NONPARAMETRIC REGRESSION WITH A PARAMETRIC SPATIAL AUTOREGRESSIVE ERROR STRUCTURE

CARLOS MARTINS-FILHO AND CHRISTOPHER F. PARMETER

Abstract. Spatial models have received considerable attention in the last decade. At the same time, researchers have also started to embrace the flexibility afforded from nonparametric methods. However, methods that allow for nonparametric aspects of models whose errors exhibit spatial dependence have only recently been explored. We propose a fully nonparametric estimator of the regression function allowing for a fully parametric spatial autoregressive error structure. This paper develops the conditions necessary for efficient estimation and showcases the estimator in both a simulated and real world setting.

1. Introduction and Overview

Nonparametric econometric methods have witnessed an increased interest in the past decade (Li & Racine 2007). There is a distinct interest in estimating a model that, under appropriate smoothness assumptions, is robust to various forms of misspecification. The asymptotic properties of these estimators, however, depend on the error structure ascribed to the model. And while estimators have been developed to handle various error covariance structures, only recently has a method been proposed that develops the asymptotic theory for a nonparametric regression estimator under various error structures (Martins-Filho & Yao 2009). However, even with this breakthrough for nonparametric regression estimators, their theorems are not currently capable of handling error structures with spatial dependence.

Given the explosion of research, both theoretical and applied, involving spatial dependence, the creation of a nonparametric estimator that attains desirable asymptotic features under general, spatial dependence conditions on the error structure are warranted. In fact, broad asymptotic properties of kernel smoothing estimators in the presence of general spatial dependence structures have only recently been studied. While nonparametric and semiparametric estimators have been developed for spatial autoregressive (SAR) error structures (see Robinson, 2006; Su and Jin, 2007; and Su, 2008) a general set of spatial dependence conditions that lead to desirable asymptotic properties has yet to be fully explored. Here, we couch the SAR error structure of Cliff & Ord (1973) into the generalized framework of broad spatial mixing conditions to provide a set of asymptotic results for the local linear estimator in the style of Martins-Filho & Yao (2009).

Additionally, we provide a consistent and asymptotically efficient two-step estimator that corrects for the presence of general spatial dependence estimators of which the Cliff-Ord model, common in applied settings, satisfies the requisite conditions. Moreover, we also show that the use of

Date: June 1, 2009.

Preliminary and Very Incomplete, not to be quoted without Authors' Permission.

residuals from a nonparametric regression produces a consistent estimator of the spatial dependence parameter using the structure in Kelijian & Prucha (1999). This paper can be viewed as a general nonparametric treatment of nonparametric regression in the presence of spatial dependence. A key contribution to our work is the recent central limit theorem of Jenish & Prucha (2009) that allows for general α or ϕ-mixing spatial processes on random fields. This result is crucial for establishing asymptotic normality of the general local linear smoother as well as obtaining a consistent estimator of the spatial dependence parameter.

The conditions imposed on the spatial process by Jenish & Prucha (2009) are weak enough that the random field can be non-stationary[1] yet still be shown to be asymptotically normal. Moreover, the conditions allow for unevenly spaced locations within the field and for general forms of sampling regions as well as allowing the random variables to form triangular arrays. Taken together this set of conditions allow us to establish the asymptotic properties of local linear estimators under very general types of spatial dependence.

The two-step estimator of Martins-Filho & Yao (2009), inspired by Ruckstuhl, Welsh & Carroll (2000), can be seen as the extension of Generalized Least Squares (GLS) in traditional linear models to the nonparametric regression setting. While their focus was on models that exhibit various forms of time-series dependence, the insights afforded can be used when spatial dependence (parametrically specified) is present. This extension results in a improvement (in terms of efficiency) in the modeling of (un)conditional expectations. We (they) focus on the local linear estimator because this technique has desirable properties, design adaptability and mini-max efficiency for example (see Fan 1992, Fan & Gijbels 1995).[2]

To examine empirically the performance of the Kelijian & Prucha (1999) generalized method of moments (GMM) SAR estimator when the residuals passed to it stem from a nonparametric regression model, as well as the comparison between the naïve local linear estimator and our proposed two-step method, we use both simulated and actual data. Our simulation results reveal that our proposed approach can successfully estimate the unknown function and the spatial dependence parameter consistently. Our empirical example showcases the use of the method when multiple covariates are present.

