Economics Letters 163 (2018) 59–64

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Nonlinear impact estimation in spatial autoregressive models

Jean-Sauveur Ay, Kassoum Ayouba *, Julie Le Gallo

CESAER, AgroSup Dijon, INRA, Université de Bourgogne Franche-Comté, 21000 Dijon, France

HIGHLIGHTS

- We extend the literature on the calculation/interpretation of impacts for SAR models.
- We compute the individual impacts for an exogenous variable introduced nonlinearly.
- Averaging these impacts smooths spatial interaction effects which may be of interest.
- We provide a graphical analysis of these impacts, with their confidence intervals.

ARTICLE INFO

Article history:
Received 17 October 2017
Received in revised form 20 November 2017
Accepted 22 November 2017
Available online 5 December 2017

JEL classification:
C11
C21

Keywords:
Spatial econometrics
Marginal impacts
Spline
Markov chain Monte Carlo

ABSTRACT

This paper extends the literature on the calculation and interpretation of impacts for spatial autoregressive models. Using a Bayesian framework, we show how the individual direct and indirect impacts associated with an exogenous variable introduced in a nonlinear way in such models can be computed, theoretically and empirically. Rather than averaging the individual impacts, we suggest to graphically analyze them along with their confidence intervals calculated from Markov chain Monte Carlo (MCMC). We also explicitly derive the form of the gap between individual impacts in the spatial autoregressive model and the corresponding model without a spatial lag and show, in our application on the Boston dataset, that it is higher for spatially highly connected observations.

1. Introduction

Spatial autoregressive (SAR) models are now widely used to analyze spatial economic interactions in various applications. For example, such a model is appropriate for the housing market since housing prices depend on prices of recently sold neighboring homes (Anselin and Lozano-Gracia, 2008). This dependence structure comes from the fact that sellers presumably use information on neighboring homes to determine the asking price. The use of a SAR model, besides providing a richer characterization of the market, has important implications for impacts calculations (for early statements of this issue, see Anselin and Le Gallo, 2006; Kim et al., 2003; Kelejian et al., 2006). Indeed, their computation and interpretation is less straightforward than in standard multiple linear a-spatial regression models: any change in an explanatory variable for a given observation not only affects the observation itself (direct impact) but also affects all other observations indirectly (indirect impact).

LeSage and Pace (2009) show how to theoretically derive these marginal impacts in SAR models. For n spatial observations, they obtain a $n \times n$ matrix of impacts for one exogenous variable. In order to have a compact representation of these impacts, they propose to report one direct impact equal to the average of the diagonal elements of the matrix of marginal impacts and one indirect impact equals to the average row sums of the non-diagonal elements of that matrix. However, when the exogenous variable of interest is introduced in a nonlinear way in the SAR model (e.g. in the form of polynomial or splines function), averaging the impacts in such a way is irrelevant.

In this paper, we extend the work of LeSage and Pace (2009) on impacts computation and estimation when the exogenous variable of interest appears in a nonlinear way in a SAR model. We also derive the form of the gap between impacts in the spatial autoregressive model and the corresponding model without a spatial lag and show in our application that it is higher for spatially highly connected observations.

* Corresponding author.
E-mail addresses: jean-sauveur.ay@inra.fr (J.-S. Ay), ayouba.kassoum@inra.fr (K. Ayouba), julie.le-gallo@agrosupdijon.fr (J. Le Gallo).

https://doi.org/10.1016/j.econlet.2017.11.031
0165-1765/

The remainder of the paper is organized as follows. The second section presents the theoretical derivation of direct and indirect impacts associated with an exogenous variable introduced in a nonlinear way in a SAR model. The third section introduces the estimation strategy which is applied on the well-known Boston housing dataset described in the fourth Section. The obtained empirical results are presented in the fifth Section. The sixth section concludes and comments on possible extensions.

2. Impacts in theory

Consider the following spatial autoregressive model[1]

$$y = \rho W y + X\beta + f(z) + \epsilon \tag{1}$$

where y is the $(n \times 1)$ dependent variable that exhibits variation across spatial observations, X is the $(n \times k)$ matrix of linear explanatory variables including an intercept term, with the associated parameters β contained in a $(k \times 1)$ vector, z is the $(n \times 1)$ variable the impact of which on y is nonlinear, and W is a specified constant $(n \times n)$ spatial weight matrix with the usual assumptions. z, our variable of interest is additively separable from the other X to simplify notations. Our approach can easily be extended to many variables introduced non-linearly with interaction effects between them. We assume that each term of the disturbance vector ϵ of dimension $(n \times 1)$ is normally and identically distributed with zero mean and variance σ^2. The scalar ρ measures the strength of the spatial dependence. $f(\cdot)$ is a linear-in-parameters function, for instance a polynomial function of degree p: $f(z) = \sum_{j=1}^{p} \gamma_j z^j$, a spline[2] function of order p and q knots $t_l, t_l, \ldots, t_q$: $f(z) = \sum_{j=1}^{p} \gamma_j z^j + \sum_{l=1}^{q} \delta_l (z - t_l)_+^p$, a B-spline function of order p and q knots: $f(z) = \sum_{j=1}^{p+q} \gamma_j B_j(z)$.[3]

Eq. (1) can be rewritten in the following reduced form

$$y = V(W)X\beta + V(W)f(z) + V(W)\epsilon \tag{2}$$

with

$$V(W) = (I_n - \rho W)^{-1} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1n} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2n} \\ V_{31} & V_{32} & V_{33} & \cdots & V_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{n1} & V_{n2} & V_{n3} & \cdots & V_{nn} \end{bmatrix}$$

and I_n the identity matrix of order n. Note that we assumed that the matrix $I_n - \rho W$ is not singular to reach the reduced form in Eq. (2).

