Spatial Econometrics & Non Parametric Econometrics

Urban Economics

Ignacio Sarmiento-Barbieri

Universidad de los Andes

February 11, 2022

Agenda

Spatial Lag Model

Spatial Autoregressive (SAR) Models

Let's consider the following model:

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

we assume that W is exogenous

If W is row standardized:

- Guarantees $|\lambda| < 1$ (Anselin, 1982)
- $[0, 1]$ Weights
- Wy Average of neighboring values
- W is no longer symmetric $\sum_j w_{ij} \neq \sum_i w_{ji}$ (complicates computation)

Spatial Lag Model

Estimation

- MLE
- 2SLS

Example: Boston House Data

- Packages

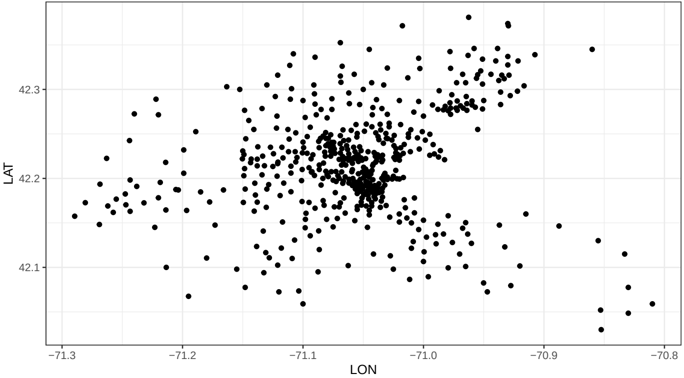
```r
require("spdep")
```

- Data

```r
data(boston)
str(boston.c)
```

```
## 'data.frame':    506 obs. of  20 variables:
##  $ TOWN    : Factor w/ 92 levels "Arlington","Ashland",..: 54 77 77 46 46 46 69 69 69 69 ...
##  $ TOWNNO  : int  0 1 1 2 2 2 3 3 3 3 ...
##  $ TRACT   : int  2011 2021 2022 2031 2032 2033 2041 2042 2043 2044 ...
##  $ LON     : num  -71 -71 -70.9 -70.9 -70.9 ...
##  $ LAT     : num  42.3 42.3 42.3 42.3 42.3 ...
##  $ MEDV    : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
##  $ CMEDV   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 22.1 16.5 18.9 ...
##  $ CRIM    : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
##  $ ZN      : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
##  $ INDUS   : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
##  $ CHAS    : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ NOX     : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...
```

Example: Boston House Data



Example: Boston House Data

```
ols<-lm(MEDV~CRIM+RM+INDUS+NOX+AGE+DIS+RAD+PTRATIO+B+LSTAT+TAX,data=boston.c)
stargazer::stargazer(ols,type="latex")
```

	Dependent variable:
	MEDV
CRIM	-0.103^{***}
	(0.033)
RM	4.074^{***}
	(0.421)
INDUS	0.018
	(0.062)
⋮	
Constant	37.308^{***}
	(5.200)
Observations	506
R^2	0.729
Adjusted R^2	0.723
Residual Std. Error	4.838 (df = 494)
F Statistic	121.004^{***} (df = 11; 494)
Note:	*p<0.1; **p<0.05; ***p<0.01

Example: Boston House Data

```r
W<-nb2listw(boston.soi, style="W", zero.policy = TRUE)
```

```
## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 506
## Number of nonzero links: 2152
## Percentage nonzero weights: 0.8405068
## Average number of links: 4.252964
##
## Weights style: W
## Weights constants summary:
##      n     nn  S0       S1       S2
## W 506 256036 506 261.3891 2071.255
```

