Nonparametric regression with spatially dependent data

Margherita Gerolimetto
University of Venice

Stefano Magrini
University of Venice

Second Draft: March 2010

Abstract

In this paper we present a new procedure for nonparametric regression in case of spatially dependent data. In particular, we modify the two-step local linear regression of Martins-Filho and Yao (2009) by introducing information on spatial dependence via a nonparametric estimate of the error covariance matrix. The finite sample performance of our proposed procedure is then shown via Monte Carlo simulations for various data generating processes and its practical usage is illustrated through an application to the familiar crime data set for 49 Columbus neighbourhoods.

Keywords
nonparametric smoothing, spatial dependence

JEL Codes
C14; C21

Address for correspondence:
Stefano Magrini
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (++39) 041 2349194
Fax: (++39) 041 2349176
e-mail: smagrini@unive.it

1 Introduction

In the analysis of cross-section data quite often researchers have to face problems of misspecifications arising from dependence across spatially organised observational units. To deal with spatial dependence, the spatial econometric literature offers a number of models, for example the Cliff-Ord type models (Cliff and Ord, 1973):

$$\begin{aligned} \mathbf{Y} &= \rho W_1 \mathbf{Y} + \beta \mathbf{X} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \epsilon \end{aligned} \tag{1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector, $\mathbf{X}$ is an $n \times p$ matrix[1], ϵ is a $n \times 1$ vector of innovations, $-1 < \rho < 1$, $-1 < \lambda < 1$, β is a $p \times 1$ vector of parameters and W_1 and W_2 are $n \times n$ spatial weights matrices whose w_{ij} elements are non negative when $i \neq j$ and zero otherwise.

Among the most common specifications derived from model (1) are the Spatial Error Model, when the constraint $\lambda = 0$ holds

$$\begin{aligned} \mathbf{Y} &= \beta \mathbf{X} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \epsilon \end{aligned} \tag{2}$$

and the Spatial Lag Model, when $\rho = 0$

$$\mathbf{Y} = \rho W_1 \mathbf{Y} + \beta \mathbf{X} + \epsilon \tag{3}$$

Asymptotic properties of various (parametric) estimates of λ, ρ and β have been studied. Anselin (1988) demonstrates the inconsistency of the Least Squares estimates of λ and β while Kelejian and Prucha (1998, 1999) show that other estimates, such as Instrumental Variables and Gaussian Maximum Likelihood, are consistent and asymptotically normal.

In this paper, we focus on the nonparametric counterparts of the Spatial Error and Spatial Lag models in (2) and (3). More precisely, we consider the following nonparametric specifications:

$$\begin{aligned} \mathbf{Y} &= m(\mathbf{X}) + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \epsilon \end{aligned} \tag{4}$$

---

[1]For simplicity's sake we assume $p = 1$. All that follows can be generalised to the multivariate case $p \geq 2$.

and

$$\mathbf{Y} = \rho W_1 \mathbf{Y} + m(\mathbf{X}) + \epsilon \tag{5}$$

where $m(\mathbf{X})$ is a smooth function. It is worth noting that the just reported models are in fact nonparametric regressions and, in particular, that model (4) is one such regression with spatial autoregressive error structure.

Nonparametric regression has become quite a standard statistical instrument when the functional form is possibly neither linear nor nonlinear of a specific type. In general, the estimate of this kind of regression can be obtained in correspondence of some fixed points by means of some smoothing methods. One of the most commonly adopted estimation techniques is the local constant estimator (LCE, hereafter), also known as Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964):

$$\hat{m}(x) = \frac{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) Y_j}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)} \tag{6}$$

where h is the bandwidth, the parameter that controls the degree of smoothness.

Under the hypothesis of independent and indentically distributed error terms, the LCE is consistent and, under further conditions, it is also asymptotically normal. In addition, Robinson (2008, 2009) derives consistency and asymptotic theory for this estimator when errors are affected by dependence of various types, among which stationary error terms on a lattice of arbitrary dimension as well as error terms generated by a spatial autoregression. It is worth noting that Robinson (2009) only considers linear forms of spatial dependence, therefore excluding the mixing ones.

As highlighted by Martins-Filho and Yao (2009), most asymptotic results for the LCE estimator in case of dependent errors are unfortunately contingent on the assumptions made on the covariance structure and it is not possible to generalize their application to different parametric structures. Stimulated by this lack of generality, attention within the nonparametric literature has focussed on estimators that, by incorporating the information contained in the error covariance structure, outperform, both asymptotically and in finite samples, traditional

3

nonparametric ones. Here, we specifically draw on the work by Martins-Filho and Yao (2009) who establish a set of sufficient conditions for the asymptotic normality of the local linear estimator (LLE)

$$\hat{m}(x) = \frac{\sum_{j=1}^{n} K(\frac{x-X_j}{h})Y_j}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})} + (x - \bar{X}_w)\frac{\sum_{j=1}^{n} K(\frac{x-X_j}{h})(X_j - \bar{X}_w)Y_j}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})(X_j - \bar{X}_w)^2} \tag{7}$$

where

$$\bar{X}_w = \frac{\sum_{j=1}^{n} K(\frac{x-X_j}{h})X_j}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})}$$

when the error correlation structure is as general as possible. Building on this, we propose a two-step procedure for nonparametric regression with spatially dependent data that does not require *a priori* parametric assumptions on spatial dependence; information on its structure is actually drawn from a nonparametric estimate of the errors spatial covariance matrix.

The structure of the paper is as follows. In the Section 2 we discuss the theoretical underpinnings of our procedure whose detailed description is given in Section 3. In Section 4 we present our finite sample experiment and the results while in Section 5 we provide an empirical illustration on how to apply our method. Details regarding the implementation of the Monte Carlo experiment are reported in the Appendix.

2   A theoretical framework for nonparametric regression with dependent errors

Within the econometric literature on nonparametric regression, several authors (for example, Xiao *et al.*, 2003; Lin and Carroll, 2000; Ruckstuhl *et al.*, 2000; Wang, 2003) study possible extensions to settings in which errors can be correlated and heteroschedastic. In all these instances, however, a specific parametric structure for the dependence must be assumed beforehand. Differently from these, Martins-Filho and Yao (2009) (hereafter MFY) develop a two-step procedure whose asymptotic validity is proved under rather general covariance structures.

More formally, MFY consider the following nonparametric regression:

$$\mathbf{Y} = m(\mathbf{X}) + \mathbf{u} \tag{8}$$

where the error term $\mathbf{u}$ is such that $E(u_i) = 0$, $\forall i = 1, ..., n$, and $E(u_i, u_j) = \omega_{ij}(\theta_0)$, $\theta_0 \in R^p$, $p < \infty$.

Under assumptions A1-A5 (Martins-Filho and Yao, 2009; page 311), MFY demonstrate in Theorem 2 the asympotic normality and convergence rate of the traditional LLE of model (8). In particular, assumptions A1, A2 and A5 provide the standard regularity conditions for the kernel function, assumption A3 guarantees, as $n \to \infty$, the convengence of the weighted average of the diagonal terms of the error covariance while assumption A4 disciplines the mixing form of dependence therein considered.

In addition, MFY observe that the traditional LLE of (8), $\hat{m}$, does not exploit the information contained in the error term correlation structure. To improve its performance they therefore suggest a two-step procedure that incorporates this information in order to yield spherical error terms. More in detail, let $\Omega(\theta_0)$ denote the $n \times n$ matrix with elements ω_{ij} and $P(\theta_0)$ be a $n \times n$ matrix such that $\Omega(\theta_0) = P(\theta_0)P(\theta_0)'$. Now, by defining the new regressand $\mathbf{Z} = P^{-1}(\theta_0)\mathbf{Y} + (I_n - P^{-1}(\theta_0))m(\mathbf{X})$, MFY replace the original regression with the following

$$\mathbf{Z} = m(\mathbf{X}) + \epsilon \tag{9}$$

where the error terms $\epsilon = P^{-1}(\theta_0)\mathbf{u}$ are now spherical by construction. The new estimator, $\check{m}(\mathbf{X})$, is simply the LLE of (9). With the additional assumption A6 (Martins-Filho and Yao, 2009; page 313), constraining the nature of the stochastic process $\mathbf{u}$ to be a linear transformation of i.i.d. processes, MFY show $\check{m}(\mathbf{X})$ to represent an improvement over $\hat{m}(\mathbf{X})$ in terms of efficiency (Theorem 3). As is usual in the literature on two-stage nonparametric regression, undersmoothing in the first stage is required in order for the bias from the first stage estimator to be smaller than the leading bias coming from the second stage.

It is particularly important to emphasise is that $\mathbf{Z}$ is not observed as it depends on $m(\mathbf{X})$ and $P(\theta_0)$ that are both unknown. To overcome this obstacle, MFY propose a feasible version of

5

the $\check{m}(\mathbf{X})$ estimator. This estimator, $\dot{m}(\mathbf{X})$, is based on an observed regressand $\dot{\mathbf{Z}} = P^{-1}(\hat{\theta})\mathbf{Y} + (I_n - P^{-1}(\hat{\theta}))\hat{m}(\mathbf{X})$ in which, compared to $\mathbf{Z}$, a pilot local linear estimate $\hat{m}(\mathbf{X})$ is used in place of $m(\mathbf{X})$ and $P(\hat{\theta})$ in place of $P(\theta_0)$. The authors also provide an asymptotic result (Theorem 4) guaranteeing that, as long as a consistent estimate $\hat{\theta}$ is plugged in the expression $P(\theta_0)$, the feasible estimator $\dot{m}(\mathbf{X})$ is asymptotically equivalent to $\check{m}(\mathbf{X})$.

Among the forms of dependence assumed to obtain these theoretical results, however, spatial dependence has not been treated. For this reason Martin-Filho and Parmeter (2009, hereafter MFP) modify MFY's assuption set. On the one hand, the error term is then supposed to possess a spatial covariance matrix such that the weighted average of the main diagonal elements converge as $n \to \infty$; on the other, spatial mixing conditions are imposed as in Jenish and Prucha (2009). Coherently with this modified assumption set, MFP derive a version of MFY's two-step estimator that is able to handle spatial dependence and prove that its feasible version, in which unobserved are substituted by estimates, is asymptotically equivalent.

