张征宇
(2011 年 7 月)

联系方式
上海社会科学院数量经济研究中心
地址：上海市淮海中路622弄7号517室
邮编：200020
电话：021-53060606-2517
Email：zyzhang@sass.org.cn

个人简介：
2009 年 6 月毕业于上海财经大学经济学院，获经济学博士学位；现任上海社会科学院数量经济研究中心助理研究员，研究兴趣为理论与应用空间计量经济学(spatial econometrics)

发表论文
1．朱平芳、张征宇、姜国麟，2011，FDI与环境规制：基于地方分权视角的实证研究，《经济研究》第6期。
2．张征宇、朱平芳，2010，地方环境支出的实证研究，《经济研究》第5期。
3．Zhang Zhengyu and Zhu Pingfang, 2010，A more efficient best three-stage least squares estimator for spatial autoregressive models, Annals of Economics and Finance, vol. 11, 153-182.

A Pairwise Difference Estimator for Partially Linear Spatial Autoregressive Models

Zhengyu Zhang

Center for Econometric Study, Shanghai Academy of Social Sciences

Email: zyzhang@sass.org.cn

Su and Jin (2010) develop for partially linear spatial autoregressive (PL-SAR) model a profile quasi-maximum likelihood based estimation procedure. More recently, Su (2011) proposes for this model a semiparametric GMM estimator. However, both of them can be computationally challenging for applied researchers and are not easy to implement in practice. In this article, we propose a computationally simple estimator for the PL-SAR model in the presence of either heteroscedastic or spatially correlated error terms. This estimator blends the essential features of both the GMM estimator for linear SAR model and the pairwise difference estimator for conventional partially linear model. Limiting distribution of the proposed estimator is established and consistent estimator for its asymptotic CV matrix is provided. Monte Carlo studies indicate that our estimator is attractive particularly when one is interested in estimating the finite-dimensional parameters in the model.

JEL Classification: C13; C14; C21;

Keywords: Spatial autoregression; Partially linear model; Pairwise difference;

1. INTRODUCTION

In this article, we consider semi-parametric estimation of a class of partially linear spatial autoregressive (PL-SAR) models, specified as

$$y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} y_j + x_{1i}' \beta_0 + m(x_{2i}) + u_i, i = 1, \cdots, n. \tag{1}$$

In model (1), n is the number of total cross sectional units; y_i is the outcome variable for i; λ_0 is the SAR parameter; $[w_{ij}]_{i,j=1,\cdots,n;i\neq j}$'s are termed as the spatial weights, which determine the structure of neighborliness among cross sectional units; $x_{li}, l = 1, 2$, is the p_{x_l} -dimensional column vector of exogenous regressors (excluding a constant term); β_0 is a p_{x_1}-dimensional vector of slope coefficients; [1] $m(\cdot)$ is a unknown function mapping from $\mathbb{R}^{p_{x_2}}$ to the real line; u_i's are zero mean error terms that are not necessarily either identically or independently distributed and allowed to exhibit a general form of spatial dependence or heteroscedasticity.[2]

In the case $m(\cdot) \equiv 0$, model (1) is simplified to a conventional linear SAR model, which has already found its widespread application in empirical analysis related to urban, environmental economics, industrial organization and regional convergence. Meanwhile, a number of methods have been proposed to estimate the parametrically specified SAR model, including the method of moments by Kelejian and Prucha (1999, 2010), the method of quasi-maximum likelihood estimation by Lee (2004), the method of two-stage least squares by Kelejian and Prucha (1998), Lee (2003), Zhang and Zhu (2010) and the generalized method of moments by Lee (2007), Lin and Lee (2010), and Liu et al. (2010). Although the SAR model and its various parametric forms have attracted a great deal of attention both theoretically and empirically, until just recently, researchers have started addressing the importance of nonparametric modeling in spatial econometrics (Baltagi and Li 2001; Pace et al. 2004; Yang et al. 2006). Su and Jin (2010) provide a detailed discussion on motivations for including both nonlinearity and nonparametric component in specifying a SAR model. In comparison with its parametric version, PL-SAR model has a more flexible functional form, thus effectively lessening the possibility of either inconsistent estimate or misleading statistical inference due to mis-specification.[3]

Several estimation procedures have been developed for model (1). Su and Jin (2010) propose a profiled quasi-maximum likelihood based estimator (PQMLE) and demonstrate that the rates of consistency for the finite-dimensional parameters in the model depend on some general features of the spatial weighting schemes. A major drawback for Su and Jin (2010)'s PQMLE seems to be that the error term u_i's in their model are required to be both homoscedastic and independently distributed. In the presence of heteroscedasticity, Lin and Lee (2010) have demonstrated that the

[1] Throughout the paper, any parameter with subscript zero represents the true parameter that generates the data.

[2] See Assumption 2 in Section 3.

[3] Other theoretical studies for regression models with partially linear structure include Robinson (1988), Chen and Khan (2001), Lee (2003) and Sun (2005), to mention a few.

QML-based estimator is usually inconsistent. More recently, Su (2011) proposes for this PL-SAR model a semiparametric GMM estimator (SPGMME). His estimation strategy contains two stages. The first stage treats the finite-dimensional parameters as given, and uses local instruments to estimate $m(\cdot)$ locally as a function of these finite-dimensional parameters. In the second stage, he uses global instruments to estimate the finite-dimensional parameters by profiling out the nonparametric component, and then recovers the estimate of the nonparametric component. As emphasized by Su (2011), SPGMME has several advantages relative to the earlier PQMLE: First, SPGMME achieves usual $\sqrt{n}$ consistency rate for the parametric component in the model whereas the consistency rate of PQMLE depends on some general features of the spatial weight matrix. Second, SPGMME is robust against both heteroscedasticity of unknown form and certain form of spatial dependence in the error terms. Third, the procedure of SPGMME can be readily extended to semi-parametric SAR panel data models or semiparametric models with endogeneity.

However, both estimators mentioned above seem to be computationally challenging for applied researchers and are not easy to implement in practice. In this article, an alternative computationally simple procedure is proposed for consistently estimating PL-SAR model (1), particularly, for simple estimation of the finite-dimensional parameters in the model. We build our estimator on a type of pairwise difference estimator (PDE) that involves pairwise differences of observations for which the regressors in the nonparametric component of the regression function are approximately equal. See Ahn and Powell (1993), Honore and Powell (2005, 2007) and Chen and Zhou (2005) for applications of PDE in various nonlinear models and microeconometric models.