In this paper, we are interested in the estimation of the partial derivative of y with respect to changes in the nonlinear variable z in our SAR model. In models containing a spatial lag, the measure of the partial derivative of the dependent variable with respect to an explanatory variable is less straightforward than in standard linear models. Indeed the standard linear regression interpretation of coefficient estimates ($\widehat{\beta}_q = \frac{\partial \hat{y}}{\partial x_q}$) as partial derivatives no longer holds in SAR model since the matrix of explanatory variable is transformed by the $n \times n$ inverse matrix $V(W)$. In such a model, any change to an explanatory variable for a given observation affects the dependent variable of the observation itself (direct impact) and potentially the dependent variable of all other observations (indirect impact) through $V(W)$. We elaborate on this observation to derive the impacts for the variable z introduced in a nonlinear way.

Starting from the reduced form in Eq. (2), the matrix of responses to a change of the nonlinear variable z on y is given by

$$\begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} & \cdots & \frac{\partial y_2}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \frac{\partial y_n}{\partial z_2} & \cdots & \frac{\partial y_n}{\partial z_n} \end{bmatrix} = \begin{bmatrix} V_{11} f_z(z_1) & V_{12} f_z(z_2) & \cdots & V_{1n} f_z(z_n) \\ V_{21} f_z(z_1) & V_{22} f_z(z_2) & \cdots & V_{2n} f_z(z_n) \\ \vdots & \vdots & \ddots & \vdots \\ V_{n1} f_z(z_1) & V_{n2} f_z(z_2) & \cdots & V_{nn} f_z(z_n) \end{bmatrix}$$

with $f_z(z)$ the derivative of $f(z)$.

Since we have an $n \times n$ matrix of impacts, the challenge is to find a way to compactly report them. In the case of an exogenous variable that enters linearly, LeSage and Pace (2009) suggest to compute the average of the main diagonal elements and the average of the off-diagonal elements of the impacts matrix to obtain a summary measure of the direct impact and the indirect impact respectively. This method of summarizing impacts is irrelevant when the effect of a variable as z can be positive and negative at different parts of it support. Therefore, we propose to plot the individual impacts (i.e. impacts of each observation) along with their confidence intervals. The individual direct (IDI) and total (ITI) impacts of the nonlinear variable z on y_i are given by[4]

$$IDI_i = \frac{\partial y_i}{\partial z_i} = V_{ii} f_z(z_i) \tag{3}$$

$$ITI_i = \frac{\partial y_i}{\partial z} = \sum_{j=1}^{n} V_{ij} f_z(z_j). \tag{4}$$

From these expressions, we can reach the individual indirect impact (III) as follows

$$III_i = ITI_i - IDI_i. \tag{5}$$

3. Estimation strategy

To get the individual impacts for z, we proceed in two steps. We first estimate the parameters of the SAR model, and then in a second step use them to compute $V(W)$ and the derivative of $f(z)$, i.e. $f_z(z)$ for each observation i. Note that, to estimate $V(W) = (I_n - \rho W)^{-1}$, we have followed an approach suggested by LeSage and Pace (2009, page 111), based on the Leontief expansion of $V(W)$, which is computationally more efficient than the brute force approach (i.e. calculating $V(W)$ *per se*) for large n: $V(W) = (I_n - \rho W)^{-1} \approx I_n + \rho W + \rho^2 W^2 + \cdots + \rho^q W^q$. We have chosen $q = 250$ to approximate $V(W)$ in our application.[5]

[1] Our approach can also be applied on spatial Durbin model.

[2] See Hastie and Tibshirani (1990) for details.

[3] Note that the B-spline function (B stands for "basis") is an extension of the spline function. The $B_j, j = 1, \ldots, p + q$ are defined recursively to avoid numerical instability.

[4] Note that our approach is also appropriate to the case of an interacted explanatory variable, e.g., $y = \rho W y + z\beta + xz\gamma + \epsilon$, where the individual direct and total impact would be respectively $\frac{\partial y_i}{\partial z_i} = V_{ii}(\beta + x_i\gamma)$ and $\frac{\partial y_i}{\partial z} = \sum_{j=1}^{n} V_{ij}(\beta + x_j\gamma)$. This logically means that the values of the variable x at each observation produce observation-level impact estimates of interest. To assess the role played by x in determining observation-level effects, we could plot these impacts estimates for sorted values of x. We thank the reviewer for pointing out this.

[5] The trace-based approach proposed by LeSage and Pace (2009, pp. 114–115) to approximate the direct and indirect impacts (defined respectively as the average of the diagonal elements of the matrix of marginal impacts and the average row sums of the non-diagonal elements of that matrix) does not apply to our case since we are not interested in aggregated measure of impacts.