The rest of the paper is organized as follows. Section 2 outlines a generic regression model with general spatial dependence and develops the necessary conditions to investigate the large sample properties of this estimator. Section 3 discusses an appropriate two-step local-linear estimator and lists additional assumptions necessary to establish the general asymptotic properties for this estimator. We also show that the SAR error structure of Cliff & Ord (1973) satisfies the spatial dependence conditions discussed in Section 2 and show that residuals from local linear estimation of our regression model can be used in the estimation procedure of Kelijian & Prucha (1999) to produce a consistent estimate of the spatial dependence parameter. Section 4 compares our estimator with

[1] In some settings the moments can be asymptotically unbounded as well.

[2] The local linear estimate also has lower bias the the more traditional local constant (Nadaraya-Watson) estimator and does not face the same boundary issues as this estimator.

several contemporary competitors and showcases its overall performance while Section 5 empirically implements the estimator in a hedonic setting to showcase the empirical merit of the estimator. Section 6 summarizes the findings of this paper and discusses avenues for future work. Technical arguments are provided in an appendix following Section 6.

2. The Local Linear Nonparametric Regression with General Parametric Spatial Dependent Error Structure

We have a sample of n observations for $\overrightarrow{y} = (Y_1, \ldots, Y_n)$, our regressand and $\overrightarrow{x} = (X_1, \ldots, X_n)$, our regressors. Our model is

$$Y_i = m(X_i) + U_i, \tag{1}$$

such that $E(U_i) = 0$ and $E(U_i U_j) = \omega_{ij}(\theta_0)$, where $\theta_0 \in \mathcal{R}^p$ and $p < \infty$. Our main objective is to estimate the regression function $m(x)$ over some grid of points in $\mathcal{R}^d$ where $d < n$. While there are numerous approaches to estimating $m(\cdot)$ (see Li & Racine 2007), we focus our attention on the local linear (LL) estimator popularized by Fan (1992) and Fan & Gijbels (1995).[3] Let $e' = (1, \mathbf{0})$, where $\mathbf{0}$ is a vector of zeros of length d. Additionally, let $1_n' = (1, \ldots, 1)$ be a vector of ones of length n and $h_n > 0$ a sequence of (possibly stochastic) bandwidths. The LL estimator is defined as

$$\check{m}(x) = e'(R_x' K_x R_x)^{-1} R_x' K_x \overrightarrow{y}, \tag{2}$$

where $R_x = (1_n, \overrightarrow{x} - 1_n x)$, $K_x = diag\left\{K_h(X_i - x)\right\}_{i=1}^n$, and $K_h(X_i - x) = h_n^{-1} K\left(h_n^{-1}(X_i - x)\right)$ is our kernel smoothing device. Standard choices for $K(\cdot)$ are the Gaussian or Epanechnikov kernels.[4] For notational simplicity we will rewrite our estimator in (2) as

$$\check{m}(x) = \sum_{i=1}^{n} A_i(x) Y_i, \tag{3}$$

where $A_i(x) = e' S_n^{-1}(x)(1, h_n^{-1}(X_i - x))' K_x$ and

$$S_n(x) = \begin{pmatrix} s_{n,0}(x) & s_{n,1}(x) \\ s_{n,1}(x) & s_{n,2}(x) \end{pmatrix}.$$

We define $s_{n,j}(x) = \sum_{i=1}^{n} h_n^{-j}(X_i - x)^j K_h(X_i - x)$. We establish the asymptotic normality of $\check{m}(x)$ under general spatial dependence following the approach of Martins-Filho & Yao (2009). This follows the traditional route of breaking the problem into two distinct pieces. First, the uniform convergence (in probability) of the components of $R_x' K_x R_x$ is established upon suitable renormalization. This follows directly from Theorem 1 in Martins-Filho & Yao (2009). Second, the asymptotic distribution of $R_x' K_x \overrightarrow{y}$ and subsequently our estimator is established, which, while similar to Theorem 2

[3] One can easily extend the insights here to the local constant estimator or the local polynomial estimator, see Masry (1996a,b).

[4] See Silverman (1986) for more on selection of kernels.

in Martins-Filho & Yao (2009), requires additional discussion given that we are considering general spatial mixing processes. As such it is stated here for completeness.

We provide a list of general assumptions that will be selectively adopted in the rest of the theorems in this paper and introduce some general notation for what follows. C is a generic constant that may take different values in $\mathcal{R}$ depending upon the setting while h_n is a sequence of bandwidths such that they are asymptotically diminishing ($h_n \to 0$) and decay slower than the sample size ($nh_n^2 \to \infty$) as $n \to \infty$.

Assumption 2.1.