Compared with other applications of PDE, the setting of our SAR model complicates the econometric analysis in two ways. First, for any SAR model, the spatially lagged variable $\overline{y}_i = \sum_{j\neq i}^{n} w_{ij}y_j$ is generally correlated with the error term, which precludes any OLS-based estimation procedures. In this position, our suggested PDE for PL-SAR model should combine the essential features of both the GMM estimator for linear SAR model and the pairwise difference estimator for conventional partially linear model. Second, for many applications of PDE, say, in the field of microeconometrics, observations are assumed to follow an i.i.d. sampling scheme. By contrast, for SAR model the observed variables, due to the presence of spatial dependence, are unlikely to be independently distributed across i, thus some technical modification is needed to adapt the asymptotic analysis that was developed for conventional PDE to this spatial context.

It is also instructive to compare our PDE procedure with Su (2011)'s SPGMME. It can be argued that our PDE shares all the above-mentioned advantages (closed form expression, $\sqrt{n}$-convergence, robustness against heteroscedasticity and correlated error terms, extendibility) of Su (2011)'s SPGMME relative to PQMLE. But while Su's SPGMME "profiles" out the nonparametric component as a function of the exogenous regressors before applying instrumental variables (IV) to estimate the finite-dimensional parameters, our PDE "eliminates" the nonparametric component by pairwise differencing, and then applies IV method to estimate the finite-dimensional parameters. By "eliminating" instead of "profiling out" the nonparametric component, our estimator seems to be computationally simpler than SPGMME particularly when one is interested in estimating the finite-dimensional parameters in the model.

The rest of the paper is organized as follows: Section 2 introduces the pairwise difference estimator and its large sample properties are established in Section 3. Section 4 reports some Monte Carlo results and Section 5 concludes. Technical details are collected in an Appendix.

2. THE ESTIMATOR

The pairwise difference estimator is based on a comparison of pairs of observations for which the regressors in the nonparametric component of the regression function, i.e., x_2, are close to each other. In an ideal situation, for two indices i and j, satisfying $x_{2i} = x_{2j}$, differencing the corresponding dependent variable y_i and y_j gives

$$\Delta_{ij}^{y} = y_i - y_j = \lambda_0(W_{ni} - W_{nj})Y_n + (x_{1i} - x_{1j})'\beta_0 + u_i - u_j, \tag{2}$$

where $Y_n = (y_1, \cdots, y_n)'$ and W_{ni} is the i-th row of W_n. Observe that for (2), the spatial lag term $(W_{ni} - W_{nj})Y_n$ is generally correlated with the error term $u_i - u_j$, which implies that any OLS-based estimator will be inconsistent. Now consider GMM estimation based on orthogonality of some instruments and the error terms. Let q_{ij}, $i, j = 1, \cdots, n$, $i < j$, be a sequence of well chosen $p_q(\geq p_{x_1} + 1)$-dimensional non-stochastic instruments, such that we have the moment condition

$$\mathbb{E}(q_{ij}(u_i - u_j)) = 0. \tag{3}$$

Bearing the condition $x_{2i} = x_{2j}$ in mind, the empirical analogue to the moment condition (3) for the whole sample can be written as

$$\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \mathbf{1}(x_{2i} = x_{2j}) \mathbb{E}(q_{ij} \Delta_{ij}^{u}(\theta_0)) = 0, \tag{4}$$

where $\theta = (\lambda, \beta')'$, $\Delta_{ij}^{u}(\theta) = \Delta_{ij}^{y} - \lambda(W_{ni} - W_{nj})Y_n - (x_{1i} - x_{1j})'\beta$. When $x_{2i} = x_{2j}$, $\Delta_{ij}^{u}(\theta_0) = \Delta_{ij}^{u} = u_i - u_j$. Then the GMM estimator of θ based on the moment condition (4) can be viewed as the solution to the following minimization problem,

$$\min_{\theta} \left[\sum_{i=1}^{n} \sum_{j>i}^{n} \mathbf{1}(x_{2i} = x_{2j}) q_{ij} \Delta_{ij}^{u}(\theta) \right]' A_n \left[\sum_{i=1}^{n} \sum_{j>i}^{n} \mathbf{1}(x_{2i} = x_{2j}) q_{ij} \Delta_{ij}^{u}(\theta) \right], \tag{5}$$

where A_n is some $k_q \times k_q$ nonnegative definite matrix and is assumed to converge to a constant matrix A. This design corresponds to Hansen (1982)'s GMM setting, which can be used to illustrate the optimal weighting issue.

However, the estimation scheme (5) can not be directly implemented. When x_{2i} is continuously distributed, there are no two such indices i, j that $x_{2i} = x_{2j}$ with probability one. A usual manner to cope with this issue is to assign each pair with indices (i, j) a weight, which declines to zero as the magnitude of the difference $\|x_{2i} - x_{2j}\|$ increases. This idea can be straightforwardly translated into replacing the indicator function $\mathbf{1}(\cdot)$ in (5) with its smoothed counterpart, a symmetric kernel function. To formally introduce our estimator, let $k(\cdot)$ be a univariate symmetric kernel satisfying $\int k(v)dv = 1$ and $k(v) = k(-v)$. $h_n = (h_{1n}, \cdots, h_{p_{x_2}n})$ is a vector of smoothing parameters. For a p_{x_2}-dimensional vector $s = (s_1, \cdots, s_{p_{x_2}})$ with all its components being absolute continuous with Lebesgue measure, define $k_{h_l}(\cdot) = h_{ln}^{-1} k(\cdot/h_{ln})$, $K_h(s) = \prod_{l=1}^{p_{x_2}} k_{h_l}(s_l)$ and $\mathbf{K}_n^{x_2}$ to be the $[n(n-1)/2] \times [n(n-1)/2]$ square matrix diag$\{K_h(x_{2i} - x_{2j})\}_{i,j=1,\cdots,n,i<j}$. For example, when $n = 3$, $\mathbf{K}_n^{x_2} = \text{diag}\{K_h(x_{21} - x_{22}), K_h(x_{21} - x_{23}), K_h(x_{22} - x_{23})\}$. Furthermore, denote $\mathbf{\Delta}_n^u(\theta)$ to be the $n(n-1)/2$-dimensional column vector of $\Delta_{ij}^u(\theta)$'s, $i, j = 1, \cdots, n, i < j$. For example, when $n = 3$, $\mathbf{\Delta}_n^u = (\Delta_{12}^u(\theta), \Delta_{13}^u(\theta), \Delta_{23}^u(\theta))'$. Similarly, let $\mathbf{Q}_n$ be the $(n(n-1)/2) \times p_q$ matrix by stacking the row vector q'_{ij}'s. With these notations, our pairwise difference estimator is the solution to the following minimization problem,

$$\widehat{\theta}_n = \arg\min_{\theta} \left[\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^u(\theta)\right]' A_n \left[\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^u(\theta)\right]. \tag{6}$$

