Journal of Econometrics 157 (2010) 18–33

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models[☆]

Liangjun Su[*], Sainan Jin

School of Economics, Singapore Management University, Singapore
Guanghua School of Management, Peking University, Beijing, PR China

ARTICLE INFO

Article history:
Available online 31 October 2009

JEL classification:
C13
C14
C21

Keywords:
Profile likelihood
Partially linear models
Quasi-maximum likelihood estimation
Spatial autoregression
Spatial dependence

ABSTRACT

We propose profile quasi-maximum likelihood estimation of spatial autoregressive models that are partially linear. The rate of convergence of the spatial parameter estimator depends on some general features of the spatial weight matrix of the model. The estimators of other finite-dimensional parameters in the model have the regular $\sqrt{n}$-rate of convergence and the estimator of the nonparametric component is consistent but with different restrictions on the choice of bandwidth parameter associated with different natures of the spatial weights. Monte Carlo simulations verify our theory and indicate that our estimators perform reasonably well in finite samples.

1. Introduction

Since Paelinck coined the term "spatial econometrics" in the early 1970s to refer to a set of methods that explicitly deals with spatial dependence and spatial heterogeneity, the field has grown rapidly. The books by Cliff and Ord (1973), Paelinck and Klaassen (1979), Anselin (1988), Cressie (1993) and Anselin and Florax (1995) contribute significantly to the development of the field. For a recent survey on the subject, see Anselin and Bera (2002).

Among the class of spatial models, spatial autoregressive (SAR) models on lattices have attracted huge attention. Various methods have been proposed to estimate the SAR models, which include the method of maximum likelihood (ML) by Ord (1975) and Smirnov and Anselin (2001), the method of moments (MM) by Kelejian and Prucha (1999, 2010), and the method of quasi-maximum likelihood estimation (QMLE) by Lee (2002b, 2004). A common feature of these methods is that they are all developed to estimate finite dimensional parameters in the SAR models which are frequently assumed to be linear. When an unknown infinite dimensional parameter is present (e.g., the regression function is of unknown form), there is a lack of guidance on the estimation and inference process.

In this paper, we consider spatial autoregressive (SAR) models on lattices when the regression function is partially specified, motivated by the following considerations. First, as was argued in Paelinck and Klaassen (1979, pp. 6–9), econometric relations in space result more often than not in highly non-linear specifications. It has well been documented in the literature that many economic variables exhibit nonlinear relationships. For example, economic inequality is associated with economic growth through an inverse-U shaped Kuznets curve. Recent study also suggests an inverse-U relationship between economic growth and environmental quality even when the spatial effect is accounted for (see Rupasingha et al., 2004). Ignoring the potential nonlinear relationship in spatial dependence models often results in inconsistent estimation of the parameters of interest and misleading conclusions.

Second, while most econometric analysis and empirical studies using the SAR models ignore potential nonlinear functional forms, there have been some considerations of flexible functional forms in the literature that try to take into account certain form of nonlinearities in the models. See, for example, van Gastel and Paelinck (1995), Baltagi and Li (2001), Pace et al. (2004) and Yang et al. (2006). Most of these papers introduce a parametric transformation (e.g., Box–cox transformation) on the response variable or/and

☆ The authors thank Qi Li for his many constructive and helpful suggestions. The very insightful comments from the two anonymous referees are also gratefully acknowledged. The authors also thank Jiti Gao for his kind help with a technical lemma in the paper, and the participants in 2005 International Symposium on Econometrics Development, Beijing for their many valuable suggestions. The authors gratefully acknowledge the financial support from the NSFC under the grant numbers 70501001 and 70601001.

* Corresponding address: School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore. Tel.: +65 6828 0386.
E-mail address: ljsu@smu.edu.sg (L. Su).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.10.033

regressors. Nevertheless, parametric functional form transformation can at most provide certain protection against some specific nonlinear forms. In the absence of a priori information and theoretical foundation, it is generally advisable to consider more flexible functional forms.

Third, as nonparametric techniques advance, more and more researchers find out the advantage of nonparametric and semiparametric methods in modeling nonlinear economic relationships (see, e.g., Yatchew, 1998). Recent researches have started addressing the importance of nonparametric modeling in spatial econometrics. For example, in modeling hedonic housing price, Gress (2004) introduced two semiparametric spatial autocorrelation models and compared them with a variety of competing parametric spatial models. He found that the semiparametric models offer more accurate and stable estimates of the regression parameters and better out-of-sample predictions than do the alternative parametric models. Basile and Gress (2004) proposed a semiparametric spatial auto-covariance specification of the growth model for the European economy and found that nonlinearities are important in regional growth in Europe even when the spatial dependence is controlled for. As a result, assuming a common linear relationship between economic growth and inputs is misleading.

Fourth, as Robinson (1988) remarked, a correctly specified parametric model affords precise inferences, a badly misspecified one, possibly seriously misleading ones, whereas nonparametric modeling is associated with both greater robustness and lesser precision. So an intermediate strategy is to apply a semiparametric form, among which partially linear models are widely used.

In this paper we extend the work of Lee (2004) and consider estimating the parameters in partially linear SAR models by the profile QMLE method. When the error term has a known density form, Staniswalis (1989) proposed estimating the nonparametric regression function by maximizing the local log-likelihood. In the case of unknown error density, we can apply the idea of the quasi-maximum likelihood (QML). Because we have both parametric and nonparametric components in our regression function, we first concentrate out the nonparametric component by expressing the nonparametric component as certain function of the parametric component and the data. Then we estimate the parametric component and recover the nonparametric component after that. Consequently, we term our estimator as a profile QML estimator. Like Lee (2004), our parametric component includes the spatial parameter, the coefficient of the linear part of the regression function, and the variance of the error term. Because the parametric component is of finite dimension, it is also called the finite dimensional parameter in the literature.

Like Lee (2004), the rates of convergence of the estimators for the finite dimensional parameters depend on some general features of the spatial weights matrix of the model. The estimator of the spatial parameter may indeed have a $\sqrt{n}$-rate of convergence and a normal limiting distribution. Nevertheless, under some circumstances, the estimator has a slow rate of convergence for some parametric components of the model, say when all elements of the spatial weights matrix tend to zero as the sample size goes to infinity. In the former case, the nonparametric component can be estimated consistently at the conventional nonparametric convergence rate. But this is not true in the latter case where more stringent conditions on the spatial weights matrix and the bandwidth parameter are required to gain consistency of the estimators for both the parametric and nonparametric components.

It is worth mentioning that the semiparametric models of Gress (2004) and Basile and Gress (2004) are special cases of our model. We can also apply our model to examine many other well-known nonlinear relationships in economics, including the relationship between economic inequality and economic growth, the relationship between economic growth and environmental inequality, the relationship between education and wages, etc.

The paper is structured as follows. In Section 2 we introduce the partially linear SAR model and the profile QMLE approach to estimate the finite and infinite dimensional parameters in the model. We make a set of assumptions in Section 3. In Section 4 we study the asymptotic properties of the profile QMLE estimators when the information matrix is nonsingular and the parametric component can be estimated at the regular $\sqrt{n}$-rate. In Section 5 we study the asymptotic properties of the profile QMLE estimators when the information matrix is singular and some of the parametric component can only be estimated at a slower rate. We conduct Monte Carlo simulations to check the performance of the proposed estimator in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

Like Kelejian and Prucha (2001), we adopt the following notation and conventions. For a matrix A_n, we denote its norm as $\|A_n\| = [\mathrm{tr}(A_nA_n')]^{1/2}$ and the (i,j)th element of A_n as $a_{n,ij}$. Similarly, for a vector a_n, $a_{n,i}$ denotes its ith element. An analogous convention is adopted for matrices and vectors that do not depend on the index n, where n is frequently suppressed. We say A_n is uniformly bounded in absolute value if $\sup_{1\le i\le n, 1\le j\le n}|a_{n,ij}| < \infty$. We say A_n is uniformly bounded in row sums (resp. column sums) if $\sup_{1\le i\le n}\sum_{j=1}^n |a_{n,ij}| \le c_a < \infty$ (resp. $\sup_{1\le j\le n}\sum_{i=1}^n |a_{n,ij}| \le c_a < \infty$).

2. Partially linear spatial autoregressive models and profile QMLE

In this paper we investigate estimation of the spatial autoregressive models:

$$Y_n = X_n\beta_0 + \mathbf{m}_0(Z_n) + \rho_0 W_n Y_n + U_n, \tag{2.1}$$

where $X_n \equiv (x_{n,1}, \ldots, x_{n,n})'$ and $Z_n \equiv (z_{n,1}, \ldots, z_{n,n})'$ are $n \times p$ and $n \times q$ matrices of regressors, respectively, W_n is a specified constant $n \times n$ spatial weight matrix, $U_n \equiv (u_1, \ldots, u_n)'$ is an n-dimensional vector of i.i.d. disturbances with zero mean and finite variance σ_0^2, $\mathbf{m}_0(Z_n) \equiv (m_0(z_{n,1}), \ldots, m_0(z_{n,n}))'$, and $m_0(\cdot)$ is an unknown function defined on $\mathbb{R}^q$. Let $\theta_0 = (\beta_0', \rho_0, \sigma_0^2)'$ be the true finite dimensional parameter vector. Denote $T_n(\rho) = I_n - \rho W_n$ for any value of ρ. It follows that

$$Y_n = T_n^{-1}\left(X_n\beta_0 + \mathbf{m}_0(Z_n) + U_n\right), \tag{2.2}$$

provided $T_n \equiv T_n(\rho_0)$ is nonsingular.

Let $U_n(\delta) = Y_n - X_n\beta - \mathbf{m}_0(Z_n) - \rho W_n Y_n$, where $\delta = (\beta', \rho)'$. In the case for which $\mathbf{m}_0(\cdot)$ is missing from the definition of $U_n(\delta)$, Lee (2002b, 2004) proposes maximizing the Gaussian quasi log likelihood

$$\log L_n(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \log|T_n(\rho)| - \frac{1}{2\sigma^2}U_n(\delta)' U_n(\delta), \tag{2.3}$$

where $\theta = (\beta', \rho, \sigma^2)'$.

Since $\mathbf{m}_0(\cdot)$ is present in Eq. (2.1), we propose estimating θ by the following two step procedure: (i) Estimate $m_0(z)$ for fixed θ, denote the resulting estimator as $m_\theta(z)$; (ii) Plug in $m_\theta(z)$ into $U_n(\delta)$ in (2.3), and obtain the QMLE estimator $\widehat{\theta}$ for θ and $m_{\widehat{\theta}}(z)$ for $m_0(z)$.

To estimate $m_0(z)$ for fixed θ in the first step, we generalize the approach of Staniswalis (1989) for likelihood-based estimation and use a method that might be called profile quasi-maximum likelihood estimation (QMLE). We give an asymptotic analysis based on the local polynomial procedure. See Fan (1992) and Fan and Gijbels (1996) for a discussion on the attractive properties of local polynomials.

Let $K(\cdot)$ denote a kernel function on $\mathbb{R}^q$ and $h = h_n$ a bandwidth sequence. Set $K_h(z) = h^{-q}K(z/h)$. Let $Y_n^*(\rho) = T_n(\rho)Y_n$ and denote its jth element as $Y^*_{n,j}(\rho)$. For the data set $\{Y_n, X_n, Z_n\}$, the rth order local polynomial regression of $Y^*_{n,i}(\rho)-x'_{n,i}\beta$ on $z_{n,i}$ can be obtained from maximizing the following profile likelihood criterion:

$$Q_n(\alpha(z)) \equiv n^{-1}\sum_{i=1}^{n} K_h\left(z_{n,i}-z\right)\log\varphi_{\sigma^2} \times \left(Y^*_{n,i}(\rho)-x'_{n,i}\beta-\sum_{0\le|\mathbf{j}|\le r}\alpha_{\mathbf{j}}h^{-|\mathbf{j}|}(z_{n,i}-z)^{\mathbf{j}}\right), \quad (2.4)$$

where $\varphi_{\sigma^2}(\cdot)$ is the pdf for a normal density with zero mean and finite variance σ^2. Here and below, we use the notation of Masry (1996): $\mathbf{j}=(j_1,\ldots,j_q)$, $|\mathbf{j}|=\sum_{i=1}^q j_i$, $z^{\mathbf{j}}=\Pi_{i=1}^q z_i$ and $\sum_{0\le|\mathbf{j}|\le r}=\sum_{k=0}^r\sum_{\substack{j_1=0\\ j_1+\cdots+j_q=k}}^{k}\cdots\sum_{j_q=0}^{k}$. For $|\mathbf{j}|=0,1,\ldots,r$, $\alpha_{\mathbf{j}}\equiv\alpha_{\mathbf{j}}(z)$ corresponds to the scaled coefficient of $(z_{n,i}-z)^{\mathbf{j}}$ in the rth order Taylor expansion of $m_0(z_{n,i})$ around z, i.e., $\alpha_{\mathbf{j}}=(h^{|\mathbf{j}|}/\mathbf{j}!)\partial^{|\mathbf{j}|}m_0(z)/(\partial^{j_1}z_1,\ldots,\partial^{j_q}z_q)$, where $\mathbf{j}!\equiv\Pi_{i=1}^q j_i!$. Further, let $\alpha(z)=(\alpha_0(z),\alpha_1(z)',\ldots,\alpha_r(z)')$, where $\alpha_l(z)$ is a collection of all the parameters $\alpha_{\mathbf{j}}(z)$ with $|\mathbf{j}|=l$ $(0\le l\le r)$ in the lexicographical order (with highest priority to last position so that $(0,0,\ldots,l)$ is the first element in the sequence and $(l,0,\ldots,0)$ is the last element). Denote the maximizer of (2.4) as $\alpha_\theta(z)=(\alpha_{0,\theta}(z),\alpha_{1,\theta}(z)',\ldots,\alpha_{r,\theta}(z)')'$, where, for given θ, $\alpha_{l,\theta}(z)$ is the collection of the estimators $\alpha_{\mathbf{j},\theta}$ of $\alpha_{\mathbf{j}}$ with $|\mathbf{j}|=l$ in the lexicographical order. Let $N_l=(l+q-1)!/(l!(q-1)!)$ be the number of distinct q-tuples $\mathbf{j}$ with $|\mathbf{j}|=l$ $(l=0,1,\ldots,r+1)$. Then $\alpha_\theta(z)$ is an $N\times 1$ vector, where $N=\sum_{l=0}^r N_l$. In the special case $r=1$, we have the local linear procedure so that the above notation can be greatly simplified: $\alpha(z)=(\alpha_0(z),\alpha_1(z)')'=(\alpha_{(0,0,\ldots,0)}(z),\alpha_{(0,0,\ldots,1)}(z),\ldots,\alpha_{(1,0,\ldots,0)}(z))'$ because $\alpha_l(z)$ is a collection of $a_{\mathbf{j}}(z)$ with $|\mathbf{j}|=l$ $(l=0,1)$; $N=N_0+N_1=1+q$.

Given θ, the maximizer $\alpha_\theta(z)$ of (2.4) can be obtained by

$$\alpha_\theta(z)=\arg\min_{\alpha}\left(Y^*_n(\rho)-X_n\beta-\overrightarrow{Z}(z)\alpha\right)' \times\mathbf{K}_h(z)\left(Y^*_n(\rho)-X_n\beta-\overrightarrow{Z}(z)\alpha\right), \quad (2.5)$$

where $\mathbf{K}_h(z)=\mathrm{diag}(K_h(z_{n,1}-z),\ldots,K_h(z_{n,n}-z))$, $\overrightarrow{Z}(z)=(\overrightarrow{Z}_1(z),\ldots,\overrightarrow{Z}_n(z))'$ and $\overrightarrow{Z}_i(z)$ is a collection of $h^{-|\mathbf{j}|}(z_{n,i}-z)^{\mathbf{j}}(|\mathbf{j}|=0,1,\ldots,r)$ in the lexicographical order. For example, if $r=1$, then $\overrightarrow{Z}_i(z)=(1,(z_{n,i}-z)'/h)'$. In general, the first element in $\alpha_\theta(z)$ represents the profile likelihood estimate of $m_0(z)$ given θ and thus we denote it as $m_\theta(z)$. Define the smoothing operator by $\mathbf{S}_n(z)=[\overrightarrow{Z}(z)'\mathbf{K}_h(z)\overrightarrow{Z}(z)]^{-1}\overrightarrow{Z}(z)'\mathbf{K}_h(z)$. Then

$$\alpha_\theta(z)=\mathbf{S}_n(z)\left(Y^*_n(\rho)-X_n\beta\right). \quad (2.6)$$

In particular, the estimator for $m_0(z)$ is given by

$$m_\theta(z)=s(z)'\left(Y^*_n(\rho)-X_n\beta\right), \quad (2.7)$$

where $s(z)'=e'\mathbf{S}_n(z)$, and $e=(1,0,\ldots,0)'$ is an $N\times 1$ vector.

When $r=0$, $\alpha_\theta(z)$ reduces to $m_\theta(z)$, and we obtain the local constant (or equivalently Nadaraya–Watson) estimator used by Robinson (1988). Nevertheless, it is well known that the bias of the local constant estimator is higher for z near the boundary than for z away from the boundary. One needs to use either a "trimming" scheme (e.g., Robinson, 1988), or some weight functions (e.g., Li and Stengos, 1996) to avoid the random denominator problem in the local constant estimation. So in the following we restrict ourselves to the case where $r\ge 1$.

In the second step, we consider maximizing the following profile likelihood

$$\log L_n(\theta,\mathbf{m}_\theta(Z_n))=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2+\log|T_n(\rho)| -\frac{1}{2\sigma^2}(T_n(\rho)Y_n-X_n\beta-\mathbf{m}_\theta(Z_n))' \times(T_n(\rho)Y_n-X_n\beta-\mathbf{m}_\theta(Z_n)), \quad (2.8)$$

where $\mathbf{m}_\theta(Z_n)=(m_\theta(z_{n,1}),\ldots,m_\theta(z_{n,n}))'$.

Let $S_n=(s(z_{n,1}),\ldots,s(z_{n,n}))'$. From the quasi log likelihood function (2.8), given ρ, the QMLE of β is

$$\widehat{\beta}(\rho)=[X'_n(I_n-S_n)'(I_n-S_n)X_n]^{-1} \times X'_n(I_n-S_n)'(I_n-S_n)T_n(\rho)Y_n, \quad (2.9)$$

and the QMLE of σ^2 is

$$\widehat{\sigma}^2(\rho)=\frac{1}{n}\left[(I_n-S_n)(T_n(\rho)Y_n-X_n\widehat{\beta}(\rho))\right]' \times\left[(I_n-S_n)(T_n(\rho)Y_n-X_n\widehat{\beta}(\rho))\right] =\frac{1}{n}Y'_nT_n(\rho)'(I_n-S_n)'M_n(I_n-S_n)T_n(\rho)Y_n, \quad (2.10)$$

where $M_n=I_n-(I_n-S_n)X_n[X'_n(I_n-S_n)'(I_n-S_n)X_n]^{-1}X'_n(I_n-S_n)'$. The concentrated log likelihood function of ρ is

$$\log L^c_n(\rho)=-\frac{n}{2}(\log(2\pi)+1)-\frac{n}{2}\log\widehat{\sigma}^2(\rho)+\log|T_n(\rho)|. \quad (2.11)$$

The QMLE $\widehat{\rho}$ of ρ maximizes the concentrated likelihood (2.11). The QMLEs of β and σ^2 are $\widehat{\beta}(\widehat{\rho})$ and $\widehat{\sigma}^2(\widehat{\rho})$ respectively. Let $\widehat{\theta}=(\widehat{\beta}',\widehat{\rho},\widehat{\sigma}^2)'$. A by-product of the above procedure is

$$\widehat{\alpha}(z)\equiv\alpha_{\widehat{\theta}}(z)=\mathbf{S}_n(z)\left(Y^*_n(\widehat{\rho})-X_n\widehat{\beta}\right), \quad (2.12)$$

the profile QMLE for the unknown function $m_0(z)$ and its scaled partial derivatives up to order r. We will study the asymptotic properties of $\widehat{\theta}$ and $\widehat{\alpha}(z)$ in subsequent sections.