(1) *Let $f_i(x)$ be the marginal density of X_i evaluated at x, with $f_i(x) < C$ $\forall i$ and x;*
(2) *$f_i^{(r)}(x)$ represents the r^{th} order derivative of $f_i(x)$ evaluated at x and $|f_i^{(1)}(x)| < C$;*
(3) *$|f_i(x) - f_i(x')| < C|x - x'|$ $\forall x$, x';*
(4) *$f_{lkijmo}(x_l, \ldots, x_o)$ denotes the joint density of $X_l, \ldots, X_o$ evaluated at $x_l, \ldots, x_o$ where $f_{lkijmo}(x_l, \ldots, x_o) < C$ $\forall$ $x_l, \ldots, x_o$;*
(5) *$\bar{f}_n(x) = n^{-1} \sum_{i=1}^{n} f_i(x) \to \bar{f}(x)$ as $n \to \infty$ where $0 < \bar{f}(x) < \infty$;*
(6) *As $n \to \infty$, $0 < inf_{x \in G} |\bar{f}(x)| < C$ for a compact set G.*

Assumption 2.2. *$K(x) : \mathcal{R} \to \mathcal{R}$ is a symmetric bounded function with bounded support S_K such that*

(1) *$\int K(x)dx = 1$;*
(2) *$\int xK(x)dx = 0$;*
(3) *$\int x^2 K(x)dx = \sigma_K^2$;*
(4) *$\forall x$, $x' \in S_K$ we have $|K(x) - K(x')| \leq C|x - x'|$.*

Assumption 2.3. *$\omega_{ij}(\theta_0)$ is the (i,j) element of $\Omega = E(UU'$ with $\omega_{ij}(\theta_0)| < C$ $\forall i,j$, $\bar{\omega}_n(\theta) = n^{-1} \sum_{i=1}^{n} \omega_{11}(\theta) \to \bar{\omega}(\theta)$ as $n \to \infty$ where $0 < \bar{\omega}(\theta) < \infty$ for every θ and $\bar{\omega}_{fn}(x, \theta) = n^{-1} \sum_{i=1}^{n} \omega_{ii} f_i(x) \to \bar{\omega}_f(x, \theta)$ as $n \to \infty$ where $0 < \bar{\omega}_f(x, \theta) < \infty$ for every x and θ.*

Before stating the remainder of our assumptions we discuss the spatial mixing processes we will be working with and the probability spaces in which they lie. Let the lattice $D \subset \mathcal{R}^w$ be infinite countable. Furthermore, let $\{X_{i,n}; i \in D_n, n \in \mathbb{N}\}$ be a triangular array of real random variables defined on a probability space $(\Lambda, \mathcal{F}, P)$ where D_n is a finite subset of D.[5] For $V \subseteq D_n$ and $W \subseteq D_n$, let $\sigma_n(V) = \sigma(X_{i,n}; \quad i \in V)$, the σ-algebra of the events $X_{i,n}$ that fall within the set V. Furthermore, define $\alpha_n(V, W) = \alpha(\sigma_n(V), \sigma_n(W))$ where

$$\alpha(\mathcal{V}, \mathcal{W}) = \sup(|P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{V}, B \in \mathcal{W}). \tag{4}$$

The α-mixing coefficients for the random field $\{X_{i,n}; i \in D_n, n \in \mathbb{N}\}$ is defined as:

$$\alpha_{k,l,n}(r) = \sup(\alpha_n(V, W), |V| \leq k, |W| \leq l, \rho(V, W) \geq r), \tag{5}$$

[5]We assume that our random field satisfies Assumption 1 of Jenish & Prucha (2009).

with k, l, r, $n \in \mathbb{N}$. In our definition $|W|$ denotes the cardinality of the set W and $\rho(V, W)$ measures the distance between sets. To explicitly note the link between the dependence of the random field on the size of the sample we also define

$$\bar{\alpha}_{k,l}(r) = \sup_n \quad \alpha_{k,l,n}(r). \tag{6}$$

Assumption 2.4. *$\{(X_i, U_i)'\}_{i=1,2,\ldots}$ is an α-mixing process whose mixing coefficients satisfy*

(1) $\lim_{k\to\infty} \sup_n \sup_{i\in D_n} \int_0^1 \alpha_{inv}^d(u) \left(\bar{Q}_{i,n}^{(k)}(u)\right)^2 du = 0$;

(2) $\sum_{m=1}^{\infty} m^{d-1}\bar{\alpha}_{k,l}(m) < \infty$ *for* $k + l \leq 4$;

(3) $\bar{\alpha}_{1,\infty}(m) = O(m^{-d-\varepsilon})$ *for some* $\varepsilon > 0$.