Given $\mathbf{\Delta}_n^u(\theta)$ being linear in θ, $\widehat{\theta}_n$ has the explicit representation,

$$\widehat{\theta}_n = \left[\mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n A_n \mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{Z}_n\right]^{-1} \mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n A_n \mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^y, \tag{7}$$

where $\mathbf{\Delta}_n^y$ is the $n(n-1)/2$-dimensional column vector of $y_i - y_j$'s, $i, j = 1, \cdots, n, i < j$, $\mathbf{Z}_n = [\mathbf{\Delta}_n^w \cdot Y_n, \mathbf{\Delta}_n^{x_1}]$, $\mathbf{\Delta}_n^w$ is the $(n(n-1)/2) \times n$ matrix by stacking the row vector $(W_{ni} - W_{nj})$'s and $\mathbf{\Delta}_n^{x_1}$ is the $(n(n-1)/2) \times p_{x_1}$ matrix by stacking the row vector $(x_{1i} - x_{1j})'$'s. Note that in the case $p_q = p_{x_1} + 1$, $\mathbf{X}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n$ is a square matrix and we get a simpler expression for $\widehat{\theta}$,

$$\widehat{\theta}_n = \left[\mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n\right]^{-1} \mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^y, \tag{8}$$

which does not depend on A_n.

For the moment, no specific form of the IV matrix $\mathbf{Q}_n$ has been given. As choice of such $\mathbf{Q}_n$ is not unique, in Section 3, we develop a general asymptotic theory for this PDE as long as $\mathbf{Q}_n$ meets some regularity conditions.[4] According to our experience in the Monte Carlo experiment, one may set $q_{ij} = (W_{ni} - W_{nj})Q_{1n}$, where the $n \times p_q$ matrix Q_{1n} is composed of linearly independent p_q columns chosen from $[X_{1n}, W_nX_{1n}, W_nX_{2n}, W_n^2X_{1n}, W_n^2X_{2n}, \cdots]$, $X_{ln} = (x_{l1}, \cdots, x_{ln})'$ for $l = 1, 2$.[5] It should be noted that here our strategy of choosing the instrument q_{ij}'s, is essentially similar to Su (2011) and other prior literature on choosing IV for a parametrically specified SAR model (e.g., Kelejian and Prucha 1998; Lin and Lee 2010). We will study the optimal choice of A_n for the given choice of $\mathbf{Q}_n$, while the optimal choice of instruments is beyond the scope of this paper.[6]

[4] See Assumption 3-(ii) in Section 3.

[5] For identification, q_{ij}'s should also be chosen to meet some rank conditions, as summarized by Assumption 6 in Section 3.

[6] Like Su (2011), the best selection of instruments q_{ij}'s is not readily available partly because both heteroscedasticity and spatial autocorrelation of unknown form is allowed in the error terms.

3. LARGE SAMPLE PROPERTIES

To analyze the asymptotic properties of our estimator, let's introduce the following regularity conditions. It should be emphasized that among these assumptions, most are also maintained by Su and Jin (2010) and Su (2011) (except Assumption 3-(ii) and 6).

ASSUMPTION 1. *(i) The spatial weights matrix $W_n = [w_{ij}]_{i,j=1,\cdots,n}$ has zero diagonals. (ii) The matrix $S_n = I_n - \lambda_0 W_n$ is nonsingular. (iii) The row and column sums of S_n^{-1} and W_n are uniformly bounded in absolute value.* [7]

Assumption 1 concerns the essential features of spatial weights matrix. Assumption 1-(i) implies that each unit is not a neighbor of itself. Assumption 1-(ii) implies that the model considered here is well defined, that is, the dependent variable Y_n is uniquely determined in terms of the disturbances conditional on the regressors. The uniform boundedness condition on W_n and S_n^{-1} is originated in a series of papers by Kelejian and Prucha, see, e.g., Kelejian and Prucha (1998, 1999), in order to limit the spatial dependence across units to a manageable degree.

ASSUMPTION 2. *$U_n = R_n \bar{\epsilon}_n$, where the components of $\bar{\epsilon}_n = (\epsilon_1, \cdots, \epsilon_n)'$ are i.i.d. with zero mean, unit variance and some finite absolute moment of order three. R_n is a $n \times n$ constant matrix whose row and column sums are uniformly bounded.*

Existence of the absolute moment of order three for ϵ_i's is assumed so that some central limit theorem of simple form can be applied. This assumption can be relaxed to the existence of some absolute moment of order higher than two. Assumption 2 allows for not only heteroscedasticity but also spatial dependence across u_i's. When R_n is a scalar matrix or a diagonal matrix, we have either i.i.d. u_i's or independently distributed but heteroscedastic u_i's. For a general R_n, it allows the error term u_i's to exhibit various scenarios of spatial dependence. For example, when $R_n = I_n + \rho_0 M_n$, U_n is assumed to follow a spatial moving average process; when $R_n = (I_n - \rho_0 M_n)^{-1}$, then U_n follows a SAR process. Note that here M_n is not necessarily required to be the same as W_n, as long as the row and column sums of M_n or $(I_n - \rho_0 M_n)^{-1}$ are uniformly bounded, as paralleling Assumption 1 for W_n and $(I_n - \lambda_0 W_n)^{-1}$.

ASSUMPTION 3. *(i) $X_n = [X_{1n}, X_{2n}]$ is a non-stochastic, $n \times (p_{x_1} + p_{x_2})$ regressor matrix whose elements are uniformly bounded. (ii) There exists a sequence of p_q-dimensional non-stochastic q_i's, $i = 1, \cdots, n$, such that $q_{ij} = q_i - q_j$, where the elements of q_i's are uniformly bounded by a constant.*

[7] The row and column sums of an $n \times n$ matrix P_n are said to be uniformly bounded if we have for all n, there exists a positive constant c independent of n with $\max_i \sum_{j=1}^n |P_{n,ij}| < c$ and $\max_j \sum_{i=1}^n |P_{n,ij}| < c$. This notion of uniform boundedness can be defined in terms of some matrix norms. The maximum column sum matrix norm $\|\cdot\|_1$ of an $n \times n$ matrix P_n is defined as $\|P_n\|_1 = \max_j \sum_i |P_{n,ij}|$, and the maximum row sum matrix norm $\|\cdot\|_\infty$ is defined as $\|P_n\|_\infty = \max_i \sum_j |P_{n,ij}|$. Thus the uniform boundedness of $\{P_n\}$ in column or row sums is equivalent to the sequence $\{\|P_n\|_1\}$ or $\{\|P_n\|_\infty\}$ being bounded.