Example: Boston House Data

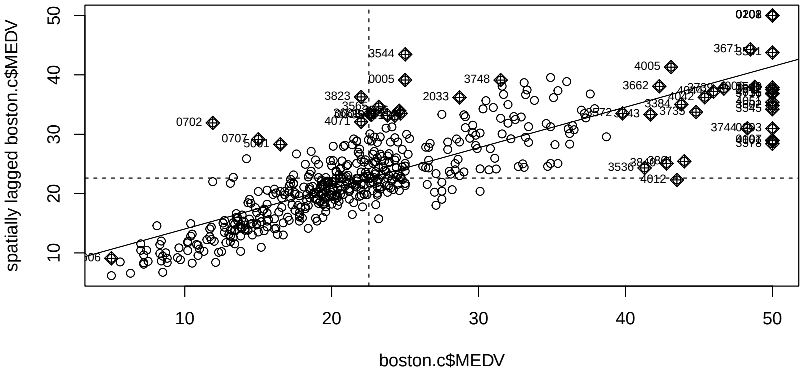
```
head(W$neighbours)
```

```
## [[1]]
## [1]  3 30 32 35
##
## [[2]]
## [1]  3  7 14 27 28 29 30
##
## [[3]]
## [1]  1  2  4  7 30
##
## [[4]]
## [1] 3 5 7
##
## [[5]]
## [1] 4 6 7
##
## [[6]]
## [1] 5
```

```
head(W$weights)
```

```
## [[1]]
## [1] 0.25 0.25 0.25 0.25
##
## [[2]]
## [1] 0.1428571 0.1428571 0.1428571 0.1428571 0.1428571 0.1428571 0.1428571
##
## [[3]]
## [1] 0.2 0.2 0.2 0.2 0.2
##
```

Example: Boston House Data



Example: Boston House Data

```
require(spatialreg)
fit.lag<-lagsarlm(MEDV~CRIM+RM+INDUS+NOX+AGE+DIS+RAD+PTRATIO+B+LSTAT+TAX,
data = boston.c, listw = W)
stargazer::stargazer(ols,fit.lag,type="latex")
```

	OLS	*spatial autoregressive*
	(1)	(2)
CRIM	-0.103^{***}	-0.056^{**}
	(0.033)	(0.028)
RM	4.074^{***}	3.817^{***}
	(0.421)	(0.356)
INDUS	0.018	0.033
	(0.062)	(0.051)
⋮		
Constant	37.308^{***}	9.949^{**}
	(5.200)	(4.625)
Observations	506	506
R^2	0.729	
Adjusted R^2	0.723	

Interpretation of Parameters

- Consider the following model for the $i - th$ observation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_r x_{ir} + \cdots + \beta_k x_{ik} \;\; i = 1, \ldots, n$$

- Recall that in OLS we have

$$\beta_1 = \frac{\partial y_i}{\partial x_{i1}}$$

- Interpretation is straight forward as long as we take into account units
- In spatial models the interpretation is less immediate and require some clarification

Interpretation of Parameters

- Lets consider the case of a Spatial Lag model with $|\lambda| < 1$

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

- Then

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

- we define

$$S(W) = (I - \lambda W)^{-1} \beta \tag{2}$$

Interpretation of Parameters

Therefore the impact of *each variable* x on y can be described through the partial derivatives $\frac{\partial E(y)}{\partial x}$ which can be arranged in the following matrix:

$$S(W) = \frac{\partial E(y)}{\partial x} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_1} & \cdots & \frac{\partial E(y_1)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_i)}{\partial x_1} & \cdots & \frac{\partial E(y_i)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_1} & \cdots & \frac{\partial E(y_n)}{\partial x_n} \end{pmatrix} \tag{3}$$

Interpretation of Parameters

On this basis, LeSage and Pace (2009) suggested the following impact measures that can be calculated for each independent variable X_i included in the model

- *Average Direct Impact*: this measure refers to the impact of changes in the $i-th$ observation of x, which we denote x_i, on y_i. This is the average of all diagonal entries in S

$$\begin{aligned} ADI &= \frac{tr(S(W))}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ii} \end{aligned} \tag{4}$$

Interpretation of Parameters

- *Average Total Impact To* an observation: this measure is related to the impact produced on one single observation y_i. For each observation this is calculated as the sum of the $i-th$ row of matrix S

$$ATIT_j = \frac{1}{n}\sum_{i=1}^{n} S(W)_{ij} \tag{5}$$

Interpretation of Parameters

- *Average Total Impact From* an observation: this measure is related to the total impact on all other observations y_i . For each observation this is calculated as the sum of the $j-th$ column of matrix S

$$ATIF_i = \frac{\sum_{j=1}^{n} S(W)_{ij}}{n} \tag{6}$$

Interpretation of Parameters

- A Global measure of the average impact obtained from the two previous measures.
- It is simply the average of all entries of matrix S

$$ATI = \frac{1}{n}\sum_{i=1}^{n} ATIT_i = \frac{1}{n}\sum_{j=1}^{n} ATIF_i \tag{7}$$

- The numerical values of the summary measures for the two forms of average total impacts are equal.
- The ATIF relates how changes in a single observation j influences all observations.
- In contrast, the ATIT considers how changes in all observations influence a single observation i.

