Partial MLE, Marginal Effects and Asymptotics for Spatial Autoregressive Nonlinear Probit Models

Anna Gloria Billé *, Samantha Leorato **

* Faculty of Economics and Management, Free University of Bozen–Bolzano, Italy
** Department of Economics and Finance, University of Rome "Tor Vergata", Italy

September 17–19, 2018

XXXIX Conferenza scientifica annuale A.I.S.Re – Bolzano (BZ)
Session: Metodi e modelli di analisi territoriale ed econometria spaziale

Outline

Spatial nonlinear models: literature and motivation

- **Spatial dependence as cross–sectional dependence**: When the data are outcomes measured at different geographical locations the assumption of independence is not plausible. E.g. *Network Economics literature*
- In **spatial econometrics**: The way by which these models are parametrized is convenient as long as we are able to exploit the information gathered not only about the observed outcomes but also on the locations of the dependent variables. E.g. *uniform boundedness assumption*.
- Theoretical papers face the added difficulties in estimating and deriving the asymptotic properties of M–type estimators, see e.g. Lee (2003, ER; 2004, EcTa), Kelejian and Prucha (2010, JoE). E.g. *Bi–directionality nature of spatial dependence*.
- Within random utility theory (McFadden, 2001, AER), **discrete choice models** have an increasing huge literature in both cross-sectional and panel data (see e.g. Wang et al., 2013, JoE; Smirnov, 2010, RSUE; Baltagi et al. 2016)
- With spatial dependence, the optimization of the objective function requires repeated calculations of $(\mathbf{I} - \rho \mathbf{W}_n)^{-1}$. E.g. Klier and McMillen (2008) - Linear approximation (Linearized GMM)

Spatial nonlinear models: literature and motivation

Main problems with MLE for spatial nonlinear models:

1. **Inconsistency** due to unknown forms of cross–sectional dependence: misspecification of the Bernoulli functional form;
2. Spatial autocorrelation implies **spatial heteroskedasticity**: typically due to a non–constant number of neighbors for each spatial unit (exception is the k–nn approach);
3. Functional forms are **highly nonlinear in parameters** (f.i. complications related to the definition of the exact likelihood function).

The likelihood function of model in (3) involve the following n–dimensional integral

$$\ell\left(\boldsymbol{\beta}, \rho, \lambda\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\right|^{\frac{1}{2}}} \int_{S_1} \int_{S_2} \cdots \int_{S_n} e^{-\frac{1}{2}\left(\mathbf{x}'\boldsymbol{\Sigma}^{-1}_{\nu(\rho,\lambda)}\mathbf{x}\right)} \qquad (1)$$

whose computation is *unfeasible for moderate–to–large sample sizes*. The elements $S_i = (a_i, b_i)$ are defined as $a_i = \mathbf{A}_\rho^{-1}\mathbf{X}\beta$ if $y_i = 0$ and $-\infty$ otherwise, and $b_i = \infty$ if $y_i = 0$ and $\mathbf{A}_\rho^{-1}\mathbf{X}\beta$ otherwise.

Spatial nonlinear models: recent literature

Recent advances in estimation of Spatial models for binary dependent v's

- Wang et al. (2013, JoE) - Partial MLE for SAE(1)–probit model.

- Approximate likelihood estimation (ProbitSpatial package in R) - see Martinetti and Geniaux (2017)

- Composite marginal estimation (uni-bivariate) (Mozharovskyi and Vogler (2016))

Spatial nonlinear models: recent literature

Recent advances in estimation of Spatial models for binary dependent v's

- Wang et al. (2013, JoE) - Partial MLE for SAE(1)–probit model.
  - Estimates not precise (large bias for both ρ and $\boldsymbol{\beta}$ in the SAE case)
  - Only considered the simpler (and less interesting) case of spatial correlation in the errors
- Approximate likelihood estimation (ProbitSpatial package in R) - see Martinetti and Geniaux (2017)
  - Particularly fast with very large datasets for SAR(1)–probit (use of sparse matrix and pivoting techniques)
  - Dense matrices? SARAR–probit model? Focus on MC
- Composite marginal estimation (uni-bivariate) (Mozharovskyi and Vogler (2016))
  - Avoid inversion and use fast Cholesky decomposition only for sparse matrix
  - Dense matrices? SARAR–probit? Asymptotics? Focus on MC

Motivation of our paper

1. PMLE estimator with for general spatial binary nonlinear models - **SARAR(1,1)–probit model**, providing the analysis of its asymptotic properties
2. Suggest a Kullback–Leibler (KL) divergence approach to define the **partition of the spatial data** that minimize the loss of statistical information
3. Extensive Monte Carlo experiment to evaluate the **finite sample properties** of the PMLE, including SARAR–probit simulation and dense matrices.
4. Definition of proper **marginal effects** definitions
5. Derivation of the **score of bivariate likelihood** (could complement and improve MV (2016))
6. Our method is computationally more intensive but procedures from MG (2017) or MV (2016) could be implemented in case of sparse weight matrices