3. Assumptions

To provide a rigorous analysis of the QMLE, we assume the following basic regularity conditions.

Assumption 1. (i) $(x_{n,i},z_{n,i})$, $i=1,\ldots,n$, are nonstochastic regressors that are uniformly bounded on $\mathcal{X}\times\mathcal{Z}$. (ii) There exist some functions $m_j(z)$ over $\mathcal{Z}$ such that for all $1\le i\le n$ and $1\le j\le p$,

$$x_{n,ij}=m_j\left(z_{n,i}\right)+\eta_{ij}, \quad (3.1)$$

and the real sequences $\{\eta_{ij}\equiv\eta_{n,ij}\}$ satisfy

$$\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\eta_i\eta'_i=\Phi_{X,X} \quad (3.2)$$

and

$$\limsup_{n\to\infty}\frac{1}{n^{1/2}\log n}\max_{1\le l\le n}\left\|\sum_{s=1}^{l}\eta_{j_s}\right\|<\infty \quad (3.3)$$

for any permutation $(j_1,\ldots,j_n)$ of $(1,\ldots,n)$, where $x_{n,ij}$ is the jth element of $x_{n,i}$ (i.e., $x_{n,i}=(x_{n,i1},\ldots,x_{n,ip})'$), $\eta_i\equiv\eta_{n,i}=(\eta_{i1},\ldots,\eta_{ip})'$, $\Phi_{X,X}$ is a positive definite matrix, and $\|\cdot\|$ denotes the Euclidean norm. (iii) $m_j(\cdot)$, $j=0,1,\ldots,p$, are $(r+1)$-times continuously differentiable and their $(r+1)$th partial derivatives are Lipschitz continuous of order one. (iv) There exists a positive density function $f(\cdot)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}v\left(z_{n,i}\right)=\int v(z)f(z)\,\mathrm{d}z \quad (3.4)$$

for any bounded continuous function $v(\cdot)$. $f(\cdot)$ is bounded away from zero on $\mathcal{Z}$.

The fixed bounded design assumption in Assumption 1(i) is made for several reasons. First, it parallels that of Kelejian and Prucha (1998, 1999, 2001, 2010), Lee (2002a,b, 2004) and Lin and Lee (2010). Second, it allows us to avoid the use of trimming factors (e.g. Robinson, 1988). Like the aforementioned literature, we allow the fixed regressors to depend on n. This is important since X_n or Z_n can include columns like W_nX_n. By writing $\eta_{ij} \equiv \eta_{n,ij}$ and $\eta_i \equiv \eta_{n,i}$, we explicitly allow for triangular array η_{ij}'s. This is important since it is possible to have $m_j(z_{n,i}) \equiv 0$ and $\eta_{ij} \equiv \eta_{n,ij} = x_{n,ij}$. Nevertheless, we will suppress the dependence of η_i's on n for notational simplicity. Assumption 1(ii) parallels Assumption 1 of Gao (1995) and Assumption 1.3.1(ii) of Härdle et al. (2000) who considered only a scalar design. If $\{\eta_i\}_{i=1}^n$ were i.i.d. with mean zero and finite variance–covariance matrix $\Phi_{X,X}$, then we could think that Assumption 1(ii) holds with probability one. Assumption 1(iii) is standard in the literature on local polynomial estimation. As in Linton (1995), Assumption 1(iv) does not preclude $\{z_{n,i}\}_{i=1}^n$ from being generated by some random mechanism. For example, if $z_{n,i}'$s were i.i.d. with density $f(\cdot)$, then (3.4) holds with probability one. So even though we focus on the fixed regressor case, our analysis holds with probability one if $\{x_{n,i}, z_{n,i}\}_{i=1}^n$ are generated randomly, and in this case, we can interpret our analysis as being conditional on $\{x_{n,i}, z_{n,i}\}_{i=1}^n$.

Assumption 2. $\{u_i\}_{i=1}^n$ are i.i.d. with mean zero and variance σ_0^2. $E(|u_1|^{4+\epsilon}) < \infty$ for some $\epsilon > 0$.

Assumption 2 rules out heteroscedasticity. When heteroscedasticity is present in the linear spatial autoregressive (SAR) models, the QMLE of Lee (2004) is generally inconsistent. For this reason, Kelejian and Prucha (2010) and Lin and Lee (2010) explore the GMM estimation of the linear SAR models with heteroscedasticity.

Assumption 3. (i) The elements $w_{n,ij}$ of W_n are at most of order l_n^{-1}, denoted by $O(1/l_n)$, uniformly in all i, j. As a normalization, $w_{n,ii} = 0$ for all i. (ii) The ratio $l_n/n \to 0$ as n goes to infinity. (iii) The matrix T_n is nonsingular. (iv) The sequences of matrices $\{W_n\}$ and $\{T_n^{-1}\}$ are uniformly bounded in both row and column sums. (v) There exists $\vartheta \in [0, 1)$ such that $l_n = O(n^{\vartheta})$.

Assumption 3(i)–(iv) concern the essential features of spatial weights matrix and they parallel Assumptions 2–5 of Lee (2004). Assumption 3(i) is always satisfied if $\{l_n\}$ is a bounded sequence. We allow $\{l_n\}$ to be divergent but at a rate smaller than n as specified in Assumption 3(ii). Assumption 3(iii) guarantees that (2.1) has an equilibrium given by (2.2). Kelejian and Prucha (1998, 1999, 2001) and Lee (2004) also assume Assumption 3(iv) which limits the spatial correlation to some degree but facilitates the study of the asymptotic properties of the spatial parameter estimators. By Horn and Johnson (1985, p. 301), $\lim\sup_n \|\rho_0 W_n\| < 1$ is sufficient to guarantee that T_n^{-1} is uniformly bounded in both row and column sums. As Lee (2004) remarked, this assumption ensures that the variance of Y_n is bounded as $n \to \infty$. Assumption 3(v) is a little bit stronger than Assumption 3(ii) and is assumed to facilitate the presentation in the proof.

Assumption 4. $\{T_n^{-1}(\rho)\}$ are uniformly bounded in either row or column sums, uniformly in ρ in a compact convex parameter space Δ. The true ρ_0 is an interior point in Δ.

By Lee (2002b, Lemma A.3), Assumption 3(iv) implies $\{T_n^{-1}(\rho)\}$ are uniformly bounded in both row and column sums uniformly in a neighborhood of ρ_0. Assumption 4 restricts the parameter space Δ to be compact and thus rules out the case where $\Delta = (-1, 1)$. As a referee kindly remarked, one can relax the compactness of Δ to establish the global consistency of our estimator but that is beyond the scope of our study. As Kelejian and Prucha (2010) noted, in the existing literature related to the Cliff–Ord models, the spatial weights matrix W_n is typically assumed to be row-normalized (say, to avoid the singularity of $I_n - \rho W_n$) so that the parameter space Δ for ρ is taken to be the interval $(-1, 1)$ and the spatial parameter is assumed not to depend on the sample size. Nevertheless, row-normalization may lead to misspecified model unless some theory suggests so. For this reason, Kelejian and Prucha (2010) allow all of the model parameters to depend on the sample size. To save space, we refer the readers to Kelejian and Prucha (2010) for more discussions on the parameter space.

Assumption 5. (i) The kernel function $K(\cdot)$ is a continuous symmetric density with compact support on $\mathbb{R}^q$. (ii) $h \propto n^{-1/\delta}$ for some $\delta > 0$ such that $nh^{2q} \to \infty$, and $nh^{4(r+1)} \to 0$.

Assumption 5 concerns the kernel and bandwidth sequence. It is pretty standard in the nonparametric literature for local linear estimation. Assumption 5(ii) requires that $r > q/2 - 1$. When $q \le 3$, we can simply choose $r = 1$ and apply the local linear procedure. When $q \ge 4$, a higher order local polynomial is required. Nevertheless, due to the "curse of dimensionality", we do not expect large q in practice.

To proceed, let $G_n = W_nT_n^{-1}$ and $R_n = G_n(X_n\beta_0 + \mathbf{m}_0(Z_n))$. Then $c_1 \equiv \lim_{n\to\infty} \mathrm{tr}(G_n)/n < \infty$, and $c_2 \equiv \lim_{n\to\infty} \mathrm{tr}((G_n + G_n')G_n)/n < \infty$ by Lemma A.5 and Facts 1–2 in the Appendix, which also implies the existence of the following limits under Assumption 1:

$$\Phi_{X,R} \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \eta_i \sum_{j=1}^n g_{n,ij}\eta_j'\beta_0 \quad \text{and}$$

$$\Phi_{R,R} \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \left(\sum_{j=1}^n g_{n,ij}\eta_j'\beta_0 \right)^2 .$$

We next make a high level assumption.

Assumption 6. $\|(I_n - S_n)G_n\mathbf{m}_0(Z_n)\|^2 = O(nh^{2(r+1)} + h^{-q})$.

Assumption 6 is a high level assumption which can be justified by assuming that $z_{n,i}'s$ are i.i.d. random variables with finite variance, where $O(nh^{2(r+1)})$ results from the squared bias by approximating $G_n\mathbf{m}_0(Z_n)$ by $S_nG_n\mathbf{m}_0(Z_n)$ and $O(h^{-q})$ reflects the variance of the approximation. In a fixed design partially linear model, Speckman (1988) made a similar higher level assumption which is justifiable by treating the regressors as being generated randomly. See, e.g., Assumptions (e)–(f) of Speckman (1988) and the discussions thereafter. In the special case where $m_j(z_{n,i}) \equiv 0$, Assumption 6 is automatically satisfied.

Assumption 7. $\Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R} > 0$.

Assumption 7 requires implicitly that the generated regressors $G_nX_n\beta_0$ and X_n, deviated from their nonparametric projection onto Z_n, are not asymptotically multicollinear. It is a sufficient identification condition of θ_0. By (A.4), $\Phi_{R,R} = 0$ and $\Phi_{X,R} = 0$ if $\vartheta > 0$. So Assumption 7 implicitly requires that $l_n = O(1)$, that is, elements of W_n cannot tend to zero uniformly as n increases to infinity.

As Lee (2004) remarked, the set of regressors $G_nX_n\beta_0$ and X_n, deviated from their nonparametric projection onto Z_n, can be linearly dependent under some circumstances, e.g., if $\beta_0 = 0$ or the columns of $G_nX_n\beta_0$ lying in the space spanned by the columns of X_n. In some cases $G_nX_n\beta_0$ and X_n are linearly independent in finite samples but they become asymptotically multicollinear (e.g., Case, 1991). This means that $\lim_{n\to\infty} n^{-1}(G_nX_n)'(I_n - S_n)'M_n(I_n - S_n)G_nX_n = 0$. In this case, we replace Assumption 7 by:

Assumption 7*. $\Phi_{R,R}-\Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R}=0$.

Denote

$$\sigma_{p,n}^2(\rho)=n^{-1}\sigma_0^2\mathrm{tr}\left\{T_n'^{-1}T_n'(\rho)T_n(\rho)T_n^{-1}\right\}. \tag{3.5}$$

When Assumption 7* holds, the parameter ρ_0 can be identified in terms of the concentrated quasi-log likelihood when $\{l_n\}$ is bounded. This is stated in the next assumption.

Assumption 8. The sequence $\{l_n\}$ is bounded and for any $\rho\neq\rho_0$,

$$\lim_{n\to\infty}\left\{\frac{1}{n}\log\left|\sigma_0^2T_n^{-1}\left(T_n^{-1}\right)'\right|-\frac{1}{n}\log\left|\sigma_{p,n}^2(\rho)T_n^{-1}(\rho)\left(T_n^{-1}(\rho)\right)'\right|\right\}\neq 0.$$

Like Lee (2004), under Assumption 7*, Assumption 8 implies that $\lim_{n\to\infty}n^{-1}\mathrm{tr}(C_n^sC_n^{s\prime})\neq 0$, where $C_n=G_n-(n^{-1}\mathrm{tr}(G_n))I_n$ and $C_n^s=C_n+C_n'$. The latter condition is necessary for the $\sqrt{n}$-consistency and asymptotic normality of the QMLE $\widehat{\theta}$ under Assumption 7*, see Lemma 4.2.

4. Asymptotic properties: The regular case

In this section we first discuss the consistency and asymptotic normality of $\widehat{\theta}$ and then the asymptotic normality of $\widehat{\alpha}(z)$.

4.1. Asymptotic property of $\widehat{\theta}$

From (2.1) and (2.2), we obtain the reduced form equation of Y_n

$$Y_n=X_n\beta_0+\mathbf{m}_0(Z_n)+\rho_0G_n(X_n\beta_0+\mathbf{m}_0(Z_n))+T_n^{-1}U_n \tag{4.1}$$

because $I_n+\rho_0G_n=T_n^{-1}$, where $G_n=W_nT_n^{-1}$. The above equation is frequently used in later derivation.

Define $Q_n(\rho)=\max_{\beta,\sigma^2}E[\log L_n(\theta,\mathbf{m}_\theta(Z_n))]$. The optimal solutions of this maximization problem are

$$\beta^*(\rho)=[X_n'(I_n-S_n)'(I_n-S_n)X_n]^{-1}X_n'(I_n-S_n)'\times(I_n-S_n)T_n(\rho)T_n^{-1}(X_n\beta_0+\mathbf{m}_0(Z_n)), \tag{4.2}$$

and the QMLE of σ^2 is

$$\begin{aligned}\sigma^{*2}(\rho)&=\frac{1}{n}E\Big\{\big[(I_n-S_n)(T_n(\rho)Y_n-X_n\beta^*(\rho))\big]'\times\big[(I_n-S_n)(T_n(\rho)Y_n-X_n\beta^*(\rho))\big]\Big\}\\&=\frac{1}{n}\{\mathbf{m}_0(Z_n)+(\rho_0-\rho)R_n\}'(I_n-S_n)'\times M_n(I_n-S_n)\{\mathbf{m}_0(Z_n)+(\rho_0-\rho)R_n\}\\&\quad+\frac{\sigma_0^2}{n}\mathrm{tr}\left\{T_n'^{-1}T_n'(\rho)T_n(\rho)T_n^{-1}\right\}.\end{aligned} \tag{4.3}$$

Consequently, we have

$$Q_n(\rho)=-\frac{n}{2}(\log(2\pi)+1)-\frac{n}{2}\log\sigma^{*2}(\rho)+\log|T_n(\rho)|. \tag{4.4}$$

To show the consistency of $\widehat{\theta}$, we follow Lee (2002b) by identifying ρ_0 based upon the maximum value of $\{Q_n(\rho)/n\}$ and showing the uniform convergence of $\{\log L_n^c(\rho)-Q_n(\rho)\}/n$ to zero on Δ.

Theorem 4.1. *Under Assumptions 1–7 or Assumptions 1–6, 7* and 8, θ_0 is globally identifiable and $\widehat{\theta}$ is consistent with θ_0.*

Next, we derive the asymptotic distribution of $\widehat{\theta}$ by using a Taylor expansion of $\partial\log L_n(\widehat{\theta})/\partial\theta$ at θ_0. Let $P_n=(I_n-S_n)'(I_n-S_n)$. The first order derivatives of $n^{-1/2}\log L_n(\theta)$ at θ_0 are

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\beta}&=\frac{1}{\sigma_0^2\sqrt{n}}X_n'P_n(\mathbf{m}_0(Z_n)+U_n),\\\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\rho}&=\frac{1}{\sigma_0^2\sqrt{n}}R_n'P_nU_n+\frac{1}{\sigma_0^2\sqrt{n}}\left\{U_n'G_n'P_nU_n-\sigma_0^2\mathrm{tr}(G_n)\right\}+\frac{1}{\sigma_0^2\sqrt{n}}(W_nY_n)'P_n\mathbf{m}_0(Z_n),\\\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\sigma^2}&=\frac{1}{2\sigma_0^4\sqrt{n}}\left\{U_n'P_nU_n-n\sigma_0^2\right\}+\frac{1}{2\sigma_0^4\sqrt{n}}\times\left\{\mathbf{m}_0(Z_n)'P_n\mathbf{m}_0(Z_n)+2\mathbf{m}_0(Z_n)'P_nU_n\right\}.\end{aligned}$$

By Lemmas A.7 and A.8 in Appendix A, $(1/\sqrt{n})A_n'P_n\mathbf{m}_0(Z_n)=o_P(1)$ for $A_n=X_n,\mathbf{m}_0(Z_n),U_n$ and W_nY_n. Consequently, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\beta}&=\frac{1}{\sigma_0^2\sqrt{n}}X_n'P_nU_n+o_P(1),\\\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\rho}&=\frac{1}{\sigma_0^2\sqrt{n}}R_n'P_nU_n+\frac{1}{\sigma_0^2\sqrt{n}}\times\left\{U_n'G_n'P_nU_n-\sigma_0^2\mathrm{tr}(G_n)\right\}+o_P(1),\\\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\theta_0)}{\partial\sigma^2}&=\frac{1}{2\sigma_0^4\sqrt{n}}\left\{U_n'P_nU_n-n\sigma_0^2\right\}+o_P(1).\end{aligned}$$

It follows from Assumptions 1–5 and Lemmas A.5–A.8 that is given in Box I.

The expected Hessian matrix is given by Eq. (4.5) in Box II. It is easy to verify that

$$E\left[\frac{1}{n}\frac{\partial\log L_n(\theta_0)}{\partial\theta}\frac{\partial\log L_n(\theta_0)}{\partial\theta'}\right]=\Sigma_\theta+\Omega_{n,\theta}+o(1), \tag{4.6}$$

where $\Omega_{n,\theta}$ is given in Box III.

Before we state Theorem 4.3, we first prove a lemma.

Lemma 4.2. *Under Assumptions 1–7 or Assumptions 1–6, 7* and 8, Σ_θ is positive definite.*

With the above lemma, we can state the second main result in this section.

Theorem 4.3. *Under Assumptions 1–7 or Assumptions 1–6, 7* and 8, $\sqrt{n}(\widehat{\theta}-\theta_0)\xrightarrow{d}N(0,\Sigma_\theta^{-1}+\Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1})$, where $\Omega_\theta=\lim_{n\to\infty}\Omega_{n,\theta}$.*

Remark 1. As in Lee (2004), the asymptotic results in Theorems 4.1 and 4.3 are valid under the two sets of conditions. In either case, the sequence $\{l_n\}$ is bounded so that c_1 is generally nonzero and Σ_θ is usually not a block diagonal matrix in (4.5). When U_n is normally distributed, $\Omega_{n,\theta}=0$. Even in this case, the estimators $(\widehat{\beta}',\widehat{\gamma})'$ and $\widehat{\sigma}$ will be generally asymptotically dependent.