Interpretation of Parameters

- *Average Indirect Impact* obtained as the difference between ATI and ADI

$$AII = ATI - ADI \tag{8}$$

- It is simply the average of all off-diagonal entries of matrix S

Introduction

- Two estimation methods can be applied to this great variety of problems: OLS and MLE
- The consistency of these estimators depends on the correct specification of the underlying parametric models
- The specification of parametric models involves making assumptions about:
  - Which variables are observable and which ones are unobservable
  - Which variables are endogenous and which ones are exogenous to the model
  - What is the set of functional relationships between these variables?
  - What distributions describe the variables

Parametric models and identification assumptions

- In experimental sciences, such as biology and physics, data arise from controlled experiments that are capable of isolating relationships between the variables of interest:
- y depends on X and ϵ, but in the experiment ϵ can be controlled.
- In economics and other social sciences, we use non-experimental data: Y is the result of both X and ϵ, and ϵ is neither observed nor controlled!
- So how do we estimate the effect of X on Y? We need assumptions.

Parametric models and identification assumptions

- We can adopt a parametric alternative: assume a structure or a model as the actual process that generates the data. For example: linear model in the parameters

$$Y_i = \alpha + \beta X_i + \epsilon_i \tag{9}$$

- With this first assumption everything we do not know about the relationship between Y and X is summarized in the parameters α and β
- **Identification problem:** can we make inference about α and β based in the observable variables Y and X?
- It depends on the assumptions we make. In the above model we need:
  - Random sample
  - $E[\epsilon|X] = 0$
  - No perfect multicollinearity

Parametric models and identification assumptions

"The goal [of identification analysis] is to learn what conclusions can and cannot be drawn from specific combinations of assumptions and data"

"We must face the fact that we cannot answer all the questions we ask ourselves"

"Empirical studies usually seek to know the value of some parameter that characterizes the population of interest. Conventional practice is to invoke assumptions strong enough to identify the exact value of that parameter. Even when these assumptions are not plausible, they are defended as necessary to be able to make an inference"

Manski, C. (1995), Identification problems in the social sciences

A more agnostic alternative: non-parametric models

- Non-parametric models don't assume any structure to define the relationship between the interest variables
- It uses methods that allow data to talk as much as possible
- Making as few assumptions as possible. At one extreme, the only acceptable assumptions are
    - The indispensables for the nonparametric model identification
- Some "reasonable" restrictions (for example, derivatives signs that are predicted by the economic theory)
- Having assumptions has its pros and cons. We'll see. . . .

Density Estimation

LAND VALUES IN CHICAGO

1857

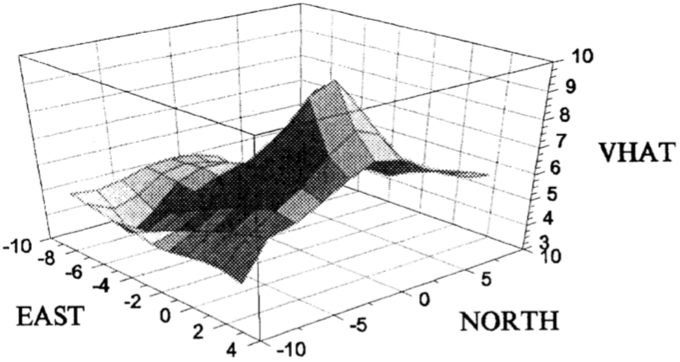


FIGURE 2

Univariate density estimation: parametric vs nonparametric methods

- Let $f = f(.)$ be the density function of the random variable X
- Let $x_1,\ x_2,\ \ldots x_n$ be a random sample of X. Then, $x_i \sim f\ iid$
- How can we estimate the density in a particular point x_0, then $f(x_0)$?

