SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Instrumental Variable Quantile Estimation of Spatial Autoregressive Models

Liangjun Su, Zhenlin Yang

August 2007

Paper No. 05-2007

Instrumental Variable Quantile Estimation of Spatial Autoregressive Models*

Liangjun Su† and Zhenlin Yang‡

August 27, 2007

Abstract

We propose an instrumental variable quantile regression (IVQR) estimator for spatial autoregressive (SAR) models. Like the GMM estimators of Lin and Lee (2006) and Kelejian and Prucha (2006), the IVQR estimator is robust against heteroscedasticity. Unlike the GMM estimators, the IVQR estimator is also robust against outliers and requires weaker moment conditions. More importantly, it allows us to characterize the heterogeneous impact of variables on different points (quantiles) of a response distribution. We derive the limiting distribution of the new estimator. Simulation results show that the new estimator performs well in finite samples at various quantile points. In the special case of median restriction, it outperforms the conventional QML estimator without taking into account of heteroscedasticity in the errors; it also outperforms the GMM estimators with or without considering the heteroscedasticity.

JEL classifications: C13, C21, C51

Key Words: Spatial Autoregressive Model; Quantile Regression; Instrumental Variable; Quasi Maximum Likelihood; GMM; Robustness.

*We thank Peter Robinson, Jean Paelinck, Bernard Fingleton, Anil Bera, and the participants of The 1st World Conference of the Spatial Econometrics Association for their helpful comments. Liangjun Su gratefully acknowledges the financial support from the NSFC (70501001, 70601001). He also thanks the School of Economics, Singapore Management University (SMU), for the hospitality during his two-month visit, and the Wharton-SMU research center, SMU, for supporting his visit. Zhenlin Yang gratefully acknowledges the support from a research grant (Grant number: C244/MSS6E012) from Singapore Management University.

†Guanghua School of Management, Peking University, Beijing 100871, China. Telephone: +86-10-6276-7444. Email: lsu@gsm.pku.edu.cn.

‡School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Telephone: +65-6828-0852. Fax: +65-6828- 0833. Email: zlyang@smu.edu.sg.

1 Introduction

Spatial dependence among the cross-sectional units has become in recent years a standard notion of economic research activities in relation to social interactions, spill-overs, copy-cat policies, externalities, etc., and has received an increasing attention by theoretical econometricians and applied researchers. Among the various models involving spatial dependence, the spatial autoregressive or SAR model is perhaps the most popular one. The SAR model has the following form:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + u_n, \tag{1.1}$$

where λ_0 is the spatial lag parameter, W_n is a known $n \times n$ spatial weight matrix, $W_n Y_n$ is the spatial lagged variable, n is the total number of spatial units, X_n is an $n \times p$ matrix with its rows $(x'_{n,i}, i = 1, \cdots, n)$ being the values of p regressors, β_0 is a p-vector of unknown regression parameters, and $u_n \equiv (u_{n,1}, \cdots, u_{n,n})'$ denotes the n-vector of random disturbance terms. In the standard SAR setting, $u_{n,i}$ is typically assumed to have a zero mean, i.e., the model is under the mean restriction.

Since $W_n Y_n$ is present on the right hand side of (1.1), the ordinary least squares (OLS) estimator is usually inconsistent. Traditionally, there are two types of estimators that have been studied and commonly used in the literature. One is the maximum likelihood (ML) or quasi maximum likelihood (QML) estimator; see, among the others, Ord (1975), Anselin (1988), Smirnov and Anselin (2001), and Lee (2002b, 2004). The other is the generalized method of moment (GMM) estimator; see, among the others, Kelejian and Prucha (1998, 1999), Lee (2003, 2007), and Liu, Lee and Bollinger (2007). Both estimators are under the assumption that the disturbances $\{u_{n,i}\}$ are independent and identically distributed (iid).

While the spatial models with iid innovations have been extensively studied and applied, researchers have started to realize that an important issue in modelling the spatial data, the heteroscedasticity, has not been adequately addressed. Lin and Lee (2006) argued that social interactions may cause the variance of the aggregated level data be inflated, and Kelejian and Prucha (2006) indicated that spatial units are often heterogeneous in important characteristics such as size. As a result, the QML estimator under iid assumption is inconsistent, and the asymptotic distribution for GMM estimation under iid assumption is not appropriate. Lin and Lee (2006) extended the GMM method to allow for heteroscedasticity for the SAR model, while Kelejian and Prucha (2006) considered the GMM estimation with heteroscedasticity to a more general model, called SARAR, where the disturbance also follows a SAR process. Clearly, these spatial models can be considered as "spatial" extensions of the usual mean regressions with or without heteroscedasticity, where the model estimation is based primarily on the restriction that the means of the error terms are zero.

Koenker and Bassett (1978) made an important extension of the standard mean regression to the quantiles of the responses, giving what is now called the `quantile regression (QR)` model, which allows a separate modelling at different points of a response distribution so that the heterogeneous impacts of explanatory variables can be characterized and differentiated at different points of a response distribution. The standard linear QR model has the form

$$Y_n = X_n\beta_{0\tau} + u_n, \tag{1.2}$$

where the τth quantile of $u_{n,i}$ is zero, and $\beta_{0\tau}$ is the so-called regression quantile that may change with the value of τ. The method of estimating the linear QR model is to minimize the average of asymmetric absolute deviations, which in the special case of $\tau = 0.5$ gives the well-known least absolute deviations (LAD) estimator. Subsequently, the QR model has been studied, extended and applied by many authors. See Koenker (2005) for an excellent exposition of the quantile regression.

While both Model (1.1) and Model (1.2) can be considered as stepping-stone models in their own fields (i.e., spatial econometrics and quantile regression), a combination of the two may open up a new and exciting research direction. In this paper we consider the estimation of the SAR model under quantile restrictions, i.e., assuming that the τth quantile of $u_{n,i}$ in (1.1) is zero (see Section 2 for details). Quantile regression is an important method for modeling heterogenous effects of variables on a response and at the same time taking into account of unobserved heterogeneity. It also permits heteroscedasticity among the disturbances. Moreover, like many other robust estimators, quantile estimators are robust and much less sensitive to outliers. Since heterogeneity, heteroscedasticity and extreme values are frequently present in spatial data, it is important to study the estimation of SAR models under quantile restrictions.

