10-2017

# Specification Test for Spatial Autoregressive Models

Liangjun SU
*Singapore Management University*, ljsu@smu.edu.sg

Xi QU
*Antai College of Economics and Management*

**Published in Journal of Business and Economic Statistics, 2017 October, Volume 35, Issue 4, Pages** 572-584
**http://doi.org/10.1080/07350015.2015.1102734**

# Specification Test for Spatial Autoregressive Models

**Liangjun Su**
School of Economics, Singapore Management University, Singapore 178903, Singapore (*ljsu@smu.edu.sg*)

**Xi Qu**
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China (*xiqu@sjtu.edu.cn*)

This article considers a simple test for the correct specification of linear spatial autoregressive models, assuming that the choice of the weight matrix $W_n$ is true. We derive the limiting distributions of the test under the null hypothesis of correct specification and a sequence of local alternatives. We show that the test is free of nuisance parameters asymptotically under the null and prove the consistency of our test. To improve the finite sample performance of our test, we also propose a residual-based wild bootstrap and justify its asymptotic validity. We conduct a small set of Monte Carlo simulations to investigate the finite sample properties of our tests. Finally, we apply the test to two empirical datasets: the vote cast and the economic growth rate. We reject the linear spatial autoregressive model in the vote cast example but fail to reject it in the economic growth rate example. Supplementary materials for this article are available online.

KEY WORDS: Generalized method of moments; Nonlinearity; Spatial autoregression; Spatial dependence; Specification test.

## 1. INTRODUCTION

There exists an enormous literature on the estimation of econometric models with spatial dependence among cross-sectional units. Among the various models involving spatial dependence, the linear spatial autoregressive (SAR) model is perhaps the most popular one. It has been widely used in regional science and urban economics to model the spillover effects. See Ord (1975), Anselin (1988), Kelejian and Prucha (1998, 1999, 2010), Smirnov and Anselin (2001), Lee (2002, 2003, 2004, 2007a), Klier and McMillen (2008), Lin and Lee (2010), Liu, Lee, and Bollinger (2010), Robinson (2010), Su and Yang (2013, 2015), and Gupta and Robinson (2015) for different estimation strategies for this kind of models. In addition, the SAR model can also be used to estimate the peer effects in social network analysis; see Lee (2007b), Goldsmith-Pinkham and Imbens (2013), Helmers and Patnam (2014), and Liu (2014).

Assuming the spatial weight matrix is true, there is still potential misspecification of functional form in linear SAR models. Motivated by this, Su and Jin (2010) considered the profile quasi-maximum likelihood estimation (QMLE) of partially linear SAR models. They demonstrated that the spatial parameters can be poorly estimated if nonlinearity is not correctly taken into account in the estimation procedure. Noting that the estimator of Su and Jin (2010) does not have an analytic form and it is not easy to implement in practice, Su (2012) proposed a nonparametric generalized method of moment (GMM) to estimate semiparametric SAR models. Zhang (2013) proposed a pairwise difference estimator for partially linear SAR models with heteroscedastic or/and spatially correlated error terms. Osipenko (2014) provided a generalized-difference-based two-stage least squares (2SLS) estimator for the linear coefficients in a partially linear SAR model with a general error structure. Despite the fact that all these researchers consider nonlinear SAR models, they still assume a specific spatial data-generating mechanism through the use of exogenous spatial weight matrix.

Another strand of research in the spatial literature tries to forgo the use of exogenous spatial weight matrix by imposing strong spatial mixing or spatial stationary structures in the data or/and to relax the linear parametric assumption on the functional form; see Conley (1999), Banerjee et al. (2004), Hallin, Lu, and Tran (2004), Majumdar et al. (2006), Lin, Li, and Gao (2009), Robinson (2011), Jenish (2012), among others. These articles allow for spatial dependence with the effect of one spatial unit on the others not necessarily being an interest in itself, unlike in the SAR model where the spatial parameter is of interest. More recently, Sun (2014) proposed a functional-coefficient SAR by allowing elements in the spatial weight matrix $W_n$ to be a common nonparametric function of certain distance measure and the exogenous regressors to enter the model with functional coefficients. So her model is of semiparametric nature too.

It is well known that a correctly specified parametric model can afford precise statistical inference, a misspecified one may offer possibly misleading inference, whereas nonparametric modeling is associated with both greater robustness and lesser precision. An intermediate strategy is to apply some semiparametric method. A natural question is whether one should use the parametric, semiparametric, or nonparametric specification in practice. This motivates the development of model specification tests.

In this article, based on nonparametric kernel smoothing techniques, we propose a consistent model specification test for SAR models, assuming the correct choice of spatial weight matrix $W_n$. Our test has several key features. First, unlike the available model specification tests that are applicable for independent data or time series data, our test is designed for SAR models where we have to take into account the spatial dependence among the cross-sectional units. To the best of our knowledge, our test is the first nonparametric test for the correct specification of parametric SAR models, assuming $W_n$ is true.

Second, our test is a kernel-based smoothing test. It has the asymptotic normal distribution under the null hypothesis and can detect Pitman local alternatives converging to the null at the usual nonparametric rate. Fan and Li (2000) showed that the kernel-based tests can be more powerful than nonsmoothing tests (e.g., Bierens 1982, 1990; Chen and Fan 1999) for the "singular" local alternatives considered by Rosenblatt (1975). The main feature of the "singular" local alternatives is that they have narrow spikes and change rapidly as the sample size $n$ increases to infinity. According to Fan and Li (2000), these "singular" local alternatives can be thought of as representing high frequency alternatives and the Pitman local alternatives as representing low frequency alternatives. Our kernel-based smoothing test has power to detect the singular local alternatives that converge to the null model at a rate faster than $n^{-1/2}$ whereas the nonsmoothing tests can only detect such "singular" alternatives that approach the null at rate $n^{-1/2}$. See Fan and Li (2000, p. 1018 and p. 1029) and Remark 3 in Su and White (2008) for more explanation on this.

Third, our test allows for both discrete and continuous regressors. As Hsiao, Li, and Racine (2007) remarked, existing kernel-based tests are limited to situations involving continuous regressors only. These tests may be generalized to admit discrete regressors by using a conventional frequency estimation method that splits the sample into different cells. Nevertheless, this sample-splitting frequency approach often results in a substantial loss of finite-sample efficiency in estimation and a great loss of power in test when the number of observations in each cell is small. This motivates them to smooth the discrete regressors in the construction of their test statistic. In this article, we follow Hsiao, Li, and Racine (2007) and propose a kernel-based test that smooths both continuous and discrete regressors. We use the least-square cross-validation (LSCV) method to select the smoothing parameters for both types of regressors and the Monte Carlo results demonstrate that the proposed test is substantially more powerful than the frequency-based kernel tests that use the sample-splitting technique.

Fourth, even though we only focus on testing the correct specification of the linear SAR models, our test can be easily generalized to test for the correct specification of nonlinear SAR models or semiparametric SAR models, assuming the spatial weight matrix is true.

The article is organized as follows. In Section 2, we introduce our hypotheses and test statistic. In Section 3, we investigate the asymptotic properties of our test under the null hypothesis as well as a sequence of local alternatives and show the consistency of the test. We propose a wild bootstrap version of our test and justify its asymptotic validity in Section 4. We provide a small set of Monte Carlo experiments to evaluate the finite sample performance of our tests in Section 5. Two empirical applications are studied in Section 6. Section 7 concludes. All proofs are relegated to the online supplementary appendix.

*Notation*. For a real matrix $A_n$, we denote its transpose as $A_n'$, its Frobenius norm as $\|A_n\|$ ($\equiv [\mathrm{tr}(A_n A_n')]^{1/2}$), and its $(i, j)$th element as $a_{n,ij}$. Here, $\mathrm{tr}(\cdot)$ is the trace operator. Similarly, for a vector $a_n$, $a_{n,i}$ denotes its $i$th element. When $A_n$ is $m_1 \times m_2$, we use $\|A_n\|_{\max}$ ($\equiv \max_{1\le i\le m_1, 1\le j\le m_2} |a_{n,ij}|$) and $\|A_n\|_1$ ($\equiv \sup_{1\le j\le m_2} \sum_{i=1}^{m_1} |a_{n,ij}|$) to denote its max norm and maximum absolute column sum norm (or 1-norm in short), respectively. Let $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and probability, respectively.

## 2. HYPOTHESIS AND TEST STATISTIC

### 2.1 The Hypothesis

Consider the following SAR model with potential nonlinearity:

$$Y_n = \mathbf{m}_n(X_n) + \rho_n^0 W_n Y_n + U_n, \tag{2.1}$$

where $\rho_n^0$ is the spatial lag parameter, $W_n$ is a specified constant $n \times n$ spatial weight matrix, $W_n Y_n$ is the spatial lagged variable, $n$ is the total number of spatial units, $X_n \equiv (x_{n,1}, \ldots, x_{n,n})'$ is an $n \times p$ matrix of exogenous regressors that do not contain the constant term, $U_n$ is an $n$-dimensional vector of zero mean independent disturbances that are not necessarily identically distributed, $\mathbf{m}_n(X_n) \equiv \left(m_n(x_{n,1}), \ldots, m_n(x_{n,n})\right)'$, and $m_n(\cdot)$ is an unknown smooth function defined on $\mathbb{R}^p$.

Under the condition that $I_n - \rho_n^0 W_n$ is nonsingular, (2.1) has the reduced form

$$Y_n = \left(I_n - \rho_n^0 W_n\right)^{-1} (\mathbf{m}_n(X_n) + U_n). \tag{2.2}$$

This reduced form will be frequently used in the derivation of the asymptotic properties of the estimator proposed below. Clearly, $m_n\left(x_{n,i}\right)$ is not the conditional mean of $y_{n,i}$ given $x_{n,i}$ unless the true spatial parameter $\rho_n^0$ is 0. In the general case, $m_n\left(x_{n,i}\right)$ plays the same role as $\beta_{n0}^0 + x_{n,i}'\beta_{n1}^0$ in the conventional linear SAR model:

$$Y_n = \mathbf{1}_n \beta_{n0}^0 + X_n'\beta_{n1}^0 + \rho_n^0 W_n Y_n + \upsilon_n, \tag{2.3}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones and $\upsilon_n$ represents the error term in the linear model.

As LeSage and Pace (2009, chap. 2.7) remarked, the interpretation of the slope coefficients in an SAR model is quite complicated, so we do not intend to interpret them. Instead, we confine ourselves to testing whether the linear specification of $m_n(\cdot)$ in (2.3) is adequate. In other words, the null hypothesis of interest is

$$\mathbb{H}_0 : m_n(x) = \beta_{n0}^0 + x'\beta_{n1}^0 \text{ for some } \beta_n^0 \equiv \left(\beta_{n0}^0, \beta_{n1}^{0\prime}\right)' \in \mathcal{B} \subset \mathbb{R}^{p+1}, \tag{2.4}$$

and the alternative hypothesis is

$$\mathbb{H}_1 : m_n(x) \neq \beta_{n0} + x'\beta_{n1} \text{ for any } \beta_n \equiv \left(\beta_{n0}, \beta_{n1}'\right)' \in \mathcal{B} \subset \mathbb{R}^{p+1}. \tag{2.5}$$

We next study the estimation of the restricted model in (2.3) and propose a test for the null hypothesis specified in (2.4).

## 2.2 Estimation

We base our test on the estimation of the restricted model (2.3) only. There are two popular types of estimators in the literature. One is the maximum likelihood estimator (MLE) or QMLE; see Anselin (1988) and Lee (2004), among others. The other is the GMM estimator; see Kelejian and Prucha (1998, 1999, 2010), Lee (2003), Lin and Lee (2010), and Liu, Lee, and Bollinger (2010). Under the correct specification of the error distribution, the MLE is efficient. When the error distribution is misspecified, the QMLE is still consistent under some regularity conditions provided that the error terms do not exhibit heteroscedasticity of unknown forms. Lin and Lee (2010) and Kelejian and Prucha (2010) demonstrated that the MLE may be inconsistent in the presence of heteroscedastic disturbances and study independently the GMM estimation of SAR models. More recently, Su and Yang (2013) studied the instrumental variable (IV) quantile estimation of (2.3), which demands less moment conditions on the error term.