Remark 2. To make a statistical inference on θ_0, we need to estimate Ω_θ and Σ_θ consistently. Let $\widehat{u}_{n,i}=y_{n,i}-x_{ni}'\widehat{\beta}-\widehat{\rho}\sum_{j=1}^n w_{n,ij}y_{n,j}-\widehat{m}(z_{n,i})$, where $\widehat{m}(z)$ is the first element in $\widehat{\alpha}(z)$ and $y_{n,i}$ is the ith element of Y_n. We can estimate σ_0^2,μ_3,μ_4 consistently by the sample moments of $\widehat{u}_{n,i}$. We denote these estimators as $\widehat{\sigma}^2$, $\widehat{\mu}_3$, and $\widehat{\mu}_4$, respectively. Then we can estimate Ω_θ consistently by $\widehat{\Omega}_n$ which equals $\Omega_{n,\theta}$ with $\sigma_0^2,\mu_3,\mu_4,G_n,R_n$ and c_1, being replaced by $\widehat{\sigma}^2,\widehat{\mu}_3,\widehat{\mu}_4,\widehat{G}_n,\widehat{R}_n$ and $\widehat{c}_1$ respectively. Here $\widehat{G}_n=W_n(I_n-\widehat{\rho}W_n)^{-1}$, $\widehat{R}_n=W_n(X_n\widehat{\beta}+\widehat{\mathbf{m}}(Z_n))^{-1}$, $\widehat{c}_1=\mathrm{tr}(\widehat{G}_n)/n$, and the ith element of $\widehat{\mathbf{m}}(Z_n)$ is given by $\widehat{m}(z_{n,i})$. It is standard to show that $\widehat{\Omega}_n$ is consistent with Ω_θ. Similarly, we can estimate Σ_θ consistently by

$$\widehat{\Sigma}_n=\begin{pmatrix}\dfrac{X_n'P_nX_n}{\widehat{\sigma}^2n}&\dfrac{X_n'P_n\widehat{R}_n}{\widehat{\sigma}^2n}&0\\\dfrac{X_n'P_n\widehat{R}_n}{\widehat{\sigma}^2n}&\dfrac{X_n'P_n\widehat{R}_n}{\widehat{\sigma}^2n}+\dfrac{\mathrm{tr}(\widehat{G}_n^2+\widehat{G}_n'P_n\widehat{G}_n)}{n}&\dfrac{\widehat{c}_1}{\widehat{\sigma}^2}\\0&\dfrac{\widehat{c}_1}{\widehat{\sigma}^2}&\dfrac{1}{2\widehat{\sigma}^4}\end{pmatrix}.$$

$$E\left[\frac{1}{n}\frac{\partial \log L_n(\theta_0)}{\partial\theta}\frac{\partial \log L_n(\theta_0)}{\partial\theta'}\right]$$
$$=\begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'P_nP_nX_n & \frac{X_n'P_nP_nR_n}{\sigma_0^2 n}+\frac{\mu_3X_n'P_n\mathrm{diag}(G_n'P_n)}{\sigma_0^4 n} & \frac{\mu_3}{2\sigma_0^6 n}X_n'P_n\mathrm{diag}(P_n) \\ * & \frac{R_n'P_nP_nR_n}{\sigma_0^2 n}+\frac{2\mu_3R_n'P_n\mathrm{diag}(G_n'P_n)}{\sigma_0^4 n}+\frac{\left(\mu_4-3\sigma_0^4\right)\sum_{i=1}^n g_{n,ii}^2}{\sigma_0^4 n}+\frac{\mathrm{tr}\left(G_n'\left(G_n'+G_n\right)\right)}{n} & \frac{\mu_3}{2\sigma_0^6 n}R_n'P_n\mathrm{diag}(P_n)+\frac{\left(\mu_4-3\sigma_0^4\right)\mathrm{tr}\left(G_n\right)}{\sigma_0^4 n}+\frac{\mathrm{tr}\left(G_n\right)}{\sigma_0^4 n} \\ * & * & \frac{1}{2\sigma_0^4}+\frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix}+o(1)$$

which is a symmetric matrix with $g_{n,ii}$ being the (i,i)th element of G_n and $\mu_j=E(u_1^j)$, $j=3$ and 4. To get the above result, we have used the facts that $E(U_n'A_nU_n)=\sigma_0^2\mathrm{tr}(A_n)$, and $E(U_n'A_nU_nU_n'B_nU_n)=(\mu_4-3\sigma_0^4)\sum_{i=1}^n a_{n,ii}b_{n,ii}+\sigma_0^4[\mathrm{tr}(A_n)\mathrm{tr}(B_n)+\mathrm{tr}(A_nB_n)+\mathrm{tr}(A_nB_n')]$.

Box I.

$$-E\left[\frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta\partial\theta'}\right]=\begin{pmatrix} \frac{X_n'P_nX_n}{\sigma_0^2 n} & \frac{X_n'P_nR_n}{\sigma_0^2 n} & \frac{X_n'P_n\mathbf{m}_0(Z_n)}{\sigma_0^4 n} \\ * & \frac{R_n'P_nR_n}{\sigma_0^2 n}+\frac{\mathrm{tr}\left(G_n^2\right)+\mathrm{tr}\left(G_n'P_nG_n\right)}{n} & \frac{\mathrm{tr}\left(G_n'P_n\right)}{\sigma_0^2 n}+\frac{R_n'P_n\mathbf{m}_0(Z_n)}{\sigma_0^4 n} \\ * & * & -\frac{1}{2\sigma_0^4}+\frac{\mathrm{tr}(P_n)}{\sigma_0^4 n}+\frac{\mathbf{m}_0'(Z_n)P_n\mathbf{m}_0(Z_n)}{\sigma_0^6 n} \end{pmatrix}$$
$$=\begin{pmatrix} \frac{1}{\sigma_0^2}\Phi_{X,X} & \frac{1}{\sigma_0^2}\Phi_{X,R} & 0 \\ \frac{1}{\sigma_0^2}\Phi_{X,R}' & \frac{1}{\sigma_0^2}\Phi_{R,R}+c_2 & \frac{c_1}{\sigma_0^2} \\ 0 & \frac{c_1}{\sigma_0^2} & \frac{1}{2\sigma_0^4} \end{pmatrix}+o(1)\equiv\Sigma_\theta+o(1) \tag{4.5}$$

Box II.

$$\Omega_{n,\theta}=\begin{pmatrix} 0 & \frac{\mu_3X_n'P_n\mathrm{diag}(G_n'P_n)}{\sigma_0^4 n} & \frac{\mu_3X_n'P_n\mathrm{diag}(P_n)}{2\sigma_0^6 n} \\ * & \frac{2\mu_3R_n'P_n\mathrm{diag}(G_n'P_n)}{\sigma_0^4 n}+\frac{\left(\mu_4-3\sigma_0^4\right)\sum_{i=1}^n g_{n,ii}^2}{\sigma_0^4 n} & \frac{\mu_3R_n'P_n\mathrm{diag}(P_n)}{2\sigma_0^6 n}+\frac{\left(\mu_4-3\sigma_0^4\right)c_1}{\sigma_0^4 n} \\ * & * & \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix}$$

Box III.

4.2. Asymptotic normality of $\widehat{\alpha}(z)$

To study the asymptotic property of $\widehat{\alpha}(z)$, we denote the true value of $\alpha(z)$ in (2.4) as $\alpha^0(z)$, an $N\times 1$ vector that is a collection of $m_0(z)$ and its scaled partial derivatives up to rth order. Also, we arrange the N_{r+1} elements of the derivatives

$$(D^{\mathbf{j}}m_0)(z)\equiv\frac{1}{\mathbf{j}!}\frac{\partial^{|\mathbf{j}|}m_0(z)}{\partial^{j_1}z_1,\ldots,\partial^{j_q}z_q} \tag{4.7}$$

($|\mathbf{j}|=r+1$) as a column vector $m_0^{(r+1)}(z)$ in the lexicographical order. Then we have the following theorem.

Theorem 4.4. *Under Assumptions 1–7 or Assumptions 1–6, 7* and 8, if* $nh^{2(r+1)+q}\to c_0\in[0,\infty)$, *then* $\sqrt{nh^q}(\widehat{\alpha}(z)-\alpha^0(z)-h^{r+1}\Pi^{-1}Bm_0^{(r+1)}(z))\xrightarrow{d}N(0,\sigma_0^2\Pi^{-1}\Gamma\Pi^{-1}/f(z))$, *where* Π, Γ *and B are defined in (Matrix) in the Appendix.*

Remark 3. Theorem 4.4 tells us that $\widehat{\alpha}(z)$ is asymptotically normally distributed as if the finite dimensional parameters $\theta=(\beta',\rho,\sigma^2)'$ are known. In particular the estimator for $m_0(z)$ and that for its partial derivatives have different rates of convergence (see the definition of $\alpha^0(z)$). In the special case where $r=1$, Π and Γ reduce to

$$\Pi=\begin{pmatrix}1 & 0' \\ 0 & \int_{\mathbb{R}^q}uu'K(u)\,\mathrm{d}u\end{pmatrix},$$
$$\Gamma=\begin{pmatrix}\int_{\mathbb{R}^q}K(u)^2\,\mathrm{d}u & 0' \\ 0 & \int_{\mathbb{R}^q}uu'K(u)^2\,\mathrm{d}u\end{pmatrix} \tag{4.8}$$

(4.8) implies that the estimator for $m_0(z)$ is asymptotically independent of the estimator for its first order partial derivatives. Furthermore, the asymptotic normal distribution given by Theorem 4.3 can be used to calculate pointwise confidence intervals for $m_0(z)$ and its derivatives. To do so, we only need to estimate $m_0^{(r+1)}(z)$ and $f(z)$ consistently given the fact that B, Π and Γ can be calculated once $K(\cdot)$ is chosen and that σ_0^2 can be estimated consistently by $\widehat{\sigma}^2$. It is easy to show that $\widehat{f}(z) = (1/nh^q)\sum_{i=1}^n K((z_{n,i}-z)/h) \to f(z)$. To estimate $m_0^{(r+1)}(z)$, one can use a higher order local polynomial regression of $y_{n,i} - x_{ni}'\widehat{\beta} - \widehat{\rho}\sum_{j=1}^n w_{n,ij}y_{n,j}$ on z_{ni}. The procedure is standard in the nonparametric literature and we omit it for brevity.

5. Asymptotic properties: The irregular case

When $\lim_{n\to\infty} l_n = \infty$, Σ_θ is singular. The singularity of the information matrix affects the convergence rate of the estimators. This is true because when $\lim_{n\to\infty} l_n = \infty$, $(1/n)\log L_n^c(\rho)$ is rather flat in ρ and the convergence of $(1/n)(\log L_n^c(\rho) - Q_n(\rho))$ to zero is too fast to be useful. Nevertheless, we follow Lee (2004) and demonstrate that with a properly adjusted rate,

$$(l_n/n)\left[\left(\log L_n^c(\rho) - \log L_n^c(\rho_0)\right) - \left(Q_n(\rho) - Q_n(\rho_0)\right)\right] \xrightarrow{p} 0 \text{ uniformly in } \Delta. \tag{5.1}$$

We consider the situation in which

$$\lim_{n\to\infty} (l_n/n)\, R_n' (I_n - S_n)' M_n (I_n - S_n) R_n = c, \text{ where } 0 \le c < \infty. \tag{5.2}$$

In this case, it is natural to assume that elements of $(I_n - S_n)R_n$ are of uniform order $O(1/\sqrt{l_n})$, implying that elements of $M_n(I_n - S_n)R_n$ are also uniformly $O(1/\sqrt{l_n})$ by Fact 2 and Lemma A.5 in the Appendix. We make the following assumption.

Assumption 9. The sequence $\{l_n\}$ is a divergent sequence, elements of $(I_n - S_n)R_n$ have the uniform order $O(1/\sqrt{l_n})$, and (5.2) holds. Under this situation, either (i) $c > 0$, or (ii) if $c = 0$,

$$\lim_{n\to\infty}\left\{\frac{l_n}{n}\log\left|\sigma_0^2 T_n^{-1} T_n'^{-1}\right| - \frac{1}{n}\log\left|\sigma_{p,n}^2(\rho)\, T_n^{-1}(\rho)\, T_n'^{-1}(\rho)\right|\right\} \neq 0$$
whenever $\rho \neq \rho_0$.

Assumption 9(ii) modifies Assumption 8 with the factor l_n that takes into account the proper convergence rate of the profile likelihood function.

5.1. Asymptotic property of $\widehat{\theta}$

The following theorem claims that the spatial parameter can be consistently estimated from the concentrated log-likelihood.

Theorem 5.1. *Suppose Assumptions 1–6 and 9 hold and* $\lim_{n\to\infty} l_n h^{2(r+1)} = 0$. *Then* $\widehat{\rho}$ *is consistent with* ρ_0.

We can also derive the asymptotic distribution of $\widehat{\rho}$ from the concentrated log-likelihood function. Once we obtain the asymptotic distribution of $\widehat{\rho}$, we can apply the delta method to derive the asymptotic distributions of the estimators of other parameters in the partially linear SAR model.

To proceed, we derive the asymptotic distribution of $\widehat{\rho}$ by using a Taylor expansion of $\partial \log L_n(\widehat{\rho})/\partial\rho = 0$ around ρ_0 to obtain

$$\sqrt{\frac{n}{l_n}}\left(\widehat{\rho} - \rho_0\right) = -\left(\frac{l_n}{n}\frac{\partial^2 \log L_n(\widetilde{\rho}_n)}{\partial\rho^2}\right)^{-1}\sqrt{\frac{l_n}{n}}\frac{\partial \log L_n(\rho_0)}{\partial\rho}, \tag{5.3}$$

where $\widetilde{\rho}_n$ is the mean value. We shall show that $(l_n/n)\partial^2 \log L_n^c(\rho)/\partial\rho^2$ converges to a nonzero scalar uniformly on Δ. Also, we can apply the central limit theorem of Lee (2004) to $\sqrt{l_n/n}\,\partial^2 \log L_n^c(\rho)/\partial\rho^2$. The result is reported in the next theorem.

Theorem 5.2. *Suppose Assumptions 1–6 and 9 hold,* $\lim_{n\to\infty} nh^{2(r+1)} = 0$ *and* $\mathrm{tr}(G_n'(S_n' + S_n - S_n'S_n)) = o(\sqrt{n/l_n})$. *Then* $\sqrt{n/l_n}(\widehat{\rho} - \rho_0) \xrightarrow{d} N(0, \sigma_\rho^2)$, *where*

$$\sigma_\rho^2 = \lim_{n\to\infty}\left[\frac{l_n}{n\sigma_0^2}R_n'M_n^\dagger R_n + \frac{l_n}{n}\mathrm{tr}\left(G_n'\left(P_nG_n + G_n'\right)\right)\right]^{-2} \times \frac{l_n}{n\sigma_0^4}\left[\sigma_0^2 R_n'M_n^\dagger R_n + \sigma_0^4\,\mathrm{tr}\left(P_nC_n\left(P_nC_n + C_n'P_n\right)\right) + 2R_n'M_n^\dagger \mathrm{diag}(P_nC_n)\mu_3\right],$$

and $M_n^\dagger = (I_n - S_n)'M_n(I_n - S_n)$.

The condition that $\lim_{n\to\infty} nh^{2(r+1)} = 0$ implies that an undersmoothing bandwidth has to be used. The assumption that $\mathrm{tr}(G_n'(S_n'+S_n-S_n'S_n)) = o(\sqrt{n/l_n})$ is a high-level assumption, which is difficult to verify. Without this assumption, we need to correct the bias of $\widehat{\rho}$ in order to get the above limiting distribution. See Step (3) in the proof Theorem 5.2. The above theorem implies that the asymptotic distribution of $\widehat{\rho}$ has the $\sqrt{n/l_n}$-rate of convergence, which is slower than the regular $\sqrt{n}$-rate as $l_n \to \infty$. As Lee (2004) remarks, some practical formulae for classical inference statistics associated with the spatial parameters are still valid after one takes into account the factor l_n in addition to the sample size n.

Theorem 5.3. *Under the conditions of Theorem 5.2,* $\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma_\beta^2)$, *and* $\sqrt{n}(\widehat{\sigma}^2 - \sigma_0^2) \xrightarrow{d} N(0, \mu_4 - \sigma_0^4)$, *where*

$$\sigma_\beta^2 = \sigma_0^2 \lim_{n\to\infty} n\left(X_n'P_nX_n\right)^{-1}X_n'P_nP_nX_n\left(X_n'P_nX_n\right)^{-1} + \sigma_\rho^2 \lim_{n\to\infty} l_n\left(X_n'P_nX_n\right)^{-1}X_n'P_nR_n\left(X_n'P_nR_n\right)'\left(X_n'P_nX_n\right)^{-1}. \tag{5.4}$$

In the special case with $\beta_0 = 0$, $\sigma_\beta^2 = \sigma_0^2 \lim_{n\to\infty} n(X_n'P_nX_n)^{-1}X_n'P_nP_nX_n(X_n'P_nX_n)^{-1}$.

Despite the slow convergence rate of $\widehat{\rho}$, Theorem 5.3 implies that both $\widehat{\beta}$ and $\widehat{\sigma}^2$ converge to their true values at the regular $\sqrt{n}$-rate. The $\sqrt{n}$-rate of convergence of $\widehat{\sigma}^2$ is not surprising, which is also obtained in Theorem 5.3 of Lee (2004). Nevertheless, Lee (2004) obtained a $\sqrt{n/h_n}$-rate of convergence for his estimator $\widehat{\beta}_n$ of β_0, where Lee's notation h_n plays the role of l_n here. To see why this occurs, we first take a close look at the asymptotic variance formula of $\sqrt{n/h_n}(\widehat{\beta}_n - \beta_0)$, which is proportional to $\lim_{n\to\infty}(n^{-1}X_n'X_n)^{-1}n^{-1}X_n'(G_nX_n\beta_0)n^{-1}(G_nX_n\beta_0)'X_n(n^{-1}X_n'X_n)^{-1}$. His result relies highly on the assumption that $n^{-1}X_n'G_nX_n\beta_0$ has a non-degenerate limit. Nevertheless, if elements $x_{n,i}$ of X_n are i.i.d. with zero mean and finite variance or behave like i.i.d. zero-mean random variables, it is easy to apply Lemma A.2 in the Appendix A and conclude that $n^{-1}X_n'(G_nX_n\beta_0) \to 0$ as $h_n \to \infty$. When this occurs, Lee would need to rescale $\widehat{\beta}_n - \beta_0$ by $\sqrt{n}$ to obtain a non-degenerate distribution and thus obtain the regular $\sqrt{n}$-rate of convergence. See Lee (2004, p. 1912) for discussions on the case where some components of $n^{-1}X_n'(G_nX_n\beta_0)$ have limit zero. In our case, because of the presence of a nonparametric object in the regression function, the data are pre-smoothed before they appear in the variance formula, resulting a zero limit for $n^{-1}X_n'P_nR_n$ when $\lim_{n\to\infty} l_n = \infty$. Instead, $(\sqrt{l_n}/n)X_n'P_nR_n$ has a non-degenerate limit under Assumption 9.