Since the spatial lagged variable is present on the right hand side of (1.1), the conventional quantile regression of Koenker and Bassett (1978) generally produces inconsistent estimates. We need to consider quantile regression with endogenous regressors by treating (1.1) as a structural equation. In this paper, we propose an instrumental quantile regression (IVQR) estimator of the SAR model by extending the method of Chernozhukov and Hansen (2006). We establish the asymptotic distribution of our IVQR estimator. Monte Carlo simulation reveals that the new estimator generally performs well in finite samples at various quantile points. In the special case of median restriction, it outperforms the conventional QML estimators without taking into account of heteroscedasticity in the errors; it also outperforms the GMM estimators with or without considering the heteroscedasticity.

To the best of our knowledge, Amemiya (1982) was the first to study the asymptotic properties of a class of two-stage median regression estimators for the structural equation model. This method was extended by Powell (1983) and Chen and Portony (1996). Recently, Abadie, Angrist and Imbens (2002) consider quantile regression methods for estimating endogenous treatment effects where the endogenous variables are dummy variables. Sakata (2006) develops an instrumental variable method to estimate structural equations based on conditional median restriction. Chernozhukov and Hansen (2005) consider modeling and identification of quantile treatment effect in the presence of endogeneity. Chernozhukov and Hansen (2006) introduce a class of instrumental variable quantile regression (IVQR) methods for structural and treatment effect models. Ma and Koenker (2006) study the quantile regression methods for recursive structural equation models. A common feature of these methods is that they are all developed to estimate a structural equation with iid data. In case of spatial data, it remains unclear whether we can develop relevant theory under quantile restrictions.

This paper is organized as follows. Section 2 introduces the model and the IVQR estimator. Section 3 studies the asymptotic properties of the IVQR estimator. Section 4 presents Monte Carlo results for the finite sample properties of the IVQR estimator, and for the comparisons with the conventional GMM and QML estimators at the special case where $\tau = 0.5$. Section 5 contains concluding remarks. All proofs are relegated to the appendix.

To proceed, we introduce some notation. Let I_n be the $n \times n$ identity matrix. For a matrix A_n, we denote its norm as $\|A_n\| = [\text{tr}(A_n A_n')]^{1/2}$, and the (i,j)th element of it as $a_{n,ij}$. Similarly, for a vector a_n, $a_{n,i}$ denotes its ith element. We say A_n is uniformly bounded in absolute value if $\sup_{1\leq i\leq n, 1\leq j\leq n} |a_{n,ij}| < \infty$. We say A_n is uniformly bounded in row sums (or column sums) if $\sup_{1\leq i\leq n} \sum_{j=1}^{n} |a_{n,ij}| \leq c_a$ (or $\sup_{1\leq j\leq n} \sum_{i=1}^{n} |a_{n,ij}| \leq c_a$) for some constant $c_a < \infty$. Let $e_{m,i}$ denote the $m \times 1$ unit vector with 1 in the ith place, $i \leq m$.

2 The Model and the Method of Estimation

2.1 The SAR Model Under Quantile Restrictions

A natural extension of the ordinary SAR model given in (1.1) is to assume the τth quantile of $u_{n,i}$ to be zero, and a natural extension of the ordinary QR model given in (1.2) is to allow a spatial lag in the model. Both extensions lead to a model of the form

$$Y_n = \lambda_{0\tau} W_n Y_n + X_n \beta_{0\tau} + u_n, \tag{2.1}$$

where the τth quantile of $u_{n,i}$ is zero for $i = 1, \cdots, n$, $\lambda_{0\tau}$ is a scalar spatial lag parameter that is τ-dependent, and $\beta_{0\tau}$ is a p-vector regression parameters that is also τ-dependent. The other quantities are defined similarly as those in Model (1.1).

This generalization can be very interesting as it allows a different degree of spatial dependence at a different point of the response distribution, i.e., it allows the spatial parameter λ_τ to be dependent on τ. At the same time, it also allows, as in the ordinary quantile regression, the impacts (β_τ) of the covariates X_n on the response Y_n to be different at the different quantile (τ) points. Taking, for example, the housing prices, while it is certainly reasonable to think that the way the price relates to the covariates at a high quantile point ($\tau = 0.9$, say) is different from that at a low quantile point ($\tau = 0.1$, say). i.e., $\beta_{0.9} \neq \beta_{0.1}$; at the same time, it should also be very reasonable to think that the way the price of a house spatially relates to the prices of its neighbors around the city center (high τ) should be different from that around a suburb (low τ), e.g., $\lambda_{0.9} \neq \lambda_{0.1}$.

Denote $S_n(\lambda) = I_n - \lambda W_n$ for any value of λ. It follows that (2.1) has the reduced form

$$Y_n = S_n^{-1}(\lambda_{0\tau})(X_n\beta_{0\tau} + u_n), \tag{2.2}$$

provided that $S_n(\lambda_{0\tau})$ is nonsingular. This reduced form will be frequently used in the derivation of the asymptotic properties of the estimator proposed below.

As reviewed in the introduction, there are many approaches to the estimation of the parameters in (1.1), among which the method of (quasi-) maximum likelihood and the (generalized) method of moments are the two most popular ones. However, no estimator has been proposed to estimate the parameters in (2.1). When $\tau = 0.5$ and the distribution of $u_{n,i}$ is symmetric, Model (2.1) becomes essentially Model (1.1). The distinction is that for the QML and GMM methods to work the disturbance term in (1.1) is usually assumed to possess finite $(4 + \eta)$th moments, whereas for the IVQR method to work the disturbance in (2.1) is only assumed to have a finite first moment. A much greater distinction between Model (1.1) and Model (2.1) is that the former is subject to the mean restriction, whereas the latter is subject to the quantile restriction in the sense that the model can be estimated separately at different quantile points, and doing so the heterogeneous impacts of the explanatory variables on different points of the distribution of a response Y can be characterized. Below we introduce the estimator based on quantile regression with endogenous regressors.

