \;\; \beta \;\; \theta]$ which leads to:

$$y = \rho W y + Z\delta + \epsilon \quad \text{or } y = (I_n - \rho W)^{-1}(Z\delta + \epsilon\,)$$

If the true value of the parameter $\boldsymbol{\rho}$ was known, let's call it $\boldsymbol{\rho}^*$, we could re-arrange the previous expression as:

$$y - \rho^* W y = Z\delta + \epsilon$$

In that case, we could obtain an estimator of δ by OLS:

$$\hat{\delta} = (Z'Z)^{-1}\, Z'(I_n - \rho^* W)y$$

Also, in this case we could also find an estimate of the noise variance parameter:

$$\widehat{\sigma^2} = \frac{1}{n} e(\rho^*) e'(\rho^*)$$

where $e(\rho^*) = \mathbf{y} - \rho^* W y - Z\hat{\delta}$

Spatial ML Estimation

These ideas **motivate** that we can **concentrate the log-likelihood** function of the model with respect δ and σ^2 and **solve the optimization first with respect** $\boldsymbol{\rho}$ and later, use this $\hat{\rho}$ to obtain ML estimates of $\hat{\delta}$ and $\hat{\sigma}^2$

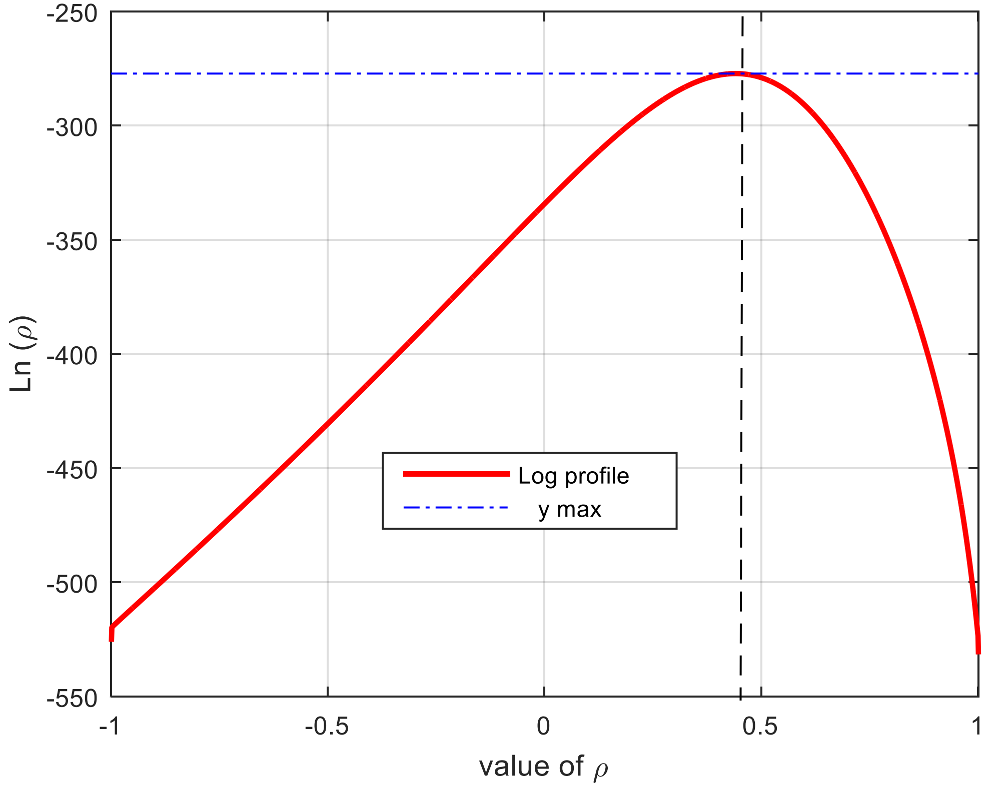
The concentrated log-likelihood function of the SAR/SDM is given by:

$$Ln\ L(\rho) = C + Ln|I - \rho W| - \frac{N}{2} LnS(\rho)$$
$$S(\rho) = e(\rho)'e(\rho) = e_0'e_0 - 2\rho e_0'e_d + \rho^2 e_d{}'e_d$$
$$e(\rho) = e_0 - \rho e_d$$
$$e_0 = y - Z\delta_0$$
$$e_d = Wy - Z\delta_d$$
$$\delta_0 = (Z'Z)^{-1}Z'y$$
$$\delta_d = (Z'Z)^{-1}Z'Wy$$

To accelerate optimization with the respect the sacalar parameter ρ Pace and Barry (1997) proposed using a grid over the feasible Interval [-1,1]

ML Estimation

Idea: Build a sufficient "precise/accurate grid" (i.e, by looping rho over the interval [-1,1] in steps of 0.001 we have 2001 points) defining the likelihood profile:



In this example the $\hat{\rho}$ =0.48. This value is later used to obtain ML estimates of

$$\hat{\delta} = \delta_0 - \hat{\rho}\delta_d \rightarrow \hat{\sigma}^2 = \frac{1}{n}S(\hat{\rho})$$

Spatial Bayesian Estimation

An important aspect of **Bayesian methodology** is the **focus on distributions for the data as well as the parameters**

Bayes's rule involves combining the data distribution embodied in **the likelihood** function **with prior distributions for the parameters** assigned by the practitioner to produce posterior distributions for the parameters

Relevant **information includes both sample data** coming from the likelihood **as well as prior or subjective information** embodied in the distributions assigned to the parameters.