Model specification

SARAR(1,1)–probit model

$$\mathbf{y}_n^* = \rho \mathbf{W}_n \mathbf{y}_n^* + \mathbf{X}_n \beta + \mathbf{u}_n, \quad \mathbf{u}_n = \lambda \mathbf{M}_n \mathbf{u}_n + \varepsilon_n, \quad \varepsilon_n \sim \mathcal{N}_n \left(\mathbf{0}_n, \sigma_\varepsilon^2 \mathbf{I} \right)$$
$$\mathbf{y}_{ni} = \mathbb{I} \left(\mathbf{y}_{ni}^* > 0 \right), \; i = 1 \ldots, n. \tag{2}$$

- $\lambda = 0$, **SAR(1)–probit model**. *Direct dependence among endogenous variable*
- $\rho = 0$, **SAE(1)–probit model**. *Intensity of spatial dependence among the shocks*
- **Identification**: (i) $\sigma_\varepsilon^2 = 1$, (ii) at least one $\beta_j, j = 1, \ldots, k$ is highly statistically significant or that $\mathbf{W}_n$ and $\mathbf{M}_n$ are substantially different for the identification of (ρ, λ)

Defining $\mathbf{A}_\rho = (\mathbf{I} - \rho \mathbf{W}_n)$ and $\mathbf{B}_\lambda = (\mathbf{I} - \lambda \mathbf{M}_n)$ we assume

Assumption (A1)

(a) All diagonal elements of $\mathbf{W}_n$ and $\mathbf{M}_n$ are zero, and $\rho \in (-1/\overline{\tau}, 1/\overline{\tau})$, $\lambda \in (-1/\overline{\tau}, 1/\overline{\tau})$ ($\overline{\tau}$ the spectral radius of either $\mathbf{W}_n$ or $\mathbf{M}_n$)
(b) Matrices $\mathbf{W}_n$ and $\mathbf{M}_n$ and $\mathbf{A}_\rho^{-1}$ and $\mathbf{B}_\lambda^{-1}$ are uniformly bounded in both row and column sum norms

Note that $(a) \Rightarrow \mathbf{A}_\rho^{-1}$ and $\mathbf{B}_\lambda^{-1}$ exist (KP,2010).

Reduced Form

Given Assumptions 3.1, the reduced form model of equation (2) is

$$\begin{aligned} \mathbf{y}_n^* &= \mathbf{A}_\rho^{-1}\mathbf{X}_n\beta + \boldsymbol{\nu}_n, \quad \boldsymbol{\nu}_n \sim \mathcal{N}_n\left(\mathbf{0}_n, \boldsymbol{\Sigma}_\nu\right) \\ \mathbf{y}_n &= \mathbb{I}_n\left(\mathbf{y}_n^* > \mathbf{0}_n\right) \end{aligned} \tag{3}$$

where $\boldsymbol{\nu}_n = \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\boldsymbol{\varepsilon}_n$ and $\boldsymbol{\Sigma}_{\nu(\rho,\lambda)} = \mathbb{E}\left[\boldsymbol{\nu}_n\boldsymbol{\nu}_n'\right] = \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\mathbf{B}_\lambda^{-1\prime}\mathbf{A}_\rho^{-1\prime}$.

- MLE is consistent if the conditional density of $\mathbf{y}_n|\mathbf{X}_n$ is correctly specified.
- Consistency can be achieved by correctly specifying the conditional expected value and the robust conditional variances, which however depends on both the unknown autoregressive coefficients, since:

$$\mathbb{E}\left[y_i \mid \mathbf{X}_n\right] = \mathrm{P}\left[y_i = 1 \mid \mathbf{X}_n\right] = \Phi\left(\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\mathbf{X}_n\}_{i.}\boldsymbol{\beta}\right)$$

$$\mathbb{V}\left[y_i \mid \mathbf{X}_n\right] = \Phi\left(\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\mathbf{X}_n\}_{i.}\boldsymbol{\beta}\right)\left[1 - \Phi\left(\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\mathbf{X}_n\}_{i.}\boldsymbol{\beta}\right)\right].$$

Partial (bivariate) MLE

Partial MLE

- In line with Wang et al. (2013), we define bivariate distributions among pairs of random variables in space
- Let assume that the couples (groups of pairs) $g = 1, \ldots, G$, with $n = 2G$ and $g = \{g_1, g_2\}$ are known and fixed
- $\mathbf{A}_\rho^{-1}$ enters in the mean and variance-covariance matrix
- $p_g = (d_1, d_2) = \mathrm{P}\left(y_{g_1} = d_1, y_{g_2} = d_2 \mid \mathbf{X}_n\right)$, where $d_1, d_2 \in \{0,1\}^2$ depends much more on the assumed $\mathbf{W}_n$