3   A new procedure for nonparametric regression with spatially dependent errors

In this Section we describe a new procedure for nonparametric regression with spatially dependent errors whose theoretical background is represented by the two-step nonparametric regression presented in the previous Section. Peculiar features of our procedure are that the spatial covariance matrix is estimated nonparametrically starting from a direct representation of spatial dependence and that the bandwidth in the second step regression is chosen by exploiting information on spatial dependence.

3.1   Nonparametric estimation of a spatial covariance matrix

A commonly adopted approach to express the elements of a generic spatial covariance matrix Ω is through a direct representation of the dependence as some function of the distance separating

locations l_i and l_j. In such an instance

$$\gamma(l_i, l_j) = \sigma^2 f(d_{ij}, \phi) \tag{10}$$

defines the spatial autocovariance function and

$$\rho(l_i, l_j) = f(d_{ij}, \phi) \tag{11}$$

defines the spatial autocorrelation function, where d_{ij} is the distance between locations i,j and $f(.)$ is a decaying function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}, \phi)| \leq 1$ with ϕ being an appropriate vector of parameters. Following this approach, the spatial covariance matrix Ω is positive definite and composed by elements ω_{ij} obtained through function $\gamma(l_i, l_j)$ in observed distances across sites. These features of matrix Ω follow directly from the stationarity and isotropy assumptions underpinning the existence of a spatial covariance function.[2]

The direct representation approach has the advantage of being parsimonious. A further important advantage is that it allows to reach a nonparametric estimate of the spatial covariance matrix thus avoiding the use of plug-in methods typical of parametric settings.

Bjørnstad and Falck (2001) (hereafter BF) propose a nonparametric estimate of the spatial covariance matrix moving from a continuous nonparametric positive semidefinite estimator of $f(d_{ij}, \phi)$ in (11), called spline correlogram. They build on the seminal work of Hall and Patil (1994) who, in turn, develop a kernel estimator of the spatial autocorrelation function $\rho(l_i, l_j)$:

$$\tilde{\rho}(s_i, s_j) = \frac{\sum_{i=1}^n \sum_{j=1}^n K(d_{ij}/a)(\hat{\rho}_{ij})}{\sum_{i=1}^n \sum_{j=1}^n K(d_{ij}/a)} \tag{12}$$

where K is a kernel function, a is a bandwidth and $\hat{\rho}_{ij}$ is the sample correlation

$$\hat{\rho}_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{1/n \sum_{l=1}^n (z_l - \bar{z})^2} \tag{13}$$

in which $\bar{z} = 1/n \sum_{l=1}^n z_l$ is the sample mean.

_____

[2]These assumptions are certainly met when Ω represents the spatial covariance matrix of homoskedastic errors of a Cliff-Ord type model.

Starting from the estimator in (12), BF (2001) opt, as a smoother, for a cubic B-spline

$$K(d/a) = \frac{1}{2} exp \left(-\frac{|d/a|}{\sqrt{2}} \right) sin \left(-\frac{|d/a|}{\sqrt{2}} + \frac{\pi}{4} \right) \tag{14}$$

where d denotes a generic measure of distance, given that this smoother adapts better to irregularly spaced data and produces a consistent estimate of the covariance function (Hyndman and Wand, 1997). However, since the estimator $\tilde{\rho}(l_i, l_j)$ must be not only consistent but also positive semidefinite, and this is not necessarily guaranteed by the estimator in equation (12), BF make use of a Fourier-filter method (Hall *et al.*, 1994). More specifically, they first calculate the Fourier transform of $\tilde{\rho}(l_i, l_j)$, then set to zero all negative excursions of the transformed function and, finally, obtain a nonparametric positive semidefinite estimate of the spatial correlation function by backtransformation.

3.2 Bandwidth choice through spatial statistics

Ellner and Seifu (2002), herafter ES, suggest a so-called Residual Spatial Autocorrelation (RSA) method to select the bandwidth h in a nonparametric regression making use of residuals spatial dependence. The logic here is rather intuitive: in case of unfitted structure, two nearby residuals tend to show a greater degree of similarity (on average) than two randomly chosen residuals. So, when residuals exhibit spatial correlation in the space defined by the independent variables, this signals that the model is not capturing all the structure in the data and a different bandwidth should be preferred.

To quantify spatial dependence, in direct analogy with the the familiar Durbin Watson test in time series data, ES make use of Moran's I statistic to test for the null hypothesis of no spatial dependence in the residuals

$$I = \frac{n \sum_{i \neq j=1}^{n} w_{ij} (e_i - \bar{e})(e_j - \bar{e})}{\sum_{i \neq j=1}^{n} w_{ij} \sum_{i=1}^{n} (e_i - \bar{e})^2} \tag{15}$$

where w_{ij} are the elements of a spatial weight matrix and e_i are the residuals with $i, j = 1, \ldots, n$.

From an operational point of view, according to the RSA method, an appropriate value for the bandwidth can therefore be chosen as follows: *i)* select a grid of possible values for

8

the bandwidth h; *ii)* compute Moran's I on the residuals from each alternative value of h; *iii)* plot these as a function of h and then choose the value of h such that I matches its expected value. In other words, if a good specification for the model is reached, a reasonable value for the bandwidth is the one that guarantees that the index does not reveal spatial dependence in the residuals.

3.3 A new procedure

Analogously to MFY and MFP, we aim at estimating model 8

$$\mathbf{Y} = m(\mathbf{X}) + \mathbf{u}$$

working under the following assumptions:

Assumption 3.1 *(i)* Let $f_i(x)$ be the marginal density of X_i evaluated at x with $f_i(x) < C$, for all i, x; *(ii)* $f_i^{(r)}(x)$ represents the r-th order derivative of $f_i(x)$ evaluated at x and $\left| f_i^{(1)}(x) \right| < C$; *(iii)* $|f_i(x) - f_i(x')| < C |x - x'|$, for all x and x' ; *(iv)* $f_{lkijmo}(x_l, \ldots, x_o)$ denotes the joint density of $X_l, \ldots, X_o$ evaluated at $x_l, \ldots, x_o$ where $f_{lkijmo}(x_l, \ldots, x_o) < C$ for all $x_l, \ldots, x_o$; *(v)* $\bar{f}_n(x) = n^{-1} \sum_i^n f_i(x) \to \bar{f}(x)$ as $n \to \infty$ where $0 < \bar{f}(x) < \infty$; *(vi)* As $n \to \infty$, $0 < inf_{x \in G |\bar{f}(x)|} < C$ for a compact set G.

Assumption 3.2 $K(x) : R \to R$ is a symmetric bounded function with bounded support S_K such that: *(i)* $\int K(x)dx = 1$; *(ii)* $\int xK(x)dx = 0$; *(iii)* $\int x^2 K(x)dx = \sigma_K^2$; *(iv)* for all $x, x' \in S_K$ we have $|K(x) - K(x')| \leq C |x - x'|$.

Assumption 3.3 Let v_{ij} be the i, j element of the spatial covariance matrix of error terms $\mathbf{u}$ defined as $V = \sigma^2 f(d_{ij}, \phi)$ where d_{ij} is the distance between locations i,j and $f(.)$ is such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}, \phi)| \leq 1$, ϕ is a vector of parameters, $|v_{ij}| < C$ for all i, j, $\bar{v}_n = n^{-1} \sum_i^n v_{ii} \to \bar{v}$ as $n \to \infty$ where $0 < \bar{v} < \infty$ and $\bar{v}_{fn}(x) = n^{-1} \sum_i^n v_{ii} f_i(x) \to \bar{v}_f(x)$ as $n \to \infty$ where $0 < \bar{v}_f(x) < \infty$ for all x.

Now, let D_n a finite subset of the infinite countable lattice $D \in R^v$, $\alpha_{k,l,n}(r)$ be the α-mixing coefficients, $\bar{\alpha}_{k,l}(r) = sup_n \alpha_{k,l,n}(r)$, $\bar{\alpha}_{1,1}(m)$ a non-increasing sequence of mixing coefficients and

$\alpha_{inv}(v) = max\{m \geq 0 : \bar{\alpha}_{1,1}(m) > v\}$.

Assumption 3.4 (x_i, u_i), $i = 1, \ldots, n$ is an α-mixing process whose mixing coefficients $\alpha_{k,l,n}(r)$ satisfy: (i) $lim_{k \to \infty} \ sup_n sup_{i \in D_n} \int_0^1 \alpha_{inv}^v(v) \left(\bar{Q}_{i,n}^{(k)}(v) \right)^2 du = 0$ where $\bar{Q}$ in an "upper-tail" quantile function as in Jenish and Prucha (2009, Assumption 2, page 88); (ii) $\sum_{m=1}^\infty m^{v-1} \bar{\alpha}_{k,l}(m) < \infty$ for $k + l < 4$; (iii) $\bar{\alpha}_{1,\infty}(m) = O(m^{-v-\epsilon})$ for some $\epsilon > 0$.

Assumption 3.5 There exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o\left(\sqrt{nh_n}\right)$ for a bandwidth sequence h_n such that $(nh_n^{-1})^{\frac{1}{2}} \bar{\alpha}_{k,l}(s_n) \to 0$ as $n \to \infty$.

Assumption 3.6 The r-th order derivative of $m(x)$, $m^{(r)}(x)$ satisfies $m^{(r)}(x) < C$ for all x and $r = 1, 2$.

Assumptions 3.1, 3.2, 3.5 and 3.6 define a standard framework for kernel regression; Assumptions 3.3 and 3.4 instead describe the form of dependence in the errors. It is worth noting that the spatial α-mixing error structure traced out through these assumptions is quite general and naturally includes Cliff-Ord type models.