The non-stochastic design assumption of X_n is made for several reasons. First, it parallels that of Kelejian and Prucha (1998, 1999, 2001, 2010), Lee (2004, 2007) and Lin and Lee (2010). Second, it allows us to avoid the use of trimming factors (e.g. Robinson 1988). As noted by Lee (2007), non-stochastic regressor design and its uniform boundedness condition are made for technical convenience. If the elements of X_n are stochastic and have unbounded ranges, conditions in Assumption 3-(i) can be replaced by some finite moment conditions. Assumption 3-(ii) provides the regularity conditions for a general $\mathbf{Q}_n$ to meet. Given the uniform boundedness condition of X_n and W_n by Assumption 1 and 3-(i), it is reasonable to construct q_i's based on some combinations of X_n and W_n. As an example, one may easily verify that the $q_{ij} = (W_{ni} - W_{nj})Q_{1n}$ defined in Section 2 satisfies these conditions, with $q_i = W_{ni}Q_{1n}$.

Assumption 4. *Suppose that $x_{2i} \in \mathcal{X}_2 \subset \mathbb{R}^{p_{x_2}}$, for $i = 1, \cdots, n$. (i) There exists a scalar function $f(\cdot)$ and a p_q-dimensional vector valued function $\varphi(\cdot) = (\varphi_1(\cdot), \cdots, \varphi_{p_q}(\cdot))'$, both of which are defined on $\mathcal{X}_2$, such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} q_{i,l}\nu(x_{2i}) = \int_{s\in\mathcal{X}^2} \nu(s)\varphi_l(s)f(s)ds \tag{9}$$

for any bounded and continuous function $\nu(\cdot)$ defined on $\mathcal{X}_2$, where q_{il} is the l-element of q_i. Furthermore, if $q_{il} = c$ for some l and for all $i = 1, \cdots, n$, $\varphi_l(\cdot) \equiv c$. Both $f(\cdot)$ and $\varphi(\cdot)$ are bounded, continuous and have bounded first order derivatives on $\mathcal{X}_2$. (ii) The function $m(x)$ is bounded, continuous and has bounded second order derivatives.

Assumption 4-(i) parallels Assumption 3 in Su (2011). Eqn. (9) is frequently seen in spatial econometrics literature involving nonparametric techniques (Su and Jin, 2010; Su, 2011), or more generally, in nonparametric regression literature with fixed regressor design (Linton 1995). Essentially, Eqn. (9) relates the fixed design to an implicit random generation mechanism. For stationary random observations, $f(\cdot)$ can be interpreted as the underlying probability density function that generates $x_{2i}, i = 1, \cdots, n$ and $\varphi(\cdot)$ can be regarded as the conditional expectation of q_i given x_{2i}. Consider three important special cases implied by this assumption . First, if $q_{il} = c$ for some l and for all $i = 1, \cdots, n$, one may regard q_{il} being generated from a degenerate distribution with all the probability concentrated on a single point c. In this case, the expectation of q_{il} conditional on any random variables will be c, which has been explicitly assumed above. Second, letting $q_{il} \equiv 1$, we have for any bounded and continuous function $\nu(\cdot)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \nu(x_{2i}) = \int_{s\in\mathcal{X}^2} \nu(s)f(s)ds, \tag{10}$$

which is consistent with the general law of large numbers. Third, if one further lets $\nu \equiv 1$, Eqn. (10) is reduced to the fact that $f(\cdot)$ integrates to one over $\mathcal{X}^2$. In Assumption 4, the components of x_{2i}'s are required to be absolute continuous with respect to Lebesgue measure. Like other literature on pairwise difference estimation (e.g., Ahn and Powell 1993; Honore and Powell 2007), this condition can be relaxed to permit discrete components of x_{2i}'s, requiring that at least one component is continuously distributed. Finally, it should be argued that even though we focus on the fixed regressor case, our analysis holds with probability one if x_{2i}'s are generated randomly, and in this case, we can interpret our analysis as being conditional on x_{2i}'s.

ASSUMPTION 5. *(i) The kernel function $k(\cdot)$ is continuous, satisfying $k(v) = k(-v)$, $\int k(v)dv = 1$, $\int |k(v)v^l|dv < \infty$ for $l = 1, 2$ and $\sup_v |k(v)| < c_k$. (ii) As $n \to \infty$, $\max_{1 \le l \le p_{x_2}} h_{ln} \to 0$, $nh_{n1} \cdots h_{np_{x_2}} \to \infty$, $n \sum_{l=1}^{p_{x_2}} h_{ln}^4 \to 0$.*

This assumption parallels Assumption 4 in Su (2011). Assumption 5-(i) concerns the choice of kernel function and Assumptions 5-(ii) concerns the choice of smoothing parameters. In the Monte Carlo simulation, we use the Gaussian kernel throughout and this kernel can be shown to meet Assumption 5-(i) automatically. Like Su (2011), the conditions in Assumption 5-(ii) imply implicitly that $p_{x_2} \le 3$. This is not too restrictive given the "curse of dimensionality" in the nonparametric literature. In the case where $p_{x_2} \ge 4$, one can apply a kernel of higher order.

ASSUMPTION 6. *The limit of $J_n = \frac{1}{n}\sum_{i=1}^n f(x_{2i})\overline{z}_i(q_i - \varphi(x_{2i}))'$ exists and has full column rank, where $\overline{z}_i = [W_{ni}S_n^{-1}(X_{1n}\beta_0 + m(X_{2n})), x_{1i}']'$, $X_{ln} = [x_{l1}, \cdots, x_{ln}]'$ for $l = 1, 2$, $m(X_{2n}) = (m(x_{21}), \cdots, m(x_{2n}))'$, $f(\cdot)$ and $\varphi(\cdot)$ has been defined in Assumption 4.*

Essentially, Assumption 6 provides a sufficient condition that allows us to identify θ uniquely. Given the expression of $\widehat{\theta}_n$ in (7), it is straightforward to show that θ is identifiable as long as the limit of $\frac{1}{n(n-1)}\mathbb{E}(\mathbf{Z}_n)'\mathbf{K}_n^{x_2}\mathbf{Q}_n$ exists and has full column rank. In the Appendix, we show that

$$\frac{1}{n(n-1)}\mathbb{E}(\mathbf{Z}_n)'\mathbf{K}_n^{x_2}\mathbf{Q}_n - J_n = o(1). \tag{11}$$

Furthermore, Assumption 1, 3-4, together with Lemma 1 imply that J_n is $O(1)$.