2.2 The IVQR Estimator of the SAR Model

Since the seminal work of Koenker and Bassett (1978), much attention has been paid to the use of quantile estimation for robust inference. Like many other robust estimators, quantile estimators are robust and much less sensitive to outliers. Also, it permits certain form of heteroscedasticity in the error terms. These properties make the Model (2.1) very interesting. The question now is how we are going to estimate Model (2.1) as it is not a standard quantile regression model.

Let $\bar{y}_{n,i}$ be the ith element of $\bar{y}_n \equiv W_nY_n$. Following Koenker and Bassett (1978), we may formulate the conventional quantile regression estimator of $(\lambda_{0\tau}, \beta_{0\tau})$ as finding the best predictor of $y_{n,i}$ given $\bar{y}_{n,i}$ and $x_{n,i}$ under the asymmetric least absolute deviation loss $\rho_\tau(u) \equiv (\tau - 1(u \leq 0))u$, where $1(\cdot)$ is the usual indicator function. In other words, we may consider estimating $(\lambda_{0\tau}, \beta_{0\tau})$ by:

$$(\overrightarrow{\lambda}_\tau, \overrightarrow{\beta}_\tau) = \arg\min_{(\lambda,\beta)} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \beta' x_{n,i}). \tag{2.3}$$

However, due to the endogeneity of $\bar{y}_{n,i}$, the above estimator is usually inconsistent.

An important development in the literature of quantile regression made by Chernozhukov and Hansen (2005, 2006) is to allow for endogeneity in the model. Our estimator is proposed based on the ideas of Chernozhukov and Hansen. To motivate our estimator, we for the moment pretend that $\{y_{n,i}, \bar{y}_{n,i}, x_{n,i}\}$ is an iid sequence, and allow the dependence between $\bar{y}_{n,i}$ and $u_{n,i}$ but not that between $x_{n,i}$ and $u_{n,i}$. This will greatly facilitate our discussions at the population level. If the τth conditional quantile of $u_{n,i}$ given $\bar{y}_{n,i}$ and $x_{n,i}$ is nonzero (which is true when there exists endogeneity), then the estimator in (2.3) may fail to be consistent because $\lambda_{0\tau}\bar{y}_{n,i}+\beta_{0\tau}'x_{n,i}$ is not the τth conditional quantile of $y_{n,i}$ given $\bar{y}_{n,i}$ and $x_{n,i}$. However, if there exists a vector of instruments $\varsigma_{n,i}$ (for $\bar{y}_{n,i}$) that it is correlated with $\bar{y}_{n,i}$ in an appropriate way but independent of $u_{n,i}$, we have under mild conditions (Chernozhukov and Hansen (2006)),

$$\Pr(y_{n,i} \leq \lambda_{0\tau}\bar{y}_{n,i} + \beta_{0\tau}'x_{n,i}|x_{n,i}, \varsigma_{n,i}) = \tau \quad \text{a.s.} \tag{2.4}$$

Eq. (2.4) simply says that 0 is a the τth quantile of $y_{n,i} - \lambda_{0\tau}\bar{y}_{n,i} - \beta_{0\tau}'x_{n,i}$ conditional on $(x_{n,i}, \varsigma_{n,i})$, almost surely for each τ. Note that the conditional probability $\Pr(y_{n,i} \leq \lambda_{0\tau}\bar{y}_{n,i} + \beta_{0\tau}'x_{n,i}|x_{n,i}, \varsigma_{n,i})$ is a measurable function of $(x_{n,i}, \varsigma_{n,i})$. Thus, following Chernozhukov and Hansen (2006), to solve Eq. (2.4) is to find $(\lambda_{0\tau}, \beta_{0\tau})$ such that 0 is a solution to the τth quantile regression of $y_{n,i} - \lambda_{0\tau}\bar{y}_{n,i} - \beta_{0\tau}'x_{n,i}$ on $(x_{n,i}', \varsigma_{n,i}')'$:

$$0 \in \arg\min_{g\in\mathcal{F}} \mathrm{E}\left[\rho_\tau(y_{n,i} - \lambda_{0\tau}\bar{y}_{n,i} - \beta_{0\tau}'x_{n,i} - g(x_{n,i}, \varsigma_{n,i}))\right], \tag{2.5}$$

where $\mathcal{F}$ is a class of measurable functions of $(x_{n,i}, \varsigma_{n,i})$ that will be suitably restricted in the finite sample applications.[1]

The arguments leading to (2.5) gives the theoretical foundation for the development of the new estimation method. In real applications, the class $\mathcal{F}$ of the measurable g functions can be restricted to be linear. Let $z_{n,i}$ be a q-vector of instrumental variables. Now, we consider a finite-sample analog of the population instrumental variable quantile regression and define the weighted quantile regression (QR) objective function as

$$Q_{n\tau}(\lambda, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \beta'x_{n,i} - \gamma'z_{n,i})v_{n,i}, \tag{2.6}$$

where $v_{n,i} > 0$ is a scalar weight. The IV $z_{n,i}$ can be formed from $x_{n,i}$ and $\varsigma_{n,i}$. The weights $v_{n,i}$ are important for nonparametric or semiparametric quantile regressions and for asymptotic efficiency considerations. Since we only focus on the parametric quantile regressions, they are not essential for our discussions below, hence $v_{n,i}$ can always be set to 1 in practice. A natural choice of the instrument matrix $Z_n = (z_{n,1}, \cdots, z_{n,n})'$ may be the matrix consisting of linearly independent columns of $\overline{X}_n \equiv W_nX_n$.

[1]This idea parallels the interpretation of the ordinary quantile regression estimator at the population level: the τth conditional quantile of $y_{n,i}$ given $x_{n,i}$ in Model (2.1) can be understood as the solution to the problem: $Q_{Y|X}(\tau) \in \arg\min_{g\in\mathcal{F}} \mathrm{E}\left[\rho_\tau(y_{n,i} - g(x_{n,i}))\right]$, where $\mathcal{F}$ is the class of measurable functions of $x_{n,i}$, restricted to the form $\beta_{0\tau}'x_{n,i}$ in finite sample applications.

