Identifying Nonlinearities in Spatial Autoregressive Models*

Nicolas Debarsy† and Vincenzo Verardi‡

June 2009-First Draft

Abstract

In spatial autoregressive models, the functional form of autocorrelation is assumed to be linear. In this paper, we propose a simple semiparametric procedure, based on the Yatchew's (1997) semiparametric partial linear model, that does not impose this restriction. Simple simulations show that this model outperforms traditional SAR estimation when nonlinearities are present. We apply this methodology to an empirical application aimed at understanding the spatial pattern of voting for independent candidates in the US presidential elections. We find that in some counties, votes for "third canidates" are non-linearly related to votes for "third candidates" in neighboring counties, which pleads in favor of strategic behavior.

*We would like to thank all our colleagues at ECARES, CRED and CERPE for useful comments.

†CERPE, Facultés Universitaires Notre Dame de la Paix de Namur. E-mail: ndebarsy@fundp.ac.be. Nicolas Debarsy is Doctoral Researcher of the FNRS and gratefully acknowledges their financial support.

‡CRED, Facultés Universitaires Notre Dame de la Paix de Namur; ECARES, CKE, Université Libre de Bruxelles. E-mail: vverardi@fundp.ac.be. Vincenzo Verardi is Associated Researcher of the FNRS and gratefully acknowledges their financial support.

Keywords: Spatial econometrics, semiparametric estimations

JEL Classification:C14, C21

1 Introduction

Spatial econometrics is the branch of econometrics that addresses the issues introduced by spatial correlation in regression analysis and hypothesis testing. Broadly speaking, these issues are mainly of two types: spatial autocorrelation and spatial heterogeneity. The models that have been proposed to deal with the first issue in the econometrics literature are either of the spatial autoregressive type (if the spatial lag of the dependent variable is among the explanatory covariates) or of the spatial error type (if the spatial dependence only takes place in the error term). The models that have been proposed to deal with the second issue are geographically weighted regressions. In this paper we focus essentially on spatial lag dependence models (SAR) as our objective is to understand how a non-linear relation between a dependent variable and its spatial lag can be estimated. We refer the reader to Anselin (1988a), Anselin et al. (2004) and Lesage and Pace (2009) for further details on spatial regression models.

The basic idea of SAR modelling is that the dependent variable (y) in a regression equation is related to values of the variable measured in the neighborhood. Unlike in time series, when one deals with spatial data no natural way of ordering exists. One thus has to model an interaction scheme between observations. This is achieved through the use of a (spatial weights) matrix W (of dimension $N \times N$, where N is the number of observations). The spatially lagged dependent variable, Wy, which provides a measure of the neighborhood value of the dependent variable, is then obtained by multiplying the interaction matrix W by the vector of y. SAR models have to be estimated by Maximum Likelihood or Generalized Method of Moment estimators (GMM) since LS estimators are biased and inconsistent due to the endogeneity of Wy .[1]

[1]The interested reader may consult Anselin (1988a) for further details.

The main scope of this paper is to present a simple econometric procedure that extends spatial autoregressive models to the case of non-linear autoregressive components. We indeed believe that the linearity assumption, standard in SAR modeling, is often too restrictive. To relax this assumption, we propose a very simple procedure. We first suggest to call on Yatchew's (1997) difference estimator to partial out the spatial autoregressive term. Having removed the spatial component, the parameters associated to the non-spatial explanatory variables can be consistently estimated using LS. Relying on these consistent estimates, it is then possible to assess the non-linear relation between y and Wy by simply running a non-parametric estimation of the fitted residuals of the differentiated regression on Wy. We furthermore propose a simple statistic to test for this supposed non-linearity.

To illustrate the usefulness and good performance of the procedure, we present some simulations considering four types of spatial autocorrelation: absent, linear, quadratic and sigmoidal.

Furthermore, a concrete research question is presented. We study whether US voters consider the votes of electors of other counties when choosing the candidate for which they vote. We then check if this relationship between counties is linear. The underlying idea is that electors can decide either to vote sincerely for their preferred candidate or vote strategically if they believe that voting sincerely might split the majority and increase the probability of success of the candidate they dislike the most. To test for this, we study whether votes for independent candidates in US counties are affected by the results of independent candidates in neighboring counties. Our results tend to confirm this "spillover effect" in at least some counties. Moreover, this externality is found to be non-linear.