To establish the limiting distribution of our estimator, let's introduce more notations. Define $\mathcal{F}_n = \text{diag}\{f(x_{21}), \cdots, f(x_{2n})\}$, $Q_n^c = [q_1 - \varphi(x_{21}), \cdots, q_n - \varphi(x_{2n})]'$, J_n in Assumption 6 then can be equivalently expressed as $\frac{1}{n}\overline{Z}_n'\mathcal{F}_n Q_n^c$. Similar arguments can be applied to show that $\Omega_n = \frac{1}{n}Q_n^{c\prime}\mathcal{F}_n R_n R_n' \mathcal{F}_n Q_n^c$ is also $O(1)$. The following proposition provides the asymptotic distribution of PDE.

PROPOSITION 1.*Suppose that Assumption 1-6 hold, then the estimator (7) has the limiting distribution $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to^d \mathcal{N}(0, \Sigma)$ where*

$$\Sigma = \lim_{n\to\infty} \left(J_n A_n J_n'\right)^{-1} J_n A_n \Omega_n A_n J_n' \left(J_n A_n J_n'\right) \tag{12}$$

provided that the limits of Ω_n and A_n exist and have full column rank.

For purposes of inference, we propose a consistent estimator for the limiting covariance matrices above. By (12), it suffices to focus on estimate of J_n and Ω_n. A consistent estimate of J_n is straightforward. Actually, in the Appendix we have shown that in proving Proposition 1, $\widehat{J}_n = \frac{1}{n(n-1)}\mathbf{Z}_n\mathbf{K}_n^{x_2}\mathbf{Q}_n \to^p \lim_{n\to\infty} J_n$. Similarly, Ω_n can be consistently estimated by

$$\widehat{\Omega}_n = \frac{1}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}[\mathbf{\Delta}_n^y - \mathbf{Z}_n\widehat{\theta}_n][\mathbf{\Delta}_n^y - \mathbf{Z}_n\widehat{\theta}_n]'\mathbf{K}_n^{x_2}\mathbf{Q}_n. \tag{13}$$

When $p_q = p_{x_1} + 1$, J_n is invertible, that is, the CV matrix Σ in (12) has a simpler form $\Sigma = \lim_{n\to\infty}(J_n\Omega_n^{-1}J_n')^{-1}$, which is independent of A_n. For a general case $p_q > p_{x_1} + 1$, given the instruments q_{ij}'s, the optimal choice of a weighting matrix A_n will be Ω_n^{-1} by the generalized Schwartz inequality, yielding $\Sigma = \lim_{n\to\infty}(J_n\Omega_n^{-1}J_n')^{-1}$ once again.

Finally, given $\sqrt{n}$-consistent estimate of the parametric components, a consistent estimate of the nonparametric component seems to be like a usual exercise, that is, $m(\cdot)$ can be estimated by,

$$\widehat{m}_n(x) = \frac{\sum_{i=1}^n \left(y_i - \widehat{\lambda}_n \sum_{j\neq i} w_{ij}y_j - x_{1i}'\widehat{\beta}_n\right) K_h(x - x_{2i})}{\sum_{i=1}^n K_h(x - x_{2i})}.[8] \tag{14}$$

4. MONTE CARLO SIMULATION

In this section, we conduct a small scale Monte Carlo experiment to evaluate the finite sample performance of the suggested PDE for PL-SAR model. A larger Monte Carlo study relating to a wider set of experiments than those described below is left for future research. The DGP is given by

$$y_i = \lambda_0 \sum_{j=1}^n w_{ij}y_j + x_{1i}\beta_{10} + x_{2i}\beta_{20} + m(z_i) + u_i, i = 1, \cdots, n, \tag{15}$$

where z_i is a scalar variable, i.i.d. drawn from the uniform distribution on $[-2, 2]$, $x_{2i} = 0.2z_i^2 - z_i + \eta_i$, where η_i is i.i.d. drawn from a standard normal distribution, and x_{1i} is a 0-1 dummy variable with $\mathbb{P}(x_{1i} = 1) = \mathbb{P}(x_{1i} = 0) = 0.5$. We set $\beta_{10} = \beta_{20} = 1, \lambda_0 = 0.5$ and consider two cases for $m(\cdot)$: (i) $m_1(z) = \exp(z)/(1 + \exp(z))$ and (ii) $m_2(z) = 1 + 0.5\sin(0.5\pi z)$, two cases for U_n: (i) $u_i = \sqrt{1 + z_i^2} \cdot \epsilon_i$ (heteroscedastic error terms) and (ii) $U_n = (I_n - 0.3 \cdot W_n)^{-1}\bar{\epsilon}_n$ (errors that are both heteroscedastic and spatially correlated), two cases for the distribution of ϵ_i's: (i) $\epsilon_i \sim \mathcal{N}(0, 1)$ and (ii) $\epsilon_i \sim \chi^2(1)/\sqrt{2}$, and two choices of instruments: (i) $(q_1^{(1)}, \cdots, q_n^{(1)})' = Q_n^{(1)} = [X_{1n}, X_{2n}, W_nX_{1n}, W_nX_{2n}, W_nZ_n]$ and (ii) $Q_n^{(2)} = [X_{1n}, X_{2n}, W_nX_{1n}, W_nX_{2n}, W_nZ_n, W_n^2X_{1n}, W_n^2X_{2n}, W_n^2Z_n]$.

Like Su and Yang (2011) and Su (2010), we generate the spatial weight matrix W_n according to the principle of *Rook* contiguity, by randomly allocating the n spatial units on a lattice of $\sqrt{n} \times \sqrt{n}$ squares, finding the neighbors for each unit, and then row normalizing. To implement the estimator, we need to choose a kernel function $k(\cdot)$ and a bandwidth sequence h_n. Throughout, we will choose a Gaussian kernel. As it is difficult to specify the optimal bandwidth sequence, we

[8] The consistency and limiting distribution of $\widehat{m}_n(x)$ can be investigated by similar arguments to Theorem 3.2 in Su (2011). However, since this paper is mainly concerned with computationally easy estimation of the model's finite-dimensional parameters, the rigorous proof of the asymptotic results about $\widehat{m}(\cdot)$ is omitted.

choose the bandwidth by a rule-of-thumb method: $h_n = \sigma_z n^{-1/3.5}$, where σ_z is the sample standard deviation of z_i's. In practice, it is recommended that we use this bandwidth as the initial smoothing parameter to obtain a preliminary consistent estimator while in the second step, we conduct the least squares cross validation method to choose the bandwidth. We have applied both methods to compute the estimators but found their performances are similar for the DGP under study. So we only report the simulation results for the former case. We have experimented with $n = 100, 196$ and 400. For each case, there are 1000 repetitions.