Following the arguments leading to (2.5) at the population level, if the finite sample objective function $Q_{n\tau}(\lambda, \beta, \gamma)$ meets certain identification conditions we expect that the estimate of γ is close to zero when (λ, β) is close to the true population values $(\lambda_{0\tau}, \beta_{0\tau})$. Now, let $\xi_{n,i} = (x'_{n,i}, z'_{n,i})'$. We define the instrumental variable quantile regression (IVQR) estimator for the SAR model as follows:

(i) for a given value of λ, run an ordinary QR of $y_{n,i} - \lambda \bar{y}_{n,i}$ on $\xi_{n,i}$ to obtain

$$(\hat{\beta}(\lambda, \tau)', \hat{\gamma}(\lambda, \tau)')' \equiv \arg\min_{(\beta,\gamma)} Q_{n\tau}(\lambda, \beta, \gamma); \tag{2.7}$$

(ii) minimize a norm of $\hat{\gamma}(\lambda, \tau)$ over λ to obtain the IVQR estimator of $\lambda_{0\tau}$, i.e.,

$$\hat{\lambda}(\tau) = \arg\min_{\gamma} \|\hat{\gamma}(\lambda, \tau)\|_{\widehat{A}} \tag{2.8}$$

where $\|\gamma\|_A = \sqrt{\gamma' A \gamma}$, and $\widehat{A} = A + o_p(1)$ for some positive definite matrix A; and finally

(iii) run an ordinary QR of $y_{n,i} - \hat{\lambda}(\tau)\bar{y}_{n,i}$ on $\xi_{n,i}$ to obtain the IVQR estimator of $\beta_{0\tau}$. That is,

$$\hat{\beta}(\tau) \equiv \hat{\beta}(\hat{\lambda}(\tau), \tau)). \tag{2.9}$$

Intuitively, to find $\hat{\lambda}(\tau)$ in Step (ii), we look for a value of λ that makes the coefficient $\hat{\gamma}(\lambda, \tau)$ of the instrumental variable as close to 0 as possible. The weight matrix $\widehat{A}$ is used for asymptotic efficiency purpose. A convenient choice is to set A equal to the inverse of the asymptotic covariance matrix of $\sqrt{n}(\hat{\gamma}(\lambda, \tau) - \gamma_0(\lambda, \tau))$ where $\gamma_0(\lambda, \tau)$ denotes the probability limit of $\hat{\gamma}(\lambda, \tau)$.

Remark 1. It is simple to implement the above IVQR procedure in practice: (i) for a given probability index τ of interest (e.g., $\tau = 0.5$ for IV median regression), define a grid of values $\{\lambda_j, j = 1, \cdots, J\}$ that lie in a compact subset of $(-1, 1)$ (say when W_n is row normalized), and run an ordinary τ-quantile regression of $y_{n,i} - \lambda \bar{y}_{n,i}$ on $(x'_{n,i}, z'_{n,i})$ with weight $v_{n,i} \equiv 1$ to obtain coefficients $(\hat{\beta}(\lambda_j, \tau), \hat{\gamma}(\lambda_j, \tau))$; and (ii) choose $\hat{\lambda}(\tau)$ as the value among $\{\lambda_j, j = 1, \cdots, J\}$ that makes $\|\hat{\gamma}(\lambda, \tau)\|_{\widehat{A}}$ closest to zero. The estimate of $\beta_{0\tau}$ is then given by $\hat{\beta}(\hat{\lambda}(\tau), \tau)$.

Remark 2. There are other approaches to obtain estimates of $(\lambda_{0\tau}, \beta_{0\tau})$ in (2.1). For example, one can follow Honoré and Hu (2004) and propose a method of moments approach that attempts to minimize $\|G^0_{n\tau}(\lambda, \beta)\|_{\widehat{A}}$ over (λ, β), where

$$G^0_{n\tau}(\lambda, \beta) = \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_{n,i} - \lambda \bar{y}_{n,i} - \beta' x_{n,i}) \xi_{n,i} v_{n,i}, \tag{2.10}$$

and $\psi_\tau(u) \equiv \partial \rho_\tau(u)/\partial u = \tau - 1(u \leq 0)$. See also Abadie (1995) in a different context. Another example is to generalize the median estimator of Sakata (2006) to our spatial context. In contrast to the IVQR approach proposed in this paper, these alternative approaches involve highly non-convex, multi-modal, and non-smooth objective functions over many parameters, which make them difficult to be implemented in practice, and thus are not considered in this paper. However, the functions $G^0_{n\tau}(\cdot,\cdot)$ and $\psi_\tau(\cdot)$ remain very important to the theoretical developments in this paper.

3 Asymptotic Properties of the IVQR Estimator

In this section, we derive the asymptotic distribution of the IVQR estimator defined above. Throughout, we denote $S_n \equiv S_n(\lambda_{0\tau})$ and $H_n \equiv W_n S_n^{-1}$. Also, we use Λ and $\mathcal{B}$ to denote the parameter spaces for λ and β, respectively, and "E" to denote the expectation operator corresponding to the true parameter values. We first provide a set of assumptions.

3.1 Assumptions

First we make some assumptions on the random disturbance terms and the spatial weight matrix.

Assumption 1. (i) The random disturbance terms $u_{n,1},\cdots,u_{n,n}$ are independent of each other. (ii) The τth quantile of $u_{n,i}$ is zero for each $i=1,\cdots,n$. (iii) The density $f_{n,i}(u)$ of $u_{n,i}$ is uniformly bounded with bounded continuous first derivatives. (iv) $\sup_n \max_{1\leq i\leq n} E|u_{n,i}| \leq \overline{\mu} < \infty$.