The structure of the paper is the following: after this short introduction we propose, in section 2, a procedure that allows estimating a semiparametric spatial autoregressive model and a test that allows understanding if assuming linearity in the autoregressive component is appropriate. In section 3 we present some modest simulations to check for the quality of fitting of the procedure and the performance of the test. In section 4 we present an empirical example on US presidential elections and we conclude in section 5.

2 Estimation method

2.1 A generalized spatial autoregressive model

The general form of a first order spatial autoregressive model is given by:

$$y_i = x_i\theta + \rho W y_i + \varepsilon_i \text{ for } i = 1, ..., N \tag{1}$$

where y_i is the value taken by the dependent variable and x_i is the row vector of characteristics of individual i. Wy_i is a measure (usually the weighted average) of y in the neighborhood $N_{(i)}$ of individual i. It is defined as $Wy_i = \sum_{j \in N_{(i)}} w_{ij} y_j$ where w_{ij} models spatial interactions between i and j. Column vector θ and the autoregressive spatial parameter ρ are the coefficients to be estimated. In equation (1) it is easy to see that this model cannot be estimated using ordinary least squares (LS) as Wy and ε are not independent. The model is then generally fitted either using two stages least squares (using the weighted average value of the x variables measured in the neighborhood of i as instruments) or, more commonly, by calling on maximum likelihood estimations. An obvious drawback of this model is that the spatial component is assumed to be linear. Hence, a unit change in Wy would imply a ρ units change in the conditional expectation of y, whatever the value of Wy. Looking at the problem from a different perspective, this would also mean that if Wy increases by 10 units, this would increase the conditional expectation of y by 10ρ units. This is probably a too restrictive assumption. It is indeed rather unlikely that an individual reacts proportionally equally to small and large changes in Wy. This problem could of course be partially tackled by adopting some transformations of the y variable (taking the log, for instance, would lead to elasticities rather than unit changes). This solution is however very limited as it would constrains the relation to be of a specific parametric form. The approach we suggest allows for a much higher flexibility by imposing only very limited assumptions on the relationship between the autoregressive component and the conditional expectation of y.

The generalized spatial autoregressive model we propose can be written as

$$y_i = x_i\theta + f(Wy_i) + \varepsilon_i \text{ for } i = 1, ..., N \tag{2}$$

As we explained in the introduction,, we believe that (2) can be easily estimated using a simple semiparametric regression model that relies on a very restricted number of assumptions.

A first assumption that we make is that Wy is drawn from a distribution with finite support (and that Wy is measured without error). Second, while the relation between y and Wy is supposed to be non-linear and of unknown form, we assume that the first derivative of f is bounded by a constant L. Finally, we assume the error ε_i are independent and identically distributed ($i.i.d$) with mean 0 and variance σ_ε^2.

Under these assumptions, and following Yatchew (1997), equation (2) can be easily fitted using a differencing semiparametric model.

2.2 Semiparametric estimation

Suppose that we rearrange the observations by sorting them in increasing order according to variable Wy (i.e. $Wy_1 \leq Wy_2 \leq ... \leq Wy_N$). By first differencing, we get:

$$y_i - y_{i-1} = (x_i - x_{i-1})\theta_{diff} + [f(Wy_i) - f(Wy_{i-1})] + (\varepsilon_i - \varepsilon_{i-1}) \text{ for } i = 2, ..., N \tag{3}$$

Increasing the number of observations (which, broadly speaking, means filling the finite support interval of Wy with new values) will cause the difference $Wy_i - Wy_{i-1}$ to shrink at a rate of about $1/N$. Since the first derivative of f is assumed bounded, we have that $|f(Wy_i) - f(Wy_{i-1})| \leq L\,|Wy_i - Wy_{i-1}|$. The shrinkage of $(Wy_i - Wy_{i-1})$ will hence induce $f(Wy_{i-1})$ to cancel out with $f(Wy_i)$. This means that reordering and differencing allows to estimate the θ parameter consistently, whatever the functional form of f, as soon as $\partial f/\partial x$ is bounded. This also means that parameter θ can be estimated consistently by a simple linear model given that the spatial component has been partialled out.