The Bayesian approach to estimation arises from some basic axioms of probability. For two random variables A and B we have that the joint probability p(A,B) can be expressed in terms of conditional probability P(A|B) or P(B|A) and the marginal probability P(A) or P(B):

$$p(A,B) = p(A|B)p(B)$$

$$p(A,B) = p(B|A)p(A)$$

Spatial Bayesian Estimation

Setting the two equal and rearranging:

$$p(A,B) = p(A|B)p(B)$$
$$p(A,B) = p(B|A)p(A)$$

$$p(B|A)p(A) = p(A|B)p(B)$$

gives rise to Baye's rule:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

For our purposes we let A $= D = \{y, X, W\}$ represent the data and $\theta = B$ denote the model parameters such that:

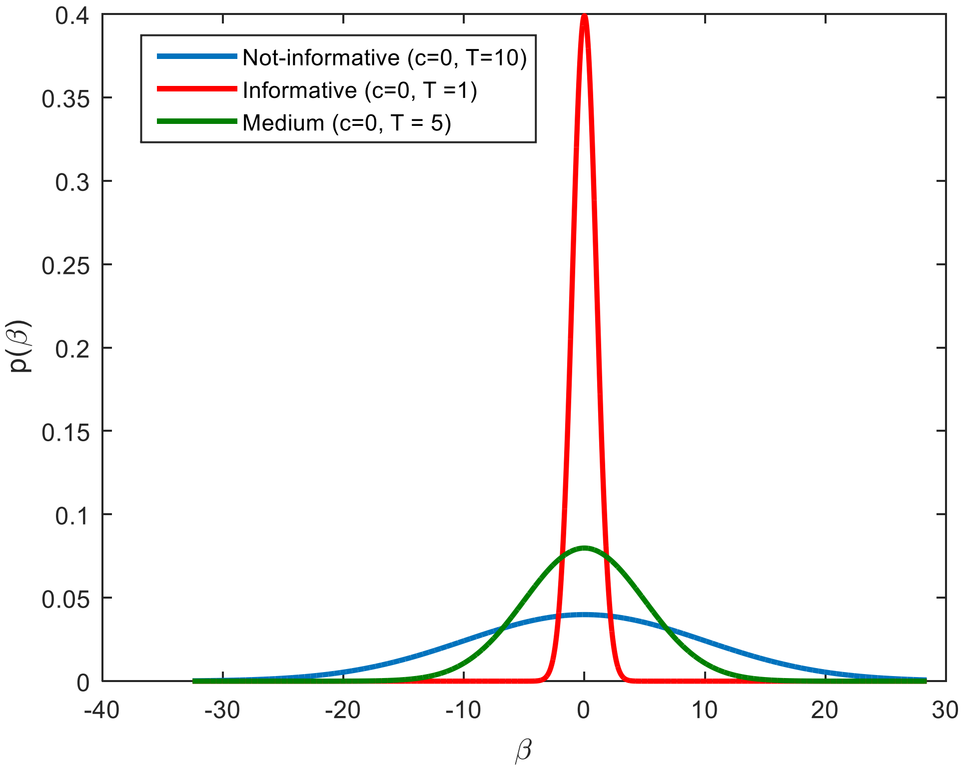
$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Key point: Bayesian modeling assumes the parameters have a prior distribution $\mathbf{p}(\boldsymbol{\theta})$ that reflects previous knowledge as well as uncertainty we have prior to observing the data.

If we know very little then this distribution should represent a vague/ambiguous probabilistic statement

Spatial Bayesian Estimation

- The shape of the prior distribution on the regressors X given by $\pi(\beta) \sim \mathrm{N}(\mathrm{c}, \sigma^2\mathrm{T})$ under different parameterization/degrees of certainty:

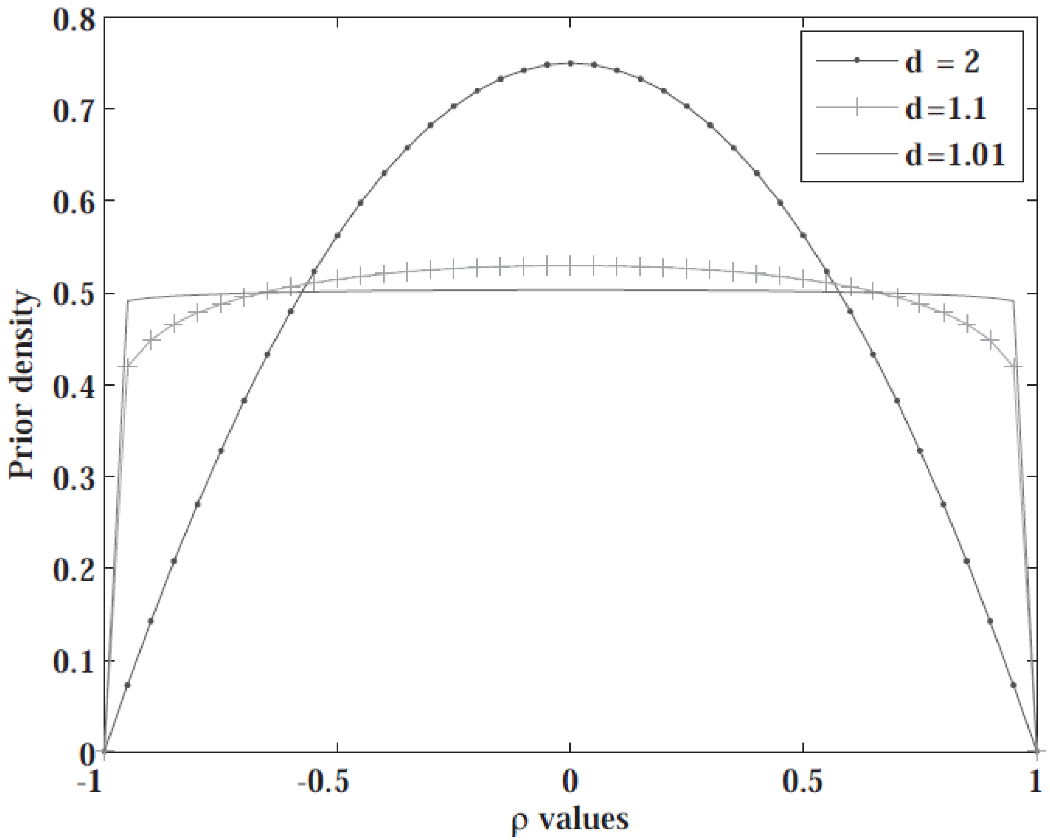


Completely diffuse (flat and non-informative) prior can be obtained setting T equal to a very large number T =10000