Assumption 2. (i) As a normalization, the diagonal elements $w_{n,ii}$ of W_n are 0 for all i. (ii) The matrix S_n is nonsingular. (iii) The sequences of matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums. (iv) $\{S_n^{-1}(\lambda)\}$ are uniformly bounded in either row or column sums, uniformly in $\lambda \in \Lambda$, where Λ is convex compact with $\lambda_{0\tau}$ in its interior. (v) The diagonal elements $h_{n,ii}$ of H_n satisfy $\lim_{n\to\infty} \min_{1\leq i\leq n} \inf_{\lambda\in\Lambda} b_{ni}(\lambda) = c_h > 0$ with $b_{ni}(\lambda) \equiv 1-(\lambda-\lambda_{0\tau})\,h_{n,ii}$.

Like Lee (2004), Assumptions 2(i)-(iv) provide the essential features of the weight matrix for the model. Assumption 2(ii) guarantees that the disturbance terms are well defined. Kelejian and Prucha (1998, 1999, 2001) and Lee (2004) also assume Assumption 2(iii) 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), $\limsup_n \|\lambda_{0\tau} W_n\| < 1$ is sufficient to guarantee that S_n^{-1} is uniformly bounded in both row and column sums. By Lee (2002b, Lemma A.3), Assumption 2(iii) implies $\{S_n^{-1}(\lambda)\}$ are uniformly bounded in both row and column sums uniformly in a neighborhood of $\lambda_{0\tau}$. Assumption 2(iv) requires this to be true uniformly in $\lambda\in\Lambda$. Assumption 2(v) restricts both W_n and the parameter space for λ. It is not as restrictive as it appears. For example, if we further assume that the elements $w_{n,ij}$ of W_n are uniformly at most of order l_n^{-1} such that $l_n\to\infty$ and $l_n/n\to 0$ (see Assumption 5* below), then by Lemma A.1 in the appendix, $h_{n,ii} = O(1/l_n) = o(1)$ so that Assumption 2(v) is automatically satisfied. One can consider relaxing Assumption 2(v) but at the cost of lengthier proofs.

For the regressors $x_{n,i}$, instruments $z_{n,i}$, weights $v_{n,i}$ and $\widehat{A}$, we make the following assumption.

Assumption 3. (i) The regressors $x_{n,i}$ are nonstochastic and uniformly bounded in absolute value, and X_n has full column rank and contains a column of ones. (ii) The instruments $z_{n,i}$ are nonstochastic and uniformly bounded in absolute value, and the instrument matrix Z_n has full column rank $q \geq 1$. (iii) The weights $v_{n,i}$ are nonnegative and uniformly bounded. (iv) $\widehat{A} = A + o_p(1)$, where A is a symmetric positive definite matrix.

Assumptions 3(i)-(ii) are standard; see Kelejian and Prucha (1998, 1999). In most applications, Z_n is composed of linearly independent columns of $(W_nX_n, W_n^2X_n, \cdots)$, where the subset contains at least the linearly independent columns of W_nX_n. The Z_n matrix chosen this way satisfy Assumption 3(ii) due to Assumptions 2(iii) and 3(i).

For identification purpose, define the population objective function as

$$Q_\tau(\lambda, \beta, \gamma) \equiv \lim_{n\to\infty} \mathrm{E}[Q_{n\tau}(\lambda, \beta, \gamma)]. \tag{3.1}$$

Let $\alpha_{0\lambda\tau} \equiv \alpha_0(\lambda, \tau) \equiv (\beta_0(\lambda, \tau)', \gamma_0(\lambda, \tau)')' \equiv \arg\min_{(\beta,\gamma)} Q_\tau(\lambda, \beta, \gamma)$. Let $G_{n\tau}(\lambda, \beta, \gamma)$ be the negative partial derivative of $Q_{n\tau}(\lambda, \beta, \gamma)$ with respect to $(\beta', \gamma')'$, i.e.,

$$G_{n\tau}(\lambda, \beta, \gamma) = \frac{1}{n}\sum_{i=1}^{n}[\psi_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \beta' x_{n,i} - \gamma' z_{n,i})]\xi_{n,i}v_{n,i}. \tag{3.2}$$

Recall the function $G^0_{n\tau}(\lambda, \beta)$ introduced in (2.10) and note that $G^0_{n\tau}(\lambda, \beta) = G_{n\tau}(\lambda, \beta, 0)$. Define

$$G^0_\tau(\lambda, \beta) = \lim_{n\to\infty} \mathrm{E}[G^0_{n\tau}(\lambda, \beta)] \quad \text{and} \quad G_\tau(\lambda, \beta, \gamma) = \lim_{n\to\infty} \mathrm{E}[G_{n\tau}(\lambda, \beta, \gamma)]. \tag{3.3}$$

We impose the following high-level assumption.

Assumption 4. Let τ be given. (i) $(\lambda_{0\tau}, \beta'_{0\tau})'$ is in the interior of a convex compact set $\Lambda \times \mathcal{B} \subset \mathbb{R}^{1+p}$. (ii) $\partial G_\tau(\lambda, \beta, \gamma)/\partial(\beta', \gamma')$ is continuous and has full rank at $(\beta'_0(\lambda, \tau), \gamma'_0(\lambda, \tau))$ uniformly in $\lambda \in \Lambda$. (iii) $\partial G^0_\tau(\lambda, \beta)/\partial(\lambda, \beta')$ is continuous and has full column rank at $(\lambda_{0\tau}, \beta'_{0\tau})$. (iv) If $G^0_\tau(\lambda^*, \beta^*) = 0$, then $\lambda^* = \lambda_{0\tau}$ and $\beta^* = \beta_{0\tau}$. (v) $\alpha_0(\lambda, \tau)$ is continuous in $\lambda \in \Lambda$.