Let $\hat{\theta}_n \equiv (\hat{\beta}_n', \hat{\rho}_n)'$ be an estimator of $\theta_n^0 \equiv (\beta_n^{0\prime}, \rho_n^0)'$ in (2.3). Let $\bar{x}_{n,i} \equiv (1, x_{n,i}')'$ and $\bar{X}_n \equiv \left(\bar{x}_{n,1}, \ldots, \bar{x}_{n,n}\right)'$. Define the restricted residual

$$\hat{V}_n \equiv Y_n - \bar{X}_n \hat{\beta}_n - \hat{\rho}_n W_n Y_n.$$

Under appropriate assumptions, we then have

$$\hat{v}_{n,i} = v_{n,i} + o_p(1), \tag{2.6}$$

where $v_{n,i} = y_{n,i} - \bar{x}_{n,i}' \beta_n^0 - \rho_n^0 \sum_{j=1}^n w_{n,ij} y_{n,j}$ by (2.3) and $v_{n,i} = u_{n,i}$ under $\mathbb{H}_0$. Consequently, we can propose a residual-based test for $\mathbb{H}_0$.

## 2.3 Test Statistic

To motivate our test statistic, we assume that $x_{n,i}$ is random with the probability density function (PDF) $f_{n,i}(\cdot)$ so that the restriction on $u_{n,i}$ turns to be $E\left(u_{n,i}|x_{n,i}\right) = 0$. Then we can consider a test statistic that was independently proposed by Fan and Li (1996) and Zheng (1996) in the independent and identically distributed (iid) framework. The test statistic is based upon $J \equiv E[v_{n,i} E\left(v_{n,i}|x_{n,i}\right) f_{n,i}\left(x_{n,i}\right)]$. Note that $J = E\{[E\left(v_{n,i}|x_{n,i}\right)]^2 f_{n,i}\left(x_{n,i}\right)\}$ is zero under $\mathbb{H}_0$ and strictly positive under $\mathbb{H}_1$. Thus it can serve as a valid candidate for testing $\mathbb{H}_0$.

The sample analog of $E[v_{n,i} E\left(v_{n,i}|x_{n,i}\right) f_{n,i}\left(x_{n,i}\right)]$ is

$$\frac{1}{n} \sum_{i=1}^n v_{n,i} E\left(v_{n,i}|x_{n,i}\right) f_{n,i}\left(x_{n,i}\right),$$

which is infeasible since neither $v_{n,i}$ nor $f_{n,i}(\cdot)$ is known. To construct a feasible test statistic, we need to replace $v_{n,i}$ and $E\left(v_{n,i}|x_{n,i}\right) f_{n,i}\left(x_{n,i}\right)$ by their consistent estimators.

We consider the case in which a subset of regressors is continuous and the remainders are discrete. As Li and Racine (2007, chap. 3) remarked, theoretically one can use a nonparametric frequency method to handle the presence of discrete regressors, but such an approach cannot be used in practice if the number of discrete cells is large relative to the sample size. In this article, we follow Racine and Li (2004) and Li and Racine (2007, 2008), and consider the kernel approach to handle both continuous and discrete data.

For clarity, write $x_{n,i} = (x_{n,i}^{c\prime}, x_{n,i}^{d\prime})'$, where $x_{n,i}^c$ denotes a $p_c \times 1$ vector of continuous regressors and $x_{n,i}^d$ denotes a $p_d \times 1$ vector of remaining discrete regressors with $p_d = p - p_c$. We assume each discrete variable in $x_{n,i}^d$ takes a finite number of discrete values. Further, we assume some discrete regressors have a natural ordering, for example, environmental conditions (excellent, good, poor) or preference ordering (like, indifference, dislike). Let $\vec{x}_{n,i}^d$ denote a $p_1 \times 1$ vector (say, the first $p_1$ components of $x_{n,i}^d$, $0 \le p_1 \le p_d$) of discrete regressors that have a natural ordering. Let $\tilde{x}_{n,i}^d$ denote the remaining $p_2 = p_d - p_1$ discrete regressors that do not have a natural ordering. Denote $x_{n,is}^c$ and $x_{n,is}^d$ as the $s$th element of $x_{n,i}^c$ and $x_{n,i}^d$, respectively ($s = 1, \ldots, p_c$ or $p_d$).

For the continuous regressor, we choose a kernel function $Q(\cdot)$ defined on $\mathbb{R}^{p_c}$ and a vector of smoothing parameters $h = \left(h_1, \ldots, h_{p_c}\right)$, where we have suppressed the dependence of $h$ and its elements on $n$ for notational simplicity. Let $Q_h(x) = (\Pi_{s=1}^{p_c} h_s^{-1}) Q\left(x_1/h_1, \ldots, x_{p_c}/h_{p_c}\right)$. Let

$$Q_{h,ij} = Q_h\left(x_{n,i}^c - x_{n,j}^c\right). \tag{2.7}$$

In practice, a frequently used choice of $Q(\cdot)$ is the product of a univariate kernel function $q(\cdot)$. In this case, $Q_{h,ij} = \Pi_{s=1}^{p_c} h_s^{-1} q((x_{n,is}^c - x_{n,js}^c)/h_s)$. Following Li and Racine (2007), we use a variation of the kernel function of Aitchison and Aitken (1976) for the unordered discrete regressor:

$$\tilde{l}\left(\tilde{x}_{n,is}^d, \tilde{x}_{n,js}^d, \lambda_s\right) = \begin{cases} 1 & \text{if } \tilde{x}_{n,is}^d = \tilde{x}_{n,js}^d \\ \lambda_s & \text{otherwise} \end{cases}, \tag{2.8}$$

and use

$$\vec{l}\left(\vec{x}_{n,is}^d, \vec{x}_{n,js}^d, \lambda_s\right) = \begin{cases} 1 & \text{if } \vec{x}_{n,is}^d = \vec{x}_{n,js}^d \\ \lambda_s^{\left|\vec{x}_{n,is}^d - \vec{x}_{n,js}^d\right|} & \text{otherwise} \end{cases} \tag{2.9}$$

for the ordered discrete regressor, where $\lambda_s \in [0, 1]$ is the smoothing parameter. In the special case of $\lambda_s = 0$, $\tilde{l}(\cdot, \cdot, \cdot)$ reduces to the usual indicator function as used in the nonparametric frequency approach. When $\lambda_s = 1$, $\tilde{x}_{n,is}^d$ or $\vec{x}_{n,is}^d$ is completely smoothed out in the sense that it will not affect the nonparametric estimation result.

Combining (2.8) and (2.9), the product kernel function for the discrete regressors is

$$\begin{aligned} L_{\lambda,ij} &\equiv L\left(x_{n,i}^d, x_{n,j}^d, \lambda\right) \\ &\equiv \left[\prod_{s=1}^{p_1} \lambda_s^{\left|\vec{x}_{n,is}^d - \vec{x}_{n,js}^d\right|}\right] \left[\prod_{s=1}^{p_2} \lambda_{s+p_1}^{1 - \mathbf{1}(\tilde{x}_{n,is}^d = \tilde{x}_{n,js}^d)}\right], \end{aligned} \tag{2.10}$$

where $\mathbf{1}(A) = 1$ if $A$ holds and 0 otherwise, and $\lambda = (\lambda_1, \ldots, \lambda_{p_d})$. Again, for notational simplicity we suppress the dependence of $\lambda$ and its elements on $n$. Combining (2.7) and (2.10), we obtain the product kernel function for all the regressors:

$$K_{h\lambda,ij} \equiv Q_{h,ij} L_{\lambda,ij}. \tag{2.11}$$

Now we can estimate $E\left(v_{n,i}|x_{n,i}\right) f_{n,i}\left(x_{n,i}\right)$ by the leave-one-out kernel estimator $\frac{1}{n} \sum_{j=1, j \ne i}^n \hat{v}_{n,j} K_{h\lambda,ij}$, where we divide by $n$ instead of $n-1$ for notational simplicity. Our test statistic is then based upon the random quantity

$$J_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \hat{v}_{n,i} \hat{v}_{n,j} K_{h\lambda,ij}. \tag{2.12}$$

We will study the asymptotic properties of $J_n$ in the next section.

It is worth mentioning that the asymptotic normal null distribution of the $J_n$-based test derived for the iid data independently by Fan and Li (1996) and Zheng (1996) only allows for continuous regressors. Hsiao, Li, and Racine (2007) studied the asymptotic properties of the $J_n$-based test for the iid data when both continuous and discrete regressors are present and the smoothing parameters are chosen by the LSCV method.

## 3. ASYMPTOTIC THEORY

In this section, we investigate the asymptotic properties of $J_n$-based test under $\mathbb{H}_0$ and a sequence of Pitman local alternatives. We also study the global consistency of the test.

### 3.1 Basic Assumptions

Following Lee (2004), let $S_n(\rho_n) = I_n - \rho_n W_n$, $G_n(\rho_n) = W_n S_n(\rho_n)^{-1}$, $S_n = S_n(\rho_n^0)$, and $G_n = G_n(\rho_n^0)$. To provide a rigorous analysis, we make the following assumptions.

*Assumption A1.* (i) Any diagonal element $w_{n,ii}$ of $W_n$ is zero. (ii) $\rho_n^0 \in (\underline{a}_n, \bar{a}_n)$ with $-\infty < \underline{a} \leq \underline{a}_n < \bar{a}_n \leq \bar{a} < \infty$. (iii) The matrix $S_n(\rho_n)$ is nonsingular for all $\rho_n \in (\underline{a}_n, \bar{a}_n)$ and sufficiently large $n$. (iv) For all sufficiently large $n$, there exists a constant $c < \infty$ such that $\|W_n\|_1 + \|W_n'\|_1 \leq c$ and $\|S_n^{-1}\|_1 + \|S_n^{-1\prime}\|_1 \leq c$.

Assumption A1 concerns essential features of the spatial weights matrix. A1(i)–(iii) parallel Assumptions 1(a)–(c) in Kelejian and Prucha (2010). A1(i) is clearly a normalization rule. A1(ii) concerns the parameter space of $\rho_n$, which may vary over the sample size. Section 2.2 of Kelejian and Prucha (2010) provides an excellent discussion on this. A1(iii) ensures that $Y_n$ defined in (2.1) has the reduced form in (2.2 ). A1(iv) limits the spatial correlation to some degree but facilitates the study of the asymptotic study. It is commonly assumed in the literature (see, e.g., Kelejian and Prucha 1998, 1999, 2001, 2010; Lee 2004).

*Assumption A2.* (i) $\{(x_{n,i}, u_{n,i})\}_i^n$ are independently distributed. (ii) $x_{n,i} = (x_{n,i}^{c\prime}, x_{n,i}^{d\prime})'$ exhibits a PDF $f_{n,i}(x^c, x^d)$ on the support $\mathcal{X}_n \equiv \mathcal{X}_n^c \times \mathcal{X}_n^d$ and $f_{n,i}(x^c, x^d)$ is continuously differentiable with respect to (wrt) its continuous argument $x^c$. (iii) $E(u_{n,i}|x_{n,i}) = 0$, $E(u_{n,i}^2) = \sigma_{n,i}^2$, $E(u_{n,i}^4) = \mu_{n,i}^4$; there exist constants $\bar{\sigma}^2 < \infty$ and $\bar{\mu}_{u4} < \infty$ such that $\sup_{i,n} \sigma_{n,i}^2 \leq \bar{\sigma}^2$ and $\sup_{i,n} \mu_{n,i}^4 \leq \bar{\mu}_{u4}$. (iv) Let $\sigma_{n,i}^2(x) = E(u_{n,i}^2|x_{n,i} = x)$, $\mu_{n,i}^4(x) = E(u_{n,i}^4|x_{n,i} = x)$, $\bar{f}_n(x) = n^{-1}\sum_{i=1}^n f_{n,i}(x)$, $\bar{\sigma}_n^2(x) = n^{-1}\sum_{i=1}^n \sigma_{n,i}^2(x) f_{n,i}(x)$, and $\bar{\mu}_n^4(x) = n^{-1}\sum_{i=1}^n \mu_{n,i}^4(x) f_{n,i}(x)$. For each $x = (x^c, x^d)$ on the support $\mathcal{X}_n$, $\bar{f}(x) = \lim_{n\to\infty} \bar{f}_n(x)$ exists, $\lim_{n\to\infty} \int [\bar{\sigma}_n^2(x)]^2 dx$ exists, $\int [\bar{\sigma}_n^2(x)]^4 dx$ and $\int [\bar{\mu}_n^4(x)]^2 dx$ are bounded above from infinity. (v) There exists a constant $\bar{\mu}_{x4} < \infty$ such that $\sup_{i,n} E||x_{n,i}||^4 < \bar{\mu}_{x4}$.