Spatial Bayesian Estimation

A common prior for the spatial lag/error term parameter is **the Beta prior** which has a distribution of probability across its range of values given by:

$$\pi(\rho) \sim \frac{1}{Beta(d,d)} \frac{(1+\rho)^{d-1}(1-\rho)^{d-1}}{2^{2d-1}}$$



Spatial Bayesian Estimation

In the expression:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

$p(\theta)$ is the prior distribution of the parameters

Usually:

$\pi(\beta) \sim \mathrm{N}(\mathrm{c}, \sigma^2 \mathrm{T}), \pi(\sigma^2) \sim IG(a, b)$ and $\rho \sim U[-1,1]$ or $\rho \sim Beta[v, d]$

$p(D|\theta)$ is the likelihood function

$p(D)$ is a fixed data distribution that can be ignored

$p(\theta|D)$ is the called the posterior distribution of θ

- it represents an "update" of the prior distribution for the parameters θ after conditioning on the sample data
- all Bayesian inference is based on the posterior density

Oftenly, bayesians work with the simplified expression:

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

Spatial Bayesian Estimation

Usually, closed/analytical expressions for $p(\theta|D)$ do not exist. Thus, Markov Chain Monte Carlo (MCMC) methods are used to approximate this distribution from which we would make inference

The idea of MCMC is that given an initial value for the parameters of the model $\theta_0 = (\beta_0, \rho_0, \sigma_0)$ we can construct a chain of parameter draws that will converge to $p(\theta|D)$ such that:

$$\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \ldots \rightarrow p(\theta|D)$$

Metropolis-Hastings

1. Propose $\theta^* \sim f(\theta_{t+1}|\theta_t) \rightarrow$ (proposal distribution f has to be symmetric (i.e, a Normal distribution with the following motion $\theta_{t+1} = \theta_t + cN[0,1]$)

2. Acceptance $\theta_{t+1} = \theta^*$ with probability:

$$\psi = \min[1, \frac{p(\theta^*|D) f(\theta_t|\theta^*)}{p(\theta_t|D) f(\theta^*|\theta_t)}]$$

Inference

- **Spillovers: Key in Regional science**

A basic definition of spillovers in a spatial context would be "**that changes occurring in one region exert impact on other regions**."

Changes in the tax rate by one spatial unit might exert an impact on tax rate setting decisions of nearby regions, a phenomenon that has been labeled tax mimicking and yardstick competition between local government (see our example below).

Situations where home **improvements made by one homeowner exert a beneficial impact on selling prices of neighboring regions**

Innovation by university researchers that diffuses to neraby firms

Air or watter pollution generated in one region spills over to nearby regions

Essentially: **"to go beyond and diffuse crossing boundaries"**

Spillovers

- Mathematically, **the notion of spillover can be thought as the derivative** $\frac{\partial y_i}{\partial x_j} \neq 0$
- This means that changes to explanatory variables in region *j* impact the dependent variable in region *i*
- In OLS model we have that $\frac{\partial y_i}{\partial x_j} = 0$

Global spillovers

- **Global spillovers**

Global spillovers arise when changes in a characteristic of one region impact all regions' outcomes.

This applies even to the region itself as impacts can pass to the neighbors and back to the own region (feedback).

Specifically, **global spillovers impact the neighbors, neighbors to the neighbors, neighbors to the neighbors to the neighbors, etc.**

Global spillovers are related to endogenous interactions passing through the dependent variable *y*.

SLM: $\boldsymbol{y} = \alpha \boldsymbol{\iota_n} + \rho \boldsymbol{W y} + \boldsymbol{X}\beta + \varepsilon$

SDM: $\boldsymbol{y} = \alpha \boldsymbol{\iota_n} + \rho \boldsymbol{W y} + \boldsymbol{X}\beta + WX\theta + \varepsilon$

They lead to a scenario where changes in one region set in motion a sequence of adjustments in (potentially) all regions in the sample such that a new long-run steady state equilibrium arises.

Local spillovers

- **Local spillovers**

Local spillovers represent a situation where the impact falls only on nearby or inmediate neighbors, dying out before they impact regions that are neighbors to the neighbors.

SLX: $\boldsymbol{y} = \alpha \boldsymbol{\iota_n} + \boldsymbol{X}\beta + WX\theta + \varepsilon$

SDEM: $\boldsymbol{y} = \alpha \boldsymbol{\iota_n} + \boldsymbol{X}\beta + WX\theta + \lambda W u + \varepsilon$

The main difference is that feedback or endogenous interactions are only possible for global spillovers.

However, depending on the structure of W it could happen that a change in a exogenous factor in a very distant place *j* affects the dependent variable of *i*

Measuring Spillovers

- Consider the SDM, which can be rewritten as:

$$\boldsymbol{y} = \alpha\boldsymbol{\iota_n} + \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\beta + WX\theta + \varepsilon$$

$$\boldsymbol{y}(I - \rho\boldsymbol{W}\boldsymbol{y}) = \boldsymbol{X}\beta + WX\theta + \varepsilon$$

$$\boldsymbol{y} = (I - \rho W)^{-1}(\boldsymbol{X}\beta + WX\theta + \varepsilon\,)$$

$$\frac{\partial y}{\partial X_k} = (I - \rho W)^{-1}\begin{pmatrix} \beta_{\boldsymbol{k}} & w_{12}\theta_{\boldsymbol{k}} & \cdots & w_{1n}\theta_{\boldsymbol{k}} \\ w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} & \cdots & w_{2n}\theta_{\boldsymbol{k}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}\theta_{\boldsymbol{k}} & w_{n2}\theta_{\boldsymbol{k}} & \cdots & \beta_{\boldsymbol{k}} \end{pmatrix}$$

