# Modelling Nonstationary Spatial Lag Models with Hidden Markov Random Fields

C.Ghiringhelli[1] F.Bartolucci [3] A.Mira [2] G.Arbia [1]

[1]Catholic University of Sacread Heart

[2]Università della Svizzera Italiana

[3]University of Perugia

Swiss Statistics Meeting, 2021

# Definition of Spatial model

The Spatial Lag Model (SLM) is defined as (Banerjee et al., 2014):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \rho\boldsymbol{W}\boldsymbol{Y} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$$

where $\boldsymbol{Y} \in \mathcal{R}^n, \boldsymbol{X} \in \mathcal{R}^{n\times p}$ are respectively the response variable and the covariates.

The spatial structure is included in the model by the hendogeneous variable $\boldsymbol{WY}$, where $\boldsymbol{W} \in \mathcal{R}^{n\times n}$ is the spatial weight matrix

$$W_{i,j} = \begin{cases} 1 & \text{if } i\text{-th and } j\text{-th observations are neighbours} \\ 0 & \text{otherwise} \end{cases}$$

then $\boldsymbol{W}$ is standardized (row-standardized, min-max standardized).

Finally, given the maximum and minimum eigenvalues of $\boldsymbol{W}$,

$\rho \in \left[\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}\right]$ is the spatial correlation parameter.

# Including heterogeneity in Spatial Lag Model

$$Y_i = \boldsymbol{X}_i\boldsymbol{\beta} + \rho \boldsymbol{W}_{i,.}\boldsymbol{Y} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2) \qquad i = 1,\dots,n.$$

$$\downarrow$$

$$Y_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \rho_i \boldsymbol{W}_{i,.}\boldsymbol{Y} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,\sigma_i^2) \qquad i = 1,\dots,n.$$

$$\downarrow$$

definition of groups/clusters $\boldsymbol{u}$ of observations with similar relationship between variables

$$Y_i = \boldsymbol{X}_i\beta(u_i) + \rho(u_i)\boldsymbol{W}_{i,.}\boldsymbol{Y} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2(u_i)) \qquad i = 1,\dots,n.$$

Let $K$ be the number of clusters, the groups are identified by a latent discrete process $\boldsymbol{u}$

$$u_i|\boldsymbol{u}_{-i},\boldsymbol{\theta} \sim Multinom(p_{1,i}(\boldsymbol{u}_{-i}),...,p_{K,i}(\boldsymbol{u}_{-i}))$$

where $p_{1,i}(\boldsymbol{u}_{-i}),...,p_{K,i}(\boldsymbol{u}_{-i})$ follows a Potts model (Besag, 1986).

# Prior specification

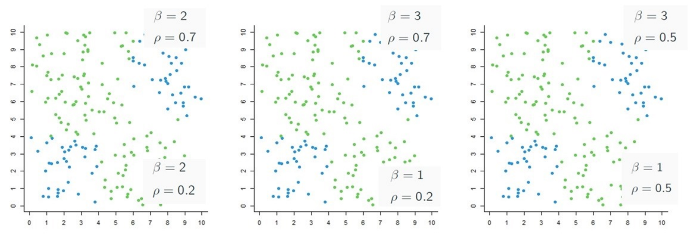
For the model parameters we assume prior distributions characterized by a high variance thus being rather vague (Lingren and Rue, 2015):

$$
\begin{aligned}
&\beta_k \overset{ind}{\sim} \mathcal{N}(0, \sigma^2_\beta), && k = 1, ..., K, \\
&\text{logit}(\rho_k) \overset{ind}{\sim} \mathcal{N}(0, \sigma^2_\rho), && k = 1, ..., K, \\
&\sigma^2_k \overset{ind}{\sim} \mathrm{inv}\Gamma(a, b), && k = 1, ..., K, \\
&\theta_{j,k} \overset{ind}{\sim} \mathcal{N}(0, \sigma^2_\theta), && j = 2, ..., K, \quad k = 1, ..., K,
\end{aligned}
$$

with hyperparameters $\sigma^2_\beta = \sigma^2_\theta = 100$, $\sigma^2_\rho = 10$, and $a = b = 0.001$.

# Simulation study

Data are simulated from Poisson process on a window 10 by 10, we propose several scenarios with different topologies and parameters values.




- The complexity of the topology implies a greater difficulty in the detection of groups,
- Differences in both parameters $\rho$ and $\beta$ results in a clear distinction between groups; variations in the spatial parameter $\rho$ alone are not sufficient for a successful cluster identification.

# A case study: Hedonic house price distribution in Boston

The results reported here refers to an MCMC procedure with a 150,000 sample chain, including the burn-in of 75,000 and a thinin of 15 samples.

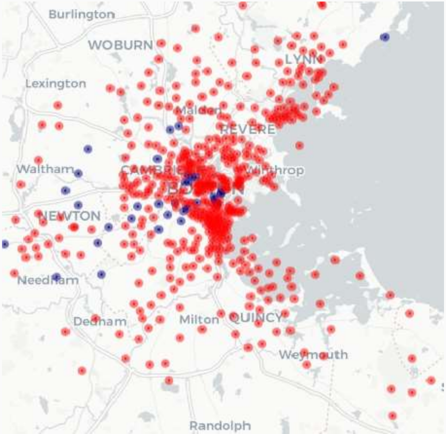
| | intercept | distance | crime | black | $\rho$ | $\sigma^2$ |
|---|---|---|---|---|---|---|
| Cluster 1 | 0.381 | 0.287 | **0.119** | **−1.223** | **0.117** | 15.205 |
| | (8.127) | (0.343) | (0.025) | (0.529) | (0.106) | (5.310) |
| Cluster 2 | **3.901** | **−0.104** | **0.007** | **0.165** | **0.645** | 13.014 |
| | (0.873) | (0.023) | (0.002) | (0.101) | (0.027) | (1.048) |

**Table 1:** In bold we highlight values statistically different from 0.

Quantiles of the dependent variable and predictors included in the model

| | Cluster 1 | | | | | Cluster 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.25 | 0.50 | 0.75 | 0.90 | 0.1 | 0.25 | 0.50 | 0.75 | 0.90 |
| house price | 39.74 | 43.30 | 48.50 | 50.00 | 50.00 | 12.70 | 16.45 | 20.60 | 24.10 | 30.10 |
| distance | 1.342 | 2.006 | 2.894 | 5.118 | 6.310 | 1.648 | 2.105 | 3.216 | 5.215 | 6.819 |
| crime | 0.017 | 0.046 | 0.520 | 1.491 | 6.191 | 0.042 | 0.083 | 0.245 | 3.736 | 11.108 |
| black ratio | 367.41 | 377.29 | 387.31 | 392.71 | 395.46 | 272.21 | 374.71 | 391.98 | 396.90 | 396.90 |

# Identified groups among the observations



Thank you

# References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B*, 48:259–302.

Lingren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63.