For notational simplicity, here and below we write $\int a_n(x)\,dx$ for $\sum_{x^d \in \mathcal{X}_n^d} \int_{\mathcal{X}_n^c} a_n(x^c, x^d)\,dx^c$, where the summation is over all possible values of $x^d$ on $\mathcal{X}_n^d$. Assumption A2 concerns essential features of the exogenous regressors and error terms in the model. Following the suggestion of a referee, we consider stochastic exogenous regressors and assume $\{(x_{n,i}, u_{n,i})\}_i^n$ follow independent but nonidentical distributions in A2(i)–(iv). A2(iii)–(iv) allow for both conditional and unconditional heteroscedasticity and kurtosis in the error terms. In the presence of heteroscedasticity, the QMLE of Lee (2004) in the linear SAR models is generally inconsistent. For this reason, Kelejian and Prucha (2010) and Lin and Lee (2010) explored the GMM estimation of the linear SAR models with heteroscedasticity. Nevertheless, they require the existence of $(4+\epsilon)$th moments of $u_{n,i}$ for some $\epsilon > 0$. In Su and Yang (2013), the IV quantile estimation of SAR models only requires the existence of the first moment of $u_{n,i}$. A2(v) imposes a moment condition on $x_{n,i}$.

*Assumption A3.* (i) The kernel function $Q(\cdot)$ is a symmetric PDF such that $Q(\cdot)$ is continuously differentiable and uniformly bounded from above by $c_Q$. (ii) As $n \to \infty$, $\|\lambda\| \to 0$, $\|h\| \to 0$, and $nh_1 \ldots h_{p_c} \to \infty$.

Assumption A3 concerns the kernel and bandwidth sequences. As Li and Racine (2007, chap. 4) remarked, in the nonparametric kernel estimation with mixed data, the optimal choice of smoothing parameters requires that $\|\lambda\|$ is of the same order as $\|h\|^2$ when $p_c \geq 1$, that is, the model has at least one continuous regressor. In this case, the optimal bandwidth rates only depend on the dimension $(p_c)$ of the continuous regressors: the optimal bandwidth rate for the continuous regressors ($h_j$'s) should be proportional to $n^{-1/(4+p_c)}$ and that for the discrete regressors ($\lambda_j$'s) should be proportional to $n^{-2/(4+p_c)}$ when a second-order kernel is used. In the absence of continuous regressors ($p_c = 0$), Ouyang, Li, and Racine (2009) showed that the optimal bandwidth rate for the discrete regressors should be proportional to $n^{-1/2}$. Below we assume $p_c \geq 1$ and remark on the case of $p_c = 0$ after Theorem 2.

### 3.2 Asymptotic Null Distribution

The following assumption is added for the asymptotic null distribution of our test statistic.

*Assumption A4.* $\hat{\theta}_n - \theta_n^0 = o_p\left(n^{-1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{-1/4}\right)$.

Assumption A4 is only imposed under the null hypothesis. Under $\mathbb{H}_0$, A4 requires that $\theta_n^0$ can be consistently estimated by $\hat{\theta}_n$ at any rate faster than $n^{1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{1/4}$. The commonly used 2SLS or GMM estimators (e.g., Kelejian and Prucha 2010; Lin and Lee 2010) typically converge to the true value at the usual $n^{1/2}$-rate and thus meet A4. The QMLE of Lee (2004) also has the $n^{1/2}$-rate of convergence in the regular case. Under some further conditions on both the bandwidths $h_j$'s and the spatial weight matrix $W_n$, the QMLE of Lee (2004) in the irregular case may converge to the true value slower than $n^{1/2}$ but faster than $n^{1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{1/4}$ and thus meets A4 too.

The following theorem studies the asymptotic null distribution of our test statistic $T_n$.

*Theorem 1.* Suppose Assumptions A1–A4 hold. Then under $\mathbb{H}_0$, $T_n \equiv n\left(\Pi_{l=1}^{p_c} h_l\right)^{1/2} J_n/\widehat{\mathbb{S}}_n \xrightarrow{d} N(0,1)$, where $\widehat{\mathbb{S}}_n^2 \equiv 2n^{-2}\Pi_{l=1}^{p_c} h_l \sum_{i=1}^n \sum_{j=1,j\neq i}^n \hat{v}_{n,i}^2 \hat{v}_{n,j}^2 K_{h\lambda,ij}^2$ is a consistent estimator of $\mathbb{S}^2 \equiv \lim_{n\to\infty} 2\int \left[\bar{\sigma}_n^2(x)\right]^2 dx \int Q^2(u)\,du$.

Proof of Theorem 1 can be found in the supplementary appendix. In the case of homoscedasticity, that is, $\sigma_{n,i}^2 = \sigma_0^2$ for all $1 \le i \le n$, we have $\mathbb{S}^2 \equiv \lim_{n\to\infty} 2\sigma_0^4 \int \bar{f}_n^2(x)\,dx \int Q^2(u)\,du$.

Let $z_\alpha$ be the upper $\alpha$-percentile of the standard normal distribution. Noting that $T_n$ is a one-sided test, we reject the null hypothesis when $T_n > z_\alpha$ at the $\alpha$-level of significance.

### 3.3 Local Power

To study the local power property, consider the following sequence of local alternatives:

$$\mathbb{H}_1(\alpha_n): m_n(x) = \beta_{n0}^0 + x'\beta_{n1}^0 + \alpha_n \delta_n(x), \tag{3.1}$$

where $\delta_n(\cdot)$ is an unknown nonlinear function defined on $\mathbb{R}^p$ such that $\lim_{n\to\infty} \int \delta_n^2(x) f_n(x)\,dx > 0$, and $\alpha_n \to 0$ is a scalar that specifies the speed at which the local alternative converges to the null model. Let $\beta_n^0 = (\beta_{n0}^0, \beta_{n1}^{0\prime})'$, $\theta_n^0 = (\beta_n^{0\prime}, \rho_n^0)'$, $\bar{\beta}_{n0}^0 = \beta_{n0}^0 + c_{n0}$, $\bar{\beta}_{n1}^0 = \beta_{n1}^0 + c_{n1}$, and $\bar{\delta}_n(x) = \delta_n(x) - \left(c_{n0} + c_{n1}'x\right)/\alpha_n$, where $c_{n\ell} = o(\alpha_n)$ for $\ell = 0, 1$. Noting that $\beta_{n0}^0 + x'\beta_{n1}^0 + \alpha_n\delta_n(x) = \bar{\beta}_{n0}^0 + x'\bar{\beta}_{n1}^0 + \alpha_n\bar{\delta}_n(x)$, $\beta_n^0$ and thus $\theta_n^0$ can only be identified up to the probability order $o(\alpha_n)$. But this fact does not affect the study of the local power property of our test.

Under $\mathbb{H}_1(\alpha_n)$, $\alpha_n$ affects the convergence speed of the parametric estimator $\hat{\theta}_n$ to $\theta_n^0$. In the following, we will choose $\alpha_n = n^{-1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{-1/4}$ to obtain the nontrivial power for our test. We add the following assumption.

*Assumption A4.* $\hat{\theta}_n - \theta_n^0 = \alpha_n \mathcal{B}_n(\alpha_n) + o_p(\alpha_n)$ with $\mathcal{B}_n \equiv \mathcal{B}_n(\alpha_n) \xrightarrow{p} \mathcal{B}$ where $\mathcal{B}$ is nonrandom.

Assumption A4* will be imposed under $\mathbb{H}_1(\alpha_n)$ for $\alpha_n = o(1)$ such that $n^{1/2}\alpha_n \to \infty$. We now check that the 2SLS estimator of $\theta_n^0$ satisfies Assumption 4* under weak conditions. Let $\tilde{X}_n = (\bar{X}_n, W_nY_n)$ and $P_{Z_n} = Z_n\left(Z_n'Z_n\right)^{-1} Z_n$, where $Z_n \equiv (\bar{X}_n, W_nX_n)$ is the chosen instrument. Denote $\delta_n(X_n) = (\delta_n\left(x_{n,1}\right), \ldots, \delta_n\left(x_{n,n}\right))'$. Then, the 2SLS estimator $\hat{\theta}_{2\text{SLS}}$ of $\theta_n^0$ is given by

$$\begin{aligned}\hat{\theta}_{2\text{SLS}} &= \left(\tilde{X}_n' P_{Z_n}\tilde{X}_n\right)^{-1}\tilde{X}_n' P_{Z_n} Y_n \\ &= \theta_n^0 + \alpha_n\left(\tilde{X}_n' P_{Z_n}\tilde{X}_n\right)^{-1}\tilde{X}_n' P_{Z_n}\delta_n(X_n) \\ &\quad + \left(\tilde{X}_n' P_{Z_n}\tilde{X}_n\right)^{-1}\tilde{X}_n' P_{Z_n} U_n \equiv \theta_n^0 + \alpha_n\mathcal{B}_n + \mathcal{V}_n,\end{aligned} \tag{3.2}$$

where $\mathcal{B}_n$ and $\mathcal{V}_n$ contribute to the bias and variance of $\hat{\theta}_{2\text{SLS}}$, respectively. In the supplementary appendix, we show explicitly that

$$\mathcal{B}_n = O_p(1) \text{ and } \mathcal{V}_n = O_p(n^{-1/2}) = o_p(\alpha_n) \tag{3.3}$$

under fairly weak conditions that are typically assumed for IV estimation of linear SAR models. That is, the 2SLS estimator can easily meet the condition in Assumption 4*. In fact, one can check that this assumption can be satisfied for the general GMM estimators of Kelejian and Prucha (2010) and Lin and Lee (2010) under weak conditions. In C.2 of the supplementary appendix, we verify A4* for Lee's (2004) QMLE under certain conditions.

To proceed, let $\xi_{n,i} \equiv \sum_{j=1}^n g_{n,ij}\bar{x}_{n,j}'\beta_n^0$, where $g_{n,ij}$ is the $(i,j)$ th element of $G_n \equiv W_nS_n^{-1}$. Define $R_{n,\bar{x},\delta} \equiv n^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E[\bar{x}_{n,i}\delta_n(x_{n,j})K_{h\lambda,ij}]$, $R_{n,\xi,\delta} \equiv n^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E[\xi_{n,i} \times \delta_n(x_{n,j})K_{h\lambda,ij}]$, and

$$R_n \equiv \begin{pmatrix} R_{n,\bar{x},\bar{x}} & R_{n,\bar{x},\xi} \\ R_{n,\bar{x},\xi}' & R_{n,\xi,\xi}\end{pmatrix} \equiv \begin{pmatrix} n^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E\left[\bar{x}_{n,i}\bar{x}_{n,i}' K_{h\lambda,ij}\right] n^{-2} \\ \sum_{i=1}^n\sum_{j=1,j\neq i}^n E\left[\bar{x}_{n,i}\xi_{n,j}K_{h\lambda,ij}\right] \\ n^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E\left[\bar{x}_{n,i}'\xi_{n,j}K_{h\lambda,ij}\right] \\ n^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E\left[\xi_{n,i}\xi_{n,j}K_{h\lambda,ij}\right]\end{pmatrix}. \tag{3.4}$$

We make the following assumption concerning the local power of our test.

*Assumption A5.* (i) $\delta_n\left(x^c, x^d\right)$ is continuously differentiable wrt its continuous argument $x^c$ such that $\int \bar{\sigma}_n^2(x)\delta_n^2(x)\bar{f}_n^2(x)\,dx < \infty$ and $\lim_{n\to\infty}\int \delta_n^2(x)\bar{f}_n^2(x)\,dx$ exists and is positive. (ii) $R \equiv \lim_{n\to\infty} R_n$ exists. (iii) $R_{\bar{x},\delta} \equiv \lim_{n\to\infty} R_{n,\bar{x},\delta}$ and $R_{\xi,\delta} \equiv \lim_{n\to\infty} R_{n,\xi,\delta}$ exist.

Under Assumptions A1–A3, we can easily argue that $R_n$, $R_{n,\bar{x},\delta}$, and $R_{n,\xi,\delta}$ are all bounded. Hence, it is reasonable to assume their limits exist in A5(ii)–(iii). We give explicit expressions for these limiting objects in the supplementary appendix (Appendix C.5).

Write $\mathcal{B}_n = (\mathcal{B}_{n1}', \mathcal{B}_{n2})'$ and $\mathcal{B} = (\mathcal{B}_1', \mathcal{B}_2)'$, where $\mathcal{B}_{n1}$ and $\mathcal{B}_1$ are $(p+1)\times 1$ vectors. The following theorem states that our test can distinguish local alternatives $\mathbb{H}_1(\alpha_n)$ at rate $\alpha_n = n^{-1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{-1/4}$.

*Theorem 2.* Suppose that Assumptions A1–A3, A4*, and A5 hold. Then under $\mathbb{H}_1(\alpha_n)$ with $\alpha_n = n^{-1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{-1/4}$, $T_n \equiv n\left(\Pi_{l=1}^{p_c} h_l\right)^{1/2} J_n/\widehat{\mathbb{S}}_n \xrightarrow{d} N(\Delta/\mathbb{S}, 1)$, where $\Delta \equiv \lim_{n\to\infty}\int \delta_n^2(x)\bar{f}_n^2(x)\,dx + \mathcal{B}'R\mathcal{B} - 2\mathcal{B}_1'R_{\bar{x},\delta} - 2\mathcal{B}_2R_{\xi,\delta}$.