We denote the joint density of $(X_i, U_i)'$ by $f_{X_i,U_i}(x_i, u_i)$, the conditional density of X_i on U_i by $f_{X_i|U_i}(x_i)$ with $f_{X_i|U_i}(x) < C$ and the conditional density of X_i, X_j given U_i, U_j by $f_{X_iX_j|U_iU_j}(x_i, x_j)$ where $f_{X_iX_j|U_iU_j}(x_i, x_j) < C$ for all x_i, x_j.

Assumption 2.5.
There exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o(\sqrt{nh_n})$ such that $(nh_n^{-1})^{1/2}\bar{\alpha}_{k,l}(s_n) \to 0$ as $n \to \infty$.

Assumption 2.6. *The r^{th} order derivative of $m(x)$, $m^{(r)}(x)$ satisfies $m^{(r)}(x) < C$ for all x and $r = 1, 2$.*

Aside from our conditions on the elements of the variance-covariance matrix of the error terms in our regression and our specification of spatial dependence, our assumptions are identical to those of Martins-Filho & Yao (2009). Assumption 2.1 requires the density of the regressors to be well behaved in the sense of being bounded. When the covariates come from heterogenous distributions we require that the average of the densities must converge. If the covariates come from the same distribution or represent a spatially stationary sequence then this assumption is automatically satisfied. Assumption 2.2 is a standard set of conditions on the behavior of the smoothing function used for nonparametric kernel regression. Assumption 2.3 ensures that a weighted average of the diagonal elements of the error covariance converge as $n \to \infty$, which is trivially met in the Cliff-Ord setting where a homoscedastic error structure is present. Under the spatial α-mixing imposed in Assumptions 2.4 and 2.5 the behavior of the process is well-defined in the sense that the central limit theorem of Jenish & Prucha (2009) can be used but are sufficiently general to include interesting types of spatial dependence, Cliff-Ord for example. Assumption 2.6 is a standard smoothness condition on the unspecified function so that Taylor approximations can be used at various points in proofs of the theorems.

3. A Two-Step Estimator

While our LL estimator discussed in the previous section has teh desirable property of being $\sqrt{nh_n}$-asymptotically normal, the fact that the parametric structure of the error term is not being

used suggests that a more efficient estimator can be constructed when this knowledge is taken into account. While various approaches to incorporating covariance structures exist, a simple, yet promising, approach is to employ a two-step procedure that transforms the errors to have a spherical variance-covariance matrix. The motivation for this approach is very simple.

Let $\Omega(\theta_0)$ be an $n \times n$ matrix with (i,j) element given by $\omega_{ij}(\theta_0)$ and $P(\theta_0)$ an $n \times n$ matrix with (i,j) element $p_{ij}(\theta_0)$ such that $\Omega(\theta_0) = P(\theta_0)P'(\theta_0)$. Let $P^{-1}(\theta_0)$ be the inverse matrix of $P(\theta_0)$ with (i,j) element $v_{ij}(\theta_0)$. Let $\overrightarrow{m}' = (m(X_1), \ldots, m(X_n))$ and $\overrightarrow{U}' = (U_1, \ldots, U_n)$. Furthermore, we define I_n to be the $n \times n$ identity matrix and $Z = P^{-1}(\theta_0)\overrightarrow{y} + (I_n - P^{-1}(\theta_0))\overrightarrow{m}$. Then we have that

$$\begin{aligned} P^{-1}(\theta_0)\overrightarrow{y} =& P^{-1}(\theta_0)\overrightarrow{m} + P^{-1}(\theta_0)\overrightarrow{U} = P^{-1}(\theta_0)\overrightarrow{m} + \overrightarrow{m} - \overrightarrow{m} + \varepsilon \\ =& (P^{-1}(\theta_0) - I_n)\overrightarrow{m} + \overrightarrow{m} + \varepsilon \\ Z =& \overrightarrow{m} + \varepsilon. \end{aligned} \tag{7}$$

Given that the components of the stochastic process $\{U_i\}_{i=1,2,\ldots}$ can be written as a linear combination of ε_j, $U_i = \sum_{j=1}^{q} p_{ij}\varepsilon_j$ where $q = 1, 2, \ldots, n$ where $\{\varepsilon_i\}_{i=1,2,\ldots,}$ is an independently and identically distributed process with mean zero and variance σ^2 the model in (7) is the standard nonparametric regression model with spherical errors. The difficulty with estimating the regression in (7) is that it is infeasible in the sense that our regressand Z is unobserved given the presence of the unknown parameters θ_0 and the regression function itself, $\overrightarrow{m}$. To make this setup feasible the unknowns must be substituted with suitable estimates. Implementation in this stage requires a preliminary first stage estimate of $\overrightarrow{m}$ be used, such as $\check{m}(x)$ and the resulting residuals, $\check{U}_i = Y_i - \check{m}(X_i)$ are used to estimated the elements of $P^{-1}(\theta_0)$, say $P^{-1}(\check{\theta})$. These estimates can then be used to construct a feasible estimate of Z, which we call $\check{Z}$, to make the connection with our use of $\check{m}$ explicit.