Table 1 summarizes the simulation results, empirical bias and empirical standard deviation of the estimators when u_i's are heteroscedastic but independent from each other. There are several main observations:

1. In the presence of heteroscedasticity, the estimators perform fairly well even in small sample $n = 100$, that is, they are computationally easy, only slightly biased while give acceptable precision.
2. Standard deviation of these estimators declines with the sample size. The magnitude of such decline is generally consistent to $\sqrt{n}$-asymptotics, which is consistent with the theoretical predictions.
3. The estimators are only slightly biased and the magnitude of such bias, if any, declines with the growing sample size. When $n = 400$, these estimators are essentially unbiased.
3. The estimators seem quite robust with respect to the distribution of the error terms, functional form of nonparametric components and choice of instruments. Their performance, in terms of either bias or standard deviation, does not seem to change much for different distributions of the error terms, functional forms of nonparametric components and choices of instruments.

Table 2 summarizes the simulation results of the estimators when u_i's are both heteroscedastic and spatially correlated. Overall, the resulting estimators perform similarly to those in Table 1.

5. CONCLUSION

In this paper we propose a semi-parametric pairwise difference estimator of partially linear SAR models where the error term may exhibit either heteroscedasticity or spatial dependence. We derive the limiting distribution of the proposed estimator and suggest a simple way to consistently estimate the asymptotic CV matrix. In comparison with other existing estimation procedures (Su and Jin 2010; Su 2011), our estimator is computationally simpler in obtaining the consistent estimate of the finite-dimensional parameters in the model. For the applied researchers who are mainly interested in estimating the parametric component of the model, our estimator seems to be more attractive. Simulation results indicate that our estimator is promising even in small samples.

APPENDIX

Lemma 1. *For any two $n \times n$ matrices B_{1n} and B_{2n} whose row and column sums are bounded uniformly bounded by a constant, (i) the row and column sums of $B_{1n}B_{2n}$ are also uniformly bounded* by a constant. *(ii) For some $n\times p$ matrices C_{1n} and C_{2n} whose elements are bounded by a constant, the elements of $B_{1n}C_{1n}$, $\frac{1}{n}C'_{1n}C_{2n}$ and $\frac{1}{n}C'_{1n}B_{1n}C_{2n}$ are uniformly bounded by a constant.*

Proof. Trivial. ■

Lemma 2. *Suppose that B_n is a $n\times n$ matrix with its column sums being uniformly bounded in absolute value, elements of the $n\times k$ matrix C_n are uniformly bounded, and the components of $U_n = (u_1,\cdots,u_n)'$ are i.i.d. with zero mean, variance σ_0^2 and finite third order absolute moment. Then, $1/\sqrt{n}C'_nB_nU_n = O_p(1)$, $1/nC'_nB_nU_n = o_p(1)$ and $1/\sqrt{n}C'_nB_nU_n \to^d \mathcal{N}\big(0, \sigma_0^2 \lim_{n\to\infty}\frac{1}{n}C'_nB_nB'_nC_n\big)$ if the limit of $\frac{1}{n}C'_nB_nB'_nC_n$ exists and is positive definite.*

Proof. See Lemma A.4 in Lin an Lee (2010). ■

Proof of Eqn (11). For ease of notation, let $p_{x_2}=1$ without loss of generality. Let $\overline{\mathbf{Z}}_n = \mathbb{E}(\mathbf{Z}_n) = [\mathbf{\Delta}_n^w S_n^{-1}(X_{1n}\beta_0 + m(X_{2n})), \mathbf{\Delta}_n^{x_1}]$. We have by symmetric kernel $k(\cdot)$ and $q_{ij} = q_i - q_j$ that

$$
\begin{aligned}
&\frac{1}{n(n-1)}\overline{\mathbf{Z}}'_n\mathbf{K}_n^{x_2}\mathbf{Q}_n = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n \overline{z}_{ij}q'_{ij}h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)\\
= \;& \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n(\overline{z}_i-\overline{z}_j)q'_{ij}h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}\overline{z}_iq'_{ij}h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)\\
= \;& \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n\overline{z}_i(q_i-q_j)'h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)\\
= \;& \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n\overline{z}_iq'_ih_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) - \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n\overline{z}_iq'_jh_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)\\
= \;& \frac{1}{n-1}\sum_{i=1}^n\overline{z}_iq'_i\frac{1}{n}\sum_{j=1}^n h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) - \frac{1}{n-1}\sum_{i=1}^n\overline{z}_i\frac{1}{n}\sum_{j=1}^n q'_jh_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)\\
= \;& \frac{1}{n-1}\sum_{i=1}^n\overline{z}_iq'_ih_n^{-1}\int k\left(\frac{x_{2i}-s}{h_n}\right)f(s)ds - \frac{1}{n-1}\sum_{i=1}^n h_n^{-1}\overline{z}_i\int\varphi'(s)k\left(\frac{x_{2i}-s}{h_n}\right)f(s)ds + o(1)\\
= \;& \frac{1}{n-1}\sum_{i=1}^n\overline{z}_iq'_i\int k\,(t)\,f(h_nt+x_{2i})dt - \frac{1}{n-1}\sum_{i=1}^n\overline{z}_i\int\varphi'(h_nt+x_{2i})k\,(t)\,f(h_nt+x_{2i})dt + o(1)\\
= \;& \frac{1}{n-1}\sum_{i=1}^n\overline{z}_iq'_i\int k\,(t)\left(f(x_{2i})+h_ntf^{(1)}(\overline{x}_{2i})\right)dt\\
- \;& \frac{1}{n-1}\sum_{i=1}^n\overline{z}_i\int\left(\varphi(x_{2i})+h_nt\varphi^{(1)}(\widetilde{x}_{2i})\right)'k\,(t)\left(f(x_{2i})+h_ntf^{(1)}(\overline{x}_{2i})\right)dt + o(1)\\
& \qquad = \frac{1}{n}\sum_{i=1}^n f(x_{2i})\overline{z}_i(q_i-\varphi(x_{2i}))' + o(1) = J_n + o(1),
\end{aligned}
$$

where $f^{(1)}(\cdot)$ and $\varphi^{(1)}(\cdot)$ are the first derivatives of f and φ respectively, both $\overline{x}_{2i}$ and $\widetilde{x}_{2i}$ lie between x_{2i} and $x_{2i}+h_nt$. ■

Proof of Proposition 1. For ease of notation, let $p_{x_2} = 1$ without loss of generality. We have

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_n - \theta_0) &= \left[\frac{1}{n(n-1)}\mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n A_n \frac{1}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{Z}_n\right]^{-1} \\ &\quad \frac{1}{n(n-1)}\mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n A_n \frac{\sqrt{n}}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\left[\mathbf{\Delta}_n^m + \mathbf{\Delta}_n^u\right],\end{aligned}$$

where $\mathbf{\Delta}_n^m$ is the $n(n-1)/2$-dimensional vector of Δ_{ij}^m's, with $\Delta_{ij}^m = m(x_{2i}) - m(x_{2j})$, for $i, j = 1, \cdots, n, i < j$. Then the desired limiting distribution of $\widehat{\theta}_n$ follows if one can establish the following results, that is, (i) $\frac{1}{n(n-1)}\mathbf{Z}_n'\mathbf{K}_n^{x_2}\mathbf{Q}_n - J_n \to^p 0$; (ii) $\frac{\sqrt{n}}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^u \to^d \mathcal{N}(0, \lim_{n\to\infty}\Omega_n)$; and (iii) $\frac{\sqrt{n}}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^m \to^p 0$.