Theorem 2 indicates that our test has power to detect local alternatives that converge to the null at the rate $\alpha_n = n^{-1/2}\left(\Pi_{l=1}^{p_c} h_l\right)^{-1/4}$ and this rate only depends on the dimension $(p_c)$ of the continuous regressor and the corresponding bandwidths ($h_l$'s). Here we implicitly assume that $p_c \ge 1$. In the special case of $p_c = 0$, it is easy to argue that our test has power to detect local alternatives that converge to the null at the $n^{-1/2}$-rate but the asymptotic variance $(\mathbb{S}^2)$ of $nJ_n$ would have a much more complicated form than the one given in Theorem 1. In this case, the proof of Lemma B.1 in the appendix breaks down because one cannot apply the central limit theorem of de Jong (1987). Instead, it is well known from the $U$-statistics literature that $T_n$ is no longer asymptotically pivotal under the null and it has an asymptotic mixture $\chi^2$ distribution under the null and local alternatives.

Let $\Delta = \Delta_1 + \Delta_2$, where $\Delta_1 \equiv \lim_{n\to\infty} \int \delta_n^2(x) \bar{f}_n^2(x)\, dx$ and $\Delta_2 \equiv \mathcal{B}' R \mathcal{B} - 2\mathcal{B}_1' R_{x,\delta} - 2\mathcal{B}_2 R_{\xi,\delta}$. Then, $\Delta_1$ is present no matter whether we need to estimate the parameter $\theta_n^0$ in the model or not, whereas $\Delta_2$ reflects the local power due to the estimation error.

## 3.4 Global Power

To study the global power property of our test, we take $\alpha_n = 1$ in ( 3.1) and study the asymptotic property of the test under the global alternative $\mathbb{H}_1(1)$, or $\mathbb{H}_1$ in short. Note that under $\mathbb{H}_1$ there does not exist $\beta_n = (\beta_{n0}, \beta_{n1}')'$ such that we can write $m_n(x) = \beta_{n0} + \beta_{n1}' x$ for almost everywhere $x$ on the support of $x_{n,i}$.

We need to replace Assumption 4* by the following assumption for the estimator $\hat{\theta}_n$.

*Assumption 4***. There exists an $O_p(1)$ object $\theta_n^\dagger$ such that $\hat{\theta}_n - \theta_n^\dagger = o_p(1)$.

Assumption 4** is weak and it essentially requires $\hat{\theta}_n$ to be $O_P(1)$. In the supplementary appendix, we show that the 2SLS estimator $\hat{\theta}_{2\mathrm{SLS}}$ meets the above condition, and

$$\hat{\theta}_{2\mathrm{SLS}} - \theta_n^\dagger = O_p(n^{-1/2}), \tag{3.5}$$

where $\theta_n^\dagger = \theta_n^\ddagger + \bar{\mathcal{B}}$, $\theta_n^\ddagger = (\beta_n^\ddagger, \rho_n^0)$, and $\beta_n^\ddagger = (\bar{X}_n' \bar{X}_n)^{-1} \bar{X}_n' \mathbf{m}_n(X_n) = O_p(1)$. Since the last element of $\bar{\mathcal{B}}$ is generally nonzero, this implies that the 2SLS estimator of the spatial parameter $\rho_n$ is generally biased under $\mathbb{H}_1$. In the supplementary appendix, we give the expression of $\bar{\mathcal{B}}$ and verify Assumption 4** for Lee's QMLE.

Write $\theta_n^\dagger = (\beta_n^{\dagger\prime}, \rho_n^\dagger)' = (\beta_{n0}^\dagger, \beta_{n1}^{\dagger\prime}, \rho_n^\dagger)'$. We redefine $\delta_n(x) = m_n(x) - \beta_{n0}^\dagger - \beta_{n1}^{\dagger\prime} x$. Apparently, $\delta_n(x)$ is a nonconstant function of $x$ under $\mathbb{H}_1$. Our test will be operational provided $\delta_n(X_n)$ and $W_n Y_n$ are not asymptotically multicollinear in the sense there exists a constant $\underline{c}_\delta > 0$ such that $\operatorname{plim}_{n\to\infty} \inf_\gamma n^{-1} \|\delta_n(X_n) - \gamma W_n Y_n\|_{\mathbf{K}_n}^2 \geq \underline{c}_\delta$, where $\|A\|_C^2 = A' C A$ and $\mathbf{K}_n = \{n^{-1} K_{h\lambda,ij}\}$. A sufficient condition for the last inequality to hold is that there does not exist $(\beta_n, \rho_n) \in \mathbb{R}^{p+1} \times \mathbb{R}$ such that

$$\mathbf{m}_n(X_n) + \rho_n^0 W_n Y_n = \bar{X}_n \beta_n + \rho_n W_n Y_n + \mathbf{e}_n$$

with $\operatorname{plim}_{n\to\infty} n^{-1} \|\mathbf{e}_n\|_{\mathbf{K}_n}^2 = 0$. That is, we cannot approximate $\mathbf{m}_n(X_n) + \rho_n^0 W_n Y_n$ by an affine function of $X_n$ and $W_n Y_n$ with an asymptotically negligible error term.

The following theorem establishes the consistency of $T_n$.

*Theorem 3.* Suppose that Assumptions A1–A3, A4**, and A5 hold. Suppose that $\operatorname{plim}_{n\to\infty} \inf_r n^{-1} \|\delta_n(X_n) - \gamma W_n Y_n\|_{\mathbf{K}_n}^2 \geq \underline{c}_\delta > 0$. Then under $\mathbb{H}_1$, $P(T_n > c_n) \to 1$ as $n \to \infty$ for any nonstochastic sequence $c_n$ with $c_n = o(n\left(\Pi_{l=1}^{p_c} h_l\right)^{1/2})$.

Theorem 3 implies that $T_n$ diverges to infinity at the rate $n\left(\Pi_{l=1}^{p_c} h_l\right)^{1/2}$ under the global alternative and thus our test has power to detect any global alternatives.

# 4. A BOOTSTRAP VERSION OF THE TEST

It is well known that nonparametric tests based on asymptotic distributions may perform poorly in finite samples. An alternative approach is to use the bootstrap approximation. In this section, we propose and analyze a wild bootstrap version of our test, following the spirit of Härdle and Mammen (1993).

We construct the bootstrap version of the test statistic $T_n$ :

$$T_n^* = \frac{\left(\Pi_{l=1}^{p_c} h_l\right)^{1/2}}{n \widehat{\mathbb{S}}_n^*} \sum_{i=1}^n \sum_{j=1, j\neq i}^n \hat{v}_{n,i}^* \hat{v}_{n,j}^* K_{h\lambda,ij}, \tag{4.1}$$

where $\hat{v}_{n,i}^*$'s are bootstrap residuals and

$$\widehat{\mathbb{S}}_n^* = \left\{ n^{-2} \left(\Pi_{l=1}^{p_c} h_l\right) \sum_{i=1}^n \sum_{j=1, j\neq i}^n \hat{v}_{n,i}^{*2} \hat{v}_{n,j}^{*2} K_{h\lambda,ij}^2 \right\}^{1/2}. \tag{4.2}$$

To construct $T_n^*$, first we generate the wild bootstrap error $v_{n,i}^* = \hat{v}_{n,i} \eta_i$, where $\eta_i$'s are iid, independent of the process $\{y_{n,i}, x_{n,i}\}$, and satisfy the conditions: $E(\eta_i) = 0$, $E\left(\eta_i^2\right) = 1$, and $\mu_{\eta 4} \equiv E\left(\eta_i^4\right) < \infty$. There are many ways to obtain such a sequence $\{\eta_i\}$. In our simulation, we draw them independently from a distribution with masses $c = \frac{1+\sqrt{5}}{2\sqrt{5}}$ and $1 - c$ at the points $\frac{1-\sqrt{5}}{2}$ and $\frac{1+\sqrt{5}}{2}$, respectively. Hence, the wild bootstrap draws each $v_{n,i}^*$ from a distribution with mean zero and variance $\hat{v}_{n,i}^2$ conditional on the data. Second, we generate the bootstrap resample $Y_n^*$ by $Y_n^* = (I_n - \hat{\rho}_n W_n)^+ (\bar{X}_n \hat{\beta}_n + v_n^*)$, where $(\cdot)^+$ denotes the Moore–Penrose generalized inverse, and have $\hat{\theta}_n^*$ as the estimator of $\hat{\theta}_n$ based on the bootstrap resample $\{Y_n^*, X_n\}$. For simplicity, we focus on the 2SLS estimator: $\hat{\theta}_n^* = \left(\tilde{X}_n^{*\prime} P_{Z_n} \tilde{X}_n^*\right)^{-1} \tilde{X}_n^{*\prime} P_{Z_n} Y_n^*$, where $P_{Z_n}$ is the same as Section 3.3 and $\tilde{X}_n^* = (\bar{X}_n, W_n Y_n^*)$. Then, we obtain the bootstrap residual $\hat{V}_n^* = Y_n^* - \tilde{X}_n^* \hat{\theta}_n^*$ to construct our $T_n^*$.

This procedure is repeated $B$ times to obtain the sequence $\{T_{n,j}^*\}_{j=1}^B$. We reject the null when $p^* = B^{-1} \sum_{j=1}^B \mathbf{1}(T_n \leq T_{n,j}^*)$ is smaller than the given level of significance.

For the asymptotic validity of the bootstrap method, we add the following assumption.

*Assumption A6.* (i) There exists a nonrandom vector $\bar{\theta}_n \equiv (\bar{\beta}_n', \bar{\rho}_n)'$ such that $\hat{\theta}_n - \bar{\theta}_n \xrightarrow{p} 0$, $\sup_n |\bar{\rho}_n| < \rho_*$, $\sup_n \left\|\rho_* W_n'\right\|_1 < 1$, $\sup_n \left\|S_n^{-1}(\bar{\rho}_n)\right\|_1 < \infty$, and $\sup_n \left\|S_n^{-1}(\bar{\rho}_n)'\right\|_1 < \infty$; (ii) $\lim_{n\to\infty} n^{-1} Z_n' Z_n$ exists and is nonsingular; (iii) $\lim_{n\to\infty} (n^{-1} Z_n' \bar{X}_n,\, n^{-1} Z_n' \bar{G}_n \bar{X}_n \bar{\beta}_n)$ exists and is of full rank; (iv) There exist a constant $c_m < \infty$ such that $\sup_{i,n} E\left|m_n\left(x_{n,i}\right)\right|^4 < c_m$.

If the linear SAR model is correctly specified or approximately correctly specified (i.e., $\alpha_n \to 0$ under $\mathbb{H}_1(\alpha_n)$), Assumption A6(i) is implied by A4 or A4* with $\bar{\theta}_n = \theta_n^0$. In the case of misspecification, A6(i) mainly requires that $\hat{\theta}_n$ should converge to some nonrandom object $\bar{\theta}_n$ in the sense that $\hat{\theta}_n - \bar{\theta}_n = o_p(1)$. To relate this to A4**, we can simply choose $\bar{\theta}_n = \theta_n^\ddagger + \bar{\mathcal{B}}$. A6(ii) is standard in the IV estimation of SAR models. A6(iii) requires that $\lim_{n\to\infty} n^{-1} Z_n' \bar{X}_n$ be of full rank, $\lim_{n\to\infty} n^{-1} Z_n' \bar{G}_n \bar{X}_n \bar{\beta}_n \neq 0$, and the latter limit be linearly independent of the former one. So it rules out the case where $\bar{\beta}_n = 0$ or $Z_n'\bar{X}_n$ and $Z_n'\bar{G}_n\bar{X}_n\bar{\beta}_n$ are asymptotically multicollinear. A6(iv) imposes some moment conditions.

*Theorem 4.* Suppose Assumptions A1–A3, A5, and A6 hold. Then, conditional on the original sample, $T_n^* \xrightarrow{d} N(0, 1)$.

Theorem 4 implies that $T_n^* \xrightarrow{d} N(0, 1)$ unconditionally no matter whether the null hypothesis holds or not for the original sample. On one hand, under $\mathbb{H}_0$, both $T_n$ and $T_n^*$ converge in distribution to $N(0, 1)$. For this reason, a test based on the bootstrap $p$-value would yield the correct asymptotic level for $T_n$. On the other hand, under $\mathbb{H}_1$, $T_n$ diverges to infinity in probability by Theorem 3 whereas Theorem 4 ensures that $T_n^* \xrightarrow{d} N(0, 1)$ unconditionally. It follows that the test based upon the bootstrap $p$-value is consistent against all global alternatives. Therefore, the wild bootstrap test is asymptotically valid.