Note that this simple estimator is rather inefficient (it has a Gaussian efficiency of only 66.7%). To increase efficiency, Yatchew (1997) suggests to use higher order differences and consider a generalization of (3) which can be written as

$$\sum_{j=0}^{D_m} d_j y_{t-j} = \left(\sum_{j=0}^{D_m} d_j x_{i-j}\right)\theta_{diff} + \sum_{j=0}^{D_m} d_j f\left(Wy_{i-j}\right) + \sum_{j=0}^{D_m} d_j \varepsilon_{i-j} \text{ for } i = D_m + 1, ..., N \quad (4)$$

where D_m ($\in \mathbb{N}_0^+$) is the order of differencing and $d_0, ..., d_m$ are differencing weights. Two conditions are imposed on $d_0, ..., d_m$. The first, that is $\sum_{j=0}^{m} d_j = 0$, ensures that the nonparametric spatial component part is partialled out and the second, that is $\sum_{j=0}^{m} d_j^2 = 1$, guarantees that the variance of the residual in (4) is σ_ε^2. Yatchew (1998) shows that, with D_m sufficiently large, the estimator approaches asymptotic efficiency. Naturally, when $D_m = 1$, $d_0 = \frac{1}{\sqrt{2}}$ and $d_1 = -\frac{1}{\sqrt{2}}$, equation (4) boils down to equation (3).

As far as inference is concerned, Yatchew (1998) shows that $\hat{\theta}_{diff}$ has the approximate sampling distribution

$$\hat{\theta}_{diff} \sim N\left(\theta, \left(1 + \frac{1}{2D_m}\right)\frac{\sigma_\varepsilon^2}{N\sigma_u^2}\right) \quad (5)$$

where σ_u^2 is the conditional variance of x given Wy. We can thus compute the standard errors of the estimated parameters in the differenced equation. For the inference associated to the variable Wy , Yatchew (1998) developed a simple test based on the comparison of the scale of the residuals of the differenced equation (s_{diff}^2) with that of the LS regression where the function f is supposed to be linear (s^2_{lin}). The underlying idea of the test is that if nonlinearity exist, a linear approximation of the tested relation will lead to an overestimation the variance of the residuals with respect to a model that explicitly models it.

The test statistic he proposes is:

$$V = \frac{\sqrt{D_m N}(s^2_{lin} - s^2_{diff})}{s^2_{diff}} \tag{6}$$

which is asymptotically distributed as a $N(0,1)$. A rejection of the null would bring evidence in favour of a non-linear relation between Wy and y.

In our setup, the estimation of s^2_{lin} cannot be based on LS as it is biased and inconsistent if an autoregressive term is present. We therefore estimate it using the residuals of a SAR model estimated by maximum likelihood (see Anselin, 1988a; Lee, 2004).

As θ_{diff} is a consistent estimator of θ, the relation between y and Wy can be assessed by running a non-parametric estimation of the fitted residuals $\tilde{\varepsilon}_i = y_i - x_i\hat{\theta}_{diff}$ on Wy. These residuals do indeed still contain the information on the spatial dependence of interest. but are cleaned of the influence of the control variables. The non-parametric estimator used here is the one developed by Nadaraya (1964) and Watson (1964) which estimates the conditional mean as a locally weighted average of y, using a kernel as a weighting function. A higher weight is awarded to points near the one where the response is being estimated and less weight to points further away. The Kernel used is Gaussian and the chosen bandwidth corresponds to the width that would minimize the mean integrated squared error if the data were Gaussian. We chose to use this non-parametric estimator rather than the well-known locally weighted polynomial regression (which is often routinely fitted after Yatchew's difference estimator) as the latter relies on local polynomial regressions which is biased and inconsistent in case of existence of an autoregressive spatial component.

To assess the performances of the estimator we propose, we present in the following section some simple simulations.

3 Simulations

The objective of this section is to present some modest simulations to, first, show how the semi-parametric model outperforms the SAR if non-linear spatial autocorrelation is present and, second, to illustrate the performance of the (modified) V-statistic in cases of absence of spatial autocorrelation, simple linear spatial autocorrelation and non-linear spatial autocorrelation.