Impacts of the system on unit 1

$$\begin{pmatrix} \frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}} \end{pmatrix} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & \frac{\partial E(y_1)}{\partial x_{2k}} & \cdots & \frac{\partial E(y_1)}{\partial x_{nk}} \\ \frac{\partial E(y_2)}{\partial x_{1k}} & \frac{\partial E(y_2)}{\partial x_{2k}} & \cdots & \frac{\partial E(y_2)}{\partial x_{nk}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_{1k}} & \frac{\partial E(y_n)}{\partial x_{2k}} & \cdots & \frac{\partial E(y_n)}{\partial x_{nk}} \end{pmatrix}$$

Impacts of unit 1 on the rest of the system

Measuring spillovers

- **Indirect effects:** the impact on the observed value of location i given a change in the explanatory variable X_k in location *j* is:

$$\frac{\partial E(y_i)}{\partial x_{jk}} = \boldsymbol{S}_k(\boldsymbol{W})_{ij}$$

where $S_k(\boldsymbol{W})_{ij}$ represents the *i*,*j*-th element of the matrix

$$\boldsymbol{S}_k(\boldsymbol{W}) = (I - \rho W)^{-1}[I\beta_{\boldsymbol{k}} + W\theta_{\boldsymbol{k}}]$$

- **Direct effects:** the impact of the expected value of region *i*, given a change in certain variable *k* for the same region *i* is given by:

$$\frac{\partial E(y_i)}{\partial x_{ik}} = \boldsymbol{S}_k(\boldsymbol{W})_{ii}$$

This impact includes the effect of **feedback loops** where observation *i* affects observation *j*, and observation *j* also affects observation *i*

Thus, a change in x_{ik} will affect the expected value of dependent variable in *i*, this will pass through the neighbors of *i (-i or j)* and back to the region itself.

Feedback effects = $\boldsymbol{S}_k(\boldsymbol{W})_{ii} - I\beta_{\boldsymbol{k}}$

Measuring spillovers

Example of Elhorst (2010):

- Suppose we have three spatial units that are arranged linearly: unit 1 is a neighbor of unit 2, unit 2 is a neighbor of both units 1 and 3, and unit 3 is a neighbor of unit 2.

$$W = \begin{pmatrix} 0 & 1 & 0 \\ w_{21} & 0 & w_{23} \\ 0 & 1 & 0 \end{pmatrix}$$

The global spatial multiplier in this system is given by:

$$= (I - \rho W)^{-1} \leftrightarrow (1 - \rho^2) \begin{pmatrix} 1 - w_{23}\rho^2 & \rho & w_{23}\rho^2 \\ \rho w_{21} & 1 & \rho w_{23} \\ \rho^2 w_{21} & \rho & 1 - w_{23}\rho^2 \end{pmatrix}$$

We can get the previous expresion by calculating the inverse analytically (it is very tedious to do it by hand)

Measuring spillovers

$$= (I - \rho W)^{-1} \leftrightarrow 1/(1-\rho^2)\begin{pmatrix} 1 - w_{23}\rho^2 & \rho & w_{23}\rho^2 \\ \rho w_{21} & 1 & \rho w_{23} \\ \rho^2 w_{21} & \rho & 1 - w_{23}\rho^2 \end{pmatrix}$$

Then, the partial derivatives of our dependent variable with respect a change in x_k are given by $\boldsymbol{S}_k(\boldsymbol{W}) = (I - \rho W)^{-1}[I\beta_{\boldsymbol{k}} + W\theta_{\boldsymbol{k}}]$

$$\begin{pmatrix} \frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}} \end{pmatrix}$$

$$= 1/(1-\rho^2)\begin{pmatrix} \beta_{\boldsymbol{k}}(1 - w_{23}\rho^2) + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}} + w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} + \rho\theta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}} + w_{23}\theta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}} + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (1 - w_{21}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \end{pmatrix}$$

Measuring spillovers

Direct effects = diagonal of previous matrix

Indirect effects = every non diagonal elements

What happens with the IEs if ρ = 0 and $\theta_{\boldsymbol{k}} = \mathbf{0}$?

$$\begin{pmatrix}\frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}}\end{pmatrix}$$

$$= 1/(1-\rho^2)\begin{pmatrix}\beta_{\boldsymbol{k}}(1-w_{23}\rho^2)+(w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}}+\theta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}}+(\rho w_{23})\theta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}}+w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}}+\rho\theta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}}+w_{23}\theta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}}+(w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}}+\theta_{\boldsymbol{k}} & (1-w_{21}\rho^2)\beta_{\boldsymbol{k}}+(\rho w_{23})\theta_{\boldsymbol{k}}\end{pmatrix}$$

$$=1/(1-\rho^2)\begin{pmatrix}\beta_{\boldsymbol{k}}(1-w_{23}0)+(w_{21}0)0 & 0\beta_{\boldsymbol{k}}+0 & (w_{23}0)\beta_{\boldsymbol{k}}+(0w_{23})0 \\ (w_{21}0)\beta_{\boldsymbol{k}}+w_{21}0 & \beta_{\boldsymbol{k}}+0 & (0w_{23})\beta_{\boldsymbol{k}}+w_{23}0 \\ (w_{21}0)\beta_{\boldsymbol{k}}+(w_{21}0)0 & 0\beta_{\boldsymbol{k}}+0 & (1-w_{21}0)\beta_{\boldsymbol{k}}+(0w_{23})0\end{pmatrix}$$

$$\begin{pmatrix}\frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}}\end{pmatrix} = \begin{pmatrix}\beta_{\boldsymbol{k}} & 0 & 0 \\ 0 & \beta_{\boldsymbol{k}} & 0 \\ 0 & 0 & \beta_{\boldsymbol{k}}\end{pmatrix}$$