## 5. MONTE CARLO SIMULATIONS

We now examine the finite sample performance of our test using Monte Carlo experiments.

### 5.1 Test With Continuous Regressors Only

To study the size behavior of our test, we generate the null models as:

DGP 1: $Y_n = \mathbf{1}_n + X_{n1} + 0.5W_nY_n + U_n$,
DGP 2: $Y_n = \mathbf{1}_n + X_{n1} + X_{n2} + 0.5W_nY_n + U_n$,

where $X_{n1} = (x_{n,11}, \ldots, x_{n,1n})'$, $X_{n2} = (x_{n,21}, \ldots, x_{n,2n})'$, $x_{n,1i}$'s are iid and each is equal to the sum of 48 independent random variables each uniformly distributed on [−0.25, 0.25], and $x_{n,2i}$'s are iid $U(-2, 2)$. According to the central limit theorem, we can treat $x_{n,1i}$ as being nearly a normal random variable with truncated support on [−12, 12]. For the error term, we generate $u_{n,i}$'s in two ways: (a) $u_{n,i}$'s are iid $N(0, 1)$; (b) $u_{n,i} = \sqrt{0.5(1 + x_{n,1i}^2)}\varepsilon_{n,i}$ in DGP 1 and $u_{n,i} = \sqrt{0.5(1 + x_{n,1i}^2 + x_{n,2i}^2)}\varepsilon_{n,i}$ in DGP 2, where $\varepsilon_{n,i}$'s are iid $N(0, 1)$. Clearly, the error terms are homoscedastic in case (a) and heteroscedastic in case (b).

To study the power behavior of our test, we generate the alternative models as:

DGP 3: $Y_n = \mathbf{1}_n + X_{n1} + 0.5W_nY_n + 0.5X_{n1}^2 + U_n$,
DGP 4: $Y_n = \mathbf{1}_n + X_{n1} + 0.5W_nY_n + 0.5\exp(X_{n1}) + U_n$,
DGP 5: $Y_n = \mathbf{1}_n + X_{n1} + X_{n2} + 0.5W_nY_n + 0.5X_{n1}^2 + U_n$,
DGP 6: $Y_n = \mathbf{1}_n + X_{n1} + X_{n2} + 0.5W_nY_n + 0.5\exp(X_{n2}) + U_n$.

To implement our test, we choose either the Gaussian kernel or the product of Gaussian kernels. As it is difficult to specify the optimal smoothing parameters in the framework of hypothesis testing, one natural choice of $h = (h_1, \ldots, h_{p_c})$ ($p_c = 1$ or 2 here) is to follow Silverman's normal reference rule of thumb (ROT) and set $h_i = s_{X_{ni}} n^{-1/(p_c+4)}$, where $s_{X_{ni}}$ is the sample standard deviation of $X_{ni}$. Alternatively, one can consider choosing $h$ by minimizing some LSCV function.

Recently, Su (2012) proposed a nonparametric GMM estimation of the semiparametric SAR models of the form $Y_n = \rho_n^0 W_nY_n + \mathbf{m}(X_n) + U_n$, where $\mathbf{m}(X_n) = (m(x_{n,1}), \ldots, m(x_{n,n}))'$, the functional form of $m(\cdot)$ is unknown, and $x_{n,i}$ can contain both continuous and discrete regressors. His estimator $\tilde{\rho}_n$ of $\rho_n$ is easy to implement due to the closed form expression. Su shows that $\sqrt{n}(\tilde{\rho}_n - \rho_n^0)$ is asymptotically normally distributed under some regularity conditions. This implies that we can conduct the LSCV by regressing $\tilde{Y}_n = Y_n - \tilde{\rho}_n W_nY_n$ on $X_n$. Let $\hat{m}_{-i}(x_{n,i})$ be the leave-one-out local constant estimator of $m(x_{n,i})$ by leaving the observation $(x_{n,i}, \tilde{y}_{n,i})$ out in the estimation procedure and by using the smoothing parameters $(h, \lambda)$. We can choose $(h, \lambda)$ to minimize the LSCV objective function

$$\mathrm{CV}(h, \lambda) = n^{-1}\sum_{i=1}^{n}\left[\tilde{y}_{n,i} - \hat{m}_{-i}(x_{n,i})\right]^2 w(x_{n,i}^c),$$

where $w(x_{n,i}^c)$ is a nonnegative weight function that has compact support. The use of a compactly supported weight function can mitigate the boundary bias problems and ensures that $\mathrm{CV}(h, \lambda)$ is well behaved asymptotically. We follow the nonparametric literature and set $w(x_{n,i}^c) = \Pi_{s=1}^{p_c}\mathbf{1}(|x_{n,s}^c - \bar{x}_s^c| \leq 2s_{X_{ni}^c})$, where $\bar{x}_s^c$ is the sample mean of $X_{ns}^c$.

To check the robustness of our tests regarding different spatial weights, we construct two sets of $W_n$'s. The first set is like Su and Yang (2015): we generate the spatial weights according to Rook contiguity, by randomly allocating the $n$ spatial units on a lattice of $5 \times \frac{n}{5}$ squares with $n$ =50, 100, and 200, finding the neighbors for each unit, and row normalizing. The second set is like Case (1991) used in Lee (2004): we set $W_n = I_{30} \otimes (1/(m-1))(l_m l_m' - I_m)$ and consider $m = 3$, 5, and 10. In all cases, we consider 1000 Monte Carlo replications and 200 bootstrap replications. We report the rejection frequencies of tests based on the asymptotic result and the bootstrap result.

Table 1 reports the empirical rejection frequencies when nominal levels are given by 0.01, 0.05, and 0.10. To see how the bootstrap test improves the small sample behaviors, for each parameter configuration, we have two sets of results: the first row of each quadrant is from the bootstrap test and the second row is from the asymptotic test. These results are based on the LSCV bandwidths. Results from Silverman's ROT smoothing parameters tend to have smaller empirical sizes, but other patterns are similar. We summarize some important findings from Table 1. First, we see that for both DGP 1 and DGP 2 with either homoscedastic or heteroscedastic errors, our bootstrap test is moderately oversized while the asymptotic test tends to be undersized. The sizes of both types of tests improve generally as the sample size increases. Second, both bootstrap and asymptotic tests are powerful in detecting any deviations from linearity in DGPs 3–6. The power is reasonably high for sample sizes as small as $n$ =100. Note that the power for the heteroscedasticity case is smaller than that for the homoscedasticity case. This is expected because in the presence of heteroscedasticity in DGPs 3–6, the signal/noise ratio is smaller than the case of homoscedasticity. In general, the larger the signal-to-noise ratio, the easier to detect any deviations from the null models, holding everything else fixed. These patterns are robust to different choices of spatial weight matrices.

Table 1. Empirical rejection frequency for DGPs 1–6

| | Rook contiguity $W_n$ | | | | | | $W_n = I_{30} \otimes (1/(m-1))(l_m l_m' - I_m)$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Homoscedasticity | | | Heteroscedasticity | | | Homoscedasticity | | | Heteroscedasticity | | |
| DGP | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 |
| | | | $n = 50$ | | | | | | $m = 3$ ($n = 90$) | | | |
| 1 | 0.013 | 0.069 | 0.140 | 0.024 | 0.076 | 0.136 | 0.025 | 0.082 | 0.154 | 0.025 | 0.093 | 0.164 |
| | 0.005 | 0.021 | 0.039 | 0.018 | 0.036 | 0.057 | 0.008 | 0.028 | 0.055 | 0.020 | 0.055 | 0.088 |
| 2 | 0.024 | 0.073 | 0.146 | 0.026 | 0.088 | 0.160 | 0.039 | 0.101 | 0.162 | 0.036 | 0.098 | 0.180 |
| | 0.006 | 0.026 | 0.052 | 0.012 | 0.040 | 0.061 | 0.009 | 0.037 | 0.066 | 0.017 | 0.047 | 0.081 |
| 3 | 0.419 | 0.672 | 0.791 | 0.464 | 0.674 | 0.787 | 0.807 | 0.925 | 0.967 | 0.794 | 0.929 | 0.967 |
| | 0.344 | 0.515 | 0.611 | 0.450 | 0.596 | 0.670 | 0.803 | 0.889 | 0.925 | 0.871 | 0.921 | 0.948 |
| 4 | 0.188 | 0.381 | 0.493 | 0.197 | 0.389 | 0.507 | 0.444 | 0.672 | 0.783 | 0.458 | 0.674 | 0.773 |
| | 0.191 | 0.303 | 0.389 | 0.264 | 0.388 | 0.452 | 0.467 | 0.601 | 0.690 | 0.565 | 0.687 | 0.745 |
| 5 | 0.277 | 0.509 | 0.645 | 0.199 | 0.400 | 0.527 | 0.663 | 0.841 | 0.903 | 0.470 | 0.697 | 0.793 |
| | 0.194 | 0.333 | 0.424 | 0.125 | 0.245 | 0.333 | 0.567 | 0.755 | 0.814 | 0.403 | 0.571 | 0.662 |
| 6 | 0.177 | 0.412 | 0.540 | 0.136 | 0.325 | 0.452 | 0.475 | 0.655 | 0.774 | 0.330 | 0.536 | 0.657 |
| | 0.099 | 0.208 | 0.305 | 0.064 | 0.155 | 0.231 | 0.311 | 0.495 | 0.590 | 0.198 | 0.366 | 0.466 |
| | | | $n = 100$ | | | | | | $m = 5(n = 150)$ | | | |
| 1 | 0.035 | 0.077 | 0.136 | 0.025 | 0.077 | 0.130 | 0.021 | 0.079 | 0.141 | 0.022 | 0.070 | 0.136 |
| | 0.016 | 0.035 | 0.041 | 0.017 | 0.040 | 0.070 | 0.009 | 0.031 | 0.053 | 0.019 | 0.041 | 0.064 |
| 2 | 0.021 | 0.085 | 0.147 | 0.029 | 0.090 | 0.166 | 0.024 | 0.083 | 0.164 | 0.024 | 0.091 | 0.175 |
| | 0.005 | 0.023 | 0.047 | 0.015 | 0.045 | 0.064 | 0.006 | 0.022 | 0.050 | 0.007 | 0.034 | 0.066 |
| 3 | 0.860 | 0.951 | 0.974 | 0.848 | 0.941 | 0.974 | 0.981 | 0.994 | 0.997 | 0.975 | 0.996 | 0.998 |
| | 0.845 | 0.912 | 0.945 | 0.887 | 0.935 | 0.957 | 0.978 | 0.993 | 0.995 | 0.990 | 0.996 | 0.998 |
| 4 | 0.486 | 0.706 | 0.809 | 0.492 | 0.690 | 0.796 | 0.725 | 0.884 | 0.943 | 0.714 | 0.868 | 0.931 |
| | 0.500 | 0.645 | 0.719 | 0.586 | 0.706 | 0.763 | 0.738 | 0.830 | 0.888 | 0.792 | 0.878 | 0.917 |
| 5 | 0.692 | 0.876 | 0.933 | 0.490 | 0.706 | 0.811 | 0.923 | 0.978 | 0.991 | 0.753 | 0.904 | 0.948 |
| | 0.605 | 0.782 | 0.859 | 0.412 | 0.578 | 0.680 | 0.897 | 0.954 | 0.976 | 0.704 | 0.836 | 0.884 |
| 6 | 0.491 | 0.705 | 0.800 | 0.327 | 0.538 | 0.672 | 0.749 | 0.878 | 0.938 | 0.535 | 0.735 | 0.817 |
| | 0.331 | 0.507 | 0.622 | 0.214 | 0.353 | 0.452 | 0.634 | 0.788 | 0.842 | 0.398 | 0.583 | 0.692 |
| | | | $n = 200$ | | | | | | $m = 10(n = 300)$ | | | |
| 1 | 0.029 | 0.070 | 0.133 | 0.024 | 0.071 | 0.133 | 0.026 | 0.072 | 0.120 | 0.020 | 0.062 | 0.124 |
| | 0.015 | 0.029 | 0.051 | 0.021 | 0.055 | 0.076 | 0.012 | 0.026 | 0.038 | 0.019 | 0.038 | 0.068 |
| 2 | 0.026 | 0.075 | 0.158 | 0.029 | 0.079 | 0.156 | 0.025 | 0.077 | 0.143 | 0.022 | 0.078 | 0.156 |
| | 0.009 | 0.026 | 0.057 | 0.010 | 0.036 | 0.077 | 0.010 | 0.027 | 0.051 | 0.010 | 0.036 | 0.073 |
| 3 | 0.997 | 0.999 | 1 | 0.996 | 0.998 | 0.999 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 0.998 | 0.999 | 0.999 | 0.998 | 0.999 | 0.999 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 0.866 | 0.959 | 0.980 | 0.864 | 0.940 | 0.974 | 0.976 | 0.994 | 0.998 | 0.975 | 0.991 | 0.995 |
| | 0.883 | 0.943 | 0.963 | 0.923 | 0.957 | 0.978 | 0.979 | 0.991 | 0.996 | 0.988 | 0.994 | 0.994 |
| 5 | 0.979 | 0.996 | 0.999 | 0.900 | 0.970 | 0.982 | 1 | 1 | 1 | 0.983 | 0.997 | 0.999 |
| | 0.972 | 0.987 | 0.995 | 0.873 | 0.946 | 0.972 | 1 | 1 | 1 | 0.983 | 0.995 | 0.998 |
| 6 | 0.893 | 0.959 | 0.976 | 0.700 | 0.861 | 0.916 | 0.982 | 0.993 | 0.998 | 0.891 | 0.957 | 0.984 |
| | 0.829 | 0.917 | 0.947 | 0.598 | 0.758 | 0.837 | 0.970 | 0.988 | 0.995 | 0.826 | 0.923 | 0.955 |

## 5.2 Test With Both Continuous and Discrete Regressors

In this subsection, we examine how our test behaves in finite samples when both continuous and discrete regressors are present in the regression. The null models are

DGP m1: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + 0.5W_nY_n + U_n$,

DGP m2: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + X_{n2}^c + 0.5W_nY_n + U_n$, where $X_{n1}^c$ and $X_{n2}^c$ are generated in the same way as $X_{n1}$ and $X_{n2}$ were generated in DGP 2, $U_n \sim N(0, I_n)$, and for $t = 1, 2$, $X_{nt}^d = \left(x_{n,t1}^d, \ldots, x_{n,tn}^d\right)'$, $P(x_{n,t1}^d = l) = 0.5$ for $l = 0, 1$.