We will establish the asymptotic normality of the feasible estimator

$$\hat{m}(x) = e'(R_x'K_xR_x)^{-1}R_x'K_x\check{Z}. \tag{8}$$

This property is achieved in two stages. First, we establish the asymptotic normality of

$$\dot{m}(x) = e'(R_x'K_xR_x)^{-1}R_x'K_x\dot{Z}, \tag{9}$$

where $\dot{Z} = P^{-1}(\theta_0)\overrightarrow{y} + (I_n - P^{-1}(\theta_0))\check{m}$, the infeasible version of Z where θ_0 is not estimated. Second, we establish the asymptotic equivalence between $\hat{m}(x)$ and $\dot{m}(x)$, i.e., $\sqrt{ng_n}(\hat{m}(x) - \dot{m}(x)) = o_p(1)$.

Our use of two nonparametric regressions requires selection of two bandwidth sequences, h_n used in the first stage and g_n used in the second stage. As is typical, we will require that we undersmooth in the first stage, $h_n/g_n \to 0$ so that the bias from the first stage is smaller than the leading bias term present in the second stage. We will require that g_n has the standard optimal rate properties of a single stage bandwidth, i.e., $g_n = O(n^{-1/(4+d)})$. Additionally, in establishing the asymptotic normality of our two-step estimator we will see that the asymptotic variance of this estimator is

smaller than the asymptotic variance of the naïve estimator. Intuitively this is consistent with the fact that we have explicitly incorporated knowledge of the covariance structure of our error structure. This is akin to Aitken's Theorem showing that the FGLS estimator is more efficient than the OLS estimator.

An interesting application of the estimator described above is to regression models where the error structure is of the Cliff-Ord type. Thus, the common spatial autoregressive model given as

$$Y_i = m(X_i) + U_i, \quad U_i = \rho M\overrightarrow{U} + \varepsilon_i, \tag{10}$$

where M is an $n \times n$ matrix with zeros along the diagonal, $|\rho| < 1$ and ε is an *iid* process, satisfies the requisite assumptions of Jenish & Prucha's (2009) central limit theorem for spatial processes and can be estimated via our two stage approach. A simple and tractable procedure to estimate ρ was provided in Kelijian & Prucha (1999) and we show that using residuals obtained via a nonparametric regression produces a consistent estimate for ρ.

4. Simulation Study

In this section we perform a series of Monte Carlo simulations to assess the performance of our two-step estimator, henceforth referred to as 2SLL, with respect to several competing estimators. We investigate 4 function specifications for $m(x)$:

(1) $m(x) = 2 + \sin(1.5x)$,
(2) $m(x) = 2 + \frac{e^{-3x}}{1+e^{-3x}}$,
(3) $m(x) = x + 2e^{-16x^2}$,
(4) $m(x) = 1.18 + 1.56x - 2.31x^2 - 0.39x^3 + 0.56x^4$.

All of our specifications for $m(\cdot)$ are nonlinear in x and twice differentiable. Figure 1 presents the shape of our functional forms. We use sample sizes of $n = 100$, 200 and 400 with 1000 replications per experiment. Our covariate is generated from $\mathcal{U}[-2, 2]$ and our error term is normally distributed with mean 0 and variance 0.1.

To introduce spatial dependence into our data we use "J ahead and J behind" specifications for P and set $J = 1$, 3, and 5. This specifies a circular world where ε_1 is dependent on ε_2 and ε_n when $J = 1$ and similarly ε_2 to ε_1 and ε_3. Furthermore, we specified P such that all nonzero elements of P are equal and all rows sum to unity. Thus, in the cases considered here each row of P has $2J$ nonzero elements which are all equal to $\frac{1}{2J}$. We allow ρ to be 0.1, 0.5 and 0.9 representing low, moderate and high levels of spatial dependence. This gives us a total of 36 experiments ($3 \times 3 \times 4$).

As a basis for comparison, we also estimate the naïve local linear estimator (LL), a misspecified linear model (L-IC) and a correctly specified nonlinear model (NL).[6] Both L-IC and NL are estimated in two stages accounting for spatial dependence following Kelijian & Prucha (1999). We compare the bias, standard deviation and root mean square error of ρ across LL, L-IC and NL on

[6]For the specification in (iv) we simply add higher order polynomial terms of x and estimate the model via ordinary least squares.