Parametric methods

Parametric methods

- They assume a particular functional form for f. :

Parametric methods

- They assume a particular functional form for f. :
- Example:

$$f(x_o) = \phi(x_o) = \frac{1}{\sigma\sqrt{2\,\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_o - \mu}{\sigma}\right)^2\right] \tag{10}$$

- The ignorance about $f(x_o)$ is limited to ignorance of the two parameters μ and σ

Parametric methods

- They assume a particular functional form for f. :
- Example:

$$f\left(x_o\right)=\phi\left(x_o\right)=\frac{1}{\sigma\sqrt{2\,\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_o-\mu}{\sigma}\right)^2\right] \tag{10}$$

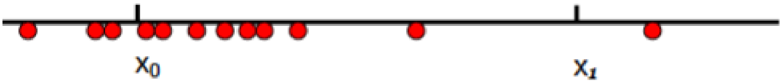
- The ignorance about $f\left(x_o\right)$ is limited to ignorance of the two parameters μ and σ
- Consistent estimators (ML) are:

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{11}$$

$$\hat{\sigma} = S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n} \tag{12}$$

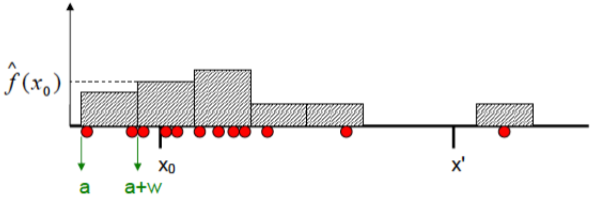
Non-parametric methods

- They seek to estimate $f(x_o))$ without assuming a particular functional form, only assuming certain regularity conditions of the density (smoothness, differentiability)
- How is the sample information interpreted?
- If we observe more data "near" x_o than x_1 we infer that $f(x_o) > f(x_1)$



A rudimentary non-parametric estimator: the histogram

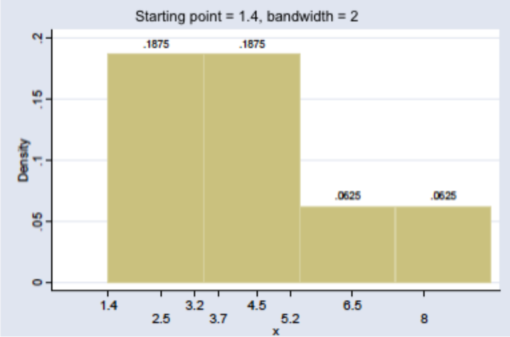
- It consists of estimating the probability within intervals through the relative frequency of observations within that interval.
- The intervals are determined from an initial point a and a bandwidth w



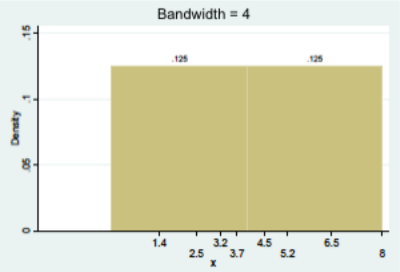
- The area of the bars is the relative frequency: $\frac{Nber\, obs.\, in\, interval}{n}$
- The height of the bars is an estimator of the density at any point in the interval

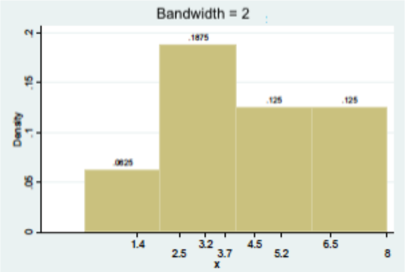
$$\hat{f}(x_o) = \frac{Nber.\, obs\ interval}{n \times w} \tag{13}$$

Histogram problems: (1) depends on starting point

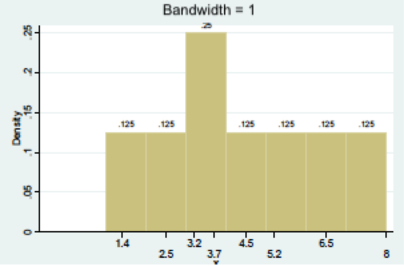


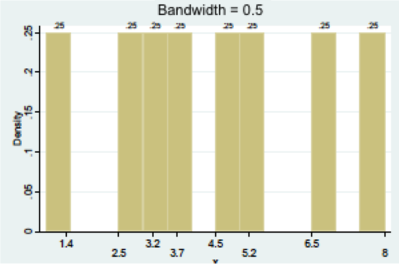
Histogram problems: (2) depends on the bandwidth