For (i), write $\mathbf{Z}_n = \mathbb{E}(\mathbf{Z}_n) + [\mathbf{\Delta}_n^w S_n^{-1}U_n, 0]$. Observing Eqn.(11), it remains to verify that $\frac{1}{n(n-1)}(\mathbf{\Delta}_n^w S_n^{-1}U_n)'\mathbf{K}_n^{x_2}\mathbf{Q}_n = o_p(1)$. This follows from Lemma 2, since by similar argument to the proof Eqn.(11) we have $\frac{1}{n(n-1)}(\mathbf{\Delta}_n^w S_n^{-1}U_n)'\mathbf{K}_n^{x_2}\mathbf{Q}_n = \frac{1}{n}\sum_{i=1}^n f(x_{2i})\widetilde{u}_i(q_i - \varphi(x_{2i}))' + o_p(1) = \frac{1}{n}(W_n S_n^{-1}U_n)'\mathcal{F}_n Q_n^c + o_p(1)$, where $\widetilde{u}_i = W_{ni}S_n^{-1}U_n$.

For (ii), similar arguments to the proof of Eqn. (11) give that

$$\begin{aligned}&\frac{\sqrt{n}}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^u = \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right)(u_i - u_j)q_{ij} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (q_i - \phi(x_{2i}))f(x_{2i})u_i + o_p(1) = \frac{1}{\sqrt{n}}Q_n^{c\prime}\mathcal{F}_n R_n\bar{\epsilon}_n + o_p(1) \to^d \mathcal{N}(0, \lim_{n\to\infty}\Omega_n)\end{aligned}$$

by Lemma 2.

For (iii),

$$\begin{aligned}&\frac{\sqrt{n}}{n(n-1)}\mathbf{Q}_n'\mathbf{K}_n^{x_2}\mathbf{\Delta}_n^m = \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n q_{ij}(m(x_{2i}) - m(x_{2j}))h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) \\ &= \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i} q_{ij}m(x_{2i})h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) \\ &= \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n q_i m(x_{2i})h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) - \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n q_j m(x_{2i})h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) \\ &= \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n q_i m(x_{2i})h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) - \frac{\sqrt{n}}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n q_i m(x_{2j})h_n^{-1}k\left(\frac{x_{2i}-x_{2j}}{h_n}\right) \\ &= \frac{\sqrt{n}}{n-1}\sum_{i=1}^n q_i\left(m(x_{2i})\int k(t)f(h_nt + x_{2i})dt - \int m(h_nt + x_{2i})k(t)f(h_nt + x_{2i})dt\right) + o(1) \\ &= \frac{-\sqrt{n}}{n-1}\sum_{i=1}^n q_i\left(\int\left(m^{(1)}(x_{2i})h_nt + \frac{1}{2}h_n^2t^2m^{(2)}(\overline{x}_{2i})\right)k(t)\big(f(x_{2i}) + h_ntf^{(1)}(\widetilde{x}_{2i})\big)dt\right) + o(1) \\ &\leq^* -h_n^2\sqrt{n}\cdot\int|k(t)t^2|dt\cdot c^2\cdot\frac{1}{n}\sum_{i=1}^n|q_i| + o(h_n^2\sqrt{n}) = o(1),\end{aligned}$$

where $m^{(1)}(\cdot)$, $m^{(2)}(\cdot)$ and $f^{(1)}(\cdot)$ are the first derivative, second derivative of $m(\cdot)$ and the first derivative of $f(\cdot)$, respectively, both $\overline{x}_{2i}$ and $\widetilde{x}_{2i}$ lie between x_{2i} and $x_{2i} + h_n t$. The inequality (*) follows from the uniform boundness of $m^{(1)}(\cdot)$, $m^{(2)}(\cdot)$, $f^{(1)}(\cdot)$ and $f(\cdot)$, e.g., by a positive constant c and Assumption 4-5. ■

REFERENCES

Ahn, H., and Powell, J.L., 1993, Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics*, 58, 3-29.

Baltagi, B. H. and Li, D., 2001, LM tests for functional form and spatial correlation, *International Regional Science Review* 24, 194-225.

Chen, S. and Khan, S., 2001, Semiparametric estimation of a partially linear censored regression model, *Econometric Theory*, 17, 567-590.

Chen, S. and Zhou, Y., 2005, A simple matching method for estimating sample selection models using experimental data, *Annals of Economics and Finance*, 6, 155-167.

Glaeser, E. L., Sacerdote, B. and Scheinkman, J. A., 1996, Crime and social interactions, *Quarterly Journal of Economics*, 111, 507-548.

Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

Honore,B. E. and Powell, J.L., 2005, Pairwise difference estimation of nonlinear models, in D. W. K. Andrews and J. H. Stock, eds., *Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 520-53.

Honore,B. E. and Powell, J.L., 2007, Pairwise Difference estimation with nonparametric control variables, *International Economic Review*, 48, 1119-1158.

Kelejian, H.H. and Prucha, I.R., 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance, *Journal of Real Estate Finance 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.

Kelejian, H.H., Prucha., I.R., 2001, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics*, 104, 219-257.

Kelejian, H. H. and Prucha, I. R., 2010, Specification and estimation of spatial autoregressive models with autoregressive and heteroscedastic disturbances, *Journal of Econometrics*, 157, 53-67.

Lee. L.F., 2003, Best spatial two stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews*, 22, 307-335.

Lee, L.F., 2004, Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica*, 72, 1899-1925.

Lee, L.F., 2007, GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics*, 137, 489-514.

Lee, S., 2003, Efficient semiparametric estimation of a partially linear quantile regression model, *Econometric theory*, 19, 1-31.

Lin, X. and Lee, L.F., 2010, GMM estimation of spatial autoregressive models with unknown heteroscedasticity, *Journal of Econometrics*, 157, 34-52.