Theorem

The joint probabilities $p_g(d_1, d_2)$ are given by:

$$p_g\left(d_1, d_2\right) = \Pr\left\{s_{g_1} Z_1 > s_{g_1} \mathbf{x}_{\rho, g_1} \boldsymbol{\beta}, s_{g_2} Z_2 > s_{g_2} \mathbf{x}_{\rho, g_2} \boldsymbol{\beta}\right\}$$

where $s_{g_i} = 2\left(d_i - 1/2\right)$, and $\mathbf{Z} = \left(Z_1, Z_2\right) \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_g\right)$, $\boldsymbol{\Sigma}_g = \begin{pmatrix} \sigma_{g_1}^2 & \sigma_{g_1, g_2} \\ \sigma_{g_1, g_2} & \sigma_{g_2}^2 \end{pmatrix}$ the $g-th$ diagonal block of $\boldsymbol{\Sigma}_\nu$

Partial (bivariate) MLE

Partial log–likelihood function

$$\ell_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}) = \frac{1}{G}\sum_{g=1}^{G} log\left(p_g\left(y_{g_1}, y_{g_2}\right)\right) \tag{4}$$

where $y_{g_1} \in \{0,1\}$, $y_{g_2} \in \{0,1\}$ and $p_g\left(y_{g_1}, y_{g_2}\right)$ is the $g-$th contribution to the pairwise likelihood function.

Score vector $\nabla(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}) = \left(\nabla_\beta(\boldsymbol{\theta})', \nabla_\rho(\boldsymbol{\theta}), \nabla_\lambda(\boldsymbol{\theta})\right)'$ is equal to

$$\nabla_\beta(\boldsymbol{\theta}) = \frac{1}{G}\sum_g \frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\boldsymbol{\beta}}{p_g\left(y_{g_1}, y_{g_2}\right)}, \qquad \nabla_\rho(\boldsymbol{\theta}) = \frac{1}{G}\sum_g \frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\rho}{p_g\left(y_{g_1}, y_{g_2}\right)}$$

$$\nabla_\lambda(\boldsymbol{\theta}) = \frac{1}{G}\sum_g \frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\lambda}{p_g\left(y_{g_1}, y_{g_2}\right)} \tag{5}$$

where $\frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\boldsymbol{\beta}}{p_g\left(y_{g_1}, y_{g_2}\right)}$, $\frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\rho}{p_g\left(y_{g_1}, y_{g_2}\right)}$ and $\frac{\partial p_g\left(y_{g_1}, y_{g_2}\right)/\partial\lambda}{p_g\left(y_{g_1}, y_{g_2}\right)}$ are the $g-$th contributions to the score with respect to $\boldsymbol{\beta}$, ρ, and λ, respectively.

Choosing couples: a Kullback–Leibler approach

How to choose which units to match in order to **minimize statistical information loss**?

- Given the set of G pairs, the PML is the exact ML of an **approximating model** where spatial dependence is only within pairs.
- Assuming the fixed rule that couples are taken from consecutive units, i.e. $(2g-1, 2g)$, each time we reorder the units according to a particular **permutation** we define a different approximating model.
- So, choosing the best set of couples (the **best approximating model**) is equivalent to choosing the *best* permutation of units, say π
- *Best* according to what **criterion**? A reasonable choice: minimize information loss $\Leftrightarrow$ minimize KL-divergence between the *true* model and the approximating model (defined by π)
- **Approximated criterion**: minimize the KL-divergence between the distributions of the latent models: minimize $KL\left(f_\theta^\pi || f_\theta\right)$ is the KL divergence between the joint distributions built on the latent variables, i.e. f_θ^π (product of bivariate normals) and f_θ (multivariate normal).

Choosing couples: a Kullback–Leibler approach

Theorem

For any $\boldsymbol{\theta} = (\beta, \rho) \in \Theta$, under model (2):

$$\arg\min_{\pi} KL\left(f_\theta^\pi \| f_\theta\right) = \arg\min_{\pi \in \mathcal{P}} \sum_{g=1}^{G} \left(b\left(\pi(2g-1), \pi(2g)\right) - \log\left(\bar{\sigma}(\pi(2g-1), \pi(2g))\right)\right) \tag{6}$$

where $b(i,j) = \sigma^*(i,j)\,\sigma(i,j) + \sigma^*(i,j)\,\sigma(i,j)$, $\bar{\sigma}(i,j) = \sigma(i,j)\,\sigma(i,j) - \sigma(i,j)\,\sigma(i,j)$, $\sigma(i,j)$ is the $(i,j)-$th component of $\boldsymbol{\Sigma}$ and $\sigma^*(i,j)$ is the $(i,j)-$th component of $\boldsymbol{\Sigma}^{-1}$.