The procedure we develop is a two-stage procedure that differs from existing ones in the fact that the consistent estimate of the variance-covariance matrix needed to yield spherical residual is obtained through a nonparametric estimator. Specifically, the variance-covariance matrix is estimated through the spatial correlogramm proposed by BF since, once the hypotheses for the existence of the spatial correlation function are met, no further assumptions on the form of spatial dependence are required.

Our procedure can be described as follows:

0. *Pilot fit*: estimate $m(\mathbf{X})$ with a local polynomial smoother, where the bandwidth is in fact a nearest neighbour smoothing parameter such that the width of each smoothing neighbourhood covers some percentage of the data. As for the degree of the polynomial, we consider $p = 0, 1$. Thus the local constant and the local linear estimators are effectively used. The output is $\hat{\mathbf{u}} = \mathbf{Y} - \hat{m}(\mathbf{X})$.

1. *Nonparametric covariance matrix estimation*: using the spline correlogram, obtain $\hat{V}$, the estimated spatial covariance matrix of $\hat{\mathbf{u}}$.

2. *Final fit*: run the following modified regression

$$\boldsymbol{\Psi} = m(\mathbf{X}) + \mathbf{e}$$

where $\boldsymbol{\Psi} = \hat{m}(\mathbf{X}) + L^{-1}\hat{\mathbf{u}}$, L is obtained by taking the Cholevsky decomposition of $\hat{V}$ and $\mathbf{e} = L^{-1}\mathbf{u}$ is a spherical error term. The bandwidth is chosen through a modified version of the RSA criterion. In particular, the adaptations we introduce to the original RSA criterion are:

- space is meant in geographical sense rather than in the variables' domain sense;

- the choice of h is made by minimising the p-value instead of solving the equation $I(h) = 0$ as originally suggested by ES; additionally, note that inference on Moran's I is based on an empirical distribution obtained by randomly reshuffling the residuals (permutation approach).

It is worth emphasising that, since Nonparametric Spatial Lag models as in equation (5) always have a Spatial Error representation

$$\mathbf{Y} = (I - \rho W_1)^{-1} m(\mathbf{X}) + (I - \rho W_1)^{-1}\epsilon$$

it follows that our procedure can be usefully applied to these models as well. In what follows, therefore, we assess the finite sample performance of the procedure generating data from both types of models.

4 Monte Carlo study

This Section describes a Monte Carlo experiment[3] that aims at showing, via simulations, the finite sample performance of our procedure in comparison with other estimation methods, both parametric and nonparametric, that do not take the presence of spatial dependence into account.

---

[3]The code has been written in Matlab (Matlab 7.7.0, R2008b).

The comparison is done in order to show the effective improvement in regression estimation results when spatial dependence is not neglected.

We carry out the Monte Carlo experiment for Spatial Error and Spatial Lag Models, considering both linear and nonlinear specifications. To begin with Spatial Error Models, the performance of all estimators is thus compared for linear specifications

$$\begin{aligned} \mathbf{Y} &= \beta\mathbf{X} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \epsilon \end{aligned} \tag{16}$$

and nonlinear ones

$$\begin{aligned} \mathbf{Y} &= m(\mathbf{X}) + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \epsilon \end{aligned} \tag{17}$$

where $\beta = 1$, $X \sim N(100,1)$, $m(X) = sin(2X) + 2e^{-2X^2}$ and $\epsilon \sim N(0,1)$. The distance matrix is based on euclidean distances computed over spatial coordinates generated as independent draws from $U(0,1000)$. The spatial weights matrix is obtained as a k-nearest neighbours contiguity matrix, where k=10% of the number of observations;[4] we label this parameter k-data.

We consider various values for λ, in particular $\lambda = 0$, 0.2, 0.4, 0.6, 0.8, in order to study situations with different levels of spatial dependence. The simulated data set length is $N = 50, 100, 200$ and the number of Monte Carlo replications is 2000.

The estimation methods we use are: our procedure, implemented both with a local constant (SNP0) and a local linear (SNP1) estimator; the traditional local constant and local linear estimators (NP0 and NP1 respectively); the Ordinary Least Squares (OLS) estimator.

In all nonparametric estimates we employ a Gaussian kernel. As far as the bandwidth parameter is concerned, we set it to 30% of the data in all traditional local regressions. In addition, we label h_p the bandwidth in the pilot estimates of both SNP0 and SNP1 and we set it to 5% of the data. As discussed in Section 2, this latter choice is motivated from the fact that, since our procedure consists of two subsequent estimates, the bandwidth of the pilot estimate needs to be small in order to avoid the mean bias to pile up.[5]

---

[4]The motivation behind the choice of this value is exaustively discussed in the Appendix.

[5]Possibles effects of the choice of the bandwith value on the estimates are investigated in the Appendix.

The last parameter to be considered is the share of nearest neighbours in the spatial weight matrix involved in the RSA criterion (k-RSA) that we set equal to k-data.[6]

The Monte Carlo experiment is then entirely repeated for Spatial Lag Models. In particular, we consider a linear data generating process

$$\mathbf{Y} = \rho W_1 \mathbf{Y} + \beta \mathbf{X} + \epsilon \tag{18}$$

and a nonlinear one

$$\mathbf{Y} = \rho W_1 \mathbf{Y} + m(\mathbf{X}) + \epsilon \tag{19}$$

where $m(X)$ and all other features of the experiment are as in the Spatial Error counterpart of the experiment.[7]

Finally, since our ultimate aim is to estimate a nonparametric regression with spatial dependence free residuals, we evaluate the performance of all estimators by applying spatial dependence diagnostics to regression residuals: using the Moran's I test, performance is then expressed through the rejections percentage of the null hypothesis of spatial independence in all residuals sets. The idea is that the better the estimate, the more dependence is captured by $\hat{m}(\mathbf{X})$, the lower the share of simulated sets in which the Moran's I null hypothesis is rejected. The I statistics significance ($\alpha = 0.05$) is established using 999 random permutations.

The results of the experiment are reported in Tables 3.1-3.12. More specifically, the first six tables show the results for linear systems: Tables 3.1, 3.3, 3.5 refer to the Spatial Error Model, while Tables 3.2, 3.4, 3.6 refer to the Spatial Lag Model. The results for nonlinear systems are instead collected in the last six tables: Spatial Error Model in Tables 3.7, 3.9, 3.11; Spatial Lag Model in Tables 3.8, 3.10, 3.12.

Overall, the results are promisingly good and, it should be emphasised, this is true both for Spatial Error and Spatial Lag models. In other words, in line with our expectations, the

---

[6]Once more, the motivation for this choice is provided in the Appendix.

[7]The only exception is that ρ only takes on values 0.2, 0.4, 0.6, 0.8 given that the case $\rho = 0$ is in fact identical to $\lambda = 0$.

procedure proposed in this paper appears to be fully capable of dealing with spatial dependence due to a lag in the dependent variable as well.

Before going through the results, however, note that the use of the output of Moran's I test to express performance can be at times misleading as the test's response depends on how the spatial weight matrix is built. More specifically, it has been shown (Anselin and Rey, 1991) that the power of the Moran's I test tends to decrease with the increase of the nearest neighbours percentage k (labelled k-test). The Appendix then describes an auxiliary Monte Carlo experiment run to determine the level of k-test such that the power is high enough for the percentage of rejections of the null hypothesis to be a proper indicator of the real performance of the estimators. The results suggest that in the present context also k-test should be set to 10%; consequently, this is the value employed throughout the experiment whenever a spatial weight matrix is required.

From all the tables it can immediately be seen that nonparametric methods (NP0, NP1, SNP0, SNP1) work in general better than OLS as they tend to reject less often the null hypothesis of spatial independence in the residuals. This holds true for both linear and nonlinear specifications, but is less evident in the former case as, for linear $m(\mathbf{X})$, the OLS represents a rival very difficult to beat. Among nonparametric methods, SNP procedures visibly outperforms the traditional polynomial regressions; for both polynomial degrees ($p = 0, 1$), SNPs appear able to detect the spatial structure and consequently to estimate $m(\mathbf{X})$ leaving no spatial dependence in the residuals.

As expected, the best performances of the SNP procedures are obtained for $0 \leq \lambda \leq$ 0.4. This is particularly evident in the nonlinear case where the percentage of rejections does not exceed 0.05 even when λ reaches 0.4. Obviously, the highest rejection percentages are found when $\lambda = 0.8$. However, it must be emphasised that the percentages achieved by SNP procedures are considerably lower than those obtained by other methods. Moreover, the case $\lambda = 0.8$ requires a special comment as it is the only one in which the performance of SNPs do not improve with the sample size. The explanation of this occurrence is that for very high values

14

of λ, spatial dependence is very likely to remain in the residuals and it becomes increasingly recognizable as the sample size gets larger.[8]

5   Empirical illustration

In this Section we present an empirical illustration based on Anselin's Columbus neighbourhood crime data set (Anselin, 1988) that consists of data on residential burglaries and vehicle thefts per thousand households ($\mathbf{Y}$), mean housing value ($\mathbf{X}_1$) and mean household income ($\mathbf{X}_2$) for 49 neighbourhoods in Columbus, Ohio, in 1980. We estimate the following nonparametric regression

$$\mathbf{Y} = m(\mathbf{X}_1, \mathbf{X}_2) + \epsilon \qquad (20)$$

using our procedure where the kernel function is the bivariate Gaussian and h_p=0.03 which is the smallest possible percentage value for the bandwidth since the corresponding nearest neighbour set consists of just one unit.

In order to implement the RSA criterion and to carry out the final diagnostics on the residuals with the Moran's I, we consider the following (row-standardised) spatial weights matrices: W_1, obtained as a k-nearest neighbour matrix where k is equal to 4 as in LeSage and Parent (2007); W_2, obtained as a k-nearest neighbour matrix where k is equal to 5, a value that corresponds to the average number of contiguous neighbourhoods; W_3, a first-order border contiguity matrix; W_4, a distance-based matrix whose elements are equal to the inverse of squared distance when regions are within a threshold corresponding to the first quartile of the distances distribution and zero otherwise.

