HAL
open science

ProbitSpatial: introduction for users

Davide Martinetti, Ghislain Geniaux

To cite this version:

Davide Martinetti, Ghislain Geniaux. ProbitSpatial: introduction for users. 9. Jean Paelinck Seminar of Spatial Econometrics, Universidad Politécnica de Cartagena. ESP., Oct 2017, Cartagena, Spain. pp.22. hal-02791267

HAL Id: hal-02791267

https://hal.inrae.fr/hal-02791267

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

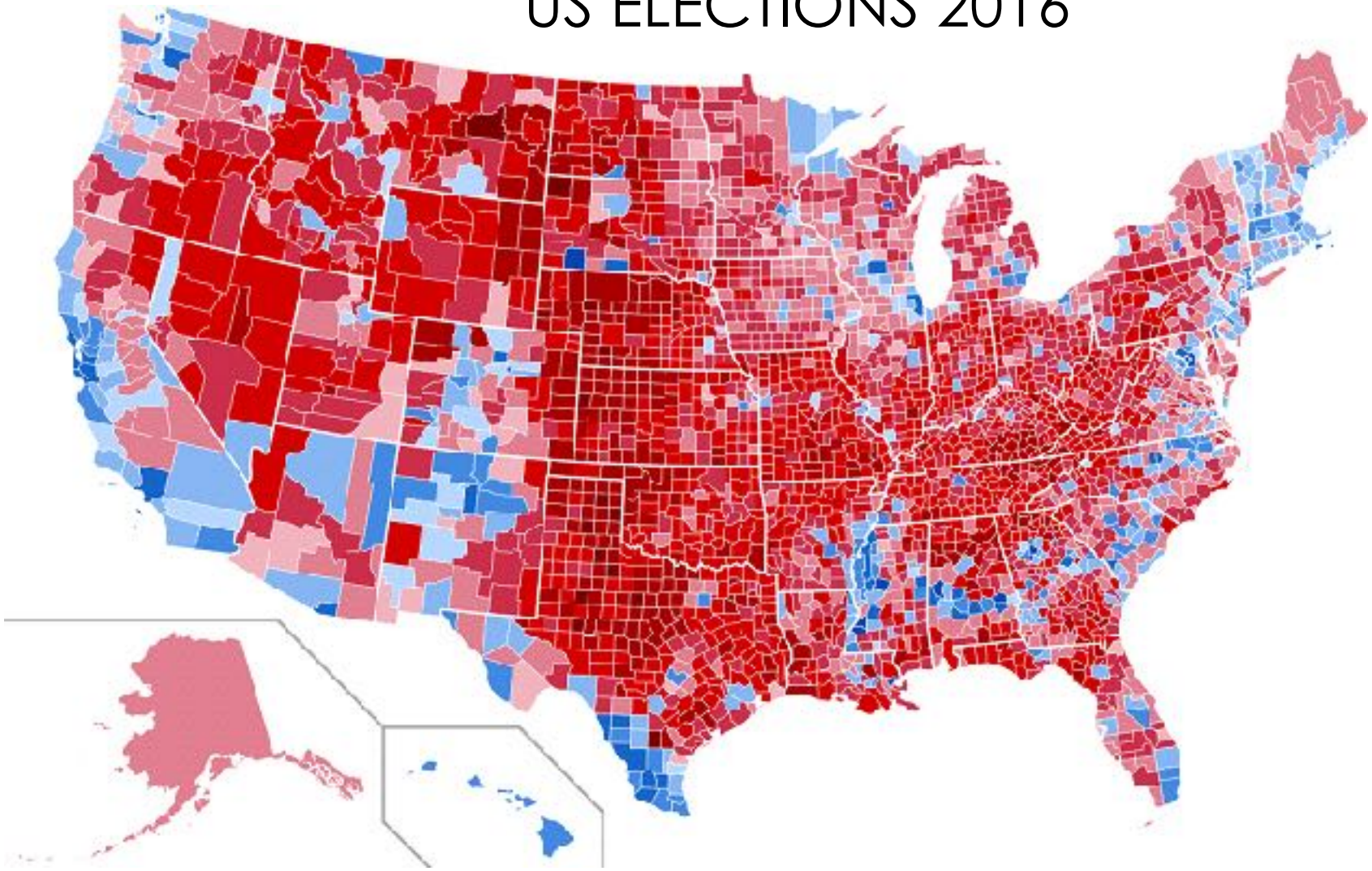
ProbitSpatial
Introduction for users

Fast and accurate spatial probit estimations

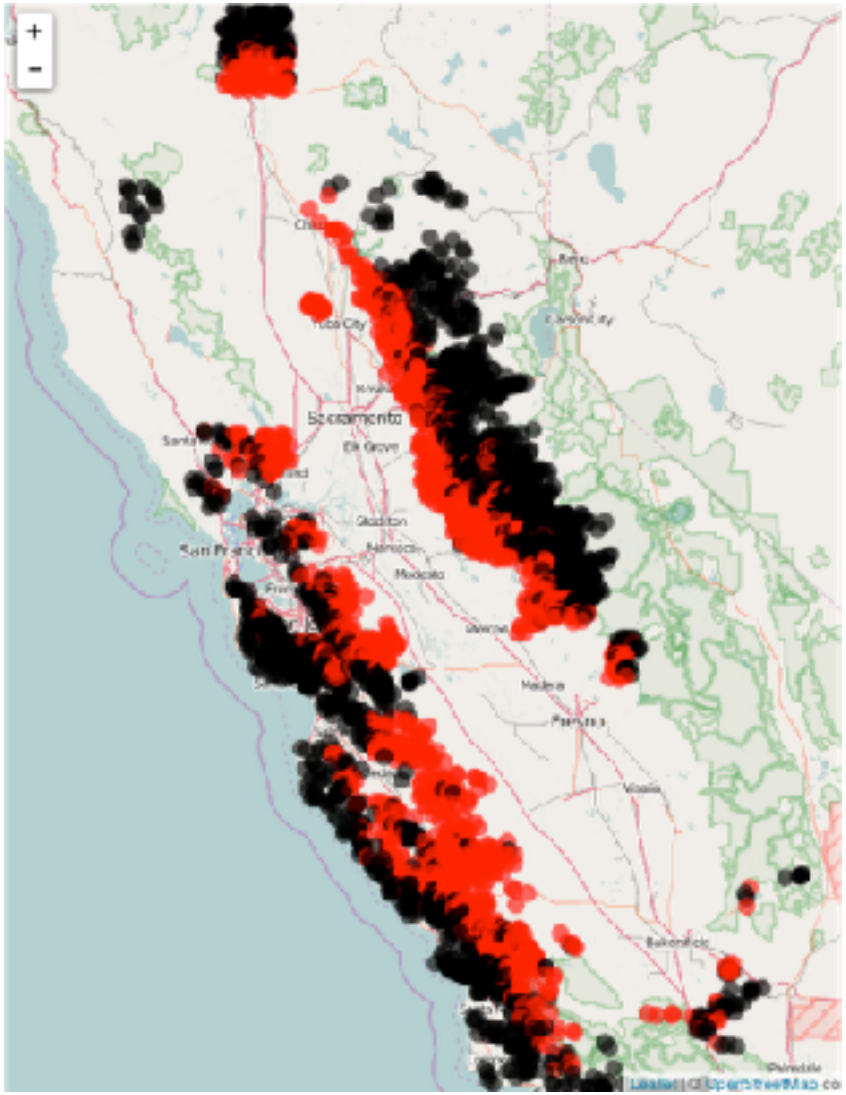
Davide Martinetti
Ghislain Geniaux

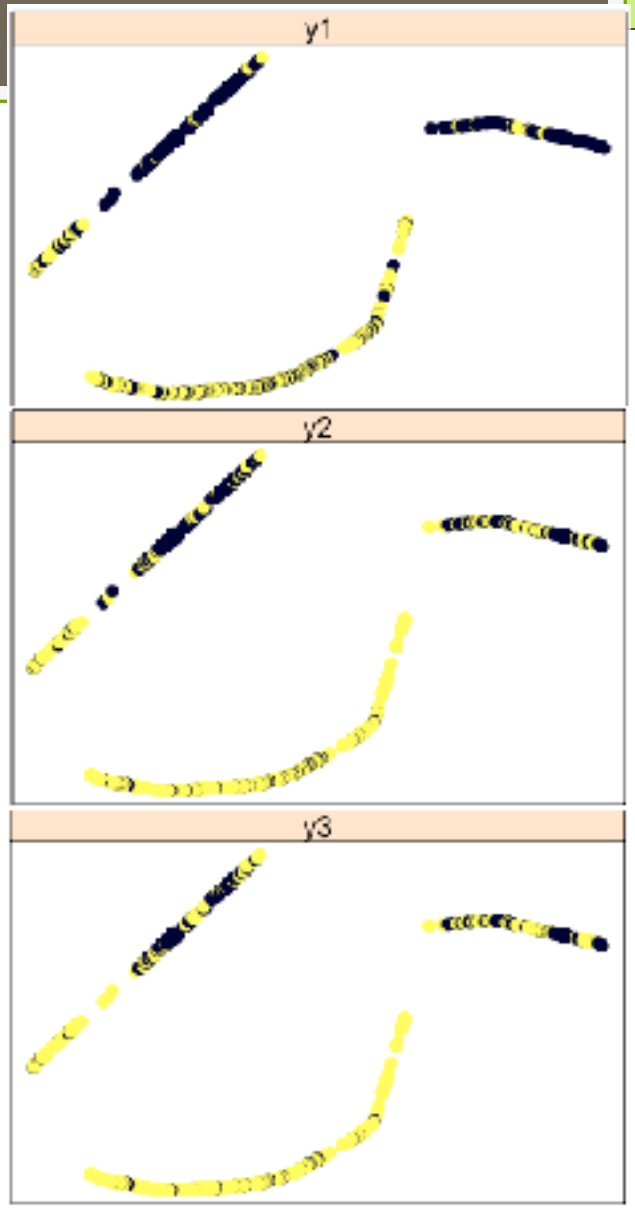
INRA - FRANCE

US ELECTIONS 2016



BLUE OAKS IN CALIFORNIA





REOPENING AFTER KATRINA

Regression of spatial data

- Spatial data tend to show some kind of **spatial pattern**
- Regression models should account for the spatial features of the data, otherwise
  - **Inefficient and inconsistent estimates (McMillen, 1992)**
  - **heteroskedastic and correlated residuals**

PROBIT model