Assumption 4(i) imposes compactness on the parameter space. Note that the objective function in the first step estimation is convex in (β, γ) for each λ. Assumption 4(ii) imposes a local identification condition for the conventional quantile regression (of $y_{n,i} - \lambda\bar{y}_{n,i}$ on $\xi_{n,i}$). This condition can further be seen as follows. Let $\alpha_{0\tau} \equiv (\beta'_{0\tau}, 0')'$ be the value of $\alpha_0(\lambda, \tau)$ at $\lambda = \lambda_{0\tau}$. Under Assumptions 1(iii) and 2(v), we have

$$\left.\frac{\partial G_\tau(\lambda, \beta, \gamma)}{\partial(\beta', \gamma')}\right|_{\substack{\beta=\beta_0(\lambda,\tau) \\ \gamma=\gamma_0(\lambda,\tau)}} = -\lim_{n\to\infty} J_{n\alpha}(\lambda, \tau), \tag{3.4}$$

where

$$J_{n\alpha}(\lambda, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[f_{n,i}\left(\frac{a_{ni}(\lambda)}{b_{ni}(\lambda)}\right)\right]\frac{\xi_{n,i}\xi'_{n,i}v_{n,i}}{b_{ni}(\lambda)}, \tag{3.5}$$

$$a_{ni}(\lambda) \equiv (\lambda - \lambda_{0\tau}) \sum_{l \neq i}^{n} h_{n,il} u_{n,l} + (\lambda - \lambda_{0\tau}) e_{n,i}' H_n X_n \beta_{0\tau} + (\alpha_0(\lambda, \tau) - \alpha_{0\tau})' \xi_{n,i} \qquad (3.6)$$

and $b_{ni}(\lambda)$ is defined in Assumption 2(v). Note that for sufficiently large n, $J_{n\alpha}(\lambda, \tau)$ is the same as $-\partial \mathrm{E}[G_{n\tau}(\lambda, \beta, \gamma)]/\partial(\beta', \gamma')$. Thus, the local identification condition of Assumption 4(ii) boils down to requiring the matrix $J_{n\alpha}(\lambda, \tau)$ to be positive definite for large enough n.

Assumption A4(iii) requires implicitly the relevance between the instruments $\xi_{n,i}$ and the endogenous variable $\bar{y}_{n,i}$. This is because under Assumptions 1(iii) and 2(v),

$$\left.\frac{\partial G_\tau^0(\lambda, \beta)}{\partial(\lambda, \beta')}\right|_{\substack{\lambda=\lambda_{0\tau} \\ \beta=\beta_{0\tau}}} = - \lim_{n\to\infty} J_n(\tau), \qquad (3.7)$$

where

$$J_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} f_{n,i}(0) \xi_{n,i} \widetilde{x}_{n,i}' v_{n,i}, \qquad (3.8)$$

with $\widetilde{x}_{n,i}' = (\sum_{l\neq i}^n h_{n,il} \mathrm{E} u_{n,l} + e_{n,i}' H_n X_n \beta_{0\tau},\ x_{n,i}')$. Thus, requiring $\partial G_\tau^0(\lambda, \beta)/\partial(\lambda, \beta')$ to have full column rank at $(\lambda_{0\tau}, \beta_{0\tau}')$ is equivalent to requiring $J_n(\tau)$ to have full column rank for large enough n, which in turn requires that $\xi_{n,i}$ be closely enough related to $\widetilde{x}_{n,i}$ and hence to $\bar{y}_{n,i}$ as the term $e_{n,i}' H_n X_n \beta_{0\tau}$ appears in the reduced-form expression for $\bar{y}_{n,i}$. Note that $J_n(\tau)$ is related to $-\partial \mathrm{E} G_{n\tau}^0(\lambda, \beta)/\partial(\lambda, \beta')$ evaluated at $(\lambda_{0\tau}, \beta_{0\tau}')$.

Noting that $G_\tau^0(\lambda_{0\tau}, \beta_{0\tau}) = 0$ by Assumption 1(ii), Assumption A4(iv) requires that $\theta_{0\tau} = (\lambda_{0\tau}, \beta_{0\tau}')'$ be the unique solution to $G_\tau^0(\lambda, \beta) = 0$. This assumption is needed for the consistency of our estimator. It is weaker than the condition that if $\mathrm{E}[G_{n\tau}^0(\lambda^*, \beta^*)] = 0$, then $\lambda^* = \lambda_{0\tau}$ and $\beta^* = \beta_{0\tau}$. The latter condition is usually satisfied in the extreme estimation for iid data or stationary time series data. See Hong and Tamer (2003) for detailed discussions on conditions under which quantile regression models with endogeneity are identified. In the study of spatial discrete-choice models, Pinkse and Slade (1998) made a similar assumption, and Pinkse, Slade and Shen (2006) assumed a slightly weaker condition.

Let $\alpha(\lambda, \tau) \equiv (\beta'(\lambda, \tau), \gamma'(\lambda, \tau))'$ be an arbitrary value of the parameter vector $(\beta', \gamma')'$ for a given λ and τ. Let $\Delta \equiv \Delta(\lambda, \tau) = \sqrt{n}(\alpha(\lambda, \tau) - \alpha_0(\lambda, \tau))$ such that $\|\Delta\| \leq M < \infty$ where the dependence of Δ on n is suppressed. Let $u_{n,i}(\lambda) = y_{n,i} - \lambda \bar{y}_{n,i} - \alpha_0(\lambda, \tau)' \xi_{n,i}$, and $u_{n,i}^*(\lambda, \Delta) = u_{n,i}(\lambda) + n^{-1/2} \Delta(\lambda, \tau)' \xi_{n,i} = y_{n,i} - \lambda \bar{y}_{n,i} - \alpha(\lambda, \tau)' \xi_{n,i}$. Let $\eta_{n,i}(\lambda) \equiv -(\psi_\tau(u_{n,i}^*(\lambda, \Delta)) - \psi_\tau(u_{n,i}(\lambda))) c_{ni}$, where $\{c_{ni}, i = 1, \cdots, n\}$ is an arbitrary bounded nonstochastic sequence and $\psi_\tau(u) = \tau - 1(u \leq 0)$ defined at (2.10). Define

$$\mathcal{S}_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{n,i}(\lambda).$$

Next, we state a high level assumption.

Assumption 5. $\mathrm{Var}(\mathcal{S}_n(\lambda)) = o(1)$ for each $\lambda \in \Lambda$.