Irrespective of the spatial weights matrix, the RSA criterion always returns the same band-width for the final regression ($h_{final} = 0.05$); a graphical representation of the obtained estimate is reported in Figure (1). It is interesting to observe that for a large section of the geometric space spanned by ($\mathbf{X}_1$) and ($\mathbf{X}_2$), the relationship appears to be weakly negative. In contrast,

---

[8]In statistical inference terms, one could say that when λ is large, it is probable that the Moran's I test is carried out under the alternative hypothesis of spatial dependence instead of under the null of independence.

the evident peak in the surface suggests the presence of a strong concentration of recorded crimes in neighbourhoods with low income households ($\mathbf{X}_2 < 12$) and middle valued houses ($\mathbf{X}_1 \in [35, 60]$).

In table 6 we report the results of the Moran's I test on the residuals of the nonparametric regression using the 4 spatial weights matrices listed above. As we can notice, the null hypothesis of spatial independence is always accepted at the nominal level of 5%.

6   Conclusions

In this paper we propose a new procedure for nonparametric regression in case of spatially dependent data, labelled SNP. Along the lines of MFY (2009), the procedure moves from a pilot estimate of the nonparametric regression with the local linear estimator and consists of two steps: firstly, we estimate the covariance matrix nonparametrically through a spline correlogram (BF, 2001); secondly, we run a modified regression exploiting the information on spatial dependence just obtained.

The original contribution provided by the procedure is twofold:

1. the estimation of the regression residuals' spatial covariance matrix, required to obtain a modified regression with spherical residuals, is obtained on the basis of a direct representation of spatial dependence;

2. the information concerning spatial dependence is exploited also when choosing the bandwidth parameter to be used in the second step regression.

We study the finite sample performance attained by the procedure through an extensive Monte Carlo experiment whose results suggest that the SNP procedure seems able to estimate nonparametric regressions quite satisfactorily, especially when the ability of traditional methods is hampered by the presence of spatial dependence. This result holds true both for Spatial Error and Spatial Lag specifications so that the SNP procedure thus appears to represent a

quite flexible tool. Moreover, as emphasised by the empirical illustration, the SNP procedure is also rather easy to implement.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrect: Kluwer.

Anselin, L. and Rey, S. (1991). Properties of Tests for Spatial Dependence in Linear-Regression Models. *Geographical Analysis*, **23**, 112-131.

Bjørnstad, O.N. and Falck, W. (2001). Nonparametric Spatial Covariance Functions: Estimation and Testing. *Enviromental and Ecological Statistics*, **8**, 53–70.

Cliff, A.D. and Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion.

Ellner, S.P, and Seifu, Y. (2002). Using Spatial Statistics to Select Model Complexity. *Journal of Computational and Graphical Statistics*, **11**, 348–369.

Florax, R.J.G.M. and Rey, S. (1995). The Impact of Misspecified Spatial Structure in Linear Regression Models. In *New directions in spatial econometrics*, Anselin L. and Florax R.J.G.M. (Eds.), Berlin: Springer-Verlag.

Hall, P. and Patil, P. (1994). Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields. *Probability Theory and Related Fields*, **99**, 399–424.

Hall, P., Fisher, N.I. and Hoffman, B. (1994). On the Nonparametric Estimation of Covariance Functions. *Annals of Statistics*, **22**, 2115–2134.

Hyndman, R.J. and Wand, M.P. (1997). Nonparametric Autocovariance Function Estimation. *Australian Journal of Statistics*, **39**, 313–324.

Jenish, N. and Prucha, I.R. (2009). Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields. *Journal of Econometric*, **150**, 86–98.

Kelejian, H.H. and Prucha, I.R. (1998). A Generalized Spatial two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate and Economics*, **17**, 99–121.

Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, **40**, 509–533.

LeSage J.P. and Parent, O. (2007). Bayesian Model Averaging for Spatial Econometric Models. *Geographical Analysis*, **39**, 241–267.

Lin, X. and Carroll, R.J. (2000). Nonparametric Function Estimator for Clustered Data when the Predictor is Measured with/without Error. *Journal of the American Statistical Association*, **95**, 520–534.

Loader, C. (1999). *Local Regression and Likelihood*. Berlin: Springer.

Martins-Filho, C. and Yao, F. (2009). Nonparametric Regression Estimation with General Parametric Error Covariance. *Journal of Multivariate Analysis*, **100**, 309–333.

Martins-Filho, C. and Parmeter, C. (2009). Nonparametric Regression Estimation with Spatial Correlation. Virginia Tech AAEC Working Paper.

Nadaraya, E.A. (1964). On Estimating Regression. *Theory of Probability and its Applications*, **9**, 141–142.

Robinson, P.M. (2008). Developments in the Analysis of Spatial Data. *Journal of the Japanese Statistical Society*, **38**, 87–96.

Robinson, P.M. (2009). Asymptotic Theory for Nonparametric Regression with Spatial data. *Journal of Econometrics, Forthcoming*.

Ruckstuhl, A.F., Welsh, A.H. and Carroll, R.J. (2000). Nonparametric Function Estimation of the Relationship between two Repeatedly Measured Variables. *Statistica Sinica*, **10**, 51–71.

Wang, N. (2003). Marginal Nonparametric Kernel Regression Accounting for within Subject Correlation. *Biometrika*, **90**, 43–52.

Watson, G.S. (1964). Smooth Regression Analysis. *Sankhya, Series A*, **26**, 359–372.

Xiao, Z., Linton, O.B., Carroll, R.J. and Mammen, E. (2003). More Efficient Local Polynomial Estimation in Nonparametric Regression with Autocorrelated Errors. *Journal of the American Statistical Association*, **98**, 980–992.

Appendix

In order to explain the value chosen for some parameters of the finite sample experiment presented in Section 4, in this Appendix we report the results of an auxiliary Monte Carlo exercise. In particular, this is done to clarify the set of criteria adopted for the choice of:

- the nearest neighbours percentage in the weight matrix used in the data generating process (k-data), in the RSA criterion (k-RSA) and in the Moran's I test for the final diagnostics (k-test)

- the bandwidth parameter in the pilot estimate (h_p) of our procedure

For this analysis we generated, within the setting of the Monte Carlo experiment in Section 4, 1000 pairs for x and y vectors via a Spatial Error Model (with both linear and nonlinear specifications) in which λ=0.4 and N=100 as well as a second set of 1000 pairs of x and y vectors via a Spatial Lag Model (again with both linear and nonlinear specifications) in which ρ=0.4 and N=100.

Firstly, our objective is to study how the percentage of rejections of the null hypothesis of no spatial dependence may vary according to k-data, k-RSA and k-test. The percentage of rejections of the Moran's I test on the residuals of OLS, NP0, NP1, SNP0, SNP1 is thus compared for various combinations of these parameters. More specifically, the values we consider for k-data are 0.05, 0.10 and 0.15, while for k-RSA and k-test the values are 0.05, 0.10, 0.15, 0.20 and 0.25.

Secondly, we explore how the choice of the bandwidth parameter h_p affects the estimator's behaviour. In particular, we consider two values for this parameter. In the first case, h_p is set to 0.30, i.e. the same value employed in NP0 and NP1 procedures. Additionally, as commonly done in two-step nonparametric estimates, we introduce a strong level of undersmoothing in the first stage by setting h_p=0.05.

Results are reported in Tables A.1-A.12. To begin with, it can be observed that when k-RSA and k-test are identical and, above all, equal to k-data, the Moran's I works with the

best information and this minimises the risk of deriving misleading conclusions in terms of estimators' performance. This is in line with the results reported by Florax and Rey (1995), according to whom the power of a test on spatial dependence increases when the connectivity function used in the test specification matches that of the data generating process.

As for the choice of k-data, it has been noted (Anselin and Rey, 1991) that the Moran's I test has lower power in more connected spatial configurations (i.e., as k-data increases). For this reason, a not too large degree of connectivity between units should allow a correct interpretation of the reported percentage of rejections in terms of the ability of the estimator, free from the possible influence of the low power of the test. The results of the present Monte Carlo excercise reveal that, in our analysis, a suitable value for k-data is 10%.

Finally, Tables A.1-A.12 also show that the performance of the estimators does not vary significantly with h_p. So, in line with what is commonly done in two-step nonparametric procedures, we set h_p=0.05.

Table 3.1: Linear Systems - Spatial Error - N=50

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.048	0.053	0.055	0.006	0.004
0.2	0.218	0.199	0.203	0.057	0.017
0.4	0.558	0.508	0.511	0.228	0.044
0.6	0.873	0.846	0.845	0.581	0.117
0.8	0.987	0.982	0.981	0.909	0.219

Table 3.2: Linear Systems - Spatial Lag - N=50

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.23	0.247	0.215	0.081	0.021
0.4	0.618	0.628	0.555	0.347	0.056
0.6	0.92	0.924	0.896	0.711	0.132
0.8	0.997	0.996	0.994	0.95	0.239

Table 3.3: Linear Systems - Spatial Error - N=100

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.058	0.055	0.054	0.009	0.002
0.2	0.229	0.217	0.221	0.064	0.008
0.4	0.589	0.543	0.568	0.266	0.031
0.6	0.887	0.874	0.872	0.653	0.097
0.8	0.996	0.993	0.994	0.947	0.226

Table 3.4: Linear Systems - Spatial Lag - N=100

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.244	0.266	0.233	0.092	0.009
0.4	0.644	0.678	0.626	0.403	0.038
0.6	0.945	0.95	0.935	0.816	0.122
0.8	0.999	0.999	0.998	0.987	0.294

Table 3.5: Linear Systems - Spatial Error - N=200

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.053	0.057	0.055	0.011	0.002
0.2	0.249	0.241	0.244	0.093	0.02
0.4	0.634	0.6	0.62	0.356	0.076
0.6	0.91	0.899	0.911	0.742	0.2
0.8	0.996	0.995	0.996	0.965	0.46

Table 3.6: Linear Systems - Spatial Lag - N=200

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.272	0.3	0.271	0.144	0.023
0.4	0.696	0.73	0.681	0.544	0.087
0.6	0.949	0.958	0.946	0.892	0.278
0.8	0.999	1	1	0.995	0.578