The asymptotic variance formula of $\sqrt{n}(\widehat{\beta} - \beta_0)$ consists of two parts. The first part is present even if we do not need to estimate ρ_0 while the second part reflects the effect of estimating ρ_0 on the estimation of β_0.

5.2. Asymptotic normality of $\widehat{\alpha}(z)$

When $\lim_{n\to\infty} l_n = \infty$, we require that $\lim_{n\to\infty} nh^{2(r+1)} = 0$ in order to ensure $\sqrt{n/l_n}$-rate of consistency of $\widehat{\rho}$. This means that undersmoothing is required in the estimation procedure. The side effect of undersmoothing is that the asymptotic bias of $\widehat{\alpha}(z)$ vanishes, as is shown in the following theorem.

Theorem 5.4. *Under the conditions of Theorem 5.2, $\sqrt{nh^q}(\widehat{\alpha}(z) - \alpha^0(z)) \xrightarrow{d} N(0, \sigma_0^2 \Pi^{-1}\Gamma\Pi^{-1}/f(z))$, where Π, Γ and B are defined in (Matrix) in the Appendix.*

Theorem 5.4 says that $\widehat{\alpha}(z)$ is asymptotically normally distributed and $\widehat{\alpha}(z)$ converges to $\alpha^0(z)$ at $\sqrt{nh^q}$-rate. Because of undersmoothing, this rate is slower than that obtained in Theorem 4.4.

6. Monte Carlo simulations

We now present Monte Carlo experiment results that illustrate the finite sample performance of the QMLE estimator. Like Lee (2004), we focus on the spatial scenario in Case (1991) with R number of districts, m members in each district, and with each neighbor of a member in a district given equal weight, i.e., $W_n = I_R \otimes B_m$, where $B_m = (l_m l_m' - I_m)/(m-1)$.

We first consider the following data generating process (DGP) with and without the nonparametric component in (2.1):

$$Y_n = X_n + \tau_0 \frac{\exp(Z_n)}{1+\exp(Z_n)} + \rho_0 W_n Y_n + U_n, \tag{6.1}$$

where $U_n \sim N(0, I_n)$, $X_n = (x_{n,1}, \ldots, x_{n,n})'$, $Z_n = (z_{n,1}, \ldots, z_{n,n})'$, and $z_{n,i}'$s are i.i.d. and each is equal to the sum of 48 independent random variables each uniformly distributed on $[-0.25, 0.25]$. According to the central limit theorem, we can treat $z_{n,i}$ as being nearly a normal random variable with truncated support on $[-12, 12]$. We generate $x_{n,i}$ as $x_{n,i} = 0.2z_{n,i}^2 - z_{n,i} + \eta_{n,i}$, where $\eta_{n,i}'$s are i.i.d., independent of $z_{n,i}'$s and generated in the same way as $z_{n,i}$. Here, $\beta_0 = \sigma_0^2 = 1$, and $m_0(z) = \tau_0 \exp(z)/(1+\exp(z))$. We will consider two choices of spatial parameter ρ_0 (0.25, 0.75), and three choices of τ_0 (0, 1, 2). $\rho_0 = 0.25$ represents relatively weak spatial dependence whereas $\rho_0 = 0.75$ represents relatively strong spatial dependence. $\tau_0 = 0$ represents the case where the regressor Z_n does not explain Y_n; and larger value of τ_0 indicates stronger nonlinear effect of Z_n on Y_n.

We choose $r = 1$ and the standardized Epanechnikov kernel: $K(z) = (3/4\sqrt{5})(1 - \frac{1}{5}z^2)1(z^2 \leq 5)$. As it is difficult to specify the optimal bandwidth sequence, we choose the bandwidth by a rule-of-thumb method: $h = s_z n^{-1/(q+4)}$, where s_z is the sample standard deviation of Z_n and $q = 1$. In practice, it is recommended that we use this bandwidth as the initial smoothing parameter to obtain a preliminary consistent estimator $(\widetilde{\beta}, \widetilde{\rho})$ of (β, ρ) in the model. In the second step, we conduct the least squares cross validation method to choose the bandwidth by regressing $Y_n - X_n\widetilde{\beta} - \widetilde{\rho} W_n Y_n$ on Z_n using the local linear procedure. We have applied both methods to obtain the estimators but found their performances are similar for the DGPs under study. So we only report the simulation results for the former case. We have experimented with two values for R and m: 10 to 40. For each case, there are 1000 repetitions.

Table 1 reports the empirical mean, finite sample standard errors, and asymptotic standard errors for the parameter estimators of $\theta_0 = (\beta_0, \rho_0, \sigma_0^2)'$ and the point estimator of $m_0(0)$. The finite sample standard errors are the empirical standard deviations of the estimators obtained in the 1000 replications; the asymptotic standard errors are calculated from the formula for the asymptotic variance–covariance matrix of these estimators and are averaged across the 1000 replications. We summarize some interesting findings from Table 1. First, we see that the biases for the estimators of β_0, σ_0^2 and $m_0(0)$ are fairly small for almost all cases and they decrease as R or m increases. Like Lee (2004), there are finite sample biases in the estimators of ρ_0 that are not negligible for small sample sizes ($m = R = 10$), but they decrease as R increases. Second, for fixed m (10 or 40), as R quadruples, we observe that the standard errors for the finite dimensional estimators are roughly halved for most cases as predicted by the theory in Section 4. For the estimator of the nonparametric component, as R quadruples, their asymptotic standard errors drops less than 50% while the decrease of their finite sample standard errors varies across cases (one reason might be that $m_0(0)$ cannot be estimated well for small sample sizes). Third, for fixed R (10 or 40), as m quadruples, the standard errors for the estimators of β_0 and σ_0^2 are also halved approximately, whereas the empirical standard errors for the estimators of ρ_0 do not vary much. As m changes the behavior of the nonparametric estimator of $m_0(0)$ is similar to the case where R changes.

For comparison purposes, we report in Table 2 the estimation result when we fit a linear SAR model of Lee (2004) to (6.1). As a referee remarks, the linear SAR model accommodates situations where the regressors include powers of $z_{n,i}$. So even though the data Y_n are generated according to (6.1), we consider fitting the data by the following two models:

$$Y_n = \beta_1 X_n + \beta_2 Z_n + \rho W_n Y_n + U_n, \quad \text{and} \tag{6.2}$$

$$Y_n = \beta_1 X_n + \beta_2 Z_n + \beta_3 Z_n^2 + \beta_4 Z_n^3 + \rho W_n Y_n + U_n. \tag{6.3}$$

Model (6.2) includes a linear term $z_{n,i}$ in the regression whereas model (6.3) includes $z_{n,i}$, $z_{n,i}^2$ and $z_{n,i}^3$. Unless $\tau_0 = 0$, either model is misspecified so that we will study the consequence of such misspecification. To save space, we only report the empirical mean and standard deviation for the estimator of ρ_0. From Table 2 we see that when the model is misspecified ($\tau_0 = 1, 2$) the biases of the estimators for ρ_0 do not decrease when either R or m increases. This indicates the inconsistency of the estimator of ρ_0 when fitting a misspecified linear model.

To check the sensitivity of the estimator to the choices of parameters other than the spatial parameter, we consider the following DGP:

$$Y_n = \beta_{01} X_{n1} + \beta_{02} X_{n2} + \mathbf{m}_0(Z_n) + 0.5 W_n Y_n + U_n, \tag{6.4}$$

where $U_n \sim N(0, I_n)$, $X_{nj} = (x_{n,1j}, \ldots, x_{n,nj})'$ for $j = 1, 2$, $Z_n = (z_{n,1}, \ldots, z_{n,n})'$. We consider generating Z_n from a SAR process: $Z_n = 0.5W_n Z_n + E_n$, where $E_n = (e_{n,1}, \ldots, e_{n,n})'$, $e_{n,i}$ are i.i.d. and each is equal to the sum of 48 independent random variables each uniformly distributed on $[-0.25, 0.25]$. For $i = 1, \ldots, n$, $x_{n,i2}$ are i.i.d. $U(0, 4)$, whereas $x_{n,i1}$ are generated analogously to $x_{n,i}$ in (6.1): $x_{n,i1} = 0.2z_{n,i}^2 - z_{n,i} + \eta_{n,i}$, where $\eta_{n,i}$ are generated like $e_{n,i}$. $\{u_{n,i}\}$, $\{e_{n,i}\}$, $\{\eta_{n,i}\}$ and $\{x_{n,i2}\}$ are mutually independent. We will consider two choices of β_{01} (0.2, 1), two choices of β_{02} (0.2, 1), and three choices of $m_0(\cdot)$:

$$\text{Case (a)}: \quad m_0(z) = -0.25z^2 + z, \tag{6.5}$$

$$\text{Case (b)}: \quad m_0(z) = 0.2444 + \phi(z; 0.2, 0.5) + 2\phi(z; 0.8, 0.25), \tag{6.6}$$

$$\text{Case (c)}: \quad m_0(z) = 1 + 1.5\sin(0.5\pi z), \tag{6.7}$$

where $\phi(z; a, b)$ is the normal density function with mean a and standard deviation b.

Table 3 reports the empirical mean and standard errors (in brackets) for the parameter estimators of $\theta_0 = (\beta_{01}, \beta_{02}, \rho_0, \sigma_0^2)'$ and the point estimator of $m_0(0)$. We summarize some findings. First, we see that the biases for the estimators of β_{01}, β_{02}, and σ_0^2 are fairly small for almost all cases and they decrease as R increases. There are finite sample biases in the estimators of ρ_0 and $m_0(0)$ (especially for Case (a)), but they decrease as R increases. Second, as R quadruples, we observe that the standard errors behave like the case for fixed m in Table 1. In particular, the standard errors of

Table 1
Empirical mean and standard errors (in brackets) of our estimators.

m	ρ_0	τ_0	$R=10$				$R=40$			
			β	ρ	σ^2	$m_0(0)$	β	ρ	σ^2	$m_0(0)$
10	0.25	0	0.995 (0.104) [0.101]	0.196 (0.144) [0.172]	0.917 (0.137) [0.130]	0.012 (0.096) [0.131]	0.998 (0.052) [0.050]	0.242 (0.059) [0.079]	0.968 (0.070) [0.068]	0.002 (0.049) [0.075]
		1	0.995 (0.104) [0.102]	0.192 (0.152) [0.181]	0.917 (0.137) [0.130]	0.547 (0.139) [0.131]	0.998 (0.051) [0.050]	0.241 (0.063) [0.083]	0.968 (0.071) [0.068]	0.507 (0.067) [0.075]
		2	0.995 (0.104) [0.102]	0.189 (0.159) [0.188]	0.917 (0.137) [0.130]	1.085 (0.212) [0.131]	0.998 (0.051) [0.050]	0.240 (0.066) [0.086]	0.968 (0.071) [0.068]	1.013 (0.096) [0.075]
	0.75	0	0.999 (0.104) [0.102]	0.730 (0.052) [0.086]	0.924 (0.139) [0.138]	0.013 (0.097) [0.131]	0.999 (0.052) [0.050]	0.747 (0.021) [0.039]	0.970 (0.072) [0.072]	0.001 (0.049) [0.075]
		1	0.999 (0.105) [0.102]	0.729 (0.054) [0.088]	0.925 (0.139) [0.138]	0.550 (0.144) [0.131]	0.999 (0.052) [0.050]	0.747 (0.022) [0.039]	0.970 (0.072) [0.072]	0.507 (0.068) [0.075]
		2	0.991 (0.104) [0.102]	0.728 (0.057) [0.090]	0.925 (0.139) [0.139]	1.091 (0.222) [0.131]	0.999 (0.052) [0.050]	0.746 (0.023) [0.040]	0.970 (0.073) [0.072]	1.028 (0.099) [0.076]
40	0.25	0	0.998 (0.051) [0.051]	0.200 (0.144) [0.177]	0.968 (0.070) [0.068]	0.011 (0.057) [0.075]	1.001 (0.026) [0.025]	0.245 (0.061) [0.081]	0.989 (0.036) [0.035]	0.005 (0.038) [0.043]
		1	0.998 (0.051) [0.050]	0.194 (0.154) [0.186]	0.968 (0.070) [0.068]	0.545 (0.118) [0.075]	1.001 (0.025) [0.025]	0.244 (0.066) [0.085]	0.989 (0.036) [0.035]	0.509 (0.061) [0.043]
		2	0.998 (0.051) [0.050]	0.188 (0.164) [0.194]	0.968 (0.070) [0.068]	1.008 (0.199) [0.075]	1.001 (0.025) [0.025]	0.245 (0.068) [0.089]	0.989 (0.036) [0.035]	1.011 (0.095) [0.043]
	0.75	0	0.999 (0.051) [0.050]	0.733 (0.049) [0.084]	0.970 (0.071) [0.069]	0.011 (0.057) [0.075]	1.001 (0.025) [0.025]	0.748 (0.019) [0.038]	0.989 (0.036) [0.035]	0.003 (0.028) [0.043]
		1	0.999 (0.051) [0.050]	0.731 (0.052) [0.085]	0.970 (0.071) [0.069]	0.545 (0.119) [0.075]	1.002 (0.025) [0.025]	0.749 (0.027) [0.038]	0.991 (0.035) [0.035]	0.507 (0.056) [0.043]
		2	0.999 (0.051) [0.050]	0.729 (0.055) [0.087]	0.970 (0.071) [0.069]	1.084 (0.201) [0.075]	1.001 (0.026) [0.025]	0.748 (0.020) [0.039]	0.989 (0.036) [0.035]	1.013 (0.078) [0.043]

Note: Numbers in round brackets are finite sample standard errors based upon 1000 replications. Numbers in square brackets are standard errors based upon asymptotic formula averaged over 1000 replications. $\beta_0 = \sigma_0^2 = 1$. $m(0) = 0, 0.5, 1$ for $\tau_0 = 0, 1$, and 2, respectively.

Table 2
Estimator of spatial dependence parameter when model (6.1) is mistaken as model (6.2) or model (6.3).

m	ρ_0	τ_0	Model (6.2)		Model (6.3)	
			$R=10$ Mean (s.e.)	$R=40$ Mean (s.e.)	$R=10$ Mean (s.e.)	$R=40$ Mean (s.e.)
10	0.25	0	0.220 (0.124)	0.246 (0.055)	0.214 (0.129)	0.245 (0.056)
		1	0.513 (0.067)	0.519 (0.031)	0.460 (0.080)	0.472 (0.036)
		2	0.699 (0.038)	0.697 (0.019)	0.651 (0.047)	0.655 (0.022)
	0.75	0	0.739 (0.045)	0.748 (0.021)	0.736 (0.046)	0.748 (0.020)
		1	0.841 (0.023)	0.842 (0.011)	0.822 (0.027)	0.826 (0.012)
		2	0.903 (0.013)	0.902 (0.007)	0.887 (0.016)	0.888 (0.007)
40	0.25	0	0.228 (0.114)	0.250 (0.046)	0.223 (0.119)	0.249 (0.049)
		1	0.674 (0.035)	0.674 (0.018)	0.639 (0.042)	0.639 (0.021)
		2	0.810 (0.022)	0.811 (0.012)	0.789 (0.026)	0.789 (0.014)
	0.75	0	0.742 (0.039)	0.749 (0.016)	0.740 (0.041)	0.749 (0.017)
		1	0.892 (0.012)	0.892 (0.006)	0.881 (0.014)	0.882 (0.007)
		2	0.939 (0.008)	0.940 (0.004)	0.932 (0.009)	0.932 (0.005)

all parametric estimators are roughly halved, which is consistent with the $\sqrt{n}$-asymptotics in Section 2.

For comparison purposes, we report in Table 4 the estimation results for the spatial parameter when we fit the data by a linear SAR model which includes either $z_{n,i}$ or $(z_{n,i}, z_{n,i}^2, z_{n,i}^3)$ (c.f. model (6.2) and (6.3)) in the regression. Note that in this case, the linear model which includes $(z_{n,i}, z_{n,i}^2, z_{n,i}^3)$ in addition to the parametric component of (6.4) is correctly specified for Case (a). From Table 4 we see that the biases of the estimators for ρ_0 do not decrease when R increases whereas the variances decrease as R increases. Interestingly, this is true even for Case (a) when $(z_{n,i}, z_{n,i}^2, z_{n,i}^3)$ is included in the model.

Table 3
Empirical mean and standard deviation (in brackets) of our estimators.