Assumption 5 restricts the degree of dependence in the data. In the special case where $\lambda = \lambda_{0\tau}$, $\mathrm{Cov}(\eta_{n,i}(\lambda), \eta_{n,j}(\lambda)) = 0$ for all $i \neq j$ by Assumption 1(i) so that it is easy to verify $\mathrm{Var}(\mathcal{S}_n(\lambda_{0\tau})) = O(n^{-1}) = o(1)$. When λ deviates from $\lambda_{0\tau}$, we can verify that this assumption can be satisfied under different primitive conditions given below.

Assumption 5*. (i) The elements $w_{n,ij}$ of W_n are uniformly at most of order l_n^{-1}, denoted by $O(1/l_n)$, such that $l_n \to \infty$ and $l_n/n \to 0$ as n goes to infinity. (ii) $\mathrm{E}u_{n,i}^2 = \sigma_{n,i}^2$ with $\sup_n \max_{1\leq i\leq n} \sigma_{n,i}^2 \leq \overline{\sigma}^2 < \infty$.

Assumption 5*(i) requires that the elements $w_{n,ij}$ of W_n tend to zero uniformly as $n \to \infty$. This assumption is reasonable when each spatial unit is affected by an infinite number of neighbors such that the effect from any individual unit is negligible but the aggregate effect is not. Assumption 5*(ii) requires the existence of the second moments of $u_{n,i}$ which together with Assumption 5*(i) ensure that $\sum_{l\neq i}^n h_{n,il}(u_{n,l} - \mathrm{E}u_{n,l}) = o_p(1)$ for each $i = 1, \cdots, n$. We show in Appendix B that Assumption 5* together with Assumptions 1-3 are sufficient for Assumption 5.

Nevertheless, Assumption 5* rules out the case where l_n does not converge to infinity, which is very important in many applications when a spatial unit is only affected by a finite number of neighbors. Following Pinkse, Shen and Slade (2007), we can control the variance of $\mathcal{S}_n(\lambda)$ by borrowing the notion of "mixing" from the time series analysis. To proceed, we divide the observations into non-overlapping groups $\mathcal{G}_{n1}, \cdots, \mathcal{G}_{nJ}$, $1 \leq J < \infty$. For each $j = 1, \cdots, J$, there are m_{nj} mutually exclusive subgroups, $\mathcal{G}_{nj1}, \cdots, \mathcal{G}_{njm_{nj}}$. Group membership of each observation can vary with the sample size n and so can the number of subgroups m_{nj} in each group j. Let n_{jt} denote the number of observations in subgroup $\mathcal{G}_{njt}$. The following assumption is adapted from Pinkse, Shen and Slade (2007).

Assumption 5.** (i) For any $j = 1, \cdots, J$, let $\mathcal{G}_n^*$, $\mathcal{G}_n^* \subset \mathcal{G}_{nj}$ be any sets for which $\forall t = 1, \cdots, m_{nj}$, if $\mathcal{G}_{njt} \cap \mathcal{G}_n^* \neq \varnothing$ then $\mathcal{G}_{njt} \cap \mathcal{G}_n^{**} = \varnothing$. Let $\mathcal{S}_n^*(\lambda) = \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{G}_n^*} \eta_{n,s}(\lambda)$ and $\mathcal{S}_n^{**}(\lambda) = \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{G}_n^{**}} \eta_{n,s}(\lambda)$. Then for each $\lambda \in \Lambda$,

$$\|\mathrm{Cov}\left(\mathcal{S}_n^*(\lambda),\ \mathcal{S}_n^{**}(\lambda)\right)\| \leq \sqrt{\mathrm{Var}\left(\mathcal{S}_n^*(\lambda)\right) \mathrm{Var}\left(\mathcal{S}_n^{**}(\lambda)\right)}\alpha_{m_{nj}},$$

for some "mixing" numbers $\alpha_{m_{nj}}$ such that $\lim_{n\to\infty} \sum_{j=1}^J m_{nj}^2 \alpha_{m_{nj}} = c_\alpha \in [0, \infty)$. (ii) For each $j = 1, \cdots, J$, $\lim_{n\to\infty} \max_{t\leq m_{nj}} n_{jt}/n = 0$.

Assumption 5**(i) requires a bound on the correlation of two quantities, each corresponding to different sets of subgroups of the same group. It is weaker than Assumption A in Pinkse, Shen and Slade (2007). For a discussion on the need of dividing observations into finite J groups, see Pinkse, Shen and Slade (2007). Assumption 5**(ii) requires that the number of observations in each subgroup is relatively small. This is needed for controlling the variance of the partial sums over each subgroup. We show in Appendix B that Assumption 5** suffices to ensure Assumption 5.

Now, define

$$\begin{aligned} \upsilon_n(\lambda) &= -\sqrt{n}\left[G_{n\tau}(\lambda,\alpha_0(\lambda,\tau)) - \mathrm{E}G_{n\tau}(\lambda,\alpha_0(\lambda,\tau))\right] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1(y_{n,i}-\lambda\bar{y}_{n,i}\le \alpha'_{0\lambda\tau}\xi_{n,i}) - \mathrm{E}(1(y_{n,i}-\lambda\bar{y}_{n,i}\le \alpha'_{0\lambda\tau}\xi_{n,i}))\right]\xi_{n,i}v_{n,i}. \end{aligned}$$

We make the following assumption.

Assumption 6. (i) $\mathrm{E}G_{n\tau}(\lambda,\alpha_0(\lambda,\tau)) - G_\tau(\lambda,\alpha_0(\lambda,\tau)) = O(n^{-1/2})$ uniformly in λ. (ii) $\sup_{\lambda\in\Lambda}\|\upsilon_n(\lambda)\| = O_p(1)$ and $\sup_{\lambda\in\Lambda}\sup_{|\lambda-\lambda^*|<\delta_n}\|\upsilon_n(\lambda)-\upsilon_n(\lambda^*)\| = o_p(1)$ for every sequence $\{\delta_n\}$ converging to zero.