1. Start with a *guess* for the value of ρ, i.e. $\tilde{\rho}$, and compute $\tilde{\boldsymbol{\Sigma}}$
2. For all (i,j), $i,j = 1, \ldots, n$ compute $b(i,j)$, $\bar{\sigma}(i,j)$ and $u(i,j) = b(i,j) - log\left(\bar{\sigma}(i,j)\right)$ using $\tilde{\boldsymbol{\Sigma}}$ and its inverse $\tilde{\boldsymbol{\Sigma}}^{-1}$
3. Build the weighted graph G, with n nodes (units) and weights equal to $-u(i,j)$ for edge (distance) $\{i,j\}$
4. Compute the maximum weighted matching, i.e. find the couples that maximize the negative of (6) (using Edmonds' *blossom algorithm*)

Unconstrained optimization

Let define $\mathbf{h} : \Re^{k+1} \to \Omega$: $\mathbf{h} \in \mathcal{C}^2$ and $\mathbf{h}\left(\mathring{\boldsymbol{\theta}}\right) = \boldsymbol{\theta}$ where $\mathring{\boldsymbol{\theta}} = \left(\mathring{\boldsymbol{\beta}}', \mathring{\rho},\right)'$ is the unconstrained vector of parameters (or working parameters) defined in $\Re^{k+1}$, then

$$\mathbf{h}\left(\mathring{\boldsymbol{\theta}}\right) : \begin{cases} \rho = \underline{\omega}_{\rho}^{-1} + \frac{\overline{\omega}_{\rho}^{-1} - \underline{\omega}_{\rho}^{-1}}{1+\exp\left(-\mathring{\rho}\right)}, \\ \boldsymbol{\beta} = \mathbf{h}_{\beta}\left(\mathring{\boldsymbol{\beta}}\right), \quad \text{for} \quad j = 1, \ldots, n \end{cases}$$

where $\left(\underline{\omega}_{\rho}, \overline{\omega}_{\rho}\right)$ are the minimum and maximum eigenvalues of the weighting matrices $\mathbf{W}_n$

$$\mathring{\nabla}\left(\mathring{\boldsymbol{\theta}}; \mathbf{y}, \mathbf{X}\right) = \mathcal{J}\left(\mathring{\boldsymbol{\theta}}; \mathbf{y}, \mathbf{X}\right)' \nabla\left(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}\right)$$

where $\mathcal{J}\left(\mathring{\boldsymbol{\theta}}; \mathbf{y}, \mathbf{X}\right)$ is the Jacobian matrix with respect to the unconstrained parameters, and is equal to

$$\mathcal{J}\left(\mathring{\boldsymbol{\theta}}\right) : \begin{cases} \mathcal{J}\left(\mathring{\rho}\right) = \frac{\left(\overline{\omega}_{\rho}^{-1} - \underline{\omega}_{\rho}^{-1}\right)\exp\left(-\mathring{\rho}\right)}{\left(1+\exp\left(-\mathring{\rho}\right)\right)^2}, \\ \mathcal{J}\left(\mathring{\boldsymbol{\beta}}\right) = \mathcal{J}\left(\boldsymbol{\beta}\right), \quad \text{for} \quad j = 1, \ldots, n. \end{cases}$$

Increasing domain asymptotics

Consistency and Asymptotic Normality of PMLE Assumptions in line with Wang et al. (2013) – assumptions (i)–(viii) Th. 1 and (i)–(iii) Th.2.

Theorem

Under the Assumptions in Wang et al. (2013)

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to \mathcal{N}\left(0, \mathbb{H}(\boldsymbol{\theta}_0)^{-1} J(\boldsymbol{\theta}_0) \mathbb{H}(\boldsymbol{\theta}_0)^{-1}\right) \tag{7}$$

where $\hat{\boldsymbol{\theta}}$ is the PMLE whereas $\mathbb{H}(\boldsymbol{\theta}_0)^{-1}$ and $\mathrm{J}(\boldsymbol{\theta}_0)$ are the Hessian and the Jacobian matrices, respectively.

- An **Approximate Partial MLE** is defined by using a truncation of the infinite series expansion

$$\begin{aligned} \mathbf{A}_\rho^{-1} &= (\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} = \mathbf{I}_n + \rho \mathbf{W}_n + \rho^2 \mathbf{W}_n^2 + \cdots + \rho^q \mathbf{W}_n^q + \ldots \\ \mathbf{B}_\lambda^{-1} &= (\mathbf{I}_n - \lambda \mathbf{M}_n)^{-1} = \mathbf{I}_n + \lambda \mathbf{M}_n + \lambda^2 \mathbf{M}_n^2 + \cdots + \lambda^q \mathbf{M}_n^q + \ldots \quad . \end{aligned} \tag{8}$$

- For the asymptotic issues of the Approximate Partial MLE we need to add some conditions on the rate of the sequence q_n

Assumption

(a) There exists a sequence $\{q_n\}$, with $\lim_{n\to\infty} q_n = \infty$, such that the matrix $\sum_{h=0}^{q_n} \rho^h \mathbf{W}_n^h$ is nonsingular for all n and for all $\rho \in (-1/\overline{\tau}, 1/\overline{\tau})$
(b) $\lim_{n\to\infty} \log n / q_n = 0$.