Linton, O., 1995, Second order approximation in the partially linear regression model, *Econometrica*, 63, 1079-1112.

Liu, X., Lee, L.F. and Bollinger, C. R., 2010, An efficient GMM estimator of spatial autoregressive models, *Journal of Econometrics*, 159, 303-319.

Pace, P. K., Barry, R., Slawson Jr., V. C. and Sirmans, C. F., 2004, Simultaneous spatial and functional form transformation. In: Anselin, L., Florax, R., Rey, S. J. (Eds.), *Advances in Spatial Econometrics*, 197-224. Springer-Verlag, Berlin.

Robinson, P. M., 1988, Root-n-consistent semiparametric regression, *Econometrica*, 56, 931-954.

Su, L., 2011, Semi-parametric GMM estimation of spatial autoregressive models, *Journal of Econometrics*, forthcoming.

Su, L. and Jin, S., 2010, Profile quasi-maximum likelihood estimation of spatial autoregressive models, *Journal of Econometrics*, 157, 18-33.

Su, L. and Yang, Z., 2011, Instrumental variable quantile estimation of spatial autoregressive models, Working paper, Singapore Management University.

Sun, Y., 2005, Semiparametric efficient estimation of partially linear quantile regression models, *Annals of Economics and Finance*, 6, 105-127.

Yang, Z., Li, C. and Tse, Y. K., 2006, Functional form and spatial dependence in dynamic panels, *Economics Letters*, 91, 138-145.

Zheng, Z. and Zhu, P., 2010, A more efficient best three stage least squares estimator of spatial autoregressive models, *Annals of Economics and Finance*, 11, 155-184.

Table 1.

PDE of Parametric Components: $u_i = \sqrt{1+z_i^2} \cdot \epsilon_i$

		$n=100$							
		$\epsilon_i \sim \mathcal{N}(0,1)$				$\epsilon_i \sim \chi^2(1)/\sqrt{2}$			
		$m_1(z)$		$m_2(z)$		$m_1(z)$		$m_2(z)$	
		$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$
λ	Bias	0.005	0.004	0.017	0.011	0.003	0.023	0.017	0.006
	Std	0.181	0.166	0.182	0.163	0.191	0.153	0.175	0.170
β_1	Bias	−0.011	−0.010	−0.011	−0.005	−0.014	−0.013	−0.002	−0.021
	Std	0.299	0.302	0.298	0.319	0.300	0.308	0.317	0.296
β_2	Bias	−0.029	−0.023	−0.023	−0.011	−0.023	−0.001	−0.010	−0.004
	Std	0.142	0.160	0.145	0.166	0.151	0.172	0.141	0.152
		$n=196$							
		$\epsilon_i \sim \mathcal{N}(0,1)$				$\epsilon_i \sim \chi^2(1)/\sqrt{2}$			
		$m_1(z)$		$m_2(z)$		$m_1(z)$		$m_2(z)$	
		$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$
λ	Bias	0.002	0.011	0.012	0.012	0.007	0.013	0.002	0.004
	Std	0.125	0.122	0.128	0.114	0.121	0.111	0.120	0.115
β_1	Bias	−0.003	−0.006	−0.008	−0.001	−0.012	−0.002	−0.022	−0.016
	Std	0.216	0.206	0.201	0.209	0.210	0.217	0.223	0.212
β_2	Bias	−0.021	−0.012	−0.017	−0.008	−0.003	−0.005	−0.007	−0.003
	Std	0.099	0.114	0.101	0.117	0.099	0.117	0.106	0.121
		$n=400$							
		$\epsilon_i \sim \mathcal{N}(0,1)$				$\epsilon_i \sim \chi^2(1)/\sqrt{2}$			
		$m_1(z)$		$m_2(z)$		$m_1(z)$		$m_2(z)$	
		$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$
λ	Bias	0.002	0.001	0.007	0.003	0.004	0.004	0.003	0.004
	Std	0.082	0.083	0.086	0.084	0.084	0.076	0.085	0.078
β_1	Bias	−0.003	−0.008	−0.003	−0.004	−0.007	−0.007	−0.004	−0.007
	Std	0.149	0.148	0.146	0.147	0.150	0.147	0.151	0.147
β_2	Bias	−0.017	−0.007	−0.008	−0.005	−0.008	−0.003	−0.003	−0.005
	Std	0.072	0.079	0.070	0.083	0.071	0.082	0.074	0.082

Table 2.
PDE of Parametric Components: $U_n = (I_n - 0.3 \cdot W_n)^{-1}\bar{\epsilon}_n$

		$\epsilon_i \sim \mathcal{N}(0,1)$				$\epsilon_i \sim \chi^2(1)/\sqrt{2}$			
		$m_1(z)$		$m_2(z)$		$m_1(z)$		$m_2(z)$	
		$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$	$Q_n^{(1)}$	$Q_n^{(2)}$
$n = 100$									
λ	Bias	0.005	0.030	0.021	0.026	0.013	0.018	0.010	0.025
	Std	0.139	0.135	0.135	0.129	0.135	0.131	0.133	0.130
β_1	Bias	-0.008	-0.009	-0.006	-0.004	-0.010	-0.003	-0.013	-0.008
	Std	0.209	0.211	0.220	0.222	0.210	0.216	0.204	0.218
β_2	Bias	-0.029	-0.018	-0.018	-0.014	-0.025	-0.020	-0.027	-0.019
	Std	0.101	0.116	0.101	0.117	0.103	0.114	0.102	0.117
$n = 196$									
λ	Bias	0.001	0.016	0.004	0.015	0.008	0.010	0.002	0.010
	Std	0.093	0.097	0.097	0.089	0.091	0.091	0.097	0.091
β_1	Bias	-0.003	-0.006	-0.006	-0.003	-0.006	-0.001	-0.002	-0.005
	Std	0.154	0.147	0.148	0.148	0.142	0.146	0.149	0.147
β_2	Bias	-0.019	-0.011	-0.012	-0.012	-0.014	-0.011	-0.011	-0.012
	Std	0.068	0.083	0.071	0.082	0.072	0.079	0.070	0.085
$n = 400$									
λ	Bias	0.001	0.003	0.003	0.003	0.005	0.007	0.004	0.004
	Std	0.065	0.067	0.064	0.066	0.063	0.065	0.067	0.068
β_1	Bias	-0.002	-0.003	-0.002	-0.005	-0.006	-0.001	-0.004	-0.001
	Std	0.100	0.103	0.103	0.104	0.103	0.106	0.105	0.105
β_2	Bias	-0.010	-0.006	-0.007	-0.001	-0.011	-0.008	-0.006	-0.003
	Std	0.048	0.059	0.047	0.057	0.049	0.059	0.051	0.058