Assumption 6(i) specifies the rate at which $\mathrm{E}G_{n\tau}(\lambda,\alpha_0(\lambda,\tau))$ converges to its limit. If the convergence holds pointwise, we can show that it must hold uniformly in λ by using the properties of the indicator function. Assumption 6(i) is automatically satisfied for iid data and stationary time series data in which case $\mathrm{E}G_{n\tau}(\lambda,\alpha_0(\lambda,\tau)) = G_\tau(\lambda,\alpha_0(\lambda,\tau))$. Assumption 6(ii) is a stochastic equicontinuity condition. Let $\xi = (x', z')'$. Consider the class of functions

$$\mathcal{M} = \left\{g(y,\bar{y},\xi,v;\lambda) = 1(y-\lambda\bar{y}-\alpha'_{0\lambda\tau}\xi\le 0)\xi v : \lambda\in\Lambda\right\}.$$

If $(y_{n,i},\bar{y}_{n,i},\xi_{n,i},v_{n,i})$ are iid with probability law P_n, it is easy to verify that $\{g(\cdot;\lambda):\lambda\in\Lambda\}$ is a Euclidean class with envelope $\overline{g}$ such that $\overline{g}(y,\bar{y},\xi,v)\equiv\|\xi v\|$ and $\int\overline{g}(y,\bar{y},\xi,v)dP_n = \mathrm{E}\|\xi v\| < \infty$. Then by Lemma 2.17 of Pakes and Pollard (1989), Assumption 6 holds for iid data. It also holds for time series data under weak data dependence conditions (e.g., Andrews (1994) and Hansen (1996)). For spatial data, we can show that Assumption 6 holds provided $\lim_{n\to\infty} l_n/\sqrt{n} = c \in (0,\infty]$. This latter condition with $c=\infty$ has been assumed in Lee (2002a) for the consistency of least squares estimation of spatial autoregressive models and in Robinson (2007) for the adaptive estimation of spatial autoregressive models. Nevertheless, it is not necessary here because there may exist other cases where Assumption 6 holds.

3.2 Asymptotic Distribution

We study the asymptotic property of the IVQR estimators defined in (2.7)-(2.9) above. Under Assumptions 1-6, we first show that the IVQR estimator $\hat{\alpha}(\lambda,\tau)$ has a Bahadur representation uniformly in λ. To do so, recall $u_{n,i}(\lambda) = y_{n,i}-\lambda\bar{y}_{n,i}-\alpha'_{0\lambda\tau}\xi_{n,i}$. Define

$$S_n(\lambda,\tau) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_\tau(u_{n,i}(\lambda))\xi_{n,i}v_{n,i}.$$

The following theorem establishes the Bahadur representation for $\hat{\alpha}(\lambda,\tau)$.

Theorem 3.1 *Suppose Assumptions 1-6 hold. Then*

$$\sqrt{n}[\hat{\alpha}(\lambda,\tau)-\alpha_0(\lambda,\tau)] = J_{n\alpha}^{-1}(\lambda,\tau)S_n(\lambda,\tau)+o_p(1) \text{ uniformly in } \lambda\in\Lambda.$$

Note that $\sup_\lambda |S_n(\lambda,\tau)| = O_p(1)$ by Lemma A.4 and $\sup_\lambda |J_{n\alpha}(\lambda,\tau)| = O(1)$ by Assumptions 1-3 and Lemma A.1. An immediate consequence of Theorem 3.1 is that $\|\hat{\alpha}(\lambda,\tau)-\alpha_0(\lambda,\tau)\| = O_p(n^{-1/2})$ uniformly in $\lambda\in\Lambda$.

Let $\theta_0(\tau)\equiv(\lambda_0(\tau),\beta_0'(\tau))'$ and let $\hat{\theta}(\tau)\equiv(\hat{\lambda}(\tau),\hat{\beta}'(\tau))'$ be its IVQR estimator. To establish the asymptotic normality of $\hat{\theta}(\tau)$, define

$$\begin{aligned} J_\lambda &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_{n,i}(0)v_{n,i}\xi_{n,i}\left[\sum_{l\neq i}^n h_{n,il}\mathrm{E}u_{n,l}+e_{n,i}'H_nX_n\beta_{0\tau}\right], \text{ and} \\ J_\alpha &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_{n,i}(0)v_{n,i}\xi_{n,i}\xi_{n,i}', \end{aligned}$$

which are, respectively, $(p+q)\times 1$ and $(p+q)\times(p+q)$. Note that J_λ is the first column of $\lim_{n\to\infty} J_n(\tau)$. Partition conformably $J_\alpha^{-1} = [\overline{J}_\beta', \overline{J}_\gamma']'$, where $\overline{J}_\beta$ and $\overline{J}_\gamma$ are $p\times(p+q)$ and $q\times(p+q)$ matrices, respectively. Then we can establish the following theorem.

Theorem 3.2 *Suppose that J_α is of full rank and Assumptions 1-6 hold. Then*

$$\sqrt{n}[\hat{\theta}(\tau)-\theta_0(\tau)] \xrightarrow{d} N(0,\Omega(A)),$$

where $\Omega(A) = Q(A)S_0Q(A)'$, $S_0 = \tau(1-\tau)\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n v_{n,i}^2\xi_{n,i}\xi_{n,i}'$, and

$$Q(A)' = \left\{\ \overline{J}_\gamma' A\overline{J}_\gamma J_\lambda(J_\lambda'\overline{J}_\gamma' A\overline{J}_\gamma J_\lambda)^{-1},\quad (I_{p+q}-J_\lambda(J_\lambda'\overline{J}_\gamma' A\overline{J}_\gamma J_\lambda)^{-1}J_\lambda'\overline{J}_\gamma' A\overline{J}_\gamma)'\overline{J}_\beta'\ \right\}.$$

The formula for the asymptotic covariance matrix of $\sqrt{n}[\hat{\theta}(\tau)-\theta_0(\tau)]$ looks complicated and it depends on the choice of the weight matrix A. In the case of just identification ($q=1