Marginal Effects

$$\frac{\partial \mathrm{P}\left(y_i = 1 \mid \mathbf{X}_n\right)}{\partial \mathbf{x}'_{.h}} |_{\bar{\mathbf{x}}} = \phi\left(\left\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\right\}_{ii}^{-1/2}\left\{\mathbf{A}_\rho^{-1}\bar{\mathbf{X}}\right\}_{i.}\boldsymbol{\beta}\right)\left\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\right\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\}_{i.}\beta_h \tag{9}$$

$$\frac{\partial \mathrm{P}\left(y_i = 1 \mid \mathbf{X}_n\right)}{\partial \mathbf{x}'_{.h}} |_{\mathbf{x}} = \phi\left(\left\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\right\}_{ii}^{-1/2}\left\{\mathbf{A}_\rho^{-1}\mathbf{X}\right\}_{i.}\boldsymbol{\beta}\right)\left\{\boldsymbol{\Sigma}_{\nu(\rho,\lambda)}\right\}_{ii}^{-1/2}\{\mathbf{A}_\rho^{-1}\}_{i.}\beta_h \tag{10}$$

where $\boldsymbol{\Sigma}_{\nu(\rho,\lambda)} = \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\mathbf{B}_\lambda^{-1\prime}\mathbf{A}_\rho^{-1\prime}$, $\bar{\mathbf{X}}$ is an n by k matrix of regressor–means, $(.)_{i.}$ considers the i–th row of the matrix inside, and $(.)_{ii}$ the i–th diagonal element of a square matrix.

- First specification explains the impact of a change in the mean of the h–th regressor, while the second is the marginal impact evaluated at each single value of $\mathbf{x}_{.h}$
- Spatial marginal effects are then split into an **average direct impact** and an **average indirect impact** (LeSage et al., 2011):
- Observation-level total effects, sorted from low-to-high values of each regressors, can be a source of **spatial heterogeneity**.

Finite sample properties

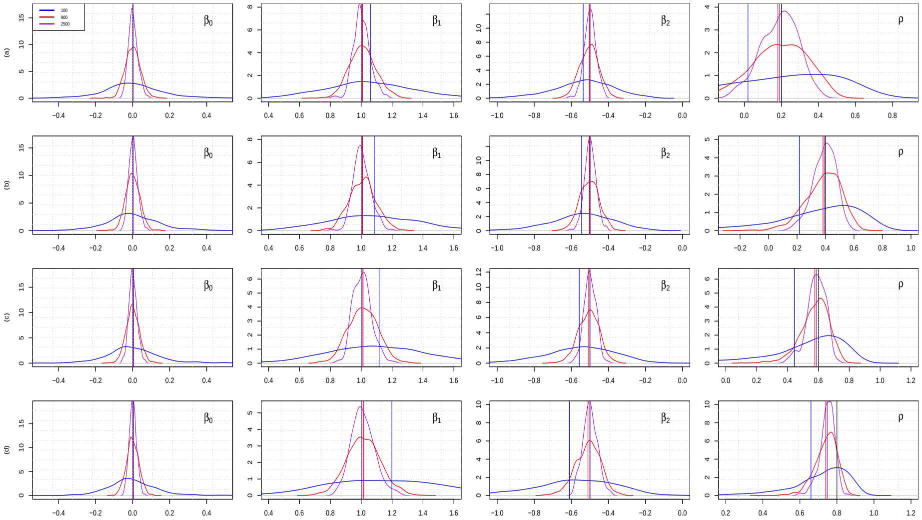
Monte Carlo experiments: SAR(1)–probit

Data generating processes (DGPs)

- $\mathbf{X}_{n\times 3} = [\mathbf{x}_{.0}, \mathbf{x}_{.1}, \mathbf{x}_{.2}]$, $\mathbf{x}_{.1} \sim \mathcal{U}(-1, 1)$, $\mathbf{x}_{.2} \sim \mathcal{N}(0, 1)$, $\boldsymbol{\beta} = (0, 1, -0.5)'$;
- $\mathbf{W}_n$: (i) k–nearest neighbor, $k = 11$ (*sparse* matrix), (ii) inverse distance–based matrix (*dense* matrix); $\rho \in [-0.8, 0.8]$; Simulation runs are $H = 1000$ each;
- Regular square lattice grids: (a) $n = 100$ (10×10), (b) $n = 900$ (30×30), (c) $n = 2500$ (50×50); Randomly generated coordinates;
- **Normalization rule**: Row–normalization, Spectral–normalization
  1) a proper parameter space for ρ
  2) the equivalence of the spatial models after normalization of the weights (Kelejian and Prucha, 2010)
- Note that the case of k–nn:
  - does not depend on "how much units are distant each others" but ensures a constant spatial statistical information (constant number of neighbors)
  - even after row–normalization, the resulting model is equivalent to the original one, which is not the case for other type of distance criteria
- **Robustness check**: Misspecification of $\mathbf{W}_n$.