We consider four alternative models:

DGP m3: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + 0.5W_nY_n + X_{n1}^{c2} + U_n$,

DGP m4: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + 0.5W_nY_n + X_{n1}^d X_{n1}^c + X_{n2}^d X_{n1}^{c2} + U_n$,

DGP m5: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + X_{n2}^c + 0.5W_nY_n + X_{n1}^{c2} + U_n$,

DGP m6: $Y_n = \mathbf{1}_n + X_{n1}^d + X_{n2}^d + X_{n1}^c + X_{n2}^c + 0.5W_nY_n + X_{n1}^d \cos\left(\pi X_{n1}^c\right) + X_{n2}^d \sin(\pi X_{n2}^c) + U_n$.

For our alternatives, DGPs m3 and m5 deviate from the null in extra squared terms of the exogenous regressor $X_{n1}^c$; DGP m4 has additional interaction terms between discrete and continuous regressors; DGP m6 uses the trigonometric function

Table 2. Empirical rejection frequency for DGPs m1–m6 under Rook contiguity weight

| | (i) LSCV $h, \lambda$ | | | (ii) LSCV $h, \lambda = 0$ | | | (iii) ROT $h, \lambda = 0$ | | | (iv) ROT $h, \lambda$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 |
| | | | | | | $n = 50$ | | | | | | |
| m1 | 0.011 | 0.031 | 0.065 | 0.006 | 0.027 | 0.054 | 0.004 | 0.017 | 0.035 | 0.003 | 0.015 | 0.031 |
| | 0.009 | 0.020 | 0.028 | 0.003 | 0.011 | 0.017 | 0.001 | 0.006 | 0.009 | 0 | 0.002 | 0.008 |
| m2 | 0.014 | 0.055 | 0.097 | 0.008 | 0.035 | 0.071 | 0.005 | 0.03 | 0.056 | 0.02 | 0.026 | 0.053 |
| | 0.007 | 0.034 | 0.062 | 0.003 | 0.014 | 0.036 | 0.002 | 0.007 | 0.023 | 0 | 0.004 | 0.015 |
| m3 | 0.460 | 0.666 | 0.747 | 0.332 | 0.567 | 0.677 | 0.345 | 0.561 | 0.671 | 0.402 | 0.616 | 0.715 |
| | 0.762 | 0.876 | 0.925 | 0.521 | 0.729 | 0.814 | 0.544 | 0.741 | 0.827 | 0.687 | 0.833 | 0.886 |
| m4 | 0.349 | 0.539 | 0.664 | 0.306 | 0.527 | 0.645 | 0.325 | 0.519 | 0.638 | 0.324 | 0.529 | 0.650 |
| | 0.390 | 0.594 | 0.681 | 0.366 | 0.554 | 0.647 | 0.368 | 0.548 | 0.635 | 0.391 | 0.572 | 0.655 |
| m5 | 0.416 | 0.626 | 0.733 | 0.267 | 0.474 | 0.590 | 0.217 | 0.434 | 0.560 | 0.336 | 0.549 | 0.652 |
| | 0.568 | 0.753 | 0.842 | 0.299 | 0.540 | 0.668 | 0.265 | 0.480 | 0.617 | 0.425 | 0.658 | 0.754 |
| m6 | 0.090 | 0.230 | 0.340 | 0.035 | 0.114 | 0.201 | 0.026 | 0.072 | 0.136 | 0.021 | 0.079 | 0.152 |
| | 0.045 | 0.170 | 0.281 | 0.013 | 0.076 | 0.139 | 0.004 | 0.031 | 0.067 | 0.003 | 0.026 | 0.064 |
| | | | | | | $n = 100$ | | | | | | |
| m1 | 0.006 | 0.037 | 0.071 | 0.008 | 0.04 | 0.073 | 0.005 | 0.026 | 0.059 | 0.005 | 0.025 | 0.044 |
| | 0.003 | 0.018 | 0.032 | 0.001 | 0.017 | 0.028 | 0 | 0.011 | 0.020 | 0 | 0.005 | 0.015 |
| m2 | 0.013 | 0.064 | 0.117 | 0.012 | 0.044 | 0.086 | 0.004 | 0.038 | 0.067 | 0.006 | 0.035 | 0.068 |
| | 0.005 | 0.024 | 0.049 | 0.004 | 0.013 | 0.032 | 0.003 | 0.008 | 0.022 | 0.001 | 0.008 | 0.019 |
| m3 | 0.845 | 0.923 | 0.948 | 0.824 | 0.920 | 0.940 | 0.809 | 0.907 | 0.939 | 0.826 | 0.908 | 0.944 |
| | 0.990 | 0.998 | 1 | 0.982 | 0.992 | 0.996 | 0.982 | 0.993 | 0.997 | 0.990 | 0.997 | 0.998 |
| m4 | 0.844 | 0.931 | 0.972 | 0.835 | 0.943 | 0.968 | 0.839 | 0.938 | 0.970 | 0.839 | 0.932 | 0.971 |
| | 0.906 | 0.951 | 0.972 | 0.903 | 0.949 | 0.968 | 0.910 | 0.949 | 0.967 | 0.913 | 0.953 | 0.970 |
| m5 | 0.882 | 0.953 | 0.974 | 0.839 | 0.930 | 0.960 | 0.799 | 0.916 | 0.955 | 0.860 | 0.931 | 0.964 |
| | 0.981 | 0.992 | 0.996 | 0.941 | 0.975 | 0.983 | 0.917 | 0.966 | 0.982 | 0.970 | 0.983 | 0.989 |
| m6 | 0.373 | 0.612 | 0.735 | 0.228 | 0.466 | 0.568 | 0.143 | 0.365 | 0.500 | 0.180 | 0.400 | 0.554 |
| | 0.295 | 0.528 | 0.649 | 0.149 | 0.352 | 0.460 | 0.078 | 0.243 | 0.367 | 0.095 | 0.268 | 0.388 |
| | | | | | | $n = 200$ | | | | | | |
| m1 | 0.012 | 0.044 | 0.081 | 0.014 | 0.044 | 0.085 | 0.005 | 0.034 | 0.071 | 0.005 | 0.032 | 0.064 |
| | 0.002 | 0.014 | 0.033 | 0.004 | 0.016 | 0.036 | 0.002 | 0.012 | 0.026 | 0.001 | 0.006 | 0.017 |
| m2 | 0.017 | 0.062 | 0.109 | 0.016 | 0.050 | 0.096 | 0.009 | 0.036 | 0.073 | 0.013 | 0.035 | 0.072 |
| | 0.004 | 0.027 | 0.049 | 0.003 | 0.020 | 0.038 | 0.003 | 0.014 | 0.028 | 0.002 | 0.013 | 0.024 |
| m3 | 0.988 | 0.996 | 0.998 | 0.988 | 0.997 | 0.998 | 0.988 | 0.995 | 0.998 | 0.989 | 0.996 | 0.998 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m4 | 1 | 1 | 1 | 1 | 1 | 1 | 0.998 | 1 | 1 | 0.997 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m5 | 0.997 | 0.999 | 0.999 | 0.998 | 0.999 | 0.999 | 0.995 | 0.999 | 0.999 | 0.996 | 0.999 | 0.999 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m6 | 0.916 | 0.973 | 0.988 | 0.823 | 0.937 | 0.965 | 0.795 | 0.917 | 0.964 | 0.831 | 0.943 | 0.978 |
| | 0.877 | 0.964 | 0.979 | 0.784 | 0.892 | 0.944 | 0.722 | 0.868 | 0.931 | 0.767 | 0.905 | 0.947 |

as in Lavergne (2001) and Hsiao, Li, and Racine (2007). This functional form is unusual in economic settings despite its wide use in the statistics literature.

We compare the size and power performances of our bootstrap (first row of each quadrant) and asymptotic (second row of each quadrant) tests for four different choices of smoothing parameters: (i) the proposed LSCV smoothing parameters $(h, \lambda)$; (ii) $h$ is chosen by the usual LSCV method and $\lambda = 0$; (iii) Silverman's ROT $h$ and $\lambda = 0$; (iv) Silverman's ROT $h$ and $\lambda$ is also chosen by some rule of thumb: $\lambda_s = s_{X^d_{ns}} n^{-2/(p_c+4)}$, where $s_{X^d_{ns}}$ is the sample standard deviation of $X^d_{ns}$. Note that tests based on (ii) and (iii) are frequency-based tests and are expected to be less powerful than those based on (i) and (iv), respectively.

Tables 2 and 3 report the empirical rejection frequencies of our test by using the above four different sets of smoothing parameters under the Rook contiguity $W_n$ and Case's $W_n$, respectively. We now summarize some important findings. First, in all cases, the test based upon the LSCV choices of both $h$ and $\lambda$ performs best in terms of size precision and power. Furthermore, one can see that the test based upon (i) is more powerful than that based upon (ii) across all cases under our investigation. Similarly, the test based upon (iv) is more powerful than that based upon (iii). In short, smoothing the discrete regressors helps improving the power of the test, and the largest improvement occurs in DGP m6 that has the unusual trigonometric function. In addition, as Hall, Li, and Racine (2007) had shown, LSCV helps to smooth out irrelevant regressors asymptotically as the LSCV-based smoothing parameters for the irrelevant components converge in probability to the upper extremities of their respective ranges. Second, the power increases as the sample size increases in all cases. Compared to the Rook contiguity