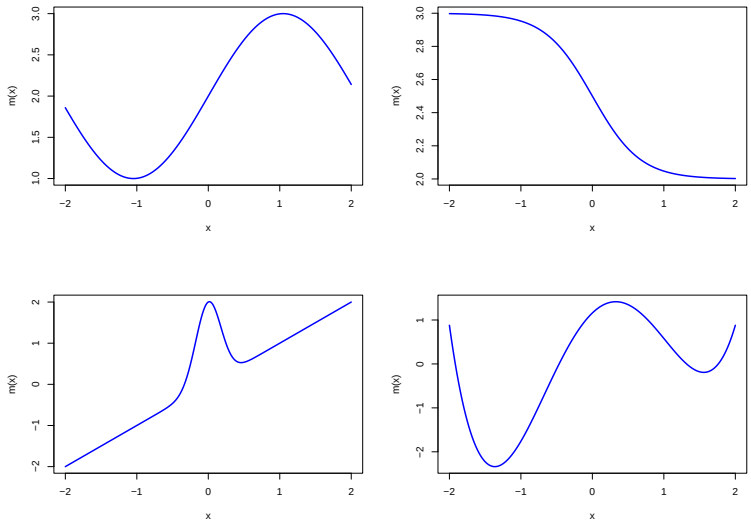

FIGURE 1. Plots of functional forms used in simulations.

the first stage estimation.[7] Additionally, on a grid of 50 equally spaced points on the inner 90% of the support of our data we obtain the average bias, standard deviation and root mean square error of our four econometric estimators.

For all simulations involving a nonparametric estimator we use the Epanechnikov kernel with bandwidths selected via Hurvich, Simonoff & Tsai's (1998) AIC_c approach.[8] For 2SLL we need to ensure that our first stage bandwidth is undersmoothed. A strategy for this purpose using data-driven methods focuses on estimation of the scale factor. A 'scale factor' refers to the unknown constant c in the formula for the optimal bandwidth in local linear regression, $h_n = cn^{-1/5}$. Cross-validation can be thought of as a method for estimating the unknown constant c, where c is independent of the sample size n.[9] This constant can then be *rescaled* for the appropriate rate necessary for the theoretical underpinnings of our model. Thus, while our bandwidth in the second stage needs to be $O(n^{-1/5})$, our bandwidth in the first stage can be $O(n^{-1/6})$.

[7]For this set of simulations we use an optimal bandwidth for LL instead of the undersmoothed bandwidth necessary to eliminate the bias in 2SLL.
[8]We use 5 'multistarts' to search for the optimal bandwidths.
[9]See Racine (1993).

5. An Application to Hedonic Modeling

6. Conclusions and Directions for Future Research

This paper has considered the asymptotic properties of the local linear nonparametric regression estimator under general spatial dependence structures. First, the general properties of the traditional nonparametric estimator were established. Second, a feasible two-stage estimator that is more efficient was suggested and studied. Additionally, we couched a common applied spatial model within our framework and established the consistency of the GMM estimation procedure commonly used in more traditional parametric settings. Overall we have provided a broad treatment of nonparametric models with general spatial dependence. Our asymptotic analysis is also highlighted by the use of recently developed central limit theorems for arrays of random fields that is novel in this setting. We provided both simulated and empirical evidence that the two stage method is feasible and easily applied. The simplicity of the two stage method allows it to be used with little to no additional cost with existing software that implements nonparametric regression and bandwidth selection.

Additional work can focus on two main areas, extending this estimator to other spatial settings such as a spatial lag or a joint spatial lag, spatial autocorrelation setup and the study of the properties of data-driven bandwidth selection procedures in the face of spatial dependence. Here it is not known if the traditional cross-validation procedures provide the same properties as when the methods are applied in more traditional time-series dependence settings. Follow up work to determine the best way to select the bandwidths in both stages is thus warranted.

References

Cliff, A. & Ord, J. (1973), *Spatial Autocorrelation*, Pion, London.

Fan, J. (1992), 'Design adaptive nonparametric regression', *Journal of the American Statistical Association* **87**, 998–1004.

Fan, J. & Gijbels, I. (1995), 'Data driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation', *Journal of the Royal Statistical Society, Series B* **57**, 371–394.

Hurvich, C. M., Simonoff, J. S. & Tsai, C. L. (1998), 'Smoohting parameter selection in nonparametric regression using an improved Akaike information criterion', *Journal of the Royal Statistical Society Series B* **60**(2), 271–293.

Jenish, N. & Prucha, I. R. (2009), 'Central limit theorems and uniform laws of large numbers for arrays of random fields', *Journal of Econometrics* **150**(1), 86–98.