Finite sample properties

PML estimates



Finite sample properties

Marginal effects with respect to $\bar{\mathbf{X}}$

	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
Regressors	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$
$\bar{\mathbf{X}}$, $\mathbf{x}_{.1}$								
Direct								
Mean	0.398	0.400	0.394	0.397	0.384	0.387	0.351	0.366
sd		0.035		0.035		0.037		0.038
Indirect								
Mean	0.098	0.099	0.251	0.253	0.530	0.515	1.207	0.969
sd		0.093		0.123		0.182		0.305
Total								
Mean	0.496	0.499	0.646	0.649	0.914	0.902	1.558	1.335
sd		0.098		0.130		0.192		0.319
$\bar{\mathbf{X}}$, $\mathbf{x}_{.2}$								
Direct								
Mean	-0.199	-0.199	-0.197	-0.197	-0.192	-0.193	-0.176	-0.184
sd		0.021		0.021		0.022		0.025
Indirect								
Mean	-0.049	-0.050	-0.126	-0.126	-0.265	-0.257	-0.603	-0.486
sd		0.047		0.063		0.095		0.157
Total								
Mean	-0.248	-0.249	-0.323	-0.322	-0.457	-0.450	-0.779	-0.670
sd		0.053		0.069		0.104		0.169

Table: Average Marginal effects summary statistics for different estimated coefficients $\hat{\rho}$.

Finite sample properties

Marginal effects with respect to **X**

	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
Regressors	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$	$m(\rho)$	$m(\hat{\rho})$
X , $\mathbf{x}_{.1}$								
Direct								
Mean	0.315	0.315	0.311	0.312	0.303	0.304	0.277	0.287
sd		0.021		0.021		0.022		0.023
Indirect								
Mean	0.077	0.078	0.198	0.199	0.419	0.404	0.953	0.757
sd		0.073		0.096		0.138		0.222
Total								
Mean	0.392	0.393	0.510	0.511	0.722	0.708	1.231	1.044
sd		0.074		0.097		0.140		0.225
X , $\mathbf{x}_{.2}$								
Direct								
Mean	-0.157	-0.156	-0.156	-0.155	-0.152	-0.151	-0.139	-0.144
sd		0.014		0.014		0.015		0.017
Indirect								
Mean	-0.039	-0.039	-0.099	-0.099	-0.209	-0.202	-0.477	-0.380
sd		0.037		0.049		0.072		0.114
Total								
Mean	-0.196	-0.196	-0.255	-0.254	-0.361	-0.353	-0.615	-0.524
sd		0.040		0.052		0.076		0.120

Table: Local Marginal effects summary statistics for different estimated coefficients $\hat{\rho}$.

Finite sample properties

Misspecification of $\mathbf{W}_n$

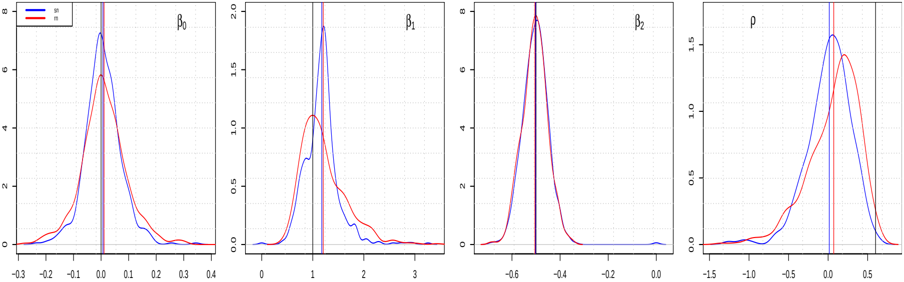


Figure: Gaussian Kernel density for the PML estimated coefficients of the SAR(1)–probit model when $\mathbf{W}_n$ is misspecified. Two cases of misspecification: (i) $\mathbf{W}_{true} = \mathbf{W}_{spectral.invdist}$ (in red), (ii) $\mathbf{W}_{true} = \mathbf{W}_{row.invdist}$ (in blue). The assumed weighting matrix is $\mathbf{W}_{k-nn}$, $n = 900$ and $\rho = 0.6$ are fixed.

Finite sample properties

Misspecification of $\mathbf{W}_n$: marginal effects

	$\bar{\mathbf{X}}$				$\mathbf{X}$			
Regressors	$m(\rho)$	$m(\hat{\rho})$	Lower	Upper	$m(\rho)$	$m(\hat{\rho})$	Lower	Upper
$\mathbf{x}_{.1}$, **Direct**								
Mean	0.399	0.467	0.242	0.731	0.301	0.350	0.184	0.557
sd		0.133				0.099		
Indirect								
Mean	0.329	0.009	-0.264	0.217	0.252	0.005	-0.207	0.165
sd		0.135				0.101		
Total								
Mean	0.728	0.476	0.351	0.616	0.553	0.355	0.284	0.419
sd		0.067				0.035		
$\mathbf{x}_{.2}$, **Direct**								
Mean	-0.199	-0.199	-0.240	-0.160	-0.151	-0.149	-0.176	-0.120
sd		0.023				0.016		
Indirect								
Mean	-0.165	-0.018	-0.165	0.076	-0.126	-0.013	-0.125	0.057
sd		0.063				0.048		
Total								
Mean	-0.364	-0.217	-0.372	-0.108	-0.276	-0.162	-0.276	-0.086
sd		0.067				0.049		

Table: Marginal effects when $\mathbf{W}_n$ is misspecified. The *true* $\mathbf{W}_n$ is based on inverse distance with spectral norm standardization, while a k-nn matrix is used for the estimation.