Table 3. Empirical rejection frequency for DGPs m1–m6 under Case's weight

| | (i) LSCV $h, \lambda$ | | | (ii) LSCV $h, \lambda = 0$ | | | (iii) ROT $h, \lambda = 0$ | | | (iv) ROT $h, \lambda$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 |
| | | | | | | $m = 3$ | | | | | | |
| m1 | 0.018 | 0.055 | 0.099 | 0.015 | 0.048 | 0.089 | 0.011 | 0.038 | 0.077 | 0.008 | 0.029 | 0.073 |
| | 0.006 | 0.024 | 0.037 | 0.005 | 0.018 | 0.027 | 0.004 | 0.015 | 0.021 | 0 | 0.011 | 0.018 |
| m2 | 0.017 | 0.066 | 0.112 | 0.015 | 0.056 | 0.106 | 0.012 | 0.039 | 0.086 | 0.007 | 0.034 | 0.077 |
| | 0.003 | 0.024 | 0.051 | 0.005 | 0.021 | 0.034 | 0.003 | 0.013 | 0.031 | 0.002 | 0.008 | 0.021 |
| m3 | 0.975 | 0.997 | 0.998 | 0.964 | 0.988 | 0.998 | 0.963 | 0.989 | 0.998 | 0.978 | 0.995 | 0.998 |
| | 0.990 | 0.995 | 0.998 | 0.974 | 0.994 | 0.996 | 0.978 | 0.994 | 0.996 | 0.992 | 0.996 | 0.998 |
| m4 | 0.894 | 0.960 | 0.978 | 0.878 | 0.964 | 0.980 | 0.890 | 0.963 | 0.976 | 0.906 | 0.965 | 0.985 |
| | 0.860 | 0.939 | 0.964 | 0.858 | 0.934 | 0.964 | 0.863 | 0.935 | 0.962 | 0.877 | 0.943 | 0.969 |
| m5 | 0.959 | 0.991 | 0.996 | 0.896 | 0.975 | 0.983 | 0.858 | 0.945 | 0.980 | 0.925 | 0.980 | 0.992 |
| | 0.970 | 0.994 | 0.996 | 0.891 | 0.969 | 0.984 | 0.835 | 0.940 | 0.974 | 0.928 | 0.984 | 0.988 |
| m6 | 0.372 | 0.565 | 0.682 | 0.235 | 0.403 | 0.492 | 0.151 | 0.330 | 0.454 | 0.172 | 0.364 | 0.498 |
| | 0.231 | 0.440 | 0.563 | 0.124 | 0.259 | 0.363 | 0.056 | 0.175 | 0.278 | 0.065 | 0.190 | 0.304 |
| | | | | | | $m = 5$ | | | | | | |
| m1 | 0.009 | 0.036 | 0.070 | 0.009 | 0.043 | 0.075 | 0.006 | 0.028 | 0.060 | 0.005 | 0.023 | 0.061 |
| | 0.002 | 0.013 | 0.029 | 0.001 | 0.011 | 0.025 | 0 | 0.004 | 0.012 | 0 | 0.003 | 0.010 |
| m2 | 0.017 | 0.045 | 0.097 | 0.011 | 0.037 | 0.075 | 0.013 | 0.033 | 0.059 | 0.009 | 0.027 | 0.057 |
| | 0.005 | 0.021 | 0.038 | 0.005 | 0.014 | 0.030 | 0.002 | 0.011 | 0.020 | 0.002 | 0.010 | 0.016 |
| m3 | 1 | 1 | 1 | 1 | 1 | 1 | 0.999 | 1 | 1 | 0.998 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m4 | 0.995 | 1 | 1 | 0.993 | 0.999 | 1 | 0.995 | 1 | 1 | 0.995 | 1 | 1 |
| | 0.997 | 1 | 1 | 0.996 | 1 | 1 | 0.996 | 1 | 1 | 0.997 | 1 | 1 |
| m5 | 1 | 1 | 1 | 0.999 | 1 | 1 | 0.998 | 0.999 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 0.996 | 1 | 1 | 1 | 1 | 1 |
| m6 | 0.730 | 0.879 | 0.927 | 0.576 | 0.767 | 0.850 | 0.489 | 0.735 | 0.833 | 0.557 | 0.784 | 0.862 |
| | 0.644 | 0.809 | 0.870 | 0.472 | 0.663 | 0.756 | 0.357 | 0.579 | 0.718 | 0.414 | 0.636 | 0.759 |
| | | | | | | $m = 10$ | | | | | | |
| m1 | 0.006 | 0.040 | 0.080 | 0.008 | 0.040 | 0.079 | 0.007 | 0.034 | 0.067 | 0.007 | 0.032 | 0.060 |
| | 0.002 | 0.012 | 0.029 | 0.001 | 0.011 | 0.029 | 0.001 | 0.007 | 0.023 | 0.001 | 0.008 | 0.018 |
| m2 | 0.024 | 0.053 | 0.110 | 0.015 | 0.055 | 0.095 | 0.015 | 0.038 | 0.077 | 0.017 | 0.038 | 0.072 |
| | 0.009 | 0.030 | 0.046 | 0.007 | 0.026 | 0.043 | 0.006 | 0.022 | 0.034 | 0.005 | 0.018 | 0.034 |
| m3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| m6 | 0.998 | 0.999 | 0.999 | 0.993 | 0.999 | 0.999 | 0.988 | 0.999 | 0.999 | 0.992 | 0.998 | 0.999 |
| | 0.997 | 0.999 | 1 | 0.990 | 0.996 | 0.999 | 0.986 | 0.996 | 0.999 | 0.988 | 0.999 | 1 |

$W_n$, powers are larger when Case's $W_n$ is adopted due to the more sparse spatial weight matrices. Third, these patterns are robust for both the bootstrap and asymptotic tests. The asymptotic tests tend to be undersized while the bootstrap tests greatly improve the size performances.

## 6. EMPIRICAL APPLICATION

In this section, we consider two empirical examples by using the specification test presented in previous sections.

### 6.1 Vote Cast Example

The first example is the study on votes cast. Voting rate is an important indicator of how the elected government or president represents the will of people. As we see from the famous map describing people's distribution of republicans and democrats in the United States, their location is spatially clustered. However, including unit level attributes (state or individual level) in the traditional regression cannot remove this correlation. Therefore, more and more spatial econometrics models have been adopted in modeling election behavior. For example, Kim, Elliott, and Wang (2003) used the U.S. county-level presidential election outcomes from 1988 to 2000 to identify spatial patterns of voting behavior. Shin and Agnew (2007) introduced the spatial analysis of Moran's I statistic to investigate the geographical dynamics of Italian electoral change from 1987 to 1994. Soares and Terron (2008) adopted an SAR model to study the voting of Lula in Brazil in 2006. All these studies are based on linear spatial models, so we want to test the model specification using the vote cast dataset from Pace and Barry (1997). In total, there are 3111 counties (or their equivalents) in the

Table 4. The estimates and test statistics of the linear SAR model for the vote data

| Regression based on the log-ratio specification | | | Regression based on the ratio specification | | |
|---|---|---|---|---|---|
| Variable | Coeff. estimate | $p$-Value | Variable | Coeff. estimate | $p$-Value |
| Constant | 0.7675 | <0.001 | Constant | −0.1006 | <0.001 |
| ln(Edu/Pop) | 0.1520 | <0.001 | Edu/Pop | 0.3350 | <0.001 |
| ln(Home/Pop) | 0.2098 | <0.001 | Home/Pop | 0.7538 | <0.001 |
| ln(Income/Pop) | −0.0877 | <0.001 | Income/Pop | −0.0081 | <0.001 |
| $W_n Y_n$ | 0.5530 | <0.001 | $W_n Y_n$ | 0.5290 | <0.001 |
| Test statistic $T_n$ | 21.45 | <0.001 | Test statistic $T_n$ | 28.17 | <0.001 |
| Bootstrap $p$-value | | <0.002 | Bootstrap $p$-value | | <0.002 |

continental United States (48 States, with Alaska, Hawaii, and D.C. excluded) from the 1990 Census that recorded votes in the 1980 presidential election. The variables in use are the total number of votes cast per county (Votes), the population in each county of 18 years of age or older (Pop), the population in each county with a 12th-grade or higher education (Edu), the number of owner-occupied housing units (Home), and the aggregate income (Income). Four counties have zero votes (La Paz at AZ, Broomfield at CO, Yellowstone National Park at MT, and Cibola at NM), so they are excluded and this results a sample size of 3107. The original data can be downloaded from *http://www.census.gov/support/USACdataDownloads.html*.

LeSage (1998) used this dataset to estimate the linear SAR model:

$$Y_n = \rho^0 W_n Y_n + \mathbf{1}_n \beta_0^0 + X_{1n}\beta_1^0 + X_{2n}\beta_2^0 + X_{3n}\beta_3^0 + v_n,$$

where the dependent variable is the voting rate Votes/Pop, the exogenous variables are log per capita education ln(Edu/Pop), log per capita homeowner ln(Home/Pop), and log per capita income ln(Home/Pop). The $W_n$ is a 3107 × 3107 spatial weight matrix based on the first-order contiguity. To check the robustness, we also consider the specification with exogenous variables being nominal values, not the log. Table 4 presents the results for the MLE, the asymptotic test, and the bootstrap test based on 500 resamples.

For both specifications of exogenous regressors, all the coefficients are highly significant, giving strong support to the parametric model. However, when we choose the Gaussian kernel and use the LSCV bandwidth $h$ to apply our test, the linear SAR model is rejected at 1% level using either the bootstrap $p$-value or the asymptotic $p$-value. This lends strong support to nonlinearity in the SAR model despite the fact such a result may not necessarily be economically interpretable.

## 6.2 Economic Growth Example

The second example is the study of economic growth rate from Ertur and Koch (2007). Knowledge accumulated in one area might depend on knowledge accumulated in other areas, especially in its neighborhoods, implying the possible existence of spatial spillover effects. Therefore, spatial econometrics models are widely used to study the technological interdependence. For example, Autant-Bernard and LeSage (2011) empirically examined spatial spillovers associated with public and private research expenditures in own- and other-industry sectors for a sample of 94 French regions. More recently, with the

Table 5. The estimates and test statistics of the linear SAR model for the growth data

| Variable | $W_{1n}$ with $w^*_{ij,n} = d_{ij}^{-2}$ for $i \neq j$ | | $W_{2n}$ with $w^*_{ij,n} = e^{-2d_{ij}}$ for $i \neq j$ | |
|---|---|---|---|---|
| | Coeff. estimate | $p$-Value | Coeff. estimate | $p$-Value |
| Constant | 0.9699 | 0.608 | 0.4801 | 0.798 |
| $\ln(s)$ | 0.8245 | <0.001 | 0.7914 | <0.001 |
| $\ln(n_p+0.05)$ | −1.4978 | 0.008 | −1.4505 | 0.009 |
| $W_n \ln(s)$ | −0.3257 | 0.075 | −0.3813 | 0.020 |
| $W_n \ln(n_p+0.05)$ | 0.5738 | 0.498 | 0.1431 | 0.856 |
| $W_n \ln(y)$ | 0.7420 | < 0.001 | 0.6630 | <0.001 |
| Test statistic $T_n$ | −0.672 | 0.749 | −0.725 | 0.766 |
| Bootstrap $p$-value | 0.664 | | 0.560 | |
| Restricted regression | | | | |
| Constant | 2.1089 | <0.001 | 2.9405 | <0.001 |
| $\ln(s) - \ln(n_p+0.05)$ | 0.8419 | <0.001 | 0.8185 | <0.001 |
| $W_n[\ln(s) - \ln(n_p+0.05)]$ | −0.2740 | 0.122 | −0.2694 | 0.098 |
| $W_n \ln(y)$ | 0.7360 | < 0.001 | 0.6440 | <0.001 |
| Test statistic $T_n$ | −0.145 | 0.558 | 0.427 | 0.665 |
| Bootstrap $p$-value | | 0.560 | | 0.260 |

NOTE: $d_{ij}$ denotes the great-circle distance between the capital cities of countries $i$ and $j$.

development of spatial panel data models such as Yu, de Jong, and Lee (2008), Lee and Yu (2010), and Yu and Lee (2010), more and more researches are based on panel data to study the technological spillovers. For example, Rho and Moon (2014) study the spatial dependence in China's regional innovation using China's province level data from 2000 to 2009. Ho, Wang, and Yu (2013) empirically examined the international spillover of economic growth through bilateral trade using a sample of 26 OECD countries over the period 1971–2005. Evans and Kim (2014) studied the spatial dynamics of growth and convergence in Korean regional incomes.

In this subsection, we want to test the cross-sectional linear SAR model specification in Ertur and Koch (2007). Their dataset covers a sample of 91 countries over the period 1960–1995 originally from Heston, Summers, and Aten (2002) obtained from Penn World Tables (PWT version 6.1). Variables in use include per worker income in 1960 ($y60$) and 1995 ($y95$), average rate of growth between 1960 and 1995 ($gy$), average investment rate of this period ($s$), and average rate of growth of working-age population ($n_p$). The dataset can be downloaded from JAE Data Archive at *http://qed.econ.queensu.ca/jae/2007-v22.6/*.

The spatial Durbin model (SDM) considered in Ertur and Koch (2007) is given by

$$Y_n = \rho^0 W_n Y_n + \mathbf{1}_n \beta_0^0 + X_n \beta^0 + W_n X_n \theta^0 + v_n,$$

where the dependent variable is log real income per worker $\ln(y95)$ and the explanatory variables include log investment rate $\ln(s)$ and log physical capital effective rate of depreciation $\ln(n_p + 0.05)$ . The associated parameters are $\beta^0 = (\beta_1^0, \beta_2^0)'$ with $X_n$ and $\theta^0 = (\theta_1^0, \theta_2^0)'$ with $W_n X_n$. A restricted regression based on the joint constraints of $\beta_1 = -\beta_2$ and $\theta_1 = -\theta_2$ is also considered in Ertur and Koch (2007) and is preferred. $W_n$ is constructed based on $d_{ij}$, the great-circle distance between the capital cities of countries $i$ and $j$.