β_0	m_0	$R = 10$					$R = 40$				
		β_1	β_2	ρ	σ^2	$m_0(0)$	β_1	β_2	ρ	σ^2	$m_0(0)$
(0.2, 0.2)	(a)	0.196	0.204	0.452	0.911	0.979	0.198	0.197	0.489	0.967	0.974
		(0.105)	(0.093)	(0.124)	(0.137)	(0.283)	(0.051)	(0.046)	(0.053)	(0.070)	(0.130)
		[0.103]	[0.088]	[0.167]	[0.135]	[0.132]	[0.050]	[0.044]	[0.075]	[0.070]	[0.076]
	(b)	0.195	0.205	0.446	1.139	1.269	0.198	0.197	0.489	1.111	1.205
		(0.114)	(0.101)	(0.132)	(0.169)	(0.332)	(0.054)	(0.047)	(0.050)	(0.075)	(0.163)
		[0.117]	[0.100]	[0.175]	[0.181]	[0.151]	[0.055]	[0.047]	[0.077]	[0.085]	[0.082]
	(c)	0.197	0.205	0.450	0.935	1.114	0.199	0.197	0.490	0.977	1.031
		(0.105)	(0.093)	(0.126)	(0.138)	(0.304)	(0.052)	(0.046)	(0.050)	(0.070)	(0.149)
		[0.105]	[0.090]	[0.167]	[0.141]	[0.135]	[0.051]	[0.044]	[0.072]	[0.070]	[0.077]
(0.2, 1)	(a)	0.196	1.004	0.465	0.911	1.041	0.198	0.997	0.492	0.968	0.972
		(0.105)	(0.093)	(0.103)	(0.137)	(0.445)	(0.052)	(0.046)	(0.042)	(0.070)	(0.195)
		[0.107]	[0.088]	[0.143]	[0.135]	[0.132]	[0.050]	[0.043]	[0.065]	[0.070]	[0.076]
	(b)	0.195	1.004	0.461	1.139	1.331	0.198	0.997	0.492	1.111	1.221
		(0.114)	(0.102)	(0.104)	(0.169)	(0.542)	(0.054)	(0.048)	(0.041)	(0.076)	(0.225)
		[0.117]	[0.101]	[0.152]	[0.180]	[0.151]	[0.055]	[0.047]	[0.067]	[0.084]	[0.082]
	(c)	0.197	1.004	0.464	0.935	1.175	0.199	0.997	0.493	0.977	1.046
		(0.105)	(0.093)	(0.100)	(0.138)	(0.483)	(0.052)	(0.046)	(0.041)	(0.070)	(0.212)
		[0.105]	[0.090]	[0.144]	[0.140]	[0.135]	[0.051]	[0.044]	[0.065]	[0.070]	[0.077]
(1, 0.2)	(a)	0.996	0.205	0.456	0.911	0.986	0.998	0.197	0.494	0.970	0.971
		(0.105)	(0.093)	(0.115)	(0.137)	(0.302)	(0.052)	(0.046)	(0.050)	(0.070)	(0.134)
		[0.108]	[0.088]	[0.157]	[0.135]	[0.132]	[0.050]	[0.043]	[0.071]	[0.069]	[0.076]
	(b)	0.995	0.205	0.457	1.138	1.249	0.998	0.197	0.493	1.110	1.198
		(0.114)	(0.101)	(0.114)	(0.168)	(0.358)	(0.054)	(0.047)	(0.042)	(0.076)	(0.154)
		[0.117]	[0.101]	[0.156]	[0.180]	[0.151]	[0.055]	[0.047]	[0.069]	[0.084]	[0.083]
	(c)	0.997	0.205	0.460	0.935	1.103	0.999	0.197	0.493	0.977	1.024
		(0.106)	(0.093)	(0.107)	(0.138)	(0.320)	(0.052)	(0.046)	(0.043)	(0.070)	(0.145)
		[0.105]	[0.090]	[0.150]	[0.140]	[0.135]	[0.051]	[0.043]	[0.067]	[0.070]	[0.077]
(1, 1)	(a)	0.996	1.004	0.467	0.911	1.045	0.998	0.997	0.494	0.968	0.975
		(0.105)	(0.093)	(0.095)	(0.137)	(0.467)	(0.052)	(0.046)	(0.040)	(0.070)	(0.197)
		[0.103]	[0.088]	[0.137]	[0.134]	[0.132]	[0.050]	[0.043]	[0.063]	[0.070]	[0.076]
	(b)	0.995	1.004	0.466	1.138	1.311	0.998	0.997	0.494	1.110	1.210
		(0.114)	(0.101)	(0.096)	(0.168)	(0.516)	(0.054)	(0.047)	(0.036)	(0.076)	(0.210)
		[0.117]	[0.101]	[0.139]	[0.179]	[0.151]	[0.055]	[0.047]	[0.062]	[0.082]	[0.083]
	(c)	0.997	1.004	0.469	0.935	1.161	0.998	0.997	0.495	0.977	1.036
		(0.106)	(0.093)	(0.090)	(0.138)	(0.471)	(0.052)	(0.046)	(0.037)	(0.070)	(0.204)
		[0.105]	[0.090]	[0.133]	[0.140]	[0.135]	[0.051]	[0.044]	[0.061]	[0.070]	[0.077]

Note: Numbers in round brackets are finite sample standard errors based upon 1000 replications. Numbers in square brackets are standard errors based upon asymptotic formula averaged over 1000 replications. $\rho_0 = 0.5$, $\sigma_0^2 = 1$, and $m_0(0) = 1$ for each case.

Table 4
Estimator of spatial dependence parameter when (6.5)–(6.7) are approximated by linear or cubic forms.

(β_{01}, β_{02})	m_0	Linear model with $z_{n,i}$		Linear model with $z_{n,i}$, $z_{n,i}^2$, $z_{n,i}^3$	
		$R = 10$ Mean (s.e.)	$R = 40$ Mean (s.e.)	$R = 10$ Mean (s.e.)	$R = 40$ Mean (s.e.)
(0.2, 0.2)	(a)	0.652 (0.055)	0.660 (0.027)	0.685 (0.049)	0.693 (0.024)
	(b)	0.691 (0.048)	0.699 (0.024)	0.726 (0.043)	0.729 (0.022)
	(c)	0.673 (0.053)	0.681 (0.025)	0.681 (0.048)	0.686 (0.025)
(0.2, 1)	(a)	0.612 (0.034)	0.615 (0.017)	0.642 (0.034)	0.645 (0.017)
	(b)	0.649 (0.036)	0.652 (0.019)	0.679 (0.034)	0.680 (0.017)
	(c)	0.633 (0.036)	0.636 (0.019)	0.634 (0.033)	0.636 (0.017)
(1, 0.2)	(a)	0.649 (0.050)	0.656 (0.025)	0.683 (0.046)	0.689 (0.022)
	(b)	0.677 (0.047)	0.684 (0.023)	0.709 (0.043)	0.712 (0.022)
	(c)	0.661 (0.049)	0.667 (0.024)	0.667 (0.045)	0.671 (0.023)
(1, 1)	(a)	0.608 (0.032)	0.610 (0.016)	0.637 (0.032)	0.639 (0.016)
	(b)	0.641 (0.035)	0.644 (0.018)	0.669 (0.033)	0.669 (0.017)
	(c)	0.625 (0.034)	0.628 (0.018)	0.626 (0.031)	0.628 (0.016)

Note: The true parameter is $\rho_0 = 0.5$.

7. Concluding remarks

This paper develops asymptotic theories for the profile quasi-maximum likelihood estimators for the parameters in partially linear SAR models. We show that the convergence rate of the estimator of the spatial parameter depends on some general features of the spatial weights matrix of the model while the estimators of other finite-dimensional parameters in the model have the regular $\sqrt{n}$-rate of convergence. The estimator of the nonparametric component is consistent under different assumptions on the spatial weights matrix and the bandwidth parameter. Simulations indicate that the proposed estimators perform reasonably well in finite samples. By allowing for the regression function to be partially specified in the SAR model, our model is applicable to many situations where we need to take into account both spatial dependence and nonlinearities simultaneously.

Several extensions are possible. First, the error terms in our partially linear model are homoscedastic, which is rather restrictive in some empirical applications. In linear SAR models with heteroscedastic errors, Lin and Lee (2010) demonstrate the inconsistency of the QMLE estimator of Lee (2004) under general cases; Kelejian and Prucha (2010) and Lin and Lee (2010) explore the GMM estimation of the linear SAR models with heteroscedasticity. We conjecture that we can extend the work of Ai and Chen (2003) to the partially specified spatial model with heteroscedastic errors. Second, as a referee notes, the specification of the spatial weights matrix is controversial. It is also interesting to introduce some unknown functional form into the spatial weights as was done by Pinkse et al. (2002) who specified elements of the spatial weights matrix to be an unknown function of some distance measure between spatial units. The simultaneous consideration of nonparametric weights and nonparametric/semiparametric functional form definitely complicates the matter to a great degree. We leave these topics for future research.

Appendix

Recall $S_n = (s(z_{n,1}), \ldots, s(z_{n,n}))'$. We will denote a typical element of S_n as $s_{n,ij}$ and a typical entry of $s(z)$ by $s(z_{n,i}, z)$, i.e., $s(z_{n,i}, z) = e'[\overrightarrow{Z}(z)'\mathbf{K}_h(z)\overrightarrow{Z}(z)]^{-1}\overrightarrow{Z}_i(z)K_h(z_{n,i} - z)$. Let $K_{iz} = K_h(z_{n,i} - z)$. Let C signify a generic constant whose exact value may vary from case to case.

Recall that we used $N_l = (l + q - 1)!/(l!(q - 1)!)$ to denote the number of distinct q-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$ and arranged the N_l q-tuples as a sequence in a lexicographical order. Like Masry (1996), let ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2r$, let $\mu_{\mathbf{j}}(K) = \int_{\mathbb{R}^q} u^{\mathbf{j}} K(u)\mathrm{d}u$, $\nu_{\mathbf{j}}(K) = \int_{\mathbb{R}^q} u^{\mathbf{j}} K^2(u)\mathrm{d}u$, and define the $N \times N$ dimensional matrices Π and Γ and $N \times N_{r+1}$ matrix B, where $N = \sum_{l=0}^{r} N_l$, by

$$\Pi = \begin{bmatrix} \Pi_{0,0} & \Pi_{0,1} & \cdots & \Pi_{0,r} \\ \Pi_{1,0} & \Pi_{1,1} & \cdots & \Pi_{1,r} \\ \vdots & \vdots & \ddots & \vdots \\ \Pi_{r,0} & \Pi_{r,1} & \cdots & \Pi_{r,r} \end{bmatrix},$$
$$\Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,r} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,r} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{r,0} & \Gamma_{r,1} & \cdots & \Gamma_{r,r} \end{bmatrix}, \qquad B = \begin{bmatrix} \Pi_{0,r+1} \\ \Pi_{1,r+1} \\ \vdots \\ \Pi_{r,r+1} \end{bmatrix}, \tag{Matrix}$$

where $\Pi_{i,j}$ and $\Gamma_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (l, m) elements are, respectively, $\mu_{\phi_i(l)+\phi_j(m)}$ and $\upsilon_{\phi_i(l)+\phi_j(m)}$. Note that the elements of the matrices Π and Γ are simply multivariate moments of the kernel K and K^2, respectively; and the matrix B depends on the kernel and the order of the local polynomial we use.

Frequently we will use two evident facts (see, e.g., Kelejian and Prucha, 1999; Lee, 2002a,b):

Fact 1. If B_{1n} and B_{2n} are $n \times n$ matrices that are uniformly bounded in row sums (resp. column sums), then their product $B_{1n}B_{2n}$ is also uniformly bounded in row sums (resp. column sums).

Fact 2. If B_{1n} is uniformly bounded in row sums (resp. column sums) and B_{2n} is a conformable matrix whose elements are uniformly $O(o_n)$, then so are the elements of $B_{1n}B_{2n}$ (resp. $B_{2n}B_{1n}$).

Appendix A. Some useful lemmas

We first state two lemmas that are used in the proof of other lemmas.

Lemma A.1 (*Abel's Inequality*)**.** *Let $\{\xi_n, n \ge 1\}$ be a sequence of real numbers such that $\sum_{n=1}^{\infty} \xi_n$ converges. Let $\{a_n\}$ be a monotone decreasing sequence of positive constants. Then $a_1(\min_{1\le k\le n} \sum_{i=1}^{k} \xi_i) \le \sum_{i=1}^{n} a_i\xi_i \le a_1(\max_{1\le k\le n} \sum_{i=1}^{k} \xi_i)$.*

The above lemma restricts the sequence $\{a_n\}$ to be non-increasing. Recently, Dragomir et al. (1998) obtained the Abel-type inequality for non-decreasing sequences. Under Assumption 1(ii), we can readily apply Abel's inequality to obtain

$$\left|\sum_{i=1}^{n} a_i\eta_{is}\right| \le \max_{1\le i\le n}|a_i| \max_{1\le k\le n}\left|\sum_{i=1}^{k}\eta_{j_i s}\right| = \max_{1\le i\le n}|a_i|\, O\left(\sqrt{n}\log n\right) \tag{A.1}$$

for $s = 1, \ldots, p$ and for some permutation $\{j_1, \ldots, j_n\}$ of $\{1, 2, \ldots, n\}$.

Lemma A.2. *Let $\upsilon_1, \ldots, \upsilon_n$ be independent random variables with means zero and finite γth moments ($\gamma \ge 2$), i.e., $\sup_{1\le j\le n} E|\upsilon_j|^{\gamma} \le C < \infty$. Assume that $\{a_{ij}, i, j = 1, \ldots, n\}$ is a sequence of real numbers such that $\sup_{1\le i,j\le n}|a_{ij}| \le O(n^{-p_1})$ for some $0 < p_1 < 1$ and $\sum_{i=1}^{n}|a_{ij}| = O(n^{p_2})$ for $p_2 \ge \max(0, 2/\gamma - p_1)$. Then $\max_{1\le j\le n}|\sum_{i=1}^{n} a_{ij}\upsilon_i| = O(n^{-(p_1-p_2)/2}\log n)$ a.s.*

The above lemma was initially proved by Liang (1999) for the case where $\{a_{ij}\}$ is a sequence of positive numbers. The positivity of a_{ij}'s is not needed once we alter his condition $\sum_{i=1}^{n} a_{ij} = O(n^{p_2})$ to $\sum_{i=1}^{n}|a_{ij}| = O(n^{p_2})$. As Härdle et al. (2000, pp. 182–183), remarked, the conclusion of Lemma A.2 remains unchanged when $\{a_{ij}\}$ is a sequence of random variables satisfying $\sup_{1\le i,j\le n}|a_{ij}| \le O(n^{-p_1})$ a.s. and $\sum_{i=1}^{n}|a_{ij}| = O(n^{p_2})$ a.s. for some $0 < p_1 < 1$ and $p_2 \ge \max(0, 2/\gamma - p_1)$. It also holds for a nonrandom sequence, e.g., $\{\eta_{is}, i = 1, \ldots, n\}, s = 1, \ldots, p$, which behaves like an i.i.d. sequence. In particular, letting $\upsilon_i = \eta_{is}, a_{ij} = s_{n,ij}$ or $s_{n,ji}, \gamma = 2\delta/(\delta - q), p_1 = (\delta - q)/\delta$, $p_2 = 0$, we obtain

$$\max_{1\le j\le n}\left|\sum_{i=1}^{n} s_{n,ij}\eta_{is}\right| = O\left(n^{-(\delta-q)/(2\delta)}\log n\right) \quad \text{and}$$
$$\max_{1\le j\le n}\left|\sum_{i=1}^{n} s_{n,ji}\eta_{is}\right| = O\left(n^{-(\delta-q)/(2\delta)}\log n\right) \tag{A.2}$$

for $s = 1, \ldots, p$, where we have used Lemma A.5 and the fact that $\max_{1\le i,j\le n}|s_{n,ij}| = O(n^{-1}h^{-q}) = O(n^{-(\delta-q)/\delta})$. Similarly, letting $\upsilon_i = \eta_{is}, a_{ij} = \widetilde{m}_{ij} \equiv m_j(z_{n,i}) - \sum_{k=1}^{n} s_{n,ik}m_j(z_{n,k}), \gamma = 2, p_1 = (r+1)/\delta, p_2 = 1 - p_1$, we obtain

$$\max_{0\le j\le p}\left|\sum_{i=1}^{n}\widetilde{m}_{ij}\eta_{is}\right| = O\left(n^{1/2-(r+1)/\delta}\log n\right) \quad \text{for } s = 1, \ldots, p, \tag{A.3}$$

where we have used the fact that $\max_{1\le i\le n}|\widetilde{m}_{ij}| = O(h^{r+1}) = O(n^{-(r+1)/\delta})$ by Lemma A.4. (Notice that (A.3) can be obtained directly by using Abel's inequality and (A.1) in particular.) Similarly, letting $\upsilon_i = \eta_{is}, a_{ij} = g_{n,ij}$ or $g_{n,ji}, \gamma = 2/\vartheta, p_1 = \vartheta, p_2 = 0$, we obtain

$$\max_{1\le j\le n}\left|\sum_{i=1}^{n} g_{n,ij}\eta_{is}\right| = O\left(\min(n^{-\vartheta/2}\log n, 1)\right) \quad \text{and}$$
$$\max_{1\le j\le n}\left|\sum_{i=1}^{n} g_{n,ji}\eta_{is}\right| = O\left(\min(n^{-\vartheta/2}\log n, 1)\right) \tag{A.4}$$

for $s = 1, \ldots, p$, where we have used the fact that G_n is uniformly bounded in both row and column sums by Lemma A.5. In particular, when $\vartheta = 0$, $\max_{1\le j\le n}|\sum_{i=1}^{n} g_{n,ji}\eta_{is}| = O(1)$ by Fact 2 and Lemma A.5.

Next, we state ten lemmas, the detailed proofs of which can be found from the long version of the paper which is available upon request.

Lemma A.3. (1) *For all* z, $\sum_{j=1}^{n} s_n(z_{n,j}, z)(z_{n,j} - z)^{\mathbf{j}} = 0$ *for* $|\mathbf{j}| = 1, \ldots, r$ *and* 1 *for* $|\mathbf{j}| = 0$. *In particular,* $\sum_{j=1}^{n} s_{n,ij} = 1$ *for all* $i = 1, \ldots, n$. (2) $\sum_{i=1}^{n} s_{n,ij} = 1 + o(1)$ *for all* $j = 1, \ldots, n$.

Lemma A.4. (1) $\max_{1\le i\le n}|\widetilde{m}_{is}| = O(h^{r+1})$, *where* $\widetilde{m}_{is} = m_s(z_{n,i}) - \sum_{j=1}^{n} s_{n,ij} m_s(z_{n,j})$ *for all* $0 \le s \le p$. (2) $\max_{1\le i\le n}|\widehat{m}_s(z_{n,i}) - m_s(z_{n,i})| = O(h^{r+1} + n^{-(\delta-q)/(2\delta)}\log n)$, *where* $\widehat{m}_s(z_{n,i}) = \sum_{j=1}^{n} s_{n,ij} x_{n,js}$ *for* $1 \le s \le p$.

Lemma A.5. (1) G_n *is uniformly bounded in both row and column sums.* (2) S_n *is uniformly bounded in row sums. It is uniformly bounded in column sums for sufficiently large* n. (3) $W_n(I_n - S_n)'$ *is uniformly bounded in column sums. It is uniformly bounded in row sums for sufficiently large* n. (4) $R_n = G_n(X_n\beta_0 + \mathbf{m}_0(Z_n))$ *is uniformly bounded.* (5) $M_n = I_n - (I_n - S_n)X_n[X_n'(I_n - S_n)'(I_n - S_n)X_n]^{-1}X_n'(I_n - S_n)'$ *is uniformly bounded in both row and column sums for sufficiently large* n. (6) $M_n^{\dagger} = (I_n - S_n)'M_n(I_n - S_n)$ *is uniformly bounded in both row and column sums for sufficiently large* n.

Lemma A.6. (1) $\mathrm{tr}(P_n)/n = 1 + o(1)$, (2) $\mathrm{tr}(G_n'P_n)/n = \mathrm{tr}(G_n')/n + o(1)$, (3) $\mathrm{tr}(P_n^2)/n = 1 + o(1)$, (4) $\mathrm{tr}(G_n'P_nG_n')/n = \mathrm{tr}(G_n'G_n')/n + o(1)$, (5) $\mathrm{tr}(G_n'P_nG_n'P_n)/n = \mathrm{tr}(G_n'G_n')/n + o(1)$, (6) $\mathrm{tr}(G_n'P_nP_nG_n)/n = \mathrm{tr}(G_n'G_n)/n + o(1)$, (7) $\mathrm{tr}(G_n'M_n^{\dagger}) = \mathrm{tr}(G_n'P_n) + O(1)$, (8) $\mathrm{tr}(G_n'M_n^{\dagger}G_n) = \mathrm{tr}(G_n'P_nG_n) + O(1)$. *Furthermore, under the assumption that* $w_{n,ij} = O(1/l_n)$ *for all* i *and* j, *we have:* (9) $(l_n/n)\mathrm{tr}(G_n'M_n^{\dagger}) = O(1)$, (10) $(l_n/n)\mathrm{tr}(G_n'M_n^{\dagger}G_n) = O(1)$, (11) $\sqrt{l_n/n}\,\mathrm{tr}(C_n'M_n^{\dagger}) = o(1)$, (12) $\mathrm{tr}[C_n'M_n^{\dagger}(C_n'M_n^{\dagger} + M_n^{\dagger}C_n)] = \mathrm{tr}(C_n'P_n(P_nC_n + C_n'P_n)) + O(1)$, (13) $\sum_{i=1}^{n}[(M_n^{\dagger}C_n)_{ii}]^2 = \sum_{i=1}^{n}((P_nC_n)_{ii})^2 + O(1/l_n) = o(n/l_n)$, *where recall* $C_n = G_n - n^{-1}\mathrm{tr}(G_n)I_n$. *If* $\lim_{n\to\infty} l_n/(nh^q) = 0$, *then* $\mathrm{tr}[C_n'M_n^{\dagger}(C_n'M_n^{\dagger} + M_n^{\dagger}C_n)] = \mathrm{tr}(C_n'C_n^{s}) + o(n/l_n)$.