Kelijian, H. J. & Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* **40**(2), 509–533.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Martins-Filho, C. & Yao, F. (2009), 'Nonparametric regression estimation with general parametric error covariance', *Journal of Multivariate Analysis* **100**(3), 309–333.

Racine, J. S. (1993), 'An efficient cross-validation algorithm for window width selection for nonparametric kernel regression', *Communications in Statistics* **22**(4), 1107–1114.

Ruckstuhl, A. F., Welsh, A. H. & Carroll, R. J. (2000), 'Nonparametric function estimation of the relationship between two repeatedly measured variables', *Statistica Sinica* **10**, 51–71.

TABLE 1. Average Bias for Kelijian & Prucha's (1999) Estimator of ρ. 1000 Monte Carlo Repetitions

		$m(x) = 2 + \sin(1.5x)$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC	0.025	0.082	0.130	0.074	0.096	0.122	0.066	0.056	0.063
	NL	0.014	0.073	0.119	0.026	0.053	0.083	0.053	0.038	0.043
$n = 100$	LL-OPT	0.018	0.072	0.118	0.051	0.074	0.101	0.064	0.048	0.054
	LL-CV	0.017	0.073	0.120	0.048	0.067	0.095	0.086	0.057	0.058
	LL-UOPT	0.019	0.072	0.118	0.055	0.078	0.105	0.063	0.049	0.055
	L-IC	0.022	0.041	0.073	0.064	0.064	0.081	0.051	0.037	0.039
	NL	0.008	0.027	0.059	0.011	0.023	0.043	0.028	0.018	0.019
$n = 200$	LL-OPT	0.013	0.031	0.062	0.032	0.039	0.058	0.042	0.027	0.027
	LL-CV	0.010	0.028	0.060	0.023	0.031	0.050	0.055	0.031	0.029
	LL-UOPT	0.015	0.033	0.064	0.039	0.044	0.062	0.043	0.029	0.030
	L-IC	0.017	0.026	0.041	0.064	0.054	0.058	0.044	0.028	0.029
	NL	0.003	0.014	0.028	0.008	0.013	0.019	0.015	0.009	0.009
$n = 400$	LL-OPT	0.007	0.017	0.032	0.025	0.024	0.030	0.026	0.015	0.015
	LL-CV	0.005	0.015	0.029	0.015	0.017	0.023	0.036	0.016	0.014
	LL-UOPT	0.009	0.018	0.033	0.033	0.030	0.036	0.029	0.017	0.017
		$m(x) = 2 + \frac{e^{-3x}}{1+e^{-3x}}$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC									
	NL									
$n = 100$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 200$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 400$	LL-OPT									
	LL-OPT									
	LL-OPT									

TABLE 2. Average Bias for Kelijian & Prucha's (1999) Estimator of ρ. 1000 Monte Carlo Repetitions

		$m(x) = x + 2e^{-16x^2}$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC									
	NL									
$n = 100$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 200$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 400$	LL-OPT									
	LL-OPT									
	LL-OPT									
		$m(x) = 1.18 + 1.56x - 2.31x^2 - 0.39x^3 + 0.56x^4$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC	0.047	0.095	0.152	0.147	0.155	0.190	0.113	0.094	0.102
	NL	0.015	0.073	0.122	0.037	0.060	0.087	0.092	0.052	0.053
$n = 100$	LL-OPT	-0.001	0.008	0.005	0.032	0.045	0.058	0.063	0.043	0.047
	LL-CV	0.012	0.052	0.082	0.043	0.060	0.083	0.090	0.056	0.056
	LL-UOPT	-0.004	-0.007	-0.020	0.033	0.043	0.053	0.060	0.043	0.048
	L-IC	0.037	0.059	0.085	0.140	0.125	0.138	0.099	0.072	0.072
	NL	0.009	0.027	0.059	0.017	0.026	0.046	0.050	0.026	0.024
$n = 200$	LL-OPT	-0.003	-0.016	-0.017	0.017	0.018	0.029	0.037	0.022	0.023
	LL-CV	0.007	0.017	0.042	0.022	0.027	0.045	0.054	0.029	0.027
	LL-UOPT	-0.007	-0.030	-0.042	0.019	0.017	0.024	0.036	0.023	0.023
	L-IC	0.034	0.044	0.057	0.136	0.115	0.117	0.092	0.062	0.060
	NL	0.004	0.015	0.029	0.011	0.015	0.020	0.028	0.013	0.011
$n = 400$	LL-OPT	-0.005	-0.015	-0.023	0.011	0.010	0.011	0.023	0.012	0.011
	LL-CV	0.003	0.010	0.020	0.014	0.015	0.021	0.032	0.015	0.013
	LL-UOPT	-0.009	-0.029	-0.046	0.013	0.009	0.007	0.022	0.012	0.011