Table 5 presents the estimation and testing results based on an SAR expression of this model that $Y_n = \rho^0 W_n Y_n + \mathbf{1}_n \beta_0^0 + \widetilde{X}_n b^0 + v_n$, where $\widetilde{X}_n = [X_n,\ W_n X_n]$ and $b^0 = (\beta^{0\prime},\ \theta^{0\prime})'$. We use the Gaussian kernel and the LSCV bandwidth $h$. Coefficients of the restricted regression are highly significant, supporting the SDM model. Both the asymptotic $p$-value and bootstrap $p$-value based on 500 bootstrap resamples exceed 10%, so we cannot reject the linear SDM specification.

## 7. CONCLUDING REMARKS

In this article, we propose a nonparametric test for correct specification of linear SAR models. We establish the asymptotic normal distributions of the test statistic under the null hypothesis and a sequence of local alternatives, and show the consistency of the test. To improve the finite sample performance of our test, we advocate a residual-based wild bootstrap procedure and justify its asymptotic validity. A small set of Monte Carlo simulations are conducted to show the test is well behaved in finite samples. This test is applied to two empirical applications. We reject the linear SAR model in the vote cast example and cannot reject the linear SDM model in the economic growth example.

Several extensions are possible. First, it is straightforward to extend our test to the case of nonlinear or panel spatial autoregressive models. Second, we conjecture that one can allow for some semiparametric specification of the SAR models and extend our test to that case. Third, it may be possible to extend our test to the SAR models with spatial errors. Fourth, it is also important to test for the correct specification of the spatial weight matrix by extending the important work of Pinkse, Slade, and Brett (2002) and Sun (2014). We leave these topics for future research.

## SUPPLEMENTARY MATERIALS

The online supplementary appendix gives the proofs of the main results in the article. It also contains some discussions on the assumptions and claims in Section 3.

## ACKNOWLEDGMENTS

The authors sincerely thank Shakeeb Khan, an associate editor, and two anonymous referees for their insightful comments and suggestions that lead to a substantial improvement of the presentation. The authors also thank Ingmar R. Prucha for his insightful comments on the early version of this article. The first author gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. The second author thanks Shanghai Pujiang Program for its financial support under Grant 14PJC065.

*[Received September 2014. Revised July 2015.]*

## REFERENCES

Aitchison, J., and Aitken, C. G. G. (1976), "Multivariate Binary Discrimination by the Kernel Method," *Biometrika*, 63, 413–420. [574]

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Amsterdam, the Netherlands: Kluwer Academic. [572,574]

Autant-Bernard, C., and LeSage, J. P. (2011), "Quantifying Knowledge Spillovers Using Spatial Econometric Models," *Journal of Regional Science*, 51, 471–496. [582]

Banerjee, S., Gelfand, A. E., Knight, J. R., and Sirmans, C. F. (2004), "Spatial Modeling of House Prices Using Normalized Distance-Weighted Sums of Stationary Processes," *Journal of Business & Economic Statistics*, 22, 206–213. [572]

Bierens, H. J. (1982), "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134. [573]

——— (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458. [573]

Case, A. C. (1991), "Spatial Patterns in Household Demand," *Econometrica*, 59, 953–965. [578]

Chen, X., and Fan, Y. (1999), "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373–401. [573]

Conley, T. G. (1999), "GMM Estimation With Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45. [572]

de Jong, P. (1987), "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277. [576]

Ertur, C., and Koch, W. (2007), "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence," *Journal of Applied Econometrics*, 22, 1033–1062. [582,583]

Evans, P., and Kim, J. U. (2014), "The Spatial Dynamics of Growth and Convergence in Korean Regional Incomes," *Applied Economics Letters*, 21, 1139–1143. [583]

Fan, Y., and Li, Q. (1996), "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865–890. [574,575]

——— (2000), "Consistent Model Specification Tests: Kernel-Based Tests versus Bierens' ICM Tests," *Econometric Theory*, 16, 1016–1041. [573]

Goldsmith-Pinkham, P., and Imbens, G. W. (2013), "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, 31, 253–264. [572]

Gupta, A., and Robinson, P. M. (2015), "Inference on Higher-order Spatial Autoregressive Models With Increasingly Many Parameters," *Journal of Econometrics*, 186, 19–31. [572]

Hall, P., Li, Q., and Racine, J. S. (2007), “Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors,” *Review of Economic Statistics*, 89, 784–789. [580]

Hallin, M., Lu, Z., and Tran, L. T. (2004), “Local Linear Spatial Regression,” *Annals of Statistics*, 32, 2469–2500. [572]

Härdle, W., and Mammen, E. (1993), “Comparing Nonparametric Versus Parametric Regression Fits,” *Annals of Statistics*, 21, 1926–1947. [577]

Helmers, C., and Patnam, M. (2014), “Does the Rotten Child Spoil His Companion? Spatial Peer Effects Among Children in Rural India,” *Quantitative Economics*, 5, 67–121. [572]

Heston, A., Summers, R., and Aten, B. (2002), *Penn World Tables Version 6.1, Downloadable Dataset*, Philadelphia, PA: Center for International Comparisons at the University of Pennsylvania. [583]

Ho, C. Y., Wang, W., and Yu, J. (2013), “Growth Spillover through Trade: A Spatial Dynamic Panel Data Approach,” *Economics Letters*, 120, 450–453. [583]

Hsiao, C., Li, Q., and Racine, J. S. (2007), “A Consistent Model Specification Test With Mixed Discrete and Continuous Data,” *Journal of Econometrics*, 140, 802–826. [573,575,580]

Jenish, N. (2012), “Nonparametric Spatial Regression Under Near-epoch Dependence,” *Journal of Econometrics*, 167, 224–239. [572]

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 Finance and Economics*, 17, 377–398. [572,574,575]

——— (1999), “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, 40, 509–533. [572,574,575]

——— (2001), “On the Asymptotic Distribution of the Moran I Test Statistic With Applications,” *Journal of Econometrics*, 104, 219–257. [575]

——— (2010), “Specification and Estimation of Spatial Autoregressive Models With Autoregressive and Heteroskedastic Disturbances,” *Journal of Econometrics*, 157, 53–67. [572,574,575,576]

Kim, J., Elliott, E., and Wang, D.-M. (2003), “A Spatial Analysis of County-level Outcomes in US Presidential Elections: 1988–2000,” *Electoral Studies*, 22, 741–761. [581]

Klier, T., and McMillen, D. P. (2008), “Clustering of Auto Supplier Plants in the United States,” *Journal of Business & Economic Statistics*, 26, 460–471. [572]

Lavergne, P. (2001), “An Equality Test Across Nonparametric Regressions,” *Journal of Econometrics*, 103, 307–344. [580]

Lee, L-F (2002), “Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models,” *Econometric Theory*, 18, 252–277. [572]

——— (2003), “Best Spatial Two-stage Least Squares Estimators for A Spatial Autoregressive Model With Autoregressive Disturbances,” *Econometric Reviews*, 22, 307–335. [572,574]

——— (2004), “Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models,” *Econometrica*, 72, 1899–1925. [572,574,575,576,578]

——— (2007a), “GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models,” *Journal of Econometrics*, 137, 489–514. [572]

——— (2007b), “Identification and Estimation of Econometric Models With Group Interactions, Contextual Factors and Fixed Effects,” *Journal of Econometrics*, 140, 333–374. [572]

Lee, L.-F., and Yu, J. (2010), “Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects,” *Journal of Econometrics*, 154, 165–185. [583]

LeSage, J. P. (1998), *Spatial Econometrics Toolbox*, University of Toledo Web Book. Available at *http://www.spatial-econometrics.com/*. [582]

LeSage, J. P., and Pace, R. K. (2009), *An Introduction to Spatial Econometrics*, New York: Chapman & Hall/CRC. [573]

Li, Q., and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, NJ: Princeton University Press. [574,575]

——— (2008), “Nonparametric Estimation of Conditional CDF and Quantile Functions With Mixed Categorical and Continuous Data,” *Journal of Business & Economic Statistics*, 26, 423–434. [574]

Lin, X., and Lee, L.-F. (2010), “GMM Estimation of Spatial Autoregressive Models With Unknown Heteroskedasticity,” *Journal of Econometrics*, 157, 34–52. [572,574,575,576]

Lin, Z., Li, D., and Gao, J. (2009), “Local Linear M-estimation in Nonparametric Spatial Regression,” *Journal of Time Series Analysis*, 30, 286–341. [572]

Liu, X. (2014), “Identification and Efficient Estimation of Simultaneous Equations Network Models,” *Journal of Business & Economic Statistics*, 32, 516–536. [572]

Liu, X., Lee, L.-F., and Bollinger, C. R. (2010), “An Efficient GMM Estimator of Spatial Autoregressive Models,” *Journal of Econometrics*, 159, 303–319. [572,574]

Majumdar, A., Munneke, H. J., Gelfand, A. E., Banerjee, S., and Sirmans, C. F. (2006), “Gradients in Spatial Response Surfaces With Application to Urban Land Values,” *Journal of Business & Economic Statistics*, 24, 77–90. [572]

Ord, J. K. (1975), “Estimation Methods for Models of Spatial Interaction,” *Journal of the American Statistical Association*, 70, 120–126. [572]

Osipenko, M. (2014), “Difference-Based Estimation of Partially Linear Spatial Autoregressive Models,” Working Paper, Humboldt-Universität zu Berlin. [572]

Ouyang, D., Li, Q., and Racine, J. S. (2009), “Nonparametric Estimation of Regression Functions With Discrete Regressors,” *Econometric Theory*, 29, 1–42. [575]

Pace, R. K., and Barry, R. (1997), “Quick Computation of Spatial Autoregressive Estimators,” *Geographical Analysis*, 33, 291–297. [581]

Pinkse, J., Slade, M. E., and Brett, C. (2002), “Spatial Price Competition: A Semiparametric Approach,” *Econometrica*, 70, 1111–1153. [583]

Racine, J., and Li, Q. (2004), “Nonparametric Estimation of Regression Function with Both Categorical and Continuous Data,” *Journal of Econometrics*, 119, 99–130. [574]

Rho, S., and Moon, I. J. (2014), “Innovation and Spillovers in China: Spatial Econometric Approach,” *Seoul Journal of Economics*, 27, 149–170. [583]

Robinson, P. M. (2010), “The Efficient Estimation of the Semiparametric Spatial Autoregressive Model,” *Journal of Econometrics*, 157, 6–17. [572]

——— (2011), “Asymptotic Theory for Nonparametric Regression With Spatial Data,” *Journal of Econometrics*, 165, 5–19. [572]

Rosenblatt, M. (1975), “A Quadratic Measure of Deviation of Two-dimensional Density Estimation and a Test of Independence,” *Annals of Statistics*, 3, 1–14. [573]

Shin, M. E., and Agnew, J. (2007), “The Geographical Dynamics of Italian Electoral Change, 1987-2001,” *Electoral Studies*, 26, 287–302. [581]

Smirnov, O., and Anselin, L. (2001), “Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach,” *Computational Statistics and Data Analysis*, 35, 301–319. [572]

Soares, G. A. D., and Terron, S. L. (2008), “Dois Lulas: A Geografia Eleitoral Da Reeleicao (Explorando Conceitos, Metodos e tenicas de Analise Geospacial),” *Opiniao Publica*, 14, 269–301. [581]

Su, L. (2012), “Semiparametric GMM Estimation of Spatial Autoregressive Models,” *Journal of Econometrics*, 167, 543–560. [572,578]

Su, L., and Jin, S. (2010), “Profile Quasi-maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models,” *Journal of Econometrics*, 157, 18–33. [572]

Su, L., and White, H. (2008), “Nonparametric Hellinger Metric Test for Conditional Independence,” *Econometric Theory*, 24, 829–864. [573]

Su, L., and Yang, Z. (2013), “Instrumental Variable Quantile Regression for Spatial Autoregressive Models,” Working Paper, School of Economics, Singapore Management University. [572,574,575]

——— (2015), “QML Estimation of Dynamic Panel Data Models With Spatial Errors,” *Journal of Econometrics*, 185, 230–258. [572,578]

Sun, Y. (2014), “Functional-Coefficient Spatial Autoregressive Models with Nonparametric Spatial Weights,” Working Paper, Department of Economics and Finance, University of Guelph. [572,583]

Yu, J., de Jong, R., and Lee, L.-F. (2008), “Quasi-maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects When Both $n$ and $T$ Are Large,” *Journal of Econometrics*, 146, 118–134. [583]

Yu, J., and Lee, L.-F. (2010), “A Spatial Dynamic Panel Data Model With Both Time and Individual Fixed Effects,” *Econometric Theory*, 26, 564–597. [583]

Zhang, Z. (2013), “A Pairwise Difference Estimator for Partially Linear Spatial Autoregressive Models,” *Spatial Economic Analysis*, 8, 176–194. [572]

Zheng, J. X. (1996), “A Consistent Test of Functional Form via Nonparametric Estimation Technique,” *Journal of Econometrics*, 75, 263–289. [574,575]