Table 3.7: Nonlinear Systems - Spatial Error - N=50

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.054	0.057	0.056	0.013	0.004
0.2	0.16	0.199	0.206	0.07	0.013
0.4	0.398	0.512	0.496	0.246	0.039
0.6	0.736	0.848	0.839	0.607	0.107
0.8	0.961	0.984	0.981	0.919	0.215

Table 3.8: Nonlinear Systems - Spatial Lag - N=50

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.231	0.234	0.216	0.088	0.017
0.4	0.567	0.569	0.543	0.328	0.051
0.6	0.868	0.897	0.872	0.689	0.12
0.8	0.989	0.99	0.99	0.944	0.225

Table 3.9: Nonlinear Systems - Spatial Error - N=100

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.05	0.055	0.049	0.01	0.003
0.2	0.16	0.21	0.213	0.073	0.009
0.4	0.405	0.546	0.551	0.28	0.026
0.6	0.765	0.865	0.876	0.683	0.089
0.8	0.975	0.992	0.992	0.954	0.224

Table 3.10: Nonlinear Systems - Spatial Lag - N=100

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.229	0.237	0.228	0.095	0.01
0.4	0.58	0.63	0.607	0.368	0.033
0.6	0.894	0.921	0.91	0.79	0.103
0.8	0.995	0.997	0.996	0.976	0.256

Table 3.11: Nonlinear Systems - Spatial Error - N=200

Moran's I empirical probability of rejections

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.054	0.053	0.056	0.02	0.002
0.2	0.175	0.227	0.247	0.107	0.012
0.4	0.45	0.597	0.607	0.389	0.066
0.6	0.791	0.896	0.904	0.773	0.194
0.8	0.976	0.995	0.994	0.973	0.462

Table 3.12: Nonlinear Systems - Spatial Lag - N=200

Moran's I empirical probability of rejections

ρ	OLS	NP0	NP1	SNP0	SNP1
0.2	0.263	0.265	0.261	0.141	0.015
0.4	0.625	0.671	0.658	0.499	0.082
0.6	0.911	0.941	0.937	0.87	0.23
0.8	0.995	0.999	0.999	0.993	0.522

Table 5.1: Moran's I test on SNP1 regression residuals – h_p=0.03

	W1	**W2**	**W3**	**W4**
I	0.0928	0.0768	0.1106	0.0774
p-value	0.1077	0.1070	0.0842	0.1119

Note: W1: 4-nearest neighbours; W2: 5-nearest neighbours; W3: first-order contiguity; W4: inverse of squared distance within threshold (1st quartile of distances)

Figure 1: SNP1 estimation
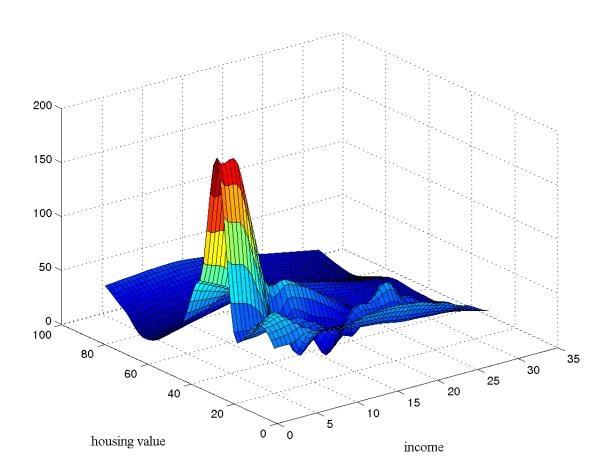
Table A.1: Spatial Error - h_p=0.05, k-data=5%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.832	0.800	0.807	0.469	**0.058**	0.622	0.794	0.802	0.532	**0.047**
0.05	0.10	0.830	0.797	0.812	0.505	0.101	0.622	0.792	0.799	0.561	0.104
0.05	0.15	0.833	0.797	0.807	0.552	0.184	0.625	0.799	0.806	0.602	0.186
0.05	0.20	0.835	0.797	0.813	0.601	0.239	0.621	0.798	0.798	0.632	0.241
0.05	0.25	0.836	0.800	0.813	0.622	0.275	0.624	0.787	0.803	0.661	0.287
0.10	0.05	0.688	0.650	0.671	0.381	0.067	0.523	0.642	0.665	0.429	0.065
0.10	0.10	0.688	0.652	0.665	0.360	0.047	0.522	0.641	0.662	0.410	0.041
0.10	0.15	0.690	0.651	0.665	0.375	0.065	0.524	0.647	0.658	0.419	0.068
0.10	0.20	0.686	0.658	0.667	0.417	0.129	0.523	0.647	0.663	0.456	0.128
0.10	0.25	0.690	0.650	0.667	0.438	0.181	0.527	0.640	0.666	0.498	0.179
0.15	0.05	0.584	0.557	0.572	0.335	0.059	0.420	0.551	0.557	0.356	0.057
0.15	0.10	0.579	0.562	0.570	0.299	0.041	0.428	0.547	0.553	0.326	0.036
0.15	0.15	0.584	0.553	0.569	0.284	0.030	0.423	0.548	0.556	0.318	0.029
0.15	0.20	0.579	0.555	0.565	0.290	0.037	0.424	0.548	0.555	0.334	0.042
0.15	0.25	0.582	0.561	0.563	0.323	0.083	0.428	0.551	0.561	0.358	0.079
0.20	0.05	0.497	0.477	0.489	0.301	0.056	0.370	0.474	0.483	0.309	0.053
0.20	0.10	0.495	0.476	0.491	0.266	0.040	0.374	0.473	0.486	0.294	0.037
0.20	0.15	0.501	0.476	0.488	0.240	0.031	0.376	0.475	0.483	0.277	0.028
0.20	0.20	0.495	0.476	0.491	0.236	0.027	0.369	0.476	0.482	0.267	0.026
0.20	0.25	0.500	0.485	0.494	0.243	0.033	0.369	0.473	0.482	0.275	0.031
0.25	0.05	0.431	0.420	0.431	0.260	0.060	0.334	0.412	0.418	0.281	0.057
0.25	0.10	0.438	0.422	0.428	0.236	0.044	0.340	0.415	0.419	0.255	0.039
0.25	0.15	0.438	0.428	0.429	0.215	0.028	0.330	0.412	0.412	0.239	0.027
0.25	0.20	0.433	0.417	0.426	0.202	0.023	0.335	0.409	0.416	0.237	0.023
0.25	0.25	0.432	0.426	0.427	0.202	0.026	0.334	0.410	0.417	0.230	0.021

Table A.2: Spatial Error - h_p=0.05, k-data=10%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.491	0.451	0.473	0.209	0.019	0.338	0.460	0.468	0.219	0.014
0.05	0.10	0.491	0.458	0.476	0.225	0.041	0.347	0.456	0.463	0.244	0.037
0.05	0.15	0.495	0.464	0.471	0.251	0.073	0.335	0.457	0.469	0.273	0.066
0.05	0.20	0.493	0.454	0.476	0.269	0.088	0.341	0.458	0.462	0.302	0.083
0.05	0.25	0.485	0.454	0.476	0.305	0.112	0.344	0.464	0.458	0.309	0.103
0.10	0.05	0.574	0.560	0.559	0.295	0.062	0.434	0.553	0.559	0.334	0.064
0.10	0.10	0.577	0.555	0.557	0.259	**0.037**	0.438	0.553	0.555	0.307	**0.031**
0.10	0.15	0.576	0.554	0.559	0.270	0.049	0.429	0.555	0.567	0.312	0.045
0.10	0.20	0.575	0.558	0.561	0.303	0.076	0.426	0.548	0.558	0.345	0.078
0.10	0.25	0.575	0.551	0.564	0.330	0.117	0.436	0.557	0.562	0.366	0.113
0.15	0.05	0.520	0.499	0.524	0.296	0.076	0.377	0.511	0.508	0.310	0.076
0.15	0.10	0.527	0.494	0.517	0.256	0.036	0.377	0.513	0.507	0.275	0.033
0.15	0.15	0.521	0.495	0.519	0.235	0.028	0.381	0.510	0.515	0.270	0.025
0.15	0.20	0.524	0.503	0.515	0.253	0.031	0.380	0.512	0.514	0.275	0.026
0.15	0.25	0.531	0.499	0.516	0.276	0.059	0.389	0.508	0.507	0.295	0.056
0.20	0.05	0.479	0.453	0.468	0.272	0.074	0.353	0.444	0.465	0.293	0.072
0.20	0.10	0.478	0.461	0.474	0.242	0.032	0.360	0.447	0.460	0.273	0.030
0.20	0.15	0.478	0.452	0.465	0.223	0.027	0.352	0.448	0.461	0.257	0.025
0.20	0.20	0.476	0.459	0.465	0.215	0.026	0.361	0.449	0.458	0.249	0.022
0.20	0.25	0.480	0.451	0.466	0.220	0.026	0.354	0.448	0.460	0.257	0.028
0.25	0.05	0.430	0.418	0.428	0.255	0.078	0.321	0.408	0.416	0.274	0.087
0.25	0.10	0.434	0.415	0.419	0.219	0.038	0.323	0.411	0.414	0.243	0.038
0.25	0.15	0.428	0.410	0.422	0.199	0.033	0.315	0.409	0.414	0.232	0.030
0.25	0.20	0.434	0.418	0.418	0.198	0.027	0.321	0.405	0.414	0.220	0.024
0.25	0.25	0.440	0.411	0.426	0.185	0.025	0.323	0.405	0.414	0.224	0.024