TABLE 3. Average Mean Square Error for Kelijian & Prucha's (1999) Estimator of ρ. 1000 Monte Carlo Repetitions

		$m(x) = 2 + \sin(1.5x)$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC	0.011	0.042	0.080	0.013	0.028	0.048	0.006	0.006	0.010
	NL	0.010	0.039	0.083	0.008	0.020	0.036	0.005	0.004	0.007
$n = 100$	LL-OPT	0.010	0.039	0.075	0.010	0.023	0.041	0.006	0.005	0.008
	LL-CV	0.010	0.040	0.078	0.010	0.022	0.039	0.012	0.007	0.009
	LL-UOPT	0.010	0.039	0.075	0.011	0.024	0.042	0.006	0.006	0.009
	L-IC	0.005	0.018	0.033	0.008	0.013	0.021	0.003	0.003	0.004
	NL	0.005	0.016	0.030	0.003	0.008	0.015	0.002	0.001	0.002
$n = 200$	LL-OPT	0.005	0.017	0.030	0.004	0.010	0.017	0.003	0.002	0.003
	LL-CV	0.005	0.016	0.030	0.004	0.009	0.016	0.005	0.002	0.003
	LL-UOPT	0.005	0.017	0.031	0.005	0.010	0.018	0.002	0.002	0.003
	L-IC	0.003	0.008	0.015	0.006	0.006	0.009	0.002	0.001	0.002
	NL	0.002	0.007	0.014	0.002	0.003	0.006	0.001	0.000	0.000
$n = 400$	LL-OPT	0.002	0.007	0.014	0.002	0.004	0.006	0.001	0.001	0.001
	LL-CV	0.002	0.007	0.014	0.002	0.004	0.006	0.002	0.001	0.001
	LL-UOPT	0.002	0.007	0.014	0.002	0.004	0.007	0.001	0.001	0.001
		$m(x) = 2 + \frac{e^{-3x}}{1+e^{-3x}}$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC									
	NL									
$n = 100$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 200$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 400$	LL-OPT									
	LL-OPT									
	LL-OPT									

TABLE 4. Average Mean Square Error for Kelijian & Prucha's (1999) Estimator of ρ. 1000 Monte Carlo Repetitions

		$m(x) = x + 2e^{-16x^2}$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC									
	NL									
$n = 100$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 200$	LL-OPT									
	LL-OPT									
	LL-OPT									
	L-IC									
	NL									
$n = 400$	LL-OPT									
	LL-OPT									
	LL-OPT									
		$m(x) = 1.18 + 1.56x - 2.31x^2 - 0.39x^3 + 0.56x^4$								
		$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
		J=1	J=3	J=5	J=1	J=3	J=5	J=1	J=3	J=5
	L-IC	0.012	0.044	0.094	0.030	0.045	0.074	0.015	0.013	0.019
	NL	0.010	0.040	0.082	0.009	0.021	0.036	0.012	0.006	0.008
$n = 100$	LL-OPT	0.009	0.024	0.036	0.008	0.017	0.025	0.006	0.005	0.007
	LL-CV	0.010	0.033	0.060	0.009	0.020	0.035	0.012	0.006	0.008
	LL-UOPT	0.009	0.022	0.032	0.008	0.016	0.023	0.006	0.005	0.007
	L-IC	0.006	0.019	0.035	0.023	0.025	0.035	0.011	0.007	0.008
	NL	0.005	0.016	0.029	0.004	0.008	0.015	0.004	0.002	0.002
$n = 200$	LL-OPT	0.005	0.012	0.018	0.003	0.007	0.012	0.002	0.001	0.002
	LL-CV	0.005	0.015	0.026	0.004	0.008	0.015	0.005	0.002	0.003
	LL-UOPT	0.005	0.012	0.017	0.003	0.007	0.011	0.002	0.001	0.002
	L-IC	0.003	0.009	0.016	0.020	0.017	0.021	0.009	0.005	0.004
	NL	0.002	0.006	0.014	0.002	0.003	0.006	0.001	0.001	0.001
$n = 400$	LL-OPT	0.002	0.006	0.010	0.002	0.003	0.005	0.001	0.001	0.001
	LL-CV	0.002	0.007	0.013	0.002	0.003	0.006	0.002	0.001	0.001
	LL-UOPT	0.002	0.006	0.010	0.002	0.003	0.005	0.001	0.001	0.001