- **Regression model** for binary dependent variable
- **Unobserved** latent variable Y* such that the observed Y=1, if Y*>0 and Y=0 otherwise

$$Y^* = X\beta + \varepsilon$$

$$P(Y^* > 0 \mid X) = \Phi(X\beta + \varepsilon)$$

- Assumptions on the disturbances ε
  - INDEPENDENT $\Phi(0, I_n)$
  - NORMALLY DISTRIBUTED

Spatial PROBIT model

SPATIAL AUTO-REGRESSIVE (SAR)

$$Y^* = \rho W Y^* + X\beta + \varepsilon$$

$$Y^* = (I_n - \rho W)^{-1}(X\beta + \varepsilon)$$

Residuals ~ N(0,Σ)with $\Sigma = (I_n - \rho W)^{-1\prime}(I_n - \rho W)^{-1}$

Spatial PROBIT model

SPATIAL ERROR MODEL (SEM)

$$Y^* = X\beta + u$$

$$u = \rho W u + \varepsilon$$

Residuals ~ N(0,Σ) with $\Sigma = (I_n - \rho W)^{-1\prime}(I_n - \rho W)^{-1}$

Spatial PROBIT model

SAR + SEM = SARAR

$$Y = \rho WY + X\beta + u$$

$$u = \lambda Mu + \varepsilon$$

Residuals ~ N(0,Σ)

$$\Sigma = (I_n - \rho W)^{-1}{}'(I_n - \lambda M)^{-1}{}'(I_n - \lambda M)^{-1}(I_n - \rho W)^{-1}$$

Spatial PROBIT estimation

Estimation problems arise due to **dependence across observations**.

ML methods implies computation of **mvn probabilities.**

- Approximate ML methods
  - E-M (`McSpatial` – McMillen '92)
  - Bayesian Gibbs Sampler (`spatialprobit`– LeSage '00)
  - RIS (a.k.a. GHK, Beron & Vijverberg '03)

- Non-likelihood methods
  - Generalised Methods of Moments (Pinkse & Slade'98)
  - Linearized GMM (Klier & McMillen '08, `McSpatial`)

Expectation-Maximization

McMillen, 1992

E-step: Y is replaced by the expected value of Y*, given estimates $(\hat{\beta}, \hat{\rho})$

M-step: new estimates by max log-likelihood

Repeat until convergence

- PROS: avoid MVN probabilities
- CONS: inversion matrix $(I - \rho W)$, slow

Gibbs sampling

LeSage, 2000

Joint distribution of Y is unknown =>

specifies the complete conditional distributions for all parameters in the model and proceeds to sample from these distributions to collect a large set of parameter

- PROS: distribution parameter => inference
- CONS: slow

Recursive importance sampling, RIS

<u>Beron and Vijverberg (2004)</u>

Direct estimation of the mvn probabilites!

Decompose Σ with Choleski, so each observation depends only on the previous one.