Lemma A.7. $n^{-1/2}A_n'P_n\mathbf{m}_0(Z_n) = o_P(1)$ *for* $A_n = X_n, \mathbf{m}_0(Z_n)$, *and* U_n.

Lemma A.8. $n^{-1/2}A_n'P_n\mathbf{m}_0(Z_n) = o_P(1)$ *for* $A_n = G_nX_n, G_n\mathbf{m}_0(Z_n)$, *and* G_nU_n.

Lemma A.9. (1) $H_{1n}(\rho) \equiv n^{-1}\mathbf{m}_0(Z_n)'M_n^{\dagger}T_n(\rho)T_n^{-1}U_n = o_P(1)$ *uniformly on* Δ; (2) $H_{2n}(\rho) \equiv n^{-1}R_n'M_n^{\dagger}T_n(\rho)T_n^{-1}U_n = o_P(1)$ *uniformly on* Δ; (3) $H_{3n}(\rho) \equiv n^{-1}U_n'(T_n^{-1})'T_n'(\rho)M_n^{\dagger}T_n(\rho)T_n^{-1}U_n = \sigma_n^2(\rho) + o_P(1)$ *uniformly on* Δ, *where* $\sigma_n^2(\rho) = n^{-1}\sigma_0^2\mathrm{tr}[(T_n^{-1})'T_n'(\rho)M_n^{\dagger}T_n(\rho)T_n^{-1}]$.

Lemma A.10. (1) $n^{-1}X_n'P_nX_n \to \Phi_{X,X}$, (2) $n^{-1}X_n'P_nR_n \to \Phi_{X,R}$, (3) $n^{-1}R_n'P_nR_n \to \Phi_{R,R}$, (4) $n^{-1}R_n'M_n^{\dagger}R_n \to \Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R}$. *Furthermore, if* $\lim_{n\to\infty} l_n = \infty$, $n^{-1}X_n'P_nR_n \to 0$, $n^{-1}R_n'P_nR_n \to 0$, *and* $n^{-1}R_n'M_n^{\dagger}R_n \to 0$.

Lemma A.11. *Under the conditions of* Theorem 5.1, (1) $T_{1n} \equiv (l_n/n)R_n'M_n^{\dagger}U_n = o_P(1)$, (2) $T_{2n} \equiv (l_n/n)R_n'M_n^{\dagger}G_nU_n = o_P(1)$, (3) $T_{3n} \equiv (l_n/n)\mathbf{m}_0(Z_n)'M_n^{\dagger}G_nU_n = o_P(1)$, (4) $T_{4n} \equiv (l_n/n)(U_n'G_n'M_n^{\dagger}U_n - \sigma_0^2\mathrm{tr}(G_n'M_n^{\dagger})) = o_P(1)$, (5) $T_{5n} \equiv (l_n/n)(U_n'G_n'M_n^{\dagger}G_nU_n - \sigma_0^2\mathrm{tr}(G_n'M_n^{\dagger}G_n)) = o_P(1)$.

Lemma A.12. *Under the conditions of* Theorem 5.1, (1) $(l_n/n)\mathbf{m}_0(Z_n)'M_n^{\dagger}\mathbf{m}_0(Z_n) = o(1)$, (2) $(l_n/n)\mathbf{m}_0(Z_n)'M_n^{\dagger}G_nG_n'M_n^{\dagger}\mathbf{m}_0(Z_n) = o(1)$, (3) $\sqrt{l_n/n}R_n'M_n^{\dagger}\mathbf{m}_0(Z_n) = o(1)$, (4) $\widehat{\sigma}^2(\rho) = \sigma_0^2 + o_P(1)$ *uniformly on* Δ.

Appendix B. Proof of the main results

Proof of Theorem 4.1. We adopt the idea of Lee (2004, 2002b) to prove the theorem. The major difference lies in the appearance of nonparametric objects in our setting. By White (1994, Theorem 3.4), it suffices to show

$$n^{-1}\left[\log L_n^c(\rho) - Q_n(\rho)\right] \xrightarrow{p} 0 \quad \text{uniformly on } \Delta, \text{ and} \tag{B.1}$$

$$\limsup_{n\to\infty}\ \max_{\rho\in N_\epsilon^c(\rho_0)} n^{-1}\left[Q_n(\rho) - Q_n(\rho_0)\right] < 0 \quad \text{for any } \epsilon > 0, \tag{B.2}$$

where $N_\epsilon^c(\rho_0)$ is the complement of an open neighborhood of ρ_0 on Δ of diameter ϵ.

By (2.11) and (4.4), $n^{-1}[\log L_n^c(\rho) - Q_n(\rho)] = -(1/2)\{\log\widehat{\sigma}^2(\rho) - \log\sigma^{*2}(\rho)\}$. To show (B.1), it is sufficient to show

$$\widehat{\sigma}^2(\rho) - \sigma^{*2}(\rho) = o_P(1) \quad \text{uniformly on } \Delta. \tag{B.3}$$

By (2.10) and (4.3), we have

$$\widehat{\sigma}^2(\rho) - \sigma^{*2}(\rho) = 2H_{1n}(\rho) + 2(\rho_0 - \rho)H_{2n}(\rho) + H_{3n}(\rho) - \sigma_n^2(\rho), \tag{B.4}$$

where $H_{in}(\rho)$ and $\sigma_n^2(\rho)$ are defined in Lemma A.9. (B.3) follows from (B.4) and Lemma A.9.

To show (B.2), write

$$\frac{1}{n}\left[Q_n(\rho) - Q_n(\rho_0)\right] = \frac{1}{n}\left[Q_{p,n}(\rho) - Q_{p,n}(\rho_0)\right] + \frac{1}{2}H_{4n}(\rho) + \frac{1}{2}H_{5n}, \tag{B.5}$$

where $\sigma_{p,n}^2(\rho) = \sigma_0^2 n^{-1}\mathrm{tr}\{T_n'^{-1}T_n'(\rho)T_n(\rho)T_n^{-1}\}$, $Q_{p,n}(\rho) = -\frac{n}{2}(\log(2\pi)+1) - \frac{n}{2}\log\sigma_{p,n}^2(\rho) + |T_n(\rho)|$, $H_{4n}(\rho) = \log\sigma_{p,n}^2(\rho) - \log\sigma^{*2}(\rho)$, and $H_{5n} = \log\sigma^{*2}(\rho_0) - \log\sigma_{p,n}^2(\rho_0)$. To show $n^{-1}[Q_{p,n}(\rho) - Q_{p,n}(\rho_0)] \le 0$ uniformly on Δ, we follow Lee (2002b) and define an auxiliary SAR process: $Y_n = \rho_0 W_nY_n + U_n$, where $U_n \sim N(0, \sigma_0^2 I_n)$. Denote the log likelihood of this process as $\log L_{a,n}(\rho, \sigma^2)$. One can verify that $Q_{p,n}(\rho) = \max_{\sigma^2} E_a[\log L_{a,n}(\rho, \sigma^2)]$, where E_a denotes expectation under the auxiliary SAR process. Consequently, for any $\rho \in \Delta$, $Q_{p,n}(\rho) \le \max_{\rho,\sigma^2} E_a[\log L_{a,n}(\rho, \sigma^2)] = E_a[\log L_{a,n}(\rho_0, \sigma_0^2)] = Q_{p,n}(\rho_0)$. Hence $n^{-1}[Q_{p,n}(\rho) - Q_{p,n}(\rho_0)] \le 0$ uniformly on Δ. To show that $H_{5n} = o(1)$, recall $M_n^{\dagger} = (I_n - S_n)'M_n(I_n - S_n)$. Then $\mathrm{tr}(M_n^{\dagger}) = \mathrm{tr}(M_n) + \mathrm{tr}(S_n'S_n) - 2\mathrm{tr}(M_nS_n)$. It is easy to show $\mathrm{tr}(M_n) = n - p$. By Lemma A.5, both M_n and S_n are uniformly bounded in both row and column sums for sufficiently large n. In addition, the elements of S_n are $O(n^{-1}h^{-q})$ uniformly. With these, it is trivial to show that $\mathrm{tr}(S_n'S_n) = O(h^{-q})$ and $\mathrm{tr}(M_nS_n) = O(h^{-q})$. Consequently, $\mathrm{tr}(M_n^{\dagger}) = n - p + O(h^{-q})$ and $\sigma_{p,n}^2(\rho_0) - \sigma^{*2}(\rho_0) = \sigma_0^2 - \sigma_0^2 n^{-1}\mathrm{tr}\{M_n^{\dagger}\} - n^{-1}\mathbf{m}_0(Z_n)'M_n^{\dagger}\mathbf{m}_0(Z_n) = O(n^{-1}h^{-q} + h^{2(r+1)}) = o(1)$, implying that $H_{5n} = o(1)$.

Next, write $\sigma_{p,n}^2(\rho) - \sigma^{*2}(\rho) = (\sigma_0^2/n)\{\mathrm{tr}[T_n'^{-1}T_n'(\rho)T_n(\rho)T_n^{-1}] - \mathrm{tr}[T_n'^{-1}T_n'(\rho)M_n^{\dagger}T_n(\rho)T_n^{-1}]\} - n^{-1}\{\mathbf{m}_0(Z_n) + (\rho_0 - \rho)R_n\}'M_n^{\dagger}\{\mathbf{m}_0(Z_n) + (\rho_0 - \rho)R_n\}$. One can show the first term of the above expression is $o(1)$ uniformly while the second term is nonnegative. Consequently,

$$\limsup_{n\to\infty}\ \max_{\rho\in N_\epsilon^c(\rho_0)} n^{-1}\left[Q_n(\rho) - Q_n(\rho_0)\right] \le 0 \quad \text{for any } \epsilon > 0. \tag{B.6}$$

To show that the above inequality holds strictly, by the compactness of $N_\epsilon^c(\rho_0)$ we assume that there exists a sequence ρ_n converging to a point $\rho^* \neq \rho_0$ such that $\lim_{n\to\infty} n^{-1}[Q_n(\rho_n) - Q_n(\rho_0)] = 0$. This would be possible only if (a) $\lim_{n\to\infty}[\sigma_{p,n}^2(\rho^*) - \sigma^{*2}(\rho^*)] = 0$, and (b) $\lim_{n\to\infty} n^{-1}[Q_{p,n}(\rho^*) - Q_{p,n}(\rho_0)] = 0$. (a) generates a contradiction to $\Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R} > 0$ by Assumption 7, Lemmas A.10 and A.12 ($n^{-1}\mathbf{m}_0(Z_n)'M_n^\dagger \mathbf{m}_0(Z_n) = o(1)$ in particular). If $\Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R} = 0$ by Assumption 7*, the contradiction follows from (b) under Assumption 8.

The consistency of $\widehat{\rho}$ and hence $\widehat{\theta}$ then follows by Theorem 3.4 of White (1994). ■

Proof of Lemma 4.2. The proof of the lemma is analogous to that of Theorem 3.2 of Lee (2002b). Let $\alpha = (\alpha_1', \alpha_2, \alpha_3)'$ be a column vector of constants such that $\Sigma_\theta \alpha = 0$. It suffices to show $\alpha = 0$. It follows from the first block of the linear system $\Sigma_\theta\alpha = 0$ that $\alpha_1 = -\Phi_{X,X}^{-1}\Phi_{X,R}\alpha_2$. The last equation of the linear system gives $\alpha_3 = -2\sigma_0^2 c_1\alpha_2$. Plugging α_1 and α_3 into the second equation of the linear system, we have

$$\left[\Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R} + \sigma_0^2\left(c_2 - 2c_1^2\right)\right]\alpha_2 = 0. \tag{B.7}$$

We can verify that $c_2 - 2c_1^2 = \lim_{n\to\infty} n^{-1}\mathrm{tr}((G_n + G_n')G_n) - 2\lim_{n\to\infty} n^{-2}[\mathrm{tr}(G_n)]^2 = \lim_{n\to\infty} n^{-1}\mathrm{tr}(C_n^s C_n^{s\prime}) \geq 0$, where recall $C_n^s = C_n + C_n'$ with $C_n = G_n - (n^{-1}\mathrm{tr}(G_n))I_n$. $\Phi_{R,R} - \Phi_{X,R}'\Phi_{X,X}^{-1}\Phi_{X,R} > 0$ when Assumption 7 holds. If instead Assumption 7* holds, then $c_2 - 2c_1^2 \neq 0$ is implied by Assumption 9. In either case, we have $\alpha_2 = 0$ and so $\alpha = 0$. ■

Proof of Theorem 4.3. By a Taylor expansion of the first order condition from maximizing (2.8), we have

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = -\left(\frac{1}{n}\frac{\partial^2 \log L_n(\widetilde{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\theta_0)}{\partial\theta}, \tag{B.8}$$

where $\widetilde{\theta}_n$ lies between $\widehat{\theta}$ and θ_0 and thus converges to θ_0 in probability by Theorem 4.1. The proof is complete if we can show

$$\frac{1}{n}\frac{\partial^2 \log L_n(\widetilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta\partial\theta'} = o_P(1) \quad \text{uniformly in } \widetilde{\theta}_n, \tag{B.9}$$

$$\frac{1}{n}\frac{\partial^2 \log L_n(\theta_0)}{\partial\theta\partial\theta'} - \Sigma_\theta = o_P(1), \quad \text{and} \tag{B.10}$$

$$\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \Sigma_\theta + \Omega_\theta). \tag{B.11}$$

To show (B.9), we show that each element of $n^{-1}\partial^2 \log L_n(\widetilde{\theta}_n)/(\partial\theta\partial\theta')$ converges to $n^{-1}\partial^2 \log L_n(\theta_0)/(\partial\theta\partial\theta')$ uniformly in probability, where

$$\frac{\partial^2 \log L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}X_n'P_nX_n,$$

$$\frac{\partial^2 \log L_n(\theta)}{\partial\sigma^2\partial\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\left(T_n(\rho)Y_n - X_n\beta\right)' P_n\left(T_n(\rho)Y_n - X_n\beta\right),$$

$$\frac{\partial^2 \log L_n(\theta)}{\partial\rho^2} = -\mathrm{tr}\left\{G_n^2(\rho)\right\} - \frac{1}{\sigma^2}Y_n'W_n'P_nW_nY_n,$$

$$\frac{\partial^2 \log L_n(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}X_n'P_n\left(T_n(\rho)Y_n - X_n\beta\right),$$

$$\frac{\partial^2 \log L_n(\theta)}{\partial\rho\partial\sigma^2} = -\frac{1}{\sigma^4}Y_n'W_nP_n\left(T_n(\rho)Y_n - X_n\beta\right),$$

$$\frac{\partial^2 \log L_n(\theta)}{\partial\beta\partial\rho} = -\frac{1}{\sigma^2}X_n'P_nW_nY_n.$$

Noting that β and $1/\sigma^2$ appear only in linear, quadratic or cubic form in $n^{-1}\partial^2\log L_n(\theta)/(\partial\theta\partial\theta')$, it is easy to show that (B.9) holds for all elements but the second derivative of $\log L_n(\theta)$ with respect to ρ. The latter is $\partial^2\log L_n(\theta)/\partial\rho^2 = -\mathrm{tr}\{G_n^2(\rho)\} - \frac{1}{\sigma^2}Y_n'W_n'(I_n - S_n)'(I_n - S_n)W_nY_n$, where $G_n(\rho) = W_nT_n^{-1}(\rho)$. By Assumptions 3(iv) and 4 and the mean value theorem, $\mathrm{tr}\{G_n^2(\widetilde{\rho}_n)\} = \mathrm{tr}(G_n^2) + 2\mathrm{tr}\{G_n^3(\widetilde{\rho}_n^*)\}(\widetilde{\rho}_n - \rho_0)$ for some $\widetilde{\rho}_n^*$ between $\widetilde{\rho}_n$ and ρ_0. Consequently,

$$\begin{aligned}&\frac{1}{n}\left(\frac{\partial^2\log L_n\left(\widetilde{\theta}_n\right)}{\partial\rho^2} - \frac{\partial^2\log L_n(\theta_0)}{\partial\rho^2}\right)\\ &\quad = -2\frac{\mathrm{tr}\left\{G_n^3\left(\widetilde{\rho}_n^*\right)\right\}}{n}\left(\widetilde{\rho}_n - \rho_0\right) + \left(\frac{1}{\sigma_0^2} - \frac{1}{\widetilde{\sigma}_n^2}\right)\frac{Y_n'W_n'P_nW_nY_n}{n}.\end{aligned} \tag{B.12}$$

Since $G_n(\rho)$ is uniformly bounded in row and column sums uniformly in a neighborhood of ρ_0 by the remark after Assumption 4, $\mathrm{tr}\{G_n^3(\widetilde{\rho}_n^*)\}/n = O(1/l_n)$, implying that the first term in (B.12) is $o_p(1)$. The second term in (B.12) is also $o_p(1)$ because we can show $Y_n'W_n'(I_n - S_n)'(I_n - S_n)W_nY_n/n = O_p(1/l_n)$. Consequently, $n^{-1}[\partial^2\log L_n(\widetilde{\theta}_n)/\partial\rho^2 - \partial^2\log L_n(\theta_0)/\partial\rho^2] = o_P(1)$.