Table A.3: Spatial Error - h_p=0.05, k-data=15%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.323	0.316	0.331	0.117	0.011	0.241	0.311	0.315	0.128	0.012
0.05	0.10	0.321	0.319	0.323	0.139	0.030	0.229	0.307	0.321	0.151	0.030
0.05	0.15	0.327	0.315	0.328	0.152	0.046	0.237	0.311	0.323	0.174	0.046
0.05	0.20	0.324	0.318	0.325	0.162	0.056	0.231	0.305	0.322	0.180	0.061
0.05	0.25	0.327	0.319	0.331	0.178	0.063	0.238	0.309	0.318	0.188	0.064
0.10	0.05	0.427	0.397	0.421	0.199	0.051	0.302	0.410	0.411	0.209	0.062
0.10	0.10	0.419	0.397	0.423	0.170	0.020	0.303	0.411	0.412	0.184	0.015
0.10	0.15	0.422	0.396	0.420	0.184	0.031	0.306	0.411	0.413	0.191	0.028
0.10	0.20	0.418	0.396	0.416	0.201	0.040	0.300	0.416	0.416	0.205	0.040
0.10	0.25	0.415	0.398	0.418	0.224	0.062	0.302	0.408	0.416	0.230	0.055
0.15	0.05	0.463	0.433	0.452	0.254	0.075	0.320	0.444	0.447	0.277	0.087
0.15	0.10	0.464	0.437	0.446	0.214	0.030	0.322	0.443	0.442	0.235	0.028
0.15	0.15	0.467	0.428	0.456	0.192	**0.022**	0.323	0.443	0.444	0.225	**0.022**
0.15	0.20	0.463	0.438	0.448	0.210	0.024	0.322	0.442	0.447	0.236	0.023
0.15	0.25	0.462	0.434	0.451	0.229	0.042	0.319	0.447	0.449	0.253	0.041
0.20	0.05	0.431	0.398	0.419	0.248	0.087	0.306	0.400	0.417	0.269	0.100
0.20	0.10	0.433	0.405	0.410	0.208	0.041	0.307	0.400	0.412	0.233	0.036
0.20	0.15	0.435	0.405	0.423	0.190	0.027	0.308	0.406	0.416	0.219	0.025
0.20	0.20	0.430	0.408	0.425	0.187	0.025	0.319	0.407	0.413	0.220	0.019
0.20	0.25	0.423	0.400	0.424	0.189	0.032	0.316	0.407	0.417	0.222	0.026
0.25	0.05	0.398	0.377	0.384	0.228	0.087	0.290	0.371	0.373	0.246	0.105
0.25	0.10	0.398	0.376	0.380	0.202	0.045	0.290	0.374	0.371	0.213	0.040
0.25	0.15	0.396	0.379	0.381	0.174	0.033	0.290	0.370	0.378	0.202	0.030
0.25	0.20	0.404	0.377	0.380	0.168	0.030	0.294	0.371	0.375	0.199	0.028
0.25	0.25	0.393	0.380	0.383	0.165	0.026	0.294	0.375	0.376	0.193	0.023

Table A.4: Spatial Error - h_p=0.30, k-data=5%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.836	0.800	0.815	0.471	**0.039**	0.625	0.798	0.803	0.529	**0.037**
0.05	0.10	0.831	0.801	0.810	0.515	0.086	0.627	0.798	0.801	0.563	0.090
0.05	0.15	0.831	0.802	0.808	0.555	0.170	0.627	0.795	0.807	0.616	0.176
0.05	0.20	0.834	0.798	0.810	0.590	0.235	0.619	0.794	0.795	0.632	0.244
0.05	0.25	0.828	0.797	0.811	0.617	0.273	0.620	0.798	0.798	0.663	0.280
0.10	0.05	0.685	0.652	0.674	0.386	0.051	0.530	0.639	0.658	0.431	0.051
0.10	0.10	0.690	0.647	0.669	0.356	0.034	0.525	0.642	0.663	0.414	0.030
0.10	0.15	0.686	0.651	0.663	0.375	0.051	0.524	0.647	0.668	0.424	0.050
0.10	0.20	0.687	0.647	0.671	0.419	0.116	0.511	0.646	0.661	0.462	0.125
0.10	0.25	0.689	0.654	0.674	0.456	0.153	0.528	0.644	0.660	0.506	0.162
0.15	0.05	0.581	0.559	0.570	0.342	0.056	0.426	0.551	0.558	0.360	0.057
0.15	0.10	0.581	0.558	0.565	0.300	0.035	0.424	0.554	0.558	0.331	0.040
0.15	0.15	0.585	0.557	0.563	0.290	0.022	0.431	0.551	0.553	0.327	0.021
0.15	0.20	0.574	0.553	0.566	0.298	0.032	0.424	0.552	0.557	0.339	0.032
0.15	0.25	0.582	0.559	0.567	0.322	0.071	0.421	0.554	0.553	0.367	0.076
0.20	0.05	0.504	0.476	0.490	0.303	0.052	0.370	0.472	0.480	0.317	0.050
0.20	0.10	0.503	0.481	0.496	0.269	0.028	0.371	0.473	0.480	0.293	0.032
0.20	0.15	0.501	0.487	0.497	0.246	0.022	0.368	0.473	0.482	0.282	0.025
0.20	0.20	0.495	0.474	0.491	0.235	0.022	0.369	0.479	0.476	0.272	0.021
0.20	0.25	0.493	0.482	0.488	0.245	0.027	0.370	0.469	0.478	0.283	0.027
0.25	0.05	0.434	0.423	0.427	0.272	0.052	0.337	0.412	0.419	0.281	0.057
0.25	0.10	0.441	0.424	0.433	0.236	0.034	0.336	0.421	0.422	0.262	0.031
0.25	0.15	0.438	0.417	0.428	0.217	0.025	0.333	0.418	0.415	0.240	0.025
0.25	0.20	0.440	0.426	0.432	0.203	0.024	0.333	0.418	0.418	0.238	0.020
0.25	0.25	0.439	0.425	0.426	0.198	0.020	0.332	0.412	0.417	0.237	0.018

Table A.5: Spatial Error - h_p=0.30, k-data=10%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.493	0.447	0.473	0.213	0.018	0.343	0.453	0.466	0.216	0.009
0.05	0.10	0.492	0.460	0.475	0.233	0.028	0.339	0.461	0.467	0.255	0.030
0.05	0.15	0.493	0.455	0.476	0.258	0.058	0.337	0.460	0.465	0.270	0.053
0.05	0.20	0.486	0.457	0.476	0.274	0.079	0.339	0.454	0.469	0.303	0.081
0.05	0.25	0.491	0.459	0.473	0.303	0.104	0.340	0.460	0.460	0.314	0.110
0.10	0.05	0.579	0.554	0.567	0.299	0.050	0.429	0.550	0.561	0.341	0.048
0.10	0.10	0.580	0.556	0.558	0.257	**0.023**	0.427	0.550	0.556	0.312	**0.016**
0.10	0.15	0.580	0.556	0.563	0.275	0.033	0.426	0.553	0.567	0.317	0.030
0.10	0.20	0.568	0.552	0.560	0.305	0.058	0.427	0.555	0.564	0.348	0.065
0.10	0.25	0.569	0.558	0.561	0.330	0.100	0.427	0.551	0.564	0.376	0.108
0.15	0.05	0.519	0.505	0.519	0.294	0.070	0.380	0.500	0.514	0.315	0.075
0.15	0.10	0.531	0.497	0.522	0.260	0.026	0.386	0.509	0.510	0.285	0.025
0.15	0.15	0.525	0.496	0.517	0.240	0.018	0.380	0.509	0.516	0.273	0.020
0.15	0.20	0.524	0.495	0.519	0.256	0.022	0.385	0.503	0.515	0.280	0.025
0.15	0.25	0.527	0.497	0.517	0.276	0.044	0.387	0.512	0.510	0.306	0.046
0.20	0.05	0.481	0.455	0.473	0.277	0.072	0.354	0.450	0.464	0.298	0.069
0.20	0.10	0.483	0.453	0.471	0.243	0.030	0.354	0.451	0.462	0.272	0.027
0.20	0.15	0.483	0.454	0.468	0.220	0.022	0.354	0.457	0.458	0.257	0.023
0.20	0.20	0.478	0.455	0.472	0.220	0.019	0.357	0.454	0.467	0.246	0.018
0.20	0.25	0.482	0.453	0.469	0.226	0.022	0.356	0.447	0.459	0.262	0.022
0.25	0.05	0.429	0.407	0.418	0.258	0.076	0.324	0.406	0.421	0.281	0.071
0.25	0.10	0.436	0.409	0.422	0.219	0.035	0.321	0.412	0.422	0.245	0.030
0.25	0.15	0.432	0.410	0.422	0.202	0.024	0.320	0.408	0.419	0.229	0.020
0.25	0.20	0.431	0.415	0.422	0.200	0.019	0.316	0.411	0.419	0.221	0.021
0.25	0.25	0.430	0.419	0.424	0.193	0.018	0.323	0.404	0.415	0.221	0.019