Four data generating processes (DGP) are considered:

a) $y_i = x_i\theta + \varepsilon_i$

b) $y_i = 0.75Wy_i + x_i\theta + \varepsilon_i$

c) $y_i = 0.75Wy_i - 0.4\left(Wy_i\right)^2 + x_i\theta + \varepsilon_i$

d) $y_i = \left(\frac{1}{1+\exp(-2Wy_i)} - 0.5\right) + x_i\theta + \varepsilon_i$

where x_i is a 1×3 vector whose elements are drawn from three independent $N(0,4)$, θ is a 3×1 vector of ones and $\varepsilon_i \sim N(0, 0.1)$. The simulated sample size is 300. Finally, we randomly generate x-coordinates from a $U[0,20]$ and y-coordinates from a $U[0,50]$. We then calculate all pairwise distances b_{ij} and generate weights as follows:

$$w_{ij} = \begin{cases} \frac{1/b_{ij}}{\sum\limits_j 1/b_{ij}} \text{ if } b_{ij} < \bar{b} \\ 0 \text{ otherwise} \end{cases}$$

where $\bar{b}$ (the threshold value above which the interaction between i and j is assumed negligible) is set to 5. By convention, $w_{ii} = 0$. All the models are fitted using a first order of differencing. Increasing the order would increase the Gaussian efficiency of the estimations.

3.1 Autocorrelation fitting

To illustrate the performance of the proposed estimation procedure, we show, in Figure 1, four graphs where we compare the non-parametric estimation (thick plain line) of the autocorrelation with the true DGP (thin dashed line). As expected, the results are unambiguous.

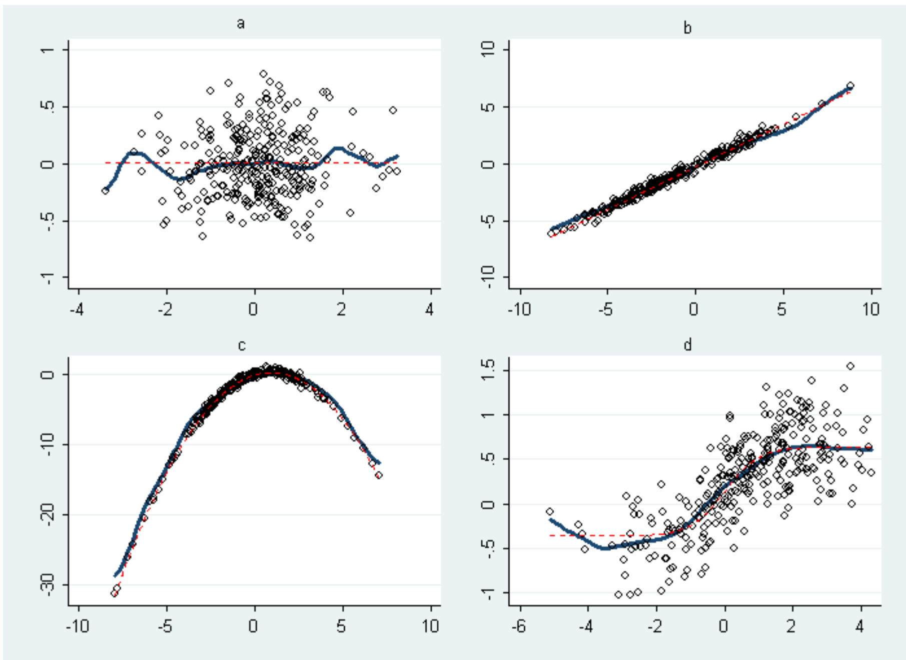

Figure 1: Non-parametric fit of spatial autocorrelation

In the case of no spatial autocorrelation (panel a), the non-parametric fit fails to find any clear relation and the non-parametric curve lies close to the horizontal line (which is the true DGP). In the three other cases (panels b, c and d), the nonparametric estimation of the autocorrelation matches well the true functional form. From Figure 1 it is also evident from the last two panels (c and d) that a linear approximation would not be satisfactory and would lead to questionable conclusions. This is especially true for the quadratic form.

3.2 Performance of the linearity test and of the estimator of the difference equation

To check for the performance of the V-statistic in detecting an eventual non-linearity, we replicate the four DGPs described above 1000 times. Each time a new error term is generated while the original design space (i.e. the space of the explanatory variables) is kept unchanged. We then compute the percentage of rejection of the null (considering a type I error of 5%) and present the results in Table 1. Again, we use a first-order differenced model to compute s^2_{diff}

Table 1: Size of test for linearity

Spatial autocorrelation	absent	linear	quadratic	sigmoidal
% Rejection	5.2%	4.8%	100%	99.2%

As expected, we observe a percentage of rejection of the null close to one for the quadratic and the sigmoidal spatial autoregressive cases. Furthermore, the size of the test is about 5% in case of absent spatial autocorrelation or linear spatial autocorrelation cases. Though we are conscious that these simulations are far too simple to provide a clear assessment of the quality and power of the test, it seems that it could be an interesting complementary tool to assess the shape of an eventual spatial autocorrelation. Note however that the test does not differentiate between absent and linear spatial autocorrelation. Hence, when the null is not rejected, we suggest to use the standard tests to check whether a linear spatial autocorrelation is present (See Anselin, 1988b; Anselin, 2001; Anselin and Florax, 1995; Anselin et al, 1996 for a review of the tests proposed in the literature).