The proof of (B.10) is straightforward by showing that linear or quadratic functions of U_n deviated from their means are all $o_P(1)$. The components of $n^{-1/2}\partial\log L_n(\theta_0)/\partial\theta$ are linear or quadratic functions of U_n. By Assumption 1, we can apply the central limit theorem for linear quadratic forms of Kelejian and Prucha (2001, Theorem 1) to obtain (B.11). ■

Proof of Theorem 4.4. Denote $\mathbf{S}_n(z_{n,i}, z)$ as a typical column of $\mathbf{S}_n(z)$, i.e., $\mathbf{S}_n(z) = (\mathbf{S}_n(z_{n,1}, z), \ldots, \mathbf{S}_n(z_{n,n}, z))$. By (2.12),

$$\begin{aligned}\widehat{\alpha}(z) &= \mathbf{S}_n(z)\left(\mathbf{m}_0(Z_n) + U_n\right) + \mathbf{S}_n(z)W_nY_n\left(\rho_0 - \widehat{\rho}\right)\\ &\quad + \mathbf{S}_n(z)X_n\left(\beta_0 - \widehat{\beta}\right).\end{aligned} \tag{B.13}$$

By Theorem 4.3 (see also the proof of Theorem 5.4), it is easy to show $\sqrt{nh^q}\mathbf{S}_n(z)W_nY_n(\rho_0 - \widehat{\rho}) = o_P(1)$ and $\sqrt{nh^q}\mathbf{S}_n(z)X_n(\beta_0 - \widehat{\beta}) = o_P(1)$. By the $(r+1)$th order Taylor expression,

$$\begin{aligned}m_0\left(z_{n,i}\right) &= \overrightarrow{Z}_i(z)'\alpha^0(z) + \sum_{|\mathbf{j}|=r+1}\left(D^{\mathbf{j}}m_0\right)(z)\left(z_{n,i} - z\right)^{\mathbf{j}}\\ &\quad + o\left(h^{r+1}\right)\end{aligned} \tag{B.14}$$

for $\|z_{n,i} - z\| \leq Ch$. Noting that $[\overrightarrow{Z}(z)'\mathbf{K}_h(z)\overrightarrow{Z}(z)]^{-1}\sum_{i=1}^n K_{iz}\overrightarrow{Z}_i(z)\overrightarrow{Z}_i(z)' = I_N$ and $\mathbf{S}_n(z_{n,i}, z) = [\overrightarrow{Z}(z)'\mathbf{K}_h(z)\overrightarrow{Z}(z)]^{-1}\overrightarrow{Z}_i(z)K_{iz}$, we have $\sqrt{nh^q}(\widehat{\alpha}(z) - \alpha^0(z)) = \sqrt{nh^q}\sum_{i=1}^n \mathbf{S}_n(z_{n,i}, z)\sum_{|\mathbf{j}|=r+1}(D^{\mathbf{j}}m_0)(z)(z_{n,i} - z)^{\mathbf{j}} + \sqrt{nh^q}\mathbf{S}_n(z)U_n + o_p(1) \equiv B_{11} + B_{12} + o_p(1)$. By Assumption 1, $n^{-1}\overrightarrow{Z}(z)'\mathbf{K}_h(z)\overrightarrow{Z}(z) = \Pi f(z) + o(1)$, and $B_{11} = \sqrt{nh^q}h^{r+1}\Pi^{-1}Bm_0^{(r+1)}(z) + o(1)$. Next it is standard to show that $B_{12} = \sqrt{nh^q}f^{-1}(z)\Pi^{-1}n^{-1}\sum_{i=1}^n \overrightarrow{Z}_i(z)K_h(z_{n,i} - z)u_i + o_p(1) \xrightarrow{d} N(0, \Pi^{-1}\Gamma\Pi^{-1}/f(z))$. ■

Proof of Theorem 5.1. Like Lee (2004), we prove the theorem in three steps: (1) We show that

$$\begin{aligned}&(l_n/n)\left[\left(\log L_n^c(\rho) - \log L_n^c(\rho_0)\right)\right.\\ &\quad \left. - \left(\log Q_n(\rho) - \log Q_n(\rho_0)\right)\right] \xrightarrow{p} 0 \quad \text{uniformly on } \Delta.\end{aligned} \tag{B.15}$$

(2) We show that $(l_n/n)(\log Q_n(\rho) - \log Q_n(\rho_0))$ is uniformly equicontinuous on Δ; (3) We show that ρ_0 is uniquely identifiable. To show (B.15), we apply the mean value theorem to obtain

$$\begin{aligned}&\frac{l_n}{n}\left[\left(\log L_n^c(\rho) - \log L_n^c(\rho_0)\right) - \left(\log Q_n(\rho) - \log Q_n(\rho_0)\right)\right]\\ &\quad = \frac{l_n}{n\widehat{\sigma}_n^2(\overline{\rho}_n)}\left\{B_n\left(\overline{\rho}_n\right) - \frac{\widehat{\sigma}_n^2(\overline{\rho}_n) - \sigma^{*2}(\overline{\rho}_n)}{\sigma^{*2}(\overline{\rho}_n)}A_n\left(\overline{\rho}_n\right)\right\}(\rho - \rho_0)\end{aligned}$$

where $\overline{\rho}_n$ lies between ρ and $\rho_0, A_n(\rho) = (\rho_0 - \rho)R_n' M_n^{\dagger} R_n + \mathbf{m}_0(Z_n)' M_n^{\dagger} R_n + \sigma_0^2 \mathrm{tr}(G_n' M_n^{\dagger} T_n(\rho) T_n^{-1})$, and $B_n(\rho) = Y_n' W_n' M_n^{\dagger} T_n(\rho) Y_n - A_n(\rho)$. Simple algebra shows that

$$\begin{aligned}(l_n/n)\, B_n(\rho) &= (l_n/n)\left(R_n' M_n^{\dagger} + 2(\rho_0-\rho) R_n' M_n^{\dagger} G_n\right.\\ &\quad \left. + \mathbf{m}_0(Z_n)' M_n^{\dagger} G_n\right) U_n + (l_n/n)\left(U_n' M_n^{\dagger} G_n U_n - \sigma_0^2 \mathrm{tr}\left(G_n' M_n^{\dagger}\right)\right)\\ &\quad + (l_n/n)(\rho_0-\rho)\left(U_n' G_n' M_n^{\dagger} G_n U_n - \sigma_0^2 \mathrm{tr}\left(G_n' M_n^{\dagger} G_n\right)\right)\\ &= o_p(1) \quad \text{uniformly on } \Delta \text{by Lemma A.11}.\end{aligned}$$

Similarly, $(l_n/n)A_n(\rho) = O(1)$ uniformly on Δ. By (B.4) and Lemma A.9, $\widehat{\sigma}_n^2(\rho) - \sigma^{*2}(\rho) = o_p(1)$ uniformly on Δ. Also, $\widehat{\sigma}_n^2(\overline{\rho}_n)$ and $\sigma^{*2}(\overline{\rho}_n)$ are bounded away from zero. Consequently, (B.15) follows.

Next, $(l_n/n)(\log Q_n(\rho) - \log Q_n(\rho_0)) = -\frac{l_n}{2}[\log \sigma^{*2}(\rho) - \log \sigma^{*2}(\rho_0)] + \frac{l_n}{n}[\log|T_n(\rho)| - \log|T_n(\rho_0)|]$. Since $T_n(\rho) = T_n + (\rho_0 - \rho)W_n$, it follows from the mean value theorem that $(l_n/n)[\log|T_n(\rho_2)| - \log|T_n(\rho_1)|] = (l_n/n)\mathrm{tr}(W_n T_n^{-1}(\overline{\rho}_{n12}))(\rho_2 - \rho_1)$, where $\overline{\rho}_{n12}$ is the mean value. Noting that $(l_n/n)\mathrm{tr}(W_n T_n^{-1}(\overline{\rho}_{n12})) = O(1)$, $(l_n/n)[\log|T_n(\rho_2)| - \log|T_n(\rho_1)|]$ is uniformly equicontinuous. Similarly, $l_n[\log \sigma^{*2}(\rho) - \log \sigma^{*2}(\rho_0)] = l_n[\sigma^{*2}(\rho) - \sigma^{*2}(\rho_0)]/\overline{\sigma}_n^{*2}(\rho)$ for some $\overline{\sigma}_n^{*2}(\rho)$ that lies between $\sigma^{*2}(\rho)$ and $\sigma^{*2}(\rho_0)$ and is uniformly bounded away from zero. In addition, simple algebra shows that $l_n[\sigma^{*2}(\rho) - \sigma^{*2}(\rho_0)] = (l_n/n)[(\rho_0 - \rho)^2(R_n' M_n^{\dagger} R_n + \sigma_0^2 \mathrm{tr}(G_n' M_n^{\dagger} G_n)) + 2(\rho_0 - \rho)(R_n' M_n^{\dagger} \mathbf{m}_0(Z_n) + \sigma_0^2 \mathrm{tr}(G_n' M_n^{\dagger}))]$, which is uniformly equicontinuous because $(l_n/n)(R_n' M_n^{\dagger} R_n + \sigma_0^2 \mathrm{tr}(G_n' M_n^{\dagger} G_n))$ and $(l_n/n)(R_n' M_n^{\dagger} \mathbf{m}_0(Z_n) + \sigma_0^2 \mathrm{tr}(G_n' M_n^{\dagger}))$ are both $O(1)$ by (5.2), Lemmas A.6 and A.12. Consequently, $\log Q_n(\rho) - \log Q_n(\rho_0)$ is uniformly equicontinuous on Δ.

To show that ρ_0 is uniquely identifiable, define $\overline{Q}_n(\rho) = -(l_n/2)[\log \sigma_{p,n}^2(\rho) - \log \sigma_{p,n}^2(\rho_0)] + (l_n/n)[\log|T_n(\rho)| - \log|T_n(\rho_0)|]$, where $\sigma_{p,n}^2(\rho)$ is defined in (3.5). Then $(l_n/n)(\log Q_n(\rho) - \log Q_n(\rho_0)) = \overline{Q}_n(\rho) - (l_n/2)[(\log \sigma^{*2}(\rho) - \log \sigma^{*2}(\rho_0)) - (\log \sigma_{p,n}^2(\rho) - \log \sigma_{p,n}^2(\rho_0))] \equiv \overline{Q}_n(\rho) - \frac{1}{2}\Delta_n(\rho)$. By the second order Taylor expansion,

$$\begin{aligned}\Delta_n(\rho) &= \frac{l_n}{\sigma^{*2}(\rho_0)}\left[\frac{\partial \sigma^{*2}(\rho_0)}{\partial \rho} - \frac{\partial \sigma_{p,n}^2(\rho_0)}{\partial \rho}\right](\rho - \rho_0)\\ &\quad + \frac{l_n}{\sigma^{*2}(\rho_0)\sigma_{p,n}^2(\rho_0)}\left[\sigma^{*2}(\rho_0) - \sigma_{p,n}^2(\rho_0)\right]\frac{\partial \sigma_{p,n}^2(\rho_0)}{\partial \rho}(\rho - \rho_0)\\ &\quad - \frac{l_n}{2\sigma^{*4}(\overline{\rho}_n)}\left\{\left(\frac{\partial \sigma^{*2}(\overline{\rho}_n)}{\partial \rho}\right)^2 - \left(\frac{\partial \sigma_{p,n}^2(\overline{\rho}_n)}{\partial \rho}\right)^2\right\}(\rho - \rho_0)^2\\ &\quad - \frac{l_n}{2\sigma^{*4}(\overline{\rho}_n)\sigma_{p,n}^4(\overline{\rho}_n)}\left\{\sigma^{*4}(\overline{\rho}_n) - \sigma_{p,n}^4(\overline{\rho}_n)\right\}\\ &\quad \times \left(\frac{\partial \sigma_{p,n}^2(\overline{\rho}_n)}{\partial \rho}\right)^2(\rho - \rho_0)^2\\ &\quad + \frac{l_n}{2\sigma^{*2}(\overline{\rho}_n)}\left\{\frac{\partial^2 \sigma^{*2}(\overline{\rho}_n)}{\partial \rho^2} - \frac{\partial^2 \sigma_{p,n}^2(\overline{\rho}_n)}{\partial \rho^2}\right\}(\rho - \rho_0)^2\\ &\quad + \frac{l_n}{2\sigma^{*2}(\overline{\rho}_n)\sigma_{p,n}^2(\overline{\rho}_n)}\left\{\sigma^{*2}(\overline{\rho}_n) - \sigma_{p,n}^2(\overline{\rho}_n)\right\}\\ &\quad \times \frac{\partial^2 \sigma_{p,n}^2(\overline{\rho}_n)}{\partial \rho^2}(\rho - \rho_0)^2\\ &\equiv \sum_{s=1}^{6} \Delta_{ns}(\rho).\end{aligned} \tag{B.16}$$

Analogous to the proof of (B.15), we can show that $\Delta_{ns}(\rho) = 0$ for $s = 2, 4, 6$. We now evaluate the other terms in (B.16). By Lemma A.12,

$$\begin{aligned}\Delta_{n1}(\rho) &= \frac{2l_n}{n\sigma^{*2}(\rho_0)} R_n' M_n^{\dagger} R_n (\rho - \rho_0)^2\\ &\quad - \frac{2l_n}{n\sigma^{*2}(\rho_0)} R_n' M_n^{\dagger} \mathbf{m}_0(Z_n)(\rho - \rho_0)\\ &= \frac{2l_n}{n\sigma^{*2}(\rho_0)} R_n' M_n^{\dagger} R_n (\rho - \rho_0)^2 + o(1) \quad \text{uniformly on } \Delta.\end{aligned}$$

Let $c_{n1}(\rho) \equiv (\sqrt{l_n}/n)(R_n' M_n^{\dagger} R_n(\rho - \rho_0) - \mathbf{m}_0(Z_n)' M_n^{\dagger} R_n)$ and $c_{n2}(\rho) \equiv (\sqrt{l_n}/n)\mathrm{tr}(G_n' M_n^{\dagger}(I_n + (\rho_0 - \rho)G_n))$. Then $c_{n1}(\rho) = o(1)$ uniformly on Δ by (5.2) and Lemma A.12, and $c_{n2}(\rho) = o(1)$ uniformly on Δ by Facts 1–2 and Lemma A.5. So $\Delta_{n3}(\rho) = -(2/\sigma^{*4}(\overline{\rho}_n))c_{n1}(\overline{\rho}_n)(c_{n1}(\overline{\rho}_n) + 2c_{n2}(\overline{\rho}_n))(\rho - \rho_0)^2 = o(1)$ uniformly on Δ. $\Delta_{n5}(\rho) = (l_n/n\sigma^{*2}(\overline{\rho}_n))R_n' M_n^{\dagger} R_n(\rho - \rho_0)^2$. Consequently,

$$\Delta_n(\rho) = \frac{3l_n}{n\sigma^{*2}(\rho_0)} R_n' M_n^{\dagger} R_n (\rho - \rho_0)^2 + o(1) \quad \text{uniformly on } \Delta,$$

and it is nonnegative for sufficiently large n when $\rho \neq \rho_0$ and $(l_n/n)R_n' M_n^{\dagger} R_n \to c \in (0, \infty)$. Whenever $\rho \neq \rho_0, \overline{Q}_n(\rho) < 0$ under Assumption 8(ii). Consequently $(l_n/n)(\log Q_n(\rho) - \log Q_n(\rho_0)) < 0$ whenever $\rho \neq \rho_0$, implying that ρ_0 is uniquely identifiable. This completes the proof. ■

Proof of Theorem 5.2. We prove the theorem in three steps: (1) Show that $(l_n/n)[\partial^2 \log L_n^c(\widetilde{\rho}_n)/\partial \rho^2 - \partial^2 \log L_n^c(\rho_0)/\partial \rho^2] = o_P(1)$ for all $\widetilde{\rho}_n = \rho_0 + o_P(1)$; (2) Show that $(l_n/n)[\partial^2 \log L_n^c(\rho_0)/\partial \rho^2 - E(\partial^2 \log L_n^c(\rho_0)/\partial \rho^2)]$ converges to zero in probability; (3) Apply the CLT of Lee (2004, Appendix A) to $\sqrt{l_n/n}\partial \log L_n^c(\rho_0)/\partial \rho$. To show (1), we obtain from (2.10) and (2.11) that

$$\frac{\partial \log L_n^c(\rho)}{\partial \rho} = \frac{1}{\widehat{\sigma}^2(\rho)} Y_n W_n' M_n^{\dagger} T_n(\rho) Y_n - \mathrm{tr}\left(W_n T_n^{-1}(\rho)\right), \quad \text{and}$$

$$\begin{aligned}\frac{\partial^2 \log L_n^c(\rho)}{\partial \rho^2} &= \frac{2}{n\widehat{\sigma}^4(\rho)}\left(Y_n W_n' M_n^{\dagger} T_n(\rho) Y_n\right)^2\\ &\quad - \frac{1}{\widehat{\sigma}^2(\rho)} Y_n W_n' M_n^{\dagger} W_n Y_n - \mathrm{tr}\left(\left[W_n T_n^{-1}(\rho)\right]^2\right).\end{aligned}$$

Note that $(l_n/n)Y_n W_n' M_n^{\dagger} W_n Y_n = (l_n/n)R_n' M_n^{\dagger} R_n + (l_n/n)U_n' G_n' M_n^{\dagger} G_n U_n + (2l_n/n)R_n' M_n^{\dagger} G_n U_n = (l_n/n)R_n' M_n^{\dagger} R_n + (l_n/n)U_n' G_n' M_n^{\dagger} G_n U_n + o_P(1)$, $(l_n/n)Y_n W_n' M_n^{\dagger} T_n(\rho) Y_n = (l_n/n)U_n' G_n' M_n^{\dagger} U_n + (\rho_0 - \rho)(l_n/n)\{R_n' M_n^{\dagger} R_n + U_n' G_n' M_n^{\dagger} G_n U_n\} + o_P(1)$, and $\widehat{\sigma}^2(\rho) = \sigma_0^2 + o_P(1)$ uniformly on Δ by Lemmas A.11 and A.12. When $\lim_{n\to\infty} l_n = \infty$, $n^{-1} Y_n W_n' M_n^{\dagger} T_n(\rho) Y_n = o_P(1)$, implying that $(l_n/n^2)(Y_n W_n' M_n^{\dagger} T_n(\rho) Y_n)^2 = o_P(1)$. So

$$\begin{aligned}\frac{l_n}{n}\frac{\partial^2 \log L_n^c(\rho)}{\partial \rho^2} &= -\frac{l_n}{n}\frac{1}{\widehat{\sigma}^2(\rho)} Y_n W_n' M_n^{\dagger} W_n Y_n\\ &\quad - \frac{l_n}{n}\mathrm{tr}\left(\left[W_n T_n^{-1}(\rho)\right]^2\right) + o_P(1)\\ &= -\frac{1}{\sigma_0^2}\frac{l_n}{n}\left\{R_n' M_n^{\dagger} R_n + U_n' G_n' M_n^{\dagger} G_n U_n\right\}\\ &\quad - \frac{l_n}{n}\mathrm{tr}\left(\left[W_n T_n^{-1}(\rho)\right]^2\right) + o_P(1).\end{aligned}$$

Noting that $(l_n/n)\mathrm{tr}(G_n^3(\rho)) = O(1)$ uniformly on Δ, we have by the mean value theorem

$$\begin{aligned}&\frac{l_n}{n}\left(\frac{\partial^2 \log L_n^c(\widetilde{\rho}_n)}{\partial \rho^2} - \frac{\partial^2 \log L_n^c(\rho_0)}{\partial \rho^2}\right)\\ &= -\frac{l_n}{n}\left\{\mathrm{tr}\left(\left[W_n' T_n^{-1}(\widetilde{\rho}_n)\right]^2\right) - \mathrm{tr}\left(\left[W_n T_n^{-1}(\rho_0)\right]^2\right)\right\} + o_P(1)\\ &= -\frac{2l_n}{n}\mathrm{tr}\left(G_n^3(\overline{\rho}_n)\right)(\widetilde{\rho}_n - \rho_0) + o_P(1) = o_P(1)\end{aligned} \tag{B.17}$$

for any $\widetilde{\rho}_n$ which converges in probability to ρ_0.