Measuring spillovers

Indirect effects = every non diagonal elements

What happens with the IEs if $\rho \neq 0$ and $\boldsymbol{\theta_k} = \mathbf{0}$? (global effects)

$$\begin{pmatrix} \frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}} \end{pmatrix}$$

$$= 1/(1-\rho^2)\begin{pmatrix} \beta_{\boldsymbol{k}}(1-w_{23}\rho^2) + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}} + w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} + \rho\theta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}} + w_{23}\theta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}} + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (1-w_{21}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \end{pmatrix}$$

$$= 1/(1-\rho^2)\begin{pmatrix} \beta_{\boldsymbol{k}}(1-w_{23}\rho^2) + (w_{21}\rho) & \rho\beta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} & (1-w_{21}\rho^2)\beta_{\boldsymbol{k}} \end{pmatrix}$$

In this case the off-diagonal elements are different from zero so a change in $\frac{\partial E(y_i)}{\partial x_{jk}} \neq 0$

Measuring spillovers

Indirect effects = every non diagonal elements

What happens with the IEs if $\rho = 0$ and $\theta_{\boldsymbol{k}} \neq \mathbf{0}$? (local effects)

$$\begin{pmatrix} \frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}} \end{pmatrix}$$
$$= 1/(1-\rho^2)\begin{pmatrix} \beta_{\boldsymbol{k}}(1-w_{23}\rho^2)+(w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}}+\theta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}}+(\rho w_{23})\theta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}}+w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}}+\rho\theta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}}+w_{23}\theta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}}+(w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}}+\theta_{\boldsymbol{k}} & (1-w_{21}\rho^2)\beta_{\boldsymbol{k}}+(\rho w_{23})\theta_{\boldsymbol{k}} \end{pmatrix}$$

$$= 1/(1-\rho^2)\begin{pmatrix} \beta_{\boldsymbol{k}} & \theta_{\boldsymbol{k}} & \mathbf{0} \\ w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} & w_{23}\theta_{\boldsymbol{k}} \\ \mathbf{0} & \theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} \end{pmatrix}$$

In this case the off-diagonal elements are different from zero so a change in $\frac{\partial E(y_i)}{\partial x_{jk}} \neq 0$

Local effects: if w_{ij} is non-zero(zero) then the effect of $x_{j\boldsymbol{k}}$ on y_i is also non zero (zero).

Measuring spillovers

The direct effects and indirect effects are different for different units in the sample

$$\begin{pmatrix} \frac{\partial E(y)}{\partial x_{1k}} & \frac{\partial E(y)}{\partial x_{2k}} & \frac{\partial E(y)}{\partial x_{3k}} \end{pmatrix}$$

$$= \begin{pmatrix} \beta_{\boldsymbol{k}}(1 - w_{23}\rho^2) + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (w_{23}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \\ (w_{21}\rho)\beta_{\boldsymbol{k}} + w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} + \rho\theta_{\boldsymbol{k}} & (\rho w_{23})\beta_{\boldsymbol{k}} + w_{23}\theta_{\boldsymbol{k}} \\ (w_{21}\rho^2)\beta_{\boldsymbol{k}} + (w_{21}\rho)\theta_{\boldsymbol{k}} & \rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} & (1 - w_{21}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23})\theta_{\boldsymbol{k}} \end{pmatrix}$$

For spatial unit *1* the response of *y* after change in the regressor x_k in spatial units 2 and 3 is given by:

$$IE(1) = \frac{\partial E(y_1)}{\partial x_{2k}} + \frac{\partial E(y_1)}{\partial x_{3k}} = [\rho\beta_{\boldsymbol{k}} + \theta_{\boldsymbol{k}} + (w_{23}\rho^2)\beta_{\boldsymbol{k}} + (\rho w_{23}\theta_{\boldsymbol{k}})]$$

For spatial unit 2:

$$IE(2) = \frac{\partial E(y_1)}{\partial x_{2k}} + \frac{\partial E(y_1)}{\partial x_{3k}} = [(w_{21}\rho)\beta_{\boldsymbol{k}} + w_{21}\theta_{\boldsymbol{k}} + (\rho w_{23})\beta_{\boldsymbol{k}} + w_{23}\theta_{\boldsymbol{k}}]$$

These magnitudes can be different depending on connectivity!

Measuring spillovers

- It can be noted that the change of each variable in each region implies n^2 potential marginal effects
- If we have K variables in our model, this implies kn^2 potential measures
- Even for small vales of n and k it may already be rather difficult to report these results compactly
- We need summary measures!

Average Direct Effects

Average Indirect Effects

Average Total Effects

Measuring spillovers

In the SDM the $\frac{\partial E(y)}{\partial x_k}$ was given by:

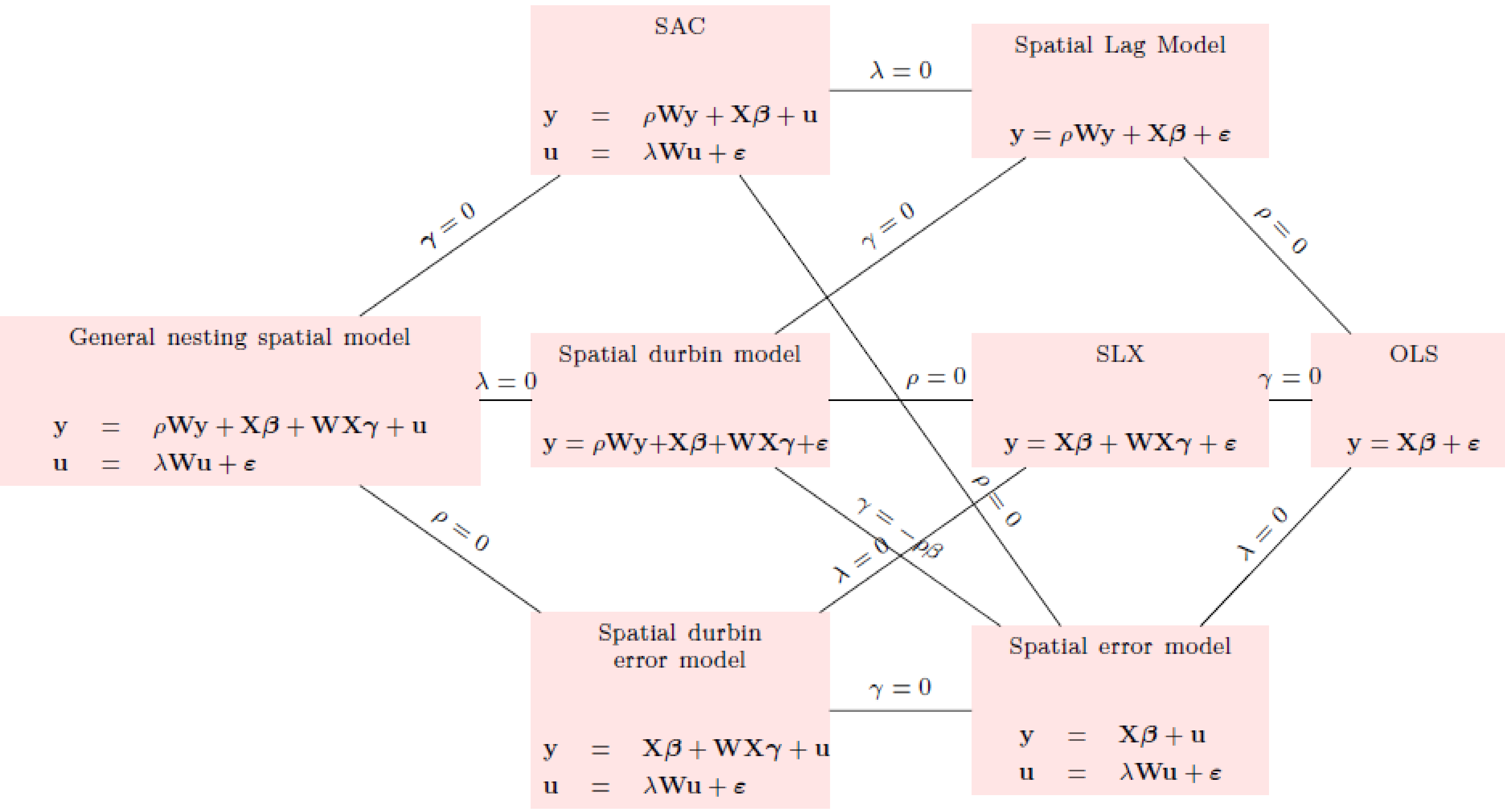
$$= (I - \rho W)^{-1} \begin{pmatrix} \beta_{\boldsymbol{k}} & w_{12}\theta_{\boldsymbol{k}} & \ldots & w_{1n}\theta_{\boldsymbol{k}} \\ w_{21}\theta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}} & \cdots & w_{2n}\theta_{\boldsymbol{k}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}\theta_{\boldsymbol{k}} & w_{n2}\theta_{\boldsymbol{k}} & \ldots & \beta_{\boldsymbol{k}} \end{pmatrix}$$

Direct Effects: are the average of the diagonal elements in previous expression.

Indirect Effects: average row/column off-diagonal elements in previous expression

To analyze its significance → simulation of the distribution of impacts by Monte Carlo methods.

Taxonomy of models



Spatial Model Selection

- Different W matrices imply very distinct measures of spatial autocorrelation, degree of connectedness, etc.
- Different models imply different interpretations for the nature of cross-regional/countries interactions
- Different models may need very different procedures of estimating spillover effects

- The Questions are:

How to select the correct W?

How to select the correct spatial functional form of the model in my analysis?

In the past → Lagrange multiplier tests or Likelihood ratio tests

In the present and future → Bayesian posterior probabilitites

Spatial Model Selection

- Posterior model probability is the key object in Bayesian inference:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Given a set of $i = 1, \ldots, m$ Bayesian models, each would be represented by a likelihood function and a prior distribution as:

$$p\left(\theta^i \middle| D, M_i\right) = \frac{p\left(D \middle| \theta^i, M\right)p\left(\theta^i \middle| M_i\right)}{p(D|M_i)}$$

Treating the posterior distribution as conditional on the model specification M_i we can apply Bayes' rule to expand terms like $p(D|M_i)$ which leads to the set of posterior model probabilities:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{p(D)}$$

This serve as the basis for inference about different models given the sample data. The term $p(D|M_i)$ is called the marginal likelihood:

$$p(D|M_i) = \int p\left(D \middle| \theta^i, M\right)p\left(\theta^i \middle| M_i\right)d\theta^i$$

Spatial Bayesian Model Selection

- We assume a set of alternative spatial models M = M_1, M_2,…, M_m as the possible candidates to explain our data **each of them based on a different spatial weight matrix W** while holding fixed other aspects of the specification such as the X or the functional form (SAR/SDM/SEM/SDEM) fixed.
- Prior probabilitites are specified for each model: $\pi(M_i)$
  - We usually set $\pi(M_i) = \frac{1}{\mathrm{m}}$ making each model equally likely a priori
- Prior distributions for the parameters: $\pi(\eta)$ with $\eta = (\alpha, \beta, \rho, \sigma)$
- The joint probability of the set of models, parameters and data takes the form:

$$p(M, \eta, D) = \pi(M)\pi(\eta|M)\pi(D|\eta, M)$$

Application of Bayes' rule produces the joint posterior for both models and parameters:

$$p(M, \eta|D) = \frac{\pi(M)\pi(\eta|M)\pi(D|\eta, M)}{\pi(D)}$$

Posterior model probabilitites requires integration over η leading to:

$$p(M|D) = \int p(M, \eta|D) d\,\eta$$

Spatial Bayesian Model Selection

Question: Do Bayesian Model Selection Works Well?