To study the consistency of the "non-spatial" estimators of the difference equation (3), we look at the distribution of the θ_{diff}s and estimate their bias and mean squared error (MSE). The results are presented in Table 2. The bias of an estimator $\hat{\theta}$ is defined as $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$ and the mean squared error is defined as $MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$. To assess the performances of $\hat{\theta}_{diff}$, we thus estimate the simulated counterpart of the bias which is $\widetilde{Bias}(\hat{\theta}_{diff}) = \bar{\theta}_{diff,s} - \theta_0$ and the simulated counterpart of the MSE which is $\widetilde{MSE} = \frac{1}{1000}\sum_{s=1}^{1000}\left(\theta_{diff,s} - \bar{\theta}_{diff,s}\right)^2 + [\widetilde{Bias}(\hat{\theta}_{diff})]^2$ (where s indexes the simulations, $\bar{\theta}_s$ is the average value of the estimated parameter over the simulations $\bar{\theta}_{diff,s} = \frac{1}{1000}\sum_{s=1}^{1000}\hat{\theta}_{diff,s}$ and θ_0 is the value of the parameter set in the simulated DGPs.

Table 2: Bias and MSE of estimated parameters in the difference equation

Spatial autocorrelation	absent	linear	quadratic	sigmoidal
Bias ($\hat{\theta}_1$)	-0.0010	-0.0015	-0.0201	-0.0001
MSE ($\hat{\theta}_1$)	0.0001	0.0001	0.0006	0.0001
Bias ($\hat{\theta}_2$)	-0.0003	-0.0003	-0.0106	-0.0006
MSE ($\hat{\theta}_2$)	0.0001	0.0001	0.0006	0.0001
Bias ($\hat{\theta}_3$)	-0.0002	-0.0009	-0.0099	-0.0007
MSE ($\hat{\theta}_3$)	0.0001	0.0001	0.0005	0.0001

Table 2 shows that the bias and MSE of the coefficients estimated once the spatial lag dependence has been partialled out are very small.

4 Application

In this section, we present a simple application to illustrate the usefulness of the procedure we propose in a concrete applied research example. More precisely, we try to understand how voters in a given county decide whether or not to vote for independent candidates in US presidential elections (in year 2000) conditional on their expectations of votes casted by the third candidate in neighboring counties. It is important to state that the election of year 2000 is particularly interesting as the third candidate, Ralph Nader, is often claimed to have deprived the democratic majority of its victory.

The hypothesis we test is that voters do not vote for the third candidate if they believe that this could help the candidate they dislike the most to win the elections. More precisely, in a (two-big-party) majority system like the one for US presidential ballot, electors often vote for a third candidate to reveal a disapproval of the political establishment. However, by doing so, they know that if the third candidate casts too many votes, their vote, by being taken away from the candidate they dislike the less, might increase the chances of the candidate they dislike the most to win the elections.

For example, this is what happened in the french presidential elections of 2002 when Le Pen, the extreme right wing candidate, obtained 16.86 percent of the votes in the first round of voting. Voters indeed voted massively for alternative candidates to shout their dissatisfaction with the two big traditional parties[2] and especially with the Socialist candidate and incumbent prime minister Lionel Jospin. However, the objection vote was so massive that it made the score of Le Pen sufficient to qualify him for the second round. Afterward, in the second round, voters decided to support the remaining candidate that they disliked the less, Jacques Chirac, who finally obtained the largest score ever in french presidential elections, winning over 82% of the vote. This fragility of electoral systems when a divided majority is facing a unified minority has been extensively discussed in the literature (see e.g. Myerson and Weber, 1993; Cox, 1997 and Myerson, 2002). We refer the reader to this literature for further details.