Next, we show (2). By the above arguments and Lemma A.6,

$$E\left(\frac{l_n}{n}\frac{\partial^2 \log L_n^c(\rho_0)}{\partial\rho^2}\right) = -\left\{\frac{l_n}{n\sigma_0^2}R_n'M_n^\dagger R_n + \frac{l_n}{n}\left[\operatorname{tr}\left(G_n'M_n^\dagger G_n\right) + \operatorname{tr}\left(G_n^2\right)\right]\right\} + o(1)$$
$$= -\left\{\frac{l_n}{n\sigma_0^2}R_n'M_n^\dagger R_n + \frac{l_n}{n}\operatorname{tr}\left(G_n'\left(P_nG_n + G_n'\right)\right)\right\} + o(1)$$
$$= O(1). \tag{B.18}$$

By the Chebyshev inequality, we can show that

$$\frac{l_n}{n}\left[\frac{\partial^2 \log L_n^c(\rho_0)}{\partial\rho^2} - E\left(\frac{\partial^2 \log L_n^c(\rho_0)}{\partial\rho^2}\right)\right] = \frac{l_n}{n\sigma_0^2}\left[U_n'G_n'M_n^\dagger G_nU_n - E\left(U_n'G_n'M_n^\dagger G_nU_n\right)\right] + o_P(1)$$
$$= o_P(1). \tag{B.19}$$

(2) follows.

We now show (3). By the proof of Lemma A.12, $\sqrt{l_n/n}\mathbf{m}_0(Z_n)'M_n^\dagger \mathbf{m}_0(Z_n)\,(\operatorname{tr}(G_n)/n) = \sqrt{l_n/n}O(nh^{2(r+1)})O(1/l_n) = O(\sqrt{n/l_n}h^{2(r+1)}) = o(1)$, and $\sqrt{l_n/n}R_n'M_n^\dagger \mathbf{m}_0(Z_n) = o(1)$. It follows that

$$\sqrt{\frac{l_n}{n}}\frac{\partial \log L_n^c(\rho_0)}{\partial\rho} = \frac{1}{\widehat{\sigma}^2(\rho_0)}\sqrt{\frac{l_n}{n}}\left\{Y_nW_n'M_n^\dagger T_nY_n - \widehat{\sigma}^2(\rho_0)\operatorname{tr}(G_n)\right\}$$
$$= \frac{1}{\widehat{\sigma}^2(\rho_0)}\sqrt{\frac{l_n}{n}}\left\{a_{1n}'U_n + U_n'a_{2n}U_n + R_n'M_n^\dagger \mathbf{m}_0(Z_n) - \mathbf{m}_0(Z_n)'M_n^\dagger \mathbf{m}_0(Z_n)\left(\operatorname{tr}(G_n)/n\right)\right\}$$
$$= \frac{1}{\sigma_0^2}\sqrt{\frac{l_n}{n}}\left\{a_{1n}'U_n + U_n'a_{2n}U_n\right\} + o_P(1), \tag{B.20}$$

where $a_{1n}' = R_n'M_n^\dagger + \mathbf{m}_0(Z_n)'M_n^\dagger G_n - 2m_0(Z_n)'M_n^\dagger(\operatorname{tr}(G_n)/n)$ and $a_{2n} = C_n'M_n^\dagger$. We verify that $E(a_{1n}'U_n) = 0$, and $E(U_n'a_{2n}U_n) = \sigma_0^2\operatorname{tr}(C_n'M_n^\dagger) = o(\sqrt{n/l_n})$ by Lemma A.6. By Lemmas A.6 and A.12, and the Cauchy–Schwartz inequality, $\sigma_{1n}^2 \equiv \operatorname{Var}(a_{1n}'U_n) = \sigma_0^2a_{1n}'a_{1n} = \sigma_0^2R_n'M_n^\dagger R_n + o(n/l_n)$, $\sigma_{2n}^2 \equiv \operatorname{Var}(U_n'a_{2n}U_n) = (\mu_4 - 3\sigma_0^4)\sum_{i=1}^n[(C_n'M_n^\dagger)_{ii}]^2 + \sigma_0^4\operatorname{tr}[C_n'M_n^\dagger(C_n'M_n^\dagger + M_n^\dagger C_n)] = \sigma_0^4\operatorname{tr}(C_n'P_n(P_nC_n + C_n'P_n)) + o(n/l_n)$, and

$$\sigma_{12n}^2 \equiv \operatorname{Cov}\left(a_{1n}'U_n, U_n'a_{2n}U_n\right) = a_{1n}'\operatorname{diag}(a_{2n})\mu_3$$
$$= \left[R_n'M_n^\dagger + \mathbf{m}_0(Z_n)'M_n^\dagger G_n - 2m_0(Z_n)'M_n^\dagger(\operatorname{tr}(G_n)/n)\right] \times \operatorname{diag}\left(C_n'M_n^\dagger\right)\mu_3$$
$$= R_n'M_n^\dagger\operatorname{diag}(P_nC_n)\mu_3 + O\left(nh^{r+1}/l_n\right) + O(1/\sqrt{l_n}),$$

because $\operatorname{diag}(M_n^\dagger C_n) = \operatorname{diag}(P_nC_n) - O(1/n)$ and elements of P_nC_n are uniformly $O(1/l_n)$. Consequently, $\operatorname{Var}(a_{1n}'U_n + U_n'a_{2n}U_n) = \{\sigma_0^2R_n'M_n^\dagger R_n + \sigma_0^4\operatorname{tr}(C_n'P_n(P_nC_n + C_n'P_n)) + R_n'M_n^\dagger\operatorname{diag}(P_nC_n)\mu_3\} + o(n/l_n) \equiv \sigma_{\rho,n}^2 + o(n/l_n)$. By the CLT for linear-quadratic forms (Lee, 2004, Appendix A), we have

$$\sqrt{\frac{l_n}{n}}\frac{\partial \log L_n^c(\rho_0)}{\partial\rho} \xrightarrow{d} N\left(0, \lim_{n\to\infty}\frac{l_n}{n\sigma_0^4}\sigma_{\rho,n}^2\right). \tag{B.21}$$

By (5.2) and (B.17)–(B.21), we have $\sqrt{n/l_n}(\widehat{\rho} - \rho_0) = [(l_n/n\sigma_0^2)R_n'M_n^\dagger R_n + (l_n/n)\operatorname{tr}(G_n'(P_nG_n + G_n'))]^{-1}\sqrt{l_n/n}\{a_{1n}'U_n + U_n'a_{2n}U_n\}/\sigma_0^2 + o_P(1) \xrightarrow{d} N(0, \sigma_\rho^2)$. ■

Proof of Theorem 5.3. By (2.9), Lemmas A.7, A.10 and A.11, Theorem 5.2, the fact that $T_n(\rho) = T_n - (\rho - \rho_0)W_n$,

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \sqrt{n}\left(X_n'P_nX_n\right)^{-1}X_n'P_nU_n + \sqrt{n}\left(X_n'P_nX_n\right)^{-1}X_n'P_n\mathbf{m}_0(Z_n) - \sqrt{\frac{n}{l_n}}(\widehat{\rho} - \rho_0)\left(X_n'P_nX_n\right)^{-1}\sqrt{l_n}X_n'P_nG_nU_n - \sqrt{\frac{n}{l_n}}(\widehat{\rho} - \rho_0)\left(X_n'P_nX_n\right)^{-1}\sqrt{l_n}X_n'P_nR_n$$
$$= \left(n^{-1}X_n'P_nX_n\right)^{-1}\frac{1}{\sqrt{n}}X_n'P_nU_n - \sqrt{\frac{n}{l_n}}(\widehat{\rho} - \rho_0)\left(X_n'P_nX_n\right)^{-1}\sqrt{l_n}X_n'P_nR_n + o_P(1). \tag{B.22}$$

It is easy to verify that $\operatorname{Cov}(n^{-1/2}X_n'P_nU_n, \sqrt{l_n/n}(a_{1n}'U_n + U_n'a_{2n}U_n)) = (\sqrt{l_n}/n)(\sigma_0^2X_n'P_na_{1n} + X_n'P_n\operatorname{diag}(a_{2n})\mu_3) = o(1)$, where a_{1n} and a_{2n} are defined after (B.20). This implies that the two dominating terms in (B.22) are asymptotically independent. Consequently, we can apply the CLT for linear-quadratic forms (Lee, 2004, Appendix A) to obtain $\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma_\beta^2)$, where σ_β^2 is given in (5.4). When $\beta_0 = 0$, noting that $\mathbf{X}_s = \mathbf{m}_s(Z_n) + \boldsymbol{\eta}_s$, it is easy to verify that the sth element of $(\sqrt{l_n}/n)X_n'P_nR_n$ is $(\sqrt{l_n}/n)\mathbf{X}_s'P_nR_n = (\sqrt{l_n}/n)\mathbf{m}_s(Z_n)'P_nG_n\mathbf{m}_0(Z_n) + (\sqrt{l_n}/n)\boldsymbol{\eta}_s'P_nG_n\mathbf{m}_0(Z_n) = o(1) + o(1) = o(1)$, by the proof of Lemma A.10(2). Consequently, the second term in (B.22) vanishes asymptotically and $\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma_0^2\lim_{n\to\infty}n(X_n'P_nX_n)^{-1}X_n'P_nP_nX_n(X_n'P_nX_n)^{-1})$.

By (2.10), $\widehat{\sigma}^2 = (1/n)U_n'M_n^\dagger U_n + (1/n)\mathbf{m}_0(Z_n)'M_n^\dagger\mathbf{m}_0(Z_n) + (2/n)\mathbf{m}_0(Z_n)'M_n^\dagger U_n + (1/n)(\widehat{\rho} - \rho_0)^2Y_n'W_n'M_n^\dagger W_nY_n - (2/n)(\widehat{\rho} - \rho_0)Y_n'W_n'M_n^\dagger T_nY_n \equiv (1/n)U_n'M_n^\dagger U_n + \sum_{s=1}^4 R_{ns}$. By Lemma A.7, $\sqrt{n}R_{n1} = o(1)$. Because $E(\sqrt{n}R_{n2}) = 0$ and $\operatorname{Var}(\sqrt{n}R_{n2}) = \sigma_0^2n^{-1}\mathbf{m}_0(Z_n)'M_n^\dagger M_n^\dagger\mathbf{m}_0(Z_n) = O(h^{2(r+1)})$, $\sqrt{n}R_{n2} = o_P(1)$. $\sqrt{n}R_{n3} = \sqrt{l_n/n}[\sqrt{n/l_n}(\widehat{\rho} - \rho_0)]^2(\sqrt{l_n}/n)Y_n'W_n'M_n^\dagger W_nY_n = \sqrt{l_n/n}O_P(1)O_P(1/\sqrt{l_n}) = o_P(1)$, and $\sqrt{n}R_{n4} = -2[\sqrt{n/l_n}(\widehat{\rho} - \rho_0)](\sqrt{l_n}/n)Y_n'W_n'M_n^\dagger W_nY_n = O_P(1)O_P(1/\sqrt{l_n}) = o_P(1)$. Furthermore,

$$n^{-1/2}U_n'M_n^\dagger U_n = n^{-1/2}U_n'P_nU_n - n^{-1/2}(n^{-1/2}U_n'P_nX_n) \times \left(n^{-1}X_n'P_nX_n\right)^{-1}n^{-1/2}X_n'P_nU_n$$
$$= n^{-1/2}U_n'P_nU_n - n^{-1/2}O_p(1)\,O(1)\,O_p(1)$$
$$= n^{-1/2}U_n'P_nU_n + o_P(1),$$

and $n^{-1/2}U_n'P_nU_n = n^{-1/2}U_n'U_n - n^{-1/2}U_n'(S_n' + S_n - S_n'S_n)U_n = n^{-1/2}U_n'U_n + o_P(1)$, because $n^{-1/2}U_n'(S_n' + S_n - S_n'S_n)U_n = n^{-1/2}\sigma_0^2\operatorname{tr}(S_n' + S_n - S_n'S_n) + o_P(1) = O(n^{-1/2}h^{-q}) + o_P(1) = o_P(1)$. Consequently, $\sqrt{n}(\widehat{\sigma}^2 - \sigma_0^2) = (1/\sqrt{n})\sum_{i=1}^n(u_i^2 - \sigma_0^2) + o_P(1) \xrightarrow{d} N(0, \mu_4 - \sigma_0^4)$. ■

Proof of Theorem 5.4. By (2.12), (B.13) and (B.14), we have

$$\sqrt{nh^q}\left(\widehat{\alpha}(z) - \alpha^0(z)\right) = \sqrt{nh^q}\sum_{i=1}^n\mathbf{S}_n(z_{it}, z)\sum_{|\mathbf{j}|=r+1}\left(D^{\mathbf{j}}m_0\right)(z)\,(z_{it} - z)^{\mathbf{j}} + \sqrt{nh^q}\mathbf{S}_n(z)\,U_n - \sqrt{nh^q}\mathbf{S}_n(z)\,X_n\left(\widehat{\beta} - \beta_0\right) - \sqrt{nh^q}\mathbf{S}_n(z)\,W_nY_n(\widehat{\rho} - \rho_0) + o_P(1)$$
$$\equiv B_{11} + B_{12} - B_{13} - B_{14} + o_P(1),$$

where B_{11} and B_{12} are analyzed in the proof of Theorem 4.4. The only difference is that now $B_{11} = \sqrt{nh^q}h^{r+1}\Pi^{-1}Bm_0^{(r+1)}(z) + o(1) = o(1)$ because $\sqrt{nh^q}h^{r+1} = o(n^{1/2}h^{r+1}) = o(1)$.

Now write $B_{13} = \widetilde{B}_{13}\sqrt{n}(\widehat{\beta} - \beta_0)$, where $\widetilde{B}_{13} = \sqrt{h^q}\mathbf{S}_n(z)X_n$. Then the s $(1 \leq s \leq p)$ column of $\widetilde{B}_{13}$ is $\widetilde{B}_{13s} = \sqrt{h^q}f(z)^{-1}\Pi^{-1}(1/n)\sum_{i=1}^{n}\overrightarrow{Z}_i(z)K_h(z_{n,i}-z)\{m_s(z_{n,i})+\eta_{is}\}(1+o(1)) = \sqrt{h^q}O(h^{r+1} + n^{-(\delta-q)/(2\delta)}\log n) = o(1)$, where the second equality follows from analogous analysis to Lemma A.3 and (A.2). For B_{14}, write $B_{14} = \sqrt{n/l_n}(\widehat{\rho} - \rho_0)\sqrt{l_n h^q}(\mathbf{S}_n(z)G_nU_n + \mathbf{S}_n(z)R_n) \equiv \sqrt{n/l_n}(\widehat{\rho}-\rho_0)\{B_{14a}+B_{14b}\}$, which is $o_P(1)$ provided $B_{14a} = o_P(1)$ and $B_{14b} = o(1)$. Note $B_{14a} = \sqrt{l_n h^q}f(z)^{-1}\Pi^{-1}\{n^{-1}\sum_{i=1}^{n}\overrightarrow{Z}_i(z)K_h(Z_i - z)\sum_{j=1}^{n}g_{n,ij}u_j\}\{1 + o(1)\}$. We verify that $E(B_{14a}) = 0$, and $E(B_{14a}B_{14a}') = O(h^q) = o(1)$. It follows by the Chebyshev inequality that $B_{14a} = o_P(1)$. Let $R_{1n} = X_n'\beta_0 + \mathbf{m}_0(Z_n)$. Then elements $r_{1n,i}$ of R_{1n} are uniformly bounded by Assumption 1, which implies that the elements of $S_nR_n = S_nG_nR_{1n}$ are also uniformly bounded by Lemma A.5 and Facts 1–2. Noting that $R_n = (I_n - S_n)R_n + S_nR_n$, we conclude that the elements $r_{n,i}$ of R_n are uniformly $O(l_n^{-1/2})$ by Assumption 9. With this, we can show that $\|B_{14b}\| = O(\sqrt{h^q}) = o(1)$. This completes the proof. ■

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795–1843.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, The Netherlands.
Anselin, L., Bera, A.K., 2002. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York.
Anselin, L., Florax, R.J.G.M., 1995. New Directions in Spatial Econometrics. Springer-Verlag, Berlin.
Baltagi, B.H., Li, D., 2001. LM tests for functional form and spatial correlation. International Regional Science Review 24, 194–225.
Basile, R., Gress, B., 2004. Semi-parametric spatial auto-covariance models of regional growth behavior in Europe. Mimeo. Dept. of Economics, UC, Riverside.
Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.
Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion Ltd., London.
Cressie, N., 1993. Statistics for Spatial Data. John Wiley & Sons, New York.
Dragomir, S.S., Pearce, C.E.M., Šunde, J., 1998. Abel-type inequalities, complex numbers and Gauss–Pólya type integral inequalities. Mathematical Communications 3, 95–101.
Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall.
Gao, J., 1995. The laws of iterated logarithm of some estimates in partially linear models. Statistics & Probability Letters 25, 153–162.
Gress, B., 2004. Using semi-parametric spatial autocorrelation models to improve hedonic housing price prediction. Mimeo. Dept. of Economics, UC, Riverside.
Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Physica-Verlag, Heidelberg.
Horn, R., Johnson, C., 1985. Matrix Analysis. John Wiley & Sons, New York.
Kelejian, H.H., 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.
Kelejian, H.H., 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., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157 (1), 53–67.
Lee, L.-F., 2002a. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.
Lee, L.-F., 2002b. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Manuscript, Dept. of Economics, Ohio State Univ.
Lee, L.-F., 2004. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.
Li, Q., Stengos, T., 1996. Semiparametric estimation of partially linear panel data models. Journal of Econometrics 71, 289–397.
Liang, H., 1999. An application of Bernstein's inequality. Econometric Theory 15, 152 (16), 619–620.
Lin, X., Lee, L.-F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157 (1), 34–52.
Linton, O., 1995. Second order approximation in the partially linear regression model. Econometrica 63, 1079–1112.
Masry, E., 1996. Multivariate local polynomial regression for time series: Uniform strong consistency rates. Journal of Time Series Analysis 17, 571–599.
Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–126.
Pace, P.K., Barry, R., Slawson Jr., V.C., Sirmans, C.F., 2004. Simultaneous spatial and functional form transformation. In: Anselin, L., Florax, R., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer-Verlag, Berlin, pp. 197–224.
Paelinck, J.H.P., Klaassen, L.H., 1979. Spatial Econometrics. Gower.
Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: A semiparametric approach. Econometrica 70, 1111–1153.
Robinson, P.M., 1988. Root-n consistent semiparametric regression. Econometrica 56, 931–954.
Rupasingha, A., Goetz, S.J., Debertin, D.L., Pagoulatos, A., 2004. The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions. Papers in Regional Science 83, 407–424.
Smirnov, O., 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.
Speckman, P., 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society. Series B 50, 413–436.
Staniswalis, J.E., 1989. On the kernel estimate of a regression function in likelihood-based models. Journal of the American Statistical Association 84, 276–283.
van Gastel, R.A.J.J., Paelinck, J.H.P., 1995. Computation of Box–Cox transform parameters: A new method and its application to spatial econometrics. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, pp. 136–155.
White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press.
Yang, Z., Li, C., Tse, Y.K., 2006. Functional form and spatial dependence in dynamic panels. Economics Letters 91, 138–145.
Yatchew, A., 1998. Nonparametric regression techniques in economics. Journal of Economics Literature 36, 669–721.