Monte Carlo experiments on the accuracy of the model selection.

- W Selection conditional to the functional form
- Functional form selection given W

Sensitivity checks:

[a] Sample Size

[b] Spatial Dependence Intensity

[c] Signal-to-noise ratio

Spatial Bayesian Model Selection

- **Experiment [A]: W Selection conditional to the functional form.**

The true model is:

$$Y_t(W_{k7}) = \alpha + \rho W_{k7} Y_t + \phi Y_{t-1} + \lambda W_{k7} Y_{t-1} + X_t \beta + W_{k7} X_t \theta + \varepsilon$$

- Model comparison is carried over alternative K-nearest neighbor's structures, $W(k) = 1, .., K$ so we estimate a number of models by means of Spatial Bayesian techniques to check if $\text{P}(W_{k7}|Y_t(W_{k7})) > \text{P}(W_{km}|Y_t(W_{k7}))$

- Sensitivity checks:

Signal-to-noise ratio: σ= 2, 5, 10.

Spatio-temporal parameter configurations:

C1: [ρ = 0.3, ϕ = 0.35, λ = -0.1] Medium Spatial Dep

C2: [ρ = 0.7, ϕ = 0.35, λ = -0.2] High Spatial Dep

C3: [ρ = 0.1, ϕ = 0.35, λ = -0.2] Low Spatial Dep

Sample Size: European NUTS-2 regions (n=272) vs Spanish Municipalities above 1,000 inhabitants (n=3032).

Spatial Bayesian Model Selection

Table: Experiment A1: Medium Level of Spatial Dependence in a DSDM

	N=272			N=3032		
Spatial Matrix	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$
knn =1	0.000	0.000	0.000	0.000	0.000	0.000
knn =2	0.000	0.000	0.000	0.000	0.000	0.000
knn =3	0.000	0.000	0.000	0.000	0.000	0.000
knn =4	0.000	0.007	0.014	0.000	0.000	0.000
knn =5	0.033	0.009	0.022	0.000	0.000	0.000
knn =6	0.161	0.122	0.148	0.010	0.020	0.020
knn =7	0.505	0.443	0.406	0.980	0.960	0.942
knn =8	0.209	0.214	0.248	0.010	0.019	0.039
knn =9	0.044	0.119	0.079	0.000	0.000	0.000
knn =10	0.036	0.044	0.044	0.000	0.000	0.000
knn =11	0.000	0.024	0.009	0.000	0.000	0.000
knn =12	0.003	0.004	0.025	0.000	0.000	0.000
knn =13	0.010	0.004	0.002	0.000	0.000	0.000
knn =14	0.000	0.001	0.000	0.000	0.000	0.000
knn =15	0.000	0.010	0.000	0.000	0.000	0.000
knn =20	0.000	0.000	0.004	0.000	0.000	0.000

Notes: Reported results correspond to $C_1 = [\rho = 0.3, \phi = 0.35, \lambda = -0.1]$. The implied signal-to-noise ratios given by $\sigma_\epsilon^2 = 2, \sigma_\epsilon^2 = 5$ and $\sigma_\epsilon^2 = 10$ are $R2 = 0.518$, $R2 = 0.31$ and $R2 = 0.194$ respectively. R-squared values do not include the contribution of spatial fixed effects.

Spatial Bayesian Model Selection

Table: Experiment A2: High Level of Spatial Dependence in a DSDM

	N=272			N=3032		
Spatial Matrix	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$
knn =1	0.000	0.000	0.000	0.000	0.000	0.000
knn =2	0.000	0.000	0.000	0.000	0.000	0.000
knn =3	0.000	0.000	0.000	0.000	0.000	0.000
knn =4	0.000	0.000	0.000	0.000	0.000	0.000
knn =5	0.000	0.000	0.000	0.000	0.000	0.000
knn =6	0.000	0.000	0.006	0.000	0.000	0.000
knn =7	1.000	1.000	0.994	1.000	1.000	1.000
knn =8	0.000	0.000	0.000	0.000	0.000	0.000
knn =9	0.000	0.000	0.000	0.000	0.000	0.000
knn =10	0.000	0.000	0.000	0.000	0.000	0.000
knn =11	0.000	0.000	0.000	0.000	0.000	0.000
knn =12	0.000	0.000	0.000	0.000	0.000	0.000
knn =13	0.000	0.000	0.000	0.000	0.000	0.000
knn =14	0.000	0.000	0.000	0.000	0.000	0.000
knn =15	0.000	0.000	0.000	0.000	0.000	0.000
knn =20	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Reported results correspond to $C_2 = [\rho = 0.7, \phi = 0.35, \lambda = -0.2]$. The implied signal-to-noise ratios given by $\sigma_\epsilon^2 = 2$,$\sigma_\epsilon^2 = 5$ and $\sigma_\epsilon^2 = 10$ are $R2 = 0.643$, $R2 = 0.459$ and $R2 = 0.331$ respectively. R-squared values do not include the contribution of spatial fixed effects.