Hence, if voters anticipate that the third candidate is collecting too many votes and could lead to their least preferred election result, they will decide to vote strategically and give their preference to a candidate that is not their sincere preference but who can stop the disliked opponent. The assumption behind this mechanism is that voters at the county level have reasonably good anticipations of the vote share each candidate will cast (in aggregate) in their State. Voters should not pay much attention to vote shares in counties outside their State since, given the peculiarities of the majority system for the US presidential elections, it is the state-aggregate votes and not the country-wide votes that matters. Relying on this, we test the possible strategic behavior of voters by estimating the following relation

$$y_i = x_i\theta + f(Wy_i) + \varepsilon_i \text{ for } i = 1, ..., N \tag{7}$$

where the variable y is the percentage of votes casted by outsider candidates and the control variables (x) are those generally considered in this type of regressions. That is to say (i) the percentage of votes casted by independents in the previous elections in

[2] 11 of the 15 running candidates obtained more than 3% of total votes each.

county i, (ii) the percentage of votes for the republican party in county i, (iii) the log of total population within the county, (iv) the per capita income within county i and (v) the ethnic composition of the county (i.e. the log of the number of blacks, whites, asians, pacific islanders and native indians and others).[3] The order of differencing used in the partial linear part of the model is 10 which allows to have a Gaussian efficiency very close to unity.

The weighting matrix considered is defined as follows: counties located in different states are assumed not to interact while within states, we assign a spatial weight proportional to the inverse of the distance between counties centroids.

For the sake of clarity, we concentrate in four states and results are reported in Figure 2. The specific choice of these counties is driven by, first, the number of counties existing in these States and, most important, by the evident non-linearity in the spatial autoregressive component that emerges from the analysis.[4] The four states we consider are North Carolina, New York, Michigan and Louisiana.

[3]Including votes for the democrats instead does not affect results.

[4]The graphs for all the other states are available from the authors upon request.