Same as GHK simulator.

- PROS: max-likelihood => LR tests
- CONS: not suitable for large n

GMM and LGMM

<u>GMM (Pinkse & Slade,1998)</u>

Minimization of generalized residuals via GMM with instruments $I + X + WX + W^2X + W^3X + \ldots$

<u>LGMM (Klier & McMillen, 2008)</u>

Linearization of GMM around convenient starting point.

- PROS: fast
- Inefficient compared to ML estimators

Our proposal

ML estimation with approximation of MVN probabilities

Trinh & Genz, 2013 or Mendell & Elston, 1974

Basic ideas

- n-dim integral ➔ Prod. of n 1-dim integrals
- Cholseky decomposition of $\Sigma = CC'$ and variable transformation $x = Cz \Leftrightarrow z = C^{-1}x$
- Iteratively recompute integral limits $a \leq Cz \leq b$ taking advantage of the triangular form of C
- Approximate z using its truncated expected values

Numerical considerations

- Sparse matrix algebra (# of neighs. ➔ # non-zeros)
- Reordering of obs. to minimize error propagation
- Reordering for faster Cholesky decomposition
- Taylor approx. $(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 \ldots$
- `Rcpp & RcppEigen`

One likelihood – four methods

<u>2 optimization strategies:</u>

- Full-likelihood with gradients;
- Conditional likelihood: compute $X^* = (I - \hat{\rho} W)^{-1} X$ conditional to $\hat{\rho}$ and fit standard probit to find β. Optimize w.r.t. $\hat{\rho}$.

<u>2 dependence structures:</u>

- Use Σ as explained earlier;
- Use precision matrix $P = \Sigma^{-1}$, since it is sparser. Lower accuracy, but lower computation time.

Some comparisons

$$Y = \rho WY + \beta_0 + X_1\beta_1 + \varepsilon$$

$$\beta = (4, -2), \rho = 0.45, X_1 \approx N(2, 4)$$

$$nneighs = 5, n = 1500$$

Method	Time (secs.)	Bias β1	Bias ρ
EM	+10000	0.99	-0.099
Gibbs	+10000	0.13	-0.053
RIS	8000	0.19	0.001
GMM	1000	-312	-0.014
Lin-GMM	10	1.19	0.100
ProbitSpatial	1.02	0.01	0.00

Going bigger

Size	Nneighs=2	Nneighs=3	Nneighs=10
100	0.164	0.192	0.234
1000	0.273	1.401	6.810
10000	2.575	5.717	
100000	28.467	127.105	
1000000	504.10		

`ProbitSpatial` package

- ProbitSpatial uniqueness
  - SEM and SARAR
  - Between 10 and 10000 times faster
  - Can digest sample sizes of 1 million
- Existing alternatives
  - In `R: McSpatial` or `spatialprobit`
  - In `Matlab: Spatial Toolbox`
- Impact
  - +1000 downloads from Rstudio since March
  - Presented at USER 2016 – Stanford University

Out-of-sample predictions

BLUP formula: Goulard et al., 2017

$$\hat{Y}_O^{BP} = \hat{Y}_O^{TC} - Q_{OO}^{-1} Q_{OS} \times \left(Y_S - \hat{Y}_S^{TC} \right)$$

$$Q = \frac{1}{\sigma^2}\left(I - \rho(W' + W) + \rho^2 W'W\right) = \begin{pmatrix} Q_{SS} & Q_{SO} \\ Q_{OS} & Q_{OO} \end{pmatrix}$$

$$\begin{pmatrix} \hat{Y}_S^{TC} \\ \hat{Y}_O^{TC} \end{pmatrix} = (I - \hat{\rho} W)^{-1} X\hat{\beta}$$

Marginal effects

LeSage and Pace, 2009

- **First problem**: interpretation of changes of X on Y is not direct: non-linear transformations due to the probit model
- **Second problem**: spatial spillover effects! Changes in a single observation influence all other observations.
- For each explanatory variable ➔ n × n matrix of partial derivatives.
- Scalar summary measures that average across the sample of observations:
  - Direct effects
  - Indirect effects
  - Total effects

Marginal effects

$$\frac{\partial \Pr(Y_i = 1)}{\partial X_{v,j}} = \Phi(\eta_i) S_{ij} \hat{\beta}_v$$

$$S = (I - \hat{\rho} W)^{-1}$$

$$\eta = S X \hat{\beta}$$

- Direct effects: mean of diagonal
- Total effects: mean of rows sum
- Indirect effects: total - direct