Spatial Bayesian Model Averaging

Table: Experiment A3: Low Level of Spatial Dependence in a DSDM

	N=272			N=3032		
Spatial Matrix	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$	$\sigma_\epsilon^2 = 2$	$\sigma_\epsilon^2 = 5$	$\sigma_\epsilon^2 = 10$
knn =1	0.020	0.063	0.003	0.000	0.000	0.000
knn =2	0.048	0.031	0.054	0.010	0.010	0.000
knn =3	0.048	0.029	0.057	0.010	0.000	0.005
knn =4	0.037	0.058	0.087	0.049	0.041	0.067
knn =5	0.075	0.063	0.068	0.061	0.084	0.060
knn =6	0.077	0.025	0.081	0.133	0.096	0.102
knn =7	0.140	0.105	0.070	0.225	0.230	0.223
knn =8	0.074	0.087	0.094	0.177	0.150	0.168
knn =9	0.069	0.065	0.066	0.119	0.128	0.068
knn =10	0.085	0.081	0.066	0.085	0.074	0.117
knn =11	0.054	0.066	0.102	0.061	0.041	0.043
knn =12	0.084	0.052	0.067	0.025	0.035	0.073
knn =13	0.063	0.029	0.038	0.004	0.072	0.023
knn =14	0.028	0.119	0.063	0.020	0.022	0.039
knn =15	0.021	0.052	0.043	0.022	0.010	0.001
knn =20	0.077	0.076	0.041	0.000	0.008	0.010

Notes: Reported results correspond to $C_3 = [\rho = 0.1, \phi = 0.35, \lambda = -0.2]$. The implied signal-to-noise ratios given by $\sigma_\epsilon^2 = 2$,$\sigma_\epsilon^2 = 5$ and $\sigma_\epsilon^2 = 10$ are $R2 = 0.533$, $R2 = 0.338$ and $R2 = 0.228$ respectively. R-squared values do not include the contribution of spatial fixed effects.

Spatial Bayesian Model Selection

- **Experiment [B]: Functional form Selection conditional to W**

The true model is a Dynamic SDM

$$Y_t = \alpha + \rho W Y_t + \phi Y_{t-1} + \lambda W Y_{t-1} + X_t \beta + W X_t \theta + \varepsilon_t$$

- Model comparison is carried over alternative spatial functional forms:

DSEM: $Y_t = \alpha + \phi Y_{t-1} + \lambda W Y_{t-1} + X_t \beta + \rho W u_t + \varepsilon_t$

DSLM: $Y_t = \alpha + \rho W Y_t + \phi Y_{t-1} + \lambda W Y_{t-1} + X_t \beta + \varepsilon_t$

DSDEM: $Y_t = \alpha + \phi Y_{t-1} + \lambda W Y_{t-1} + X_t \beta + W X_t \theta + \rho W u_t + \varepsilon_t$

- Sensitivity checks:

Signal-to-noise ratio: σ= 2, 5, 10.

Spatio-temporal parameter configurations:

C1: [ρ = 0.3, ϕ = 0.35, λ = -0.1] Medium Spatial Dep

C2: [ρ = 0.7, ϕ = 0.35, λ = -0.2] High Spatial Dep

C3: [ρ = 0.1, ϕ = 0.35, λ = -0.2] Low Spatial Dep

Sample Size: European NUTS-2 regions (n=272) vs Spanish Municipalities above 1,000 inhabitants (n=3032).

Spatial Bayesian Model Selection

Table: Experiment B1: Baseline Level of Spatial Dependence

	$\sigma_\epsilon^2 = 2$				$\sigma_\epsilon^2 = 5$				$\sigma_\epsilon^2 = 10$			
$W_{N=272}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.83	0.17	0.00	0.00
knn=6	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.82	0.18	0.00	0.00
knn=7	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.83	0.17	0.00	0.00
knn=8	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.82	0.18	0.00	0.00
knn=9	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.82	0.18	0.00	0.00
knn=10	0.96	0.04	0.00	0.00	0.95	0.05	0.00	0.00	0.83	0.17	0.00	0.00
$W_{N=3032}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
knn=6	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
knn=7	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
knn=8	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
knn=9	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
knn=10	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00

Spatial Bayesian Model Selection

Table: Experiment B2: High Level of Spatial Dependence

	$\sigma_\epsilon^2 = 2$				$\sigma_\epsilon^2 = 5$				$\sigma_\epsilon^2 = 10$			
$W_{N=272}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	0.996	0.004	0.000	0.000	0.788	0.212	0.000	0.000	0.735	0.265	0.000	0.000
knn=6	0.996	0.004	0.000	0.000	0.788	0.212	0.000	0.000	0.735	0.265	0.000	0.000
knn=7	0.996	0.004	0.000	0.000	0.788	0.212	0.000	0.000	0.735	0.265	0.000	0.000
knn=8	0.996	0.004	0.000	0.000	0.789	0.211	0.000	0.000	0.725	0.275	0.000	0.000
knn=9	0.996	0.004	0.000	0.000	0.789	0.211	0.000	0.000	0.724	0.276	0.000	0.000
knn=10	0.996	0.004	0.000	0.000	0.789	0.211	0.000	0.000	0.724	0.276	0.000	0.000
$W_{N=3032}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=6	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=7	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=8	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=9	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=10	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00

Spatial Bayesian Model Selection

Table: Experiment B3: Low Level of Spatial Dependence

	$\sigma_\epsilon^2 = 2$				$\sigma_\epsilon^2 = 5$				$\sigma_\epsilon^2 = 10$			
$W_{N=272}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	0.99	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.89	0.11	0.00	0.00
knn=6	0.99	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.89	0.11	0.00	0.00
knn=7	0.99	0.01	0.00	0.00	0.98	0.02	0.00	0.00	0.90	0.10	0.00	0.00
knn=8	0.99	0.01	0.00	0.00	0.97	0.03	0.00	0.00	0.90	0.10	0.00	0.00
knn=9	0.99	0.01	0.00	0.00	0.97	0.03	0.00	0.00	0.90	0.10	0.00	0.00
knn=10	0.99	0.01	0.00	0.00	0.97	0.03	0.00	0.00	0.90	0.10	0.00	0.00
$W_{N=3032}$	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM	SDM	SLM	SDEM	SEM
knn=5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=6	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=7	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=8	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=9	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
knn=10	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00

Spatial Bayesian Model Selection

- **Summing up**
- Bayesian model selection in the context of dynamic spatial panel data models is highly accurate (this also holds for the cross-sections or static spatial panels)

- Accuracy increases with sample size, with a higher signal to noise ratio and with the intensity of spatial dependence

- Bayesian model selection performs better finding true models among alternative spatial specifications than across alternative spatial weight matrices