Choice of pairs

$n = 900$	$\beta_0 = 0$	$\beta_1 = 1$	$\beta_2 = -0.5$	$\rho = 0.6$
	Default pairs			
Mean	0.004	1.091	-0.497	0.433
Median	-0.003	1.044	-0.496	0.591
sd	0.076	0.181	0.054	0.528
RMSE	0.076	0.203	0.054	0.554
	max-matching pairs, $\tilde{\rho} = 0.2$			
Mean	0.004	1.088	-0.499	0.397
Median	-0.003	1.071	-0.497	0.505
sd	0.061	0.140	0.054	0.339
RMSE	0.061	0.165	0.054	0.395
	max-matching pairs, $\tilde{\rho} = 0.6$			
Mean	0.004	1.088	-0.498	0.401
Median	-0.003	1.065	-0.497	0.504
sd	0.059	0.139	0.054	0.338
RMSE	0.059	0.164	0.054	0.392

Table: Summary statistics for the PML estimates of the SAR(1)–probit coefficients using alternative choices of pairs.

Finite sample properties

Monte Carlo experiments: SARAR(1,1)–probit

Data generating processes (DGPs)

- $\mathbf{X}_{n\times 3} = [\mathbf{x}_{.0}, \mathbf{x}_{.1}, \mathbf{x}_{.2}]$, $\mathbf{x}_{.1} \sim \mathcal{U}(-1,1)$, $\mathbf{x}_{.2} \sim \mathcal{N}(0,1)$, $\boldsymbol{\beta} = (0, 1, -0.5)'$;
- $\mathbf{W}_n$: k–nearest neighbor, $k = 11$; $\rho = 0.6$
- $\mathbf{M}_n$: queen contiguity matrix; $\lambda \in \{0.8, 0.6, 0.4, 0.2\}$;
- $n = 900$, simulation runs are $H = 200$ each;
- **Robustness check**: dense matrix for $\mathbf{M}_n$ results in higher variance but only $\hat{\lambda}$ is affected.

Finite sample properties

PML estimates: SARAR(1,1)–probit

$n = 900$									
True Value	Mean	Median	*sd*	RMSE	True Value	Mean	Median	*sd*	RMSE
$\beta_0 = 0.0$	0.006	-0.001	0.178	0.178	$\beta_0 = 0.0$	0.010	-0.002	0.115	0.116
$\beta_1 = 1.0$	0.983	0.992	0.198	0.199	$\beta_1 = 1.0$	1.008	1.001	0.144	0.145
$\beta_2 = -0.5$	-0.488	-0.484	0.106	0.106	$\beta_2 = -0.5$	-0.498	-0.484	0.085	0.085
$\rho = 0.6$	0.527	0.611	0.323	0.332	$\rho = 0.6$	0.542	0.583	0.255	0.261
$\lambda = \mathbf{0.8}$	0.658	0.717	0.246	0.284	$\lambda = \mathbf{0.6}$	0.531	0.561	0.220	0.231
True Value	Mean	Median	*sd*	RMSE	True Value	Mean	Median	*sd*	RMSE
$\beta_0 = 0.0$	0.005	0.002	0.071	0.071	$\beta_0 = 0.0$	0.003	-0.000	0.052	0.052
$\beta_1 = 1.0$	1.014	1.010	0.123	0.123	$\beta_1 = 1.0$	1.019	1.008	0.112	0.113
$\beta_2 = -0.5$	-0.501	-0.489	0.074	0.074	$\beta_2 = -0.5$	-0.501	-0.497	0.062	0.062
$\rho = 0.6$	0.557	0.589	0.192	0.197	$\rho = 0.6$	0.564	0.592	0.150	0.155
$\lambda = \mathbf{0.4}$	0.355	0.376	0.224	0.229	$\lambda = \mathbf{0.2}$	0.165	0.162	0.233	0.236

Empirical Application

Data set in LeSage et al. (2011): evaluate which factors have influenced decisions of establishments in reopening in the aftermath of Hurricane Katrina

- Avoid Zero–distance problems we reduced the sample size from 673 to 658 observations.
- Spatial effects are accounted for to consider potential network effects among these decisions trough the associated utility function
- Coherently with their analysis,
  - SAR(1)–probit model is estimated for three different time horizons: (a) 0–3 months, (b) 0–6 months, (c) 0–12 months. Within each time horizon, firms' decisions are supposed to be simultaneous
  - The weighting matrix: k–nearest neighbor approach with $k = 11$ for time horizon (a) and $k = 15$ for time horizons (b), (c).
- Sampling: $\boldsymbol{\varepsilon}_b \sim \mathcal{N}(0, \mathbf{I})$

$$\mathbf{y}_b^* = \hat{\mathbf{A}}_\rho^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{A}}_\rho^{-1}\boldsymbol{\varepsilon}_b$$
$$\mathbf{y}_b = \mathbb{I}_n(\mathbf{y}_b^* > \mathbf{0}) \tag{11}$$

where $b = 1, \dots, B$ with $B = 200$ samples. Standard deviations are obtained using the distribution of the new estimates $\hat{\boldsymbol{\theta}}_b$ over the B samples.