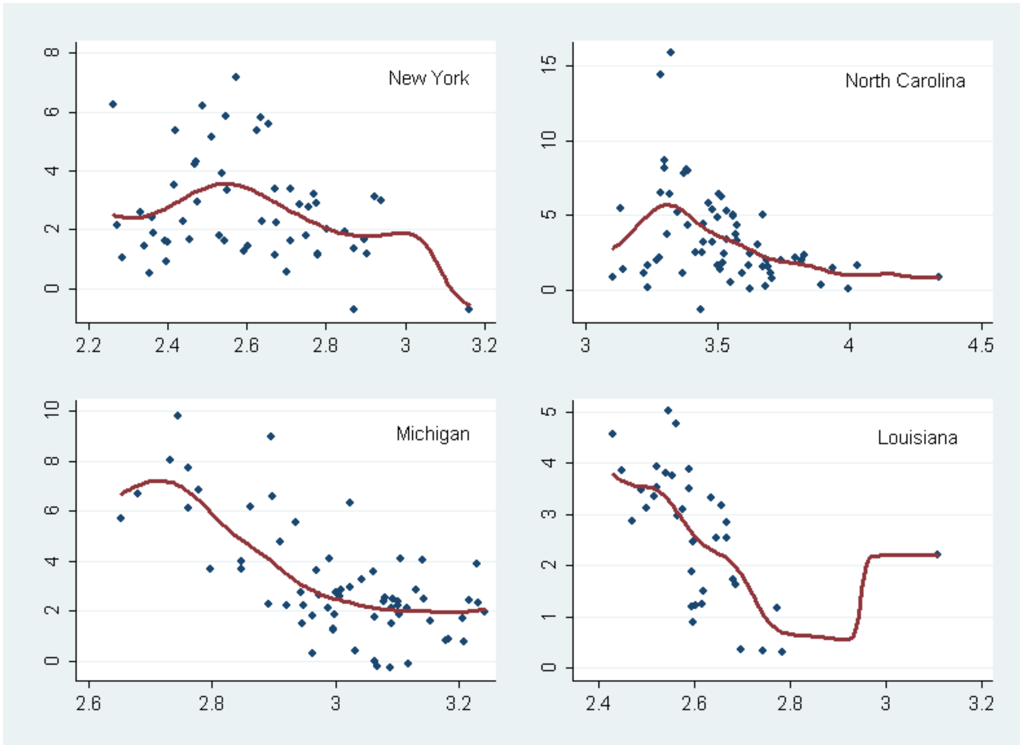

Figure 2: Nonlinear SAR by county

For the State of New York, we clearly observe a non-linear effect of the spatially lag dependent variable. Indeed, when Wy is small (below 2.6%) an increase in the vote share for independents in the neighborhood induces an increase of the vote share for independents in the considered county. This is probably because voters think that their vote will strengthen the message conveyed by the neighbors on the dissatisfaction with the political establishment. However, when the votes for outsiders become larger in the neighborhood (around 2.6 percent of the total votes),the "votes for change" start decreasing. This is probably due to the fact that voters realize that they should vote strategically to prevent the candidate they dislike the most to win the elections. A similar feature can be observed in North Carolina and Michigan. We decided to present the case of Louisiana as well, but for a different reason than the shape of the non-parametric function. If we observe the data cloud and the regression fit,we see that the slope of the function is relatively linear and steep up to a value of Wy of around

2.8. Then it becomes awkwardly flat and goes through the single outlying point. If instead of fitting the non-parametric line, we would have fitted a line estimated using a standard SAR model, the outlying county would not have distorted the end of the curve as here but it would have affected the slope of the autoregressive line upwards leading to an underestimation of the autoregressive effect. This also means that this method is probably better suited than SAR in case of existence of outliers. We however think that this is somehow out of the scope of the paper and is presented here as anecdotal evidence.

As far as inference is concerned, the V-statistics is larger than 2 (in absolute value) in all states. Its value is 3.06 for New York, 2.85 for North Carolina, 20.60 for Michigan and 61.79 for Louisiana. This clearly rejects the linearity assumption in all the cases considered, which confirms the visual inspection of the graphs.

5 Conclusion

In spatial econometrics, one of the most commonly used model is the so called spatial autoregressive (SAR). This model assumes a linear relation between what happens in a location and in the neighborhood. However, supposing this linearity is probably too restrictive. For this reason, we propose a simple semiparametric model that allows estimating consistently the parameters associated to the covariates, for all possible forms of spatial autocorrelation. Furthermore, it also allows to estimate nonparametrically the shape of the effect of the endogenous spatial lag by fitting the residuals from the partialled out equation with the endogenous spatial lag. A simple test is moreover proposed to check the relevance of estimating a nonlinear SAR model. We then present some simple simulations and an empirical application to highlight the usefulness of the procedure. Though we are conscious that still a lot of work remains to be done to generalize SAR models to alternative situations than the ones presented here, we believe that approaching autoregressive models from this alternative perspective is promising.

References

[1] Anselin L. (1988a), *Spatial Econometrics, Methods and Models*, Kluwer Academic Publishers, Dordrecht.

[2] Anselin L. (1988b), "Lagrange Multiplier Tests Diagnostics for Spatial Dependence and Spatial Heterogeneity", *Geographical Analysis*, 20: 1-17.

[3] Anselin L. (2001), "Rao's Score Test in Spatial Econometrics", *Journal of Statistical Planning and Inference*, 97: 113-139.

[4] Anselin L., Bera A.K., Florax R., Yoon M.J. (1996), "Simple diagnostic tests for spatial dependence", *Regional Science and Urban Economics*, 26(1): 77-104.

[5] Anselin L., Florax R. (1995), "Small sample properties of tests for spatial dependence in regression models: some further results", in: L. Anselin and R. Florax, Eds), New Directions in Spatial Econometrics, Springer, Berlin.

[6] Anselin L, Florax R, Rey S. (2004), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, Berlin.

[7] Cox, G. (1997), *Making Votes Count*, Cambridge University Press, Cambridge.

[8] Lee. L.F. (2004), "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models", *Econometrica*, 72: 1899-1925.

[9] Lesage J., Pace K. (2009), *Introduction to Spatial Econometrics*, CRC Press/Taylor and Francis Group, London.

[10] Myerson, R. (2002). "Comparison of Scoring Rules in Poisson Voting Games." *Journal of Economic Theory*, 103: 219-251.

[11] Myerson, R. and Weber, R. (1993), "A Theory of Voting Equilibria," *American Political Science Review*, 87: 102-114.

[12] Nadaraya, E. A. (1964), “On Estimating Regression”, *Theory of Probability and its Applications* 9 (1): 141-142.

[13] Yatchew, A. (1997), “An elementary estimator of the partial linear model” *Economics Letters* 57(2): 135-143.

[14] Yatchew, A. (1998), “Nonparametric Regression Techniques in Economics”, *Journal of Economic Literature* 36(2): 669-721.

[15] Watson, G. S. (1964), “Smooth regression analysis”, *Sankhya*, Ser. A, 359-372.