Table A.6: Spatial Error - h_p=0.30, k-data=15%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.327	0.313	0.325	0.121	0.010	0.238	0.309	0.321	0.135	0.010
0.05	0.10	0.326	0.321	0.325	0.136	0.024	0.234	0.312	0.321	0.154	0.019
0.05	0.15	0.318	0.317	0.324	0.151	0.033	0.235	0.311	0.317	0.174	0.034
0.05	0.20	0.326	0.315	0.326	0.154	0.043	0.234	0.310	0.316	0.177	0.051
0.05	0.25	0.322	0.313	0.328	0.177	0.052	0.234	0.310	0.315	0.191	0.052
0.10	0.05	0.421	0.398	0.426	0.201	0.045	0.301	0.406	0.417	0.217	0.047
0.10	0.10	0.420	0.391	0.419	0.175	0.015	0.310	0.408	0.407	0.185	0.012
0.10	0.15	0.431	0.398	0.425	0.181	0.028	0.295	0.406	0.413	0.192	0.023
0.10	0.20	0.425	0.403	0.419	0.201	0.038	0.304	0.411	0.411	0.206	0.039
0.10	0.25	0.422	0.399	0.416	0.222	0.053	0.304	0.408	0.413	0.225	0.053
0.15	0.05	0.457	0.434	0.449	0.262	0.074	0.322	0.440	0.447	0.274	0.083
0.15	0.10	0.454	0.427	0.449	0.213	0.030	0.325	0.438	0.441	0.235	0.026
0.15	0.15	0.465	0.437	0.452	0.197	**0.015**	0.322	0.443	0.447	0.224	**0.017**
0.15	0.20	0.464	0.435	0.450	0.215	0.017	0.320	0.442	0.443	0.234	0.018
0.15	0.25	0.459	0.431	0.447	0.221	0.032	0.326	0.442	0.444	0.254	0.035
0.20	0.05	0.429	0.409	0.417	0.247	0.075	0.310	0.400	0.411	0.261	0.085
0.20	0.10	0.428	0.407	0.421	0.207	0.032	0.313	0.403	0.421	0.239	0.032
0.20	0.15	0.428	0.398	0.428	0.198	0.017	0.306	0.406	0.406	0.217	0.016
0.20	0.20	0.433	0.402	0.417	0.185	0.015	0.310	0.406	0.410	0.212	0.015
0.20	0.25	0.432	0.403	0.416	0.186	0.018	0.308	0.404	0.415	0.221	0.023
0.25	0.05	0.392	0.377	0.385	0.229	0.078	0.294	0.371	0.375	0.254	0.093
0.25	0.10	0.397	0.378	0.382	0.201	0.035	0.291	0.370	0.372	0.215	0.031
0.25	0.15	0.401	0.378	0.380	0.186	0.018	0.285	0.370	0.373	0.199	0.014
0.25	0.20	0.397	0.379	0.391	0.171	0.017	0.285	0.371	0.378	0.192	0.013
0.25	0.25	0.399	0.377	0.389	0.171	0.014	0.286	0.372	0.378	0.189	0.015

Table A.7: Spatial Lag - h_p=0.05, k-data=5%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.878	0.899	0.870	0.643	**0.059**	0.822	0.862	0.844	0.627	**0.058**
0.05	0.10	0.880	0.901	0.867	0.670	0.103	0.819	0.860	0.847	0.664	0.106
0.05	0.15	0.880	0.901	0.868	0.717	0.175	0.820	0.869	0.852	0.696	0.190
0.05	0.20	0.880	0.900	0.866	0.747	0.219	0.819	0.866	0.846	0.715	0.262
0.05	0.25	0.880	0.900	0.874	0.762	0.272	0.822	0.864	0.849	0.735	0.304
0.10	0.05	0.750	0.782	0.732	0.546	0.061	0.678	0.739	0.717	0.525	0.067
0.10	0.10	0.749	0.775	0.732	0.518	0.045	0.689	0.747	0.710	0.504	0.049
0.10	0.15	0.748	0.778	0.735	0.542	0.064	0.685	0.742	0.711	0.516	0.074
0.10	0.20	0.746	0.783	0.732	0.581	0.130	0.684	0.740	0.707	0.554	0.157
0.10	0.25	0.750	0.772	0.739	0.606	0.178	0.685	0.748	0.711	0.586	0.208
0.15	0.05	0.648	0.670	0.630	0.473	0.068	0.572	0.623	0.595	0.449	0.066
0.15	0.10	0.649	0.672	0.624	0.435	0.055	0.580	0.615	0.599	0.414	0.043
0.15	0.15	0.649	0.667	0.631	0.427	0.043	0.573	0.625	0.593	0.393	0.033
0.15	0.20	0.645	0.671	0.628	0.437	0.050	0.581	0.621	0.593	0.412	0.046
0.15	0.25	0.650	0.671	0.634	0.466	0.101	0.583	0.623	0.594	0.439	0.089
0.20	0.05	0.568	0.576	0.556	0.397	0.062	0.492	0.534	0.515	0.376	0.063
0.20	0.10	0.576	0.577	0.552	0.363	0.049	0.501	0.534	0.515	0.358	0.043
0.20	0.15	0.581	0.571	0.550	0.340	0.039	0.495	0.529	0.508	0.328	0.035
0.20	0.20	0.574	0.568	0.551	0.332	0.037	0.497	0.526	0.511	0.319	0.031
0.20	0.25	0.571	0.568	0.554	0.348	0.044	0.497	0.530	0.516	0.328	0.039
0.25	0.05	0.500	0.514	0.481	0.358	0.067	0.429	0.453	0.451	0.325	0.062
0.25	0.10	0.503	0.511	0.484	0.324	0.051	0.431	0.460	0.446	0.309	0.043
0.25	0.15	0.496	0.506	0.486	0.304	0.038	0.433	0.455	0.449	0.289	0.036
0.25	0.20	0.500	0.510	0.485	0.296	0.038	0.431	0.455	0.448	0.282	0.033
0.25	0.25	0.499	0.506	0.489	0.296	0.031	0.433	0.456	0.449	0.280	0.030

Table A.8: Spatial Lag - h_p=0.05, k-data=10%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.555	0.587	0.531	0.308	0.030	0.493	0.531	0.510	0.295	0.020
0.05	0.10	0.550	0.588	0.530	0.330	0.046	0.489	0.533	0.505	0.318	0.036
0.05	0.15	0.547	0.587	0.534	0.354	0.070	0.492	0.533	0.515	0.335	0.061
0.05	0.20	0.550	0.591	0.533	0.378	0.095	0.492	0.532	0.510	0.366	0.083
0.05	0.25	0.548	0.579	0.530	0.401	0.122	0.489	0.531	0.508	0.382	0.125
0.10	0.05	0.641	0.680	0.627	0.436	0.084	0.589	0.642	0.614	0.421	0.063
0.10	0.10	0.639	0.674	0.635	0.405	**0.036**	0.597	0.643	0.611	0.394	**0.041**
0.10	0.15	0.642	0.681	0.628	0.423	0.051	0.598	0.641	0.599	0.409	0.056
0.10	0.20	0.637	0.679	0.625	0.455	0.085	0.592	0.638	0.610	0.441	0.096
0.10	0.25	0.648	0.670	0.626	0.478	0.121	0.596	0.641	0.611	0.461	0.138
0.15	0.05	0.601	0.621	0.579	0.409	0.080	0.524	0.570	0.552	0.388	0.080
0.15	0.10	0.595	0.624	0.579	0.386	0.039	0.539	0.580	0.553	0.360	0.038
0.15	0.15	0.594	0.626	0.575	0.364	0.027	0.530	0.571	0.556	0.349	0.031
0.15	0.20	0.605	0.631	0.583	0.376	0.031	0.535	0.571	0.552	0.358	0.035
0.15	0.25	0.594	0.627	0.573	0.406	0.057	0.531	0.570	0.556	0.374	0.067
0.20	0.05	0.543	0.563	0.520	0.383	0.086	0.484	0.508	0.498	0.342	0.068
0.20	0.10	0.547	0.555	0.524	0.343	0.039	0.481	0.508	0.503	0.314	0.033
0.20	0.15	0.544	0.560	0.519	0.325	0.034	0.486	0.514	0.496	0.305	0.031
0.20	0.20	0.545	0.555	0.517	0.318	0.033	0.484	0.510	0.497	0.297	0.029
0.20	0.25	0.537	0.559	0.518	0.322	0.033	0.481	0.509	0.507	0.301	0.035
0.25	0.05	0.488	0.506	0.472	0.360	0.083	0.420	0.457	0.446	0.315	0.075
0.25	0.10	0.490	0.503	0.466	0.321	0.048	0.420	0.461	0.446	0.292	0.038
0.25	0.15	0.494	0.503	0.471	0.312	0.032	0.417	0.456	0.438	0.280	0.035
0.25	0.20	0.489	0.505	0.473	0.304	0.028	0.418	0.454	0.440	0.273	0.032
0.25	0.25	0.486	0.508	0.465	0.299	0.025	0.419	0.462	0.443	0.267	0.025

Table A.9: Spatial Lag - h_p=0.05, k-data=15%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.366	0.382	0.361	0.179	0.016	0.331	0.360	0.344	0.175	0.014
0.05	0.10	0.367	0.385	0.370	0.204	0.030	0.328	0.359	0.343	0.199	0.025
0.05	0.15	0.373	0.386	0.363	0.210	0.049	0.332	0.357	0.343	0.208	0.046
0.05	0.20	0.369	0.391	0.356	0.229	0.055	0.330	0.360	0.338	0.219	0.056
0.05	0.25	0.369	0.389	0.358	0.240	0.075	0.331	0.357	0.342	0.230	0.062
0.10	0.05	0.469	0.499	0.452	0.267	0.045	0.421	0.449	0.451	0.275	0.049
0.10	0.10	0.467	0.496	0.451	0.237	0.021	0.419	0.452	0.444	0.241	0.021
0.10	0.15	0.466	0.496	0.457	0.256	0.034	0.425	0.456	0.449	0.255	0.031
0.10	0.20	0.467	0.501	0.452	0.271	0.043	0.418	0.451	0.448	0.265	0.049
0.10	0.25	0.472	0.505	0.451	0.278	0.066	0.420	0.452	0.452	0.278	0.070
0.15	0.05	0.516	0.533	0.501	0.348	0.078	0.456	0.495	0.488	0.337	0.089
0.15	0.10	0.517	0.536	0.503	0.296	0.036	0.459	0.493	0.493	0.298	0.035
0.15	0.15	0.513	0.536	0.507	0.284	**0.023**	0.458	0.493	0.491	0.288	**0.022**
0.15	0.20	0.517	0.538	0.499	0.289	0.029	0.460	0.496	0.499	0.299	0.026
0.15	0.25	0.519	0.541	0.504	0.307	0.039	0.458	0.502	0.494	0.302	0.039
0.20	0.05	0.476	0.506	0.459	0.346	0.087	0.425	0.460	0.458	0.310	0.085
0.20	0.10	0.477	0.502	0.462	0.291	0.043	0.422	0.465	0.458	0.281	0.033
0.20	0.15	0.479	0.505	0.462	0.279	0.023	0.422	0.456	0.459	0.264	0.022
0.20	0.20	0.478	0.510	0.467	0.276	0.026	0.425	0.460	0.456	0.263	0.024
0.20	0.25	0.476	0.506	0.465	0.278	0.031	0.417	0.458	0.460	0.270	0.025
0.25	0.05	0.444	0.459	0.430	0.324	0.092	0.383	0.419	0.404	0.303	0.100
0.25	0.10	0.452	0.462	0.430	0.295	0.047	0.386	0.418	0.416	0.272	0.042
0.25	0.15	0.444	0.464	0.433	0.272	0.029	0.387	0.416	0.409	0.253	0.028
0.25	0.20	0.450	0.457	0.432	0.266	0.028	0.385	0.418	0.409	0.251	0.023
0.25	0.25	0.450	0.470	0.430	0.264	0.021	0.381	0.418	0.415	0.238	0.018