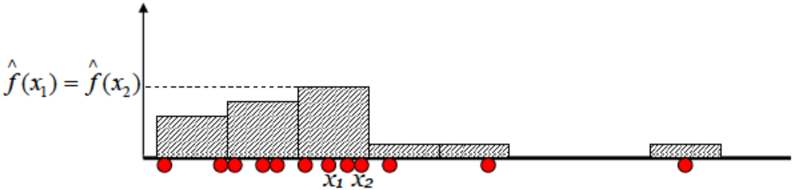
Histogram problems: (2) depends on the bandwidth





Histogram problems: (3) is discontinuous at the ends of the interval

- Note that $\hat{f}\ (x_1) = \hat{f}\ (x_2)$
- But $\hat{f}\ (x_2 + \epsilon) = \frac{1}{4}\hat{f}\ (x_2)$ for any $\epsilon > 0$



Another non-parametric estimator: the naive estimator

Discrete case

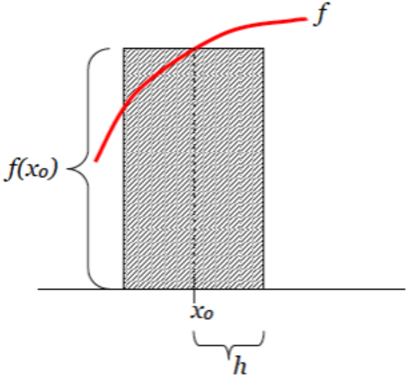
- X is a discrete VA, *iid*
- Objective: estimate $\Pr\left(X = x_o\right) = f\left(x_o\right)$
- The naive estimator comes from:

$$\hat{f}\left(x_o\right) = \frac{\#\, x_i = x_o}{n} = \frac{1}{n}\sum_{i=1}^{n} I\left(x_i = x_o\right) \tag{14}$$

Another non-parametric estimator: the naive estimator

Continuous case

- X is a continuous VA ($Pr\left(X = x0\right) = 0$) we evaluate the probability that X is "close" to x_0.
- We say that x_i is close to x_0 if x_i belongs to the interval $(x_o - \ h, \ x_o + \ h)$



Another non-parametric estimator: the naive estimator

- Formally:

$$2h \times f(x_0) \approx \int_{x_o-h}^{x_o+h} f(z)\,dz = \Pr\left[x_o - h < X < x_o + h\right] \tag{15}$$

$$\approx \frac{\#\, x_i \, \epsilon \, (x_0 - h, x_o + h)}{n} \tag{16}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I\left(x_i \, \epsilon \, (x_0 - h, x_o + h)\right) \tag{17}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I\left(x_0 - h < x_i < x_o + h\right) \tag{18}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I\left(-1 < \frac{x_i - x_o}{h} < 1\right) \tag{19}$$

(20)

Another non-parametric estimator: the naive estimator

- Then:

$$2h \ \times f\left(x_0\right) \rightarrow f(x_o) \cong \frac{1}{2h \times n} \sum_{i=1}^{n} I \left(-1 < \frac{x_i - x_o}{h} < 1 \right) \tag{21}$$

- $\hat{f}(x_o) \equiv \frac{1}{h \times n} \sum_{i=1}^{n} \frac{1}{2} I \ \left(-1 < \frac{x_i - x_o}{h} < 1\right)$ is an estimator of the height of the rectangle in the previous graph

Further Readings

- Arbia, G. (2014). A primer for spatial econometrics with applications in R. Palgrave Macmillan. (Chapter 2 and 3)
- Anselin, Luc, & Anil K Bera. 1998. “Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics.” Statistics Textbooks and Monographs 155. MARCEL DEKKER AG: 237–90.
- Anselin, L. (1982). A note on small sample properties of estimators in a first-order spatial autoregressive model. Environment and Planning A, 14(8), 1023-1030.
- Hsiao, C. (1983), “Identification”, Handbook of Econometrics, Vol.1, Ch. 4. Li, Q. and J. S. Racine (2007). Nonparametric Econometrics. Princeton University Press.
- LeSage, J. (1999). The Theory and Practice of Spatial Econometrics. Mimeo
- LeSage, J. & Pace, R. K. (2009). Introduction to spatial econometrics. Chapman and Hall/CRC.
- Mansky (1995), “Identification problems in the social sciences”, Harvard University Press.
- Matzkin, R. (2007), Nonparametric Identification, Handbook of Econometrics, Vol. 6B, Ch 73.