Empirical Application

	PMLE			Bayes		
Impacts	First	Second	Third	First	Second	Third
Direct						
flood depth	-0.038	-0.027	-0.022	-0.048	-0.028	-0.020
log(median income)	0.141	0.058	0.062	0.212	0.078	0.111
small size	-0.094	-0.054	-0.052	-0.080	-0.028	-0.050
large size	-0.100	-0.107	-0.092	-0.095	-0.094	-0.082
low status customers	-0.126	-0.108	-0.111	-0.095	-0.086	-0.074
high status customers	0.009	-0.002	-0.052	0.025	0.010	-0.023
sole proprietorship	0.155	0.070	0.017	0.160	0.091	0.033
national chain	0.016	-0.024	-0.134	0.020	0.074	-0.029
Indirect						
flood depth	-0.037	-0.041	-0.040	-0.030	-0.034	-0.027
log(median income)	0.140	0.088	0.113	0.128	0.097	0.154
small size	-0.093	-0.082	-0.094	-0.050	-0.035	-0.072
large size	-0.099	-0.163	-0.167	-0.061	-0.121	-0.116
low status customers	-0.125	-0.164	-0.202	-0.058	-0.110	-0.102
high status customers	0.009	-0.002	-0.095	0.015	0.012	-0.034
sole proprietorship	0.154	0.107	0.031	0.099	0.118	0.050
national chain	0.016	-0.036	-0.244	0.012	0.100	-0.037
Total						
flood depth	-0.075	-0.068	-0.062	-0.078	-0.062	-0.048
log(median income)	0.282	0.146	0.175	0.340	0.174	0.265
small size	-0.188	-0.136	-0.146	-0.130	-0.063	-0.122
large size	-0.200	-0.270	-0.259	-0.156	-0.251	-0.199
low status customers	-0.250	-0.272	-0.313	-0.153	-0.195	-0.176
high status customers	0.019	-0.004	-0.147	0.040	0.023	-0.057
sole proprietorship	0.309	0.176	0.048	0.259	0.209	0.083
national chain	0.033	-0.060	-0.378	0.032	0.174	-0.067

Table: Marginal Effects respect to $\mathbf{X}$ for the first, second and third time horizons of the data set Katrina.

Conclusions

- Introduction of the PML estimator for **SARAR(1,1)–probit** models (extending the approach in Wang et al. (2013)) and its approximation based on truncation of the covariance matrix.
- **Complete asymptotic analysis** of both PMLE and Approximate PMLE
- Introduction of **average and local marginal effects**
- **KL-divergence-based method** for the choice of the pairs
- The method is feasible also with **dense weight matrices**
- **Finite-sample performances** are generally good under correct specification, even for very small samples: the sample distribution of $\hat{\rho}$ has higher variability and asymmetry, tending to disappear in larger samples.
- Estimates of β and the direct effects are quite robust to **misspecification**.

Future developments

- Weight matrices and identification issues of the SARAR(1,1)–probit model
- **Other forms of misspecification**
- More on the analysis of the finite sample performance in the **dense weight matrix case** and normalization
- **Computational issues** with very large samples
- Extension to **panel data**

Thank you for your attention!

Some References

- Baltagi, Badi, Peter H. Egger, and Michaela Kesina (2016), Bayesian Spatial Bivariate Panel Probit Estimation, in Badi H. Baltagi , James P. Lesage , R. Kelley Pace (ed.) Spatial Econometrics: Qualitative and Limited Dependent Variables (Advances in Econometrics, Volume 37) Emerald Group Publishing Limited, pp. 119-144.
- Beron, Kurt J., Murdoch, James C. and Wim P.M. Vijverberg (2003) Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol, The Review of Economics and Statistics, 85 (2), 286-297.
- Kelejian, Harry H and Prucha, Ingmar R (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Journal of Econometrics, 157 (1): 53–67.
- Klier, Thomas and McMillen, Daniel P. (2008) Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples, Journal of Business & Economic Statistics, 26 (4), DOI 10.1198/073500107000000188.
- Lee, Lung–fei (2003) Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances, Econometric Reviews, 22 (4), 307-335, DOI: 10.1081/ETC-120025891.
- Lee, Lung–fei (2004) Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models, Econometrica, 72 (6), 1899-1925.

Some References

LeSage, James P. et al. (2011) New Orleans business recovery in the aftermath of Hurricane Katrina, Journal of the Royal Statistical Society: Series A, 174 (4), 10071027.

Martinetti, Davide and Geniaux, Ghislain (2017) Approximate likelihood estimation of spatial probit models, Regional Science and Urban Economics, 64, 30–45.

Mozharovskyi, Pavlo and Vogler, Jan (2016) Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples, Economics Letters, 148, 87–90.

Smirnov, Oleg A. (2010) Modeling spatial discrete choice, Regional Science and Urban Economics, 40, 292–298.

Wang, Honglin, Iglesias, Emma M. and Wooldridge, Jeffrey M. (2013) Partial maximum likelihood estimation of spatial probit models, Journal of Econometrics, 172, 77–89.