Table A.10: Spatial Lag - h_p=0.30, k-data=5%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.879	0.899	0.866	0.653	**0.041**	0.819	0.865	0.846	0.647	**0.049**
0.05	0.10	0.876	0.901	0.869	0.684	0.085	0.818	0.868	0.848	0.669	0.100
0.05	0.15	0.877	0.902	0.867	0.726	0.159	0.820	0.864	0.847	0.703	0.185
0.05	0.20	0.880	0.902	0.870	0.755	0.221	0.821	0.868	0.851	0.724	0.246
0.05	0.25	0.874	0.901	0.869	0.775	0.270	0.821	0.863	0.839	0.739	0.298
0.10	0.05	0.751	0.781	0.738	0.556	0.059	0.687	0.739	0.709	0.537	0.050
0.10	0.10	0.753	0.776	0.736	0.529	0.034	0.692	0.744	0.712	0.505	0.037
0.10	0.15	0.751	0.776	0.731	0.557	0.064	0.688	0.740	0.710	0.528	0.064
0.10	0.20	0.753	0.776	0.736	0.597	0.123	0.688	0.739	0.713	0.570	0.140
0.10	0.25	0.753	0.779	0.739	0.622	0.175	0.681	0.748	0.710	0.594	0.192
0.15	0.05	0.649	0.672	0.628	0.479	0.051	0.571	0.624	0.597	0.451	0.062
0.15	0.10	0.644	0.670	0.630	0.464	0.039	0.574	0.624	0.595	0.425	0.035
0.15	0.15	0.650	0.669	0.630	0.432	0.030	0.579	0.631	0.592	0.408	0.031
0.15	0.20	0.653	0.663	0.629	0.447	0.040	0.582	0.622	0.600	0.426	0.035
0.15	0.25	0.646	0.667	0.630	0.478	0.073	0.576	0.619	0.594	0.450	0.089
0.20	0.05	0.578	0.575	0.549	0.406	0.055	0.490	0.527	0.514	0.379	0.053
0.20	0.10	0.580	0.573	0.550	0.364	0.041	0.495	0.526	0.516	0.361	0.031
0.20	0.15	0.575	0.580	0.551	0.354	0.032	0.497	0.532	0.519	0.335	0.027
0.20	0.20	0.577	0.576	0.554	0.346	0.030	0.505	0.528	0.517	0.321	0.023
0.20	0.25	0.581	0.579	0.554	0.359	0.030	0.502	0.521	0.512	0.335	0.029
0.25	0.05	0.500	0.510	0.481	0.360	0.053	0.436	0.461	0.453	0.339	0.053
0.25	0.10	0.501	0.511	0.483	0.332	0.044	0.428	0.452	0.447	0.301	0.038
0.25	0.15	0.497	0.508	0.489	0.309	0.032	0.429	0.456	0.450	0.293	0.034
0.25	0.20	0.498	0.504	0.485	0.310	0.023	0.435	0.456	0.447	0.298	0.025
0.25	0.25	0.500	0.508	0.485	0.302	0.018	0.433	0.463	0.450	0.283	0.022

Table A.11: Spatial Lag - h_p=0.30, k-data=10%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.543	0.585	0.531	0.316	0.019	0.492	0.530	0.511	0.308	0.020
0.05	0.10	0.552	0.583	0.523	0.334	0.029	0.492	0.533	0.508	0.319	0.028
0.05	0.15	0.553	0.590	0.536	0.353	0.053	0.491	0.536	0.511	0.343	0.051
0.05	0.20	0.547	0.587	0.530	0.381	0.081	0.497	0.532	0.512	0.375	0.079
0.05	0.25	0.556	0.584	0.535	0.404	0.118	0.487	0.532	0.510	0.393	0.121
0.10	0.05	0.642	0.680	0.628	0.453	0.072	0.600	0.645	0.616	0.439	0.059
0.10	0.10	0.638	0.670	0.625	0.411	**0.026**	0.600	0.637	0.606	0.408	**0.029**
0.10	0.15	0.642	0.676	0.626	0.434	0.038	0.598	0.637	0.608	0.416	0.043
0.10	0.20	0.641	0.679	0.625	0.454	0.068	0.597	0.639	0.607	0.442	0.081
0.10	0.25	0.639	0.676	0.628	0.485	0.109	0.598	0.640	0.606	0.465	0.119
0.15	0.05	0.597	0.620	0.581	0.419	0.066	0.533	0.575	0.552	0.391	0.071
0.15	0.10	0.591	0.618	0.578	0.393	0.027	0.536	0.573	0.554	0.367	0.026
0.15	0.15	0.602	0.625	0.575	0.372	0.021	0.529	0.569	0.561	0.354	0.023
0.15	0.20	0.598	0.625	0.583	0.384	0.025	0.540	0.573	0.553	0.359	0.026
0.15	0.25	0.603	0.624	0.579	0.408	0.050	0.527	0.576	0.546	0.386	0.056
0.20	0.05	0.548	0.557	0.521	0.369	0.067	0.484	0.509	0.500	0.340	0.065
0.20	0.10	0.546	0.557	0.525	0.344	0.025	0.478	0.511	0.499	0.321	0.028
0.20	0.15	0.543	0.559	0.523	0.326	0.021	0.484	0.510	0.495	0.304	0.021
0.20	0.20	0.547	0.561	0.524	0.320	0.022	0.478	0.511	0.500	0.302	0.019
0.20	0.25	0.544	0.555	0.516	0.324	0.024	0.487	0.509	0.502	0.308	0.025
0.25	0.05	0.488	0.501	0.467	0.358	0.072	0.418	0.457	0.446	0.317	0.066
0.25	0.10	0.492	0.504	0.472	0.330	0.029	0.426	0.458	0.441	0.298	0.031
0.25	0.15	0.490	0.504	0.476	0.313	0.027	0.426	0.463	0.444	0.283	0.025
0.25	0.20	0.488	0.509	0.466	0.303	0.015	0.423	0.453	0.441	0.274	0.022
0.25	0.25	0.479	0.500	0.466	0.304	0.017	0.428	0.460	0.444	0.269	0.020

Table A.12: Spatial Lag - h_p=0.30, k-data=15%

Moran's I empirical probability of rejections

k-test	k-RSA	linear					nonlinear				
		OLS	NP0	NP1	SNP0	SNP1	OLS	NP0	NP1	SNP0	SNP1
0.05	0.05	0.370	0.386	0.355	0.180	0.009	0.331	0.353	0.343	0.172	0.009
0.05	0.10	0.374	0.389	0.364	0.212	0.022	0.338	0.359	0.344	0.195	0.023
0.05	0.15	0.373	0.388	0.361	0.220	0.041	0.333	0.360	0.337	0.211	0.036
0.05	0.20	0.371	0.382	0.360	0.227	0.051	0.331	0.356	0.341	0.219	0.043
0.05	0.25	0.377	0.384	0.365	0.244	0.063	0.330	0.361	0.343	0.223	0.061
0.10	0.05	0.468	0.499	0.454	0.268	0.048	0.422	0.454	0.451	0.270	0.045
0.10	0.10	0.461	0.498	0.449	0.240	0.016	0.425	0.451	0.450	0.250	0.011
0.10	0.15	0.469	0.500	0.456	0.253	0.024	0.421	0.456	0.448	0.255	0.023
0.10	0.20	0.469	0.503	0.445	0.267	0.031	0.426	0.447	0.450	0.263	0.036
0.10	0.25	0.470	0.496	0.455	0.282	0.056	0.427	0.452	0.454	0.278	0.063
0.15	0.05	0.517	0.544	0.505	0.344	0.088	0.466	0.493	0.494	0.340	0.090
0.15	0.10	0.519	0.535	0.508	0.304	0.029	0.459	0.496	0.495	0.298	0.031
0.15	0.15	0.522	0.543	0.514	0.284	**0.020**	0.457	0.491	0.495	0.293	**0.017**
0.15	0.20	0.517	0.537	0.502	0.294	0.019	0.460	0.497	0.493	0.301	0.019
0.15	0.25	0.520	0.538	0.500	0.307	0.037	0.458	0.496	0.495	0.306	0.038
0.20	0.05	0.472	0.501	0.465	0.344	0.093	0.422	0.461	0.459	0.316	0.089
0.20	0.10	0.478	0.505	0.460	0.285	0.034	0.424	0.460	0.462	0.277	0.033
0.20	0.15	0.475	0.502	0.466	0.276	0.019	0.426	0.464	0.458	0.271	0.016
0.20	0.20	0.478	0.508	0.461	0.275	0.018	0.423	0.463	0.453	0.266	0.015
0.20	0.25	0.470	0.509	0.466	0.284	0.023	0.420	0.457	0.451	0.265	0.020
0.25	0.05	0.449	0.457	0.432	0.327	0.091	0.384	0.414	0.408	0.294	0.088
0.25	0.10	0.448	0.461	0.432	0.289	0.041	0.387	0.414	0.415	0.271	0.033
0.25	0.15	0.453	0.463	0.425	0.275	0.020	0.389	0.421	0.412	0.252	0.020
0.25	0.20	0.448	0.464	0.431	0.269	0.019	0.388	0.422	0.411	0.250	0.015
0.25	0.25	0.446	0.473	0.431	0.267	0.016	0.382	0.420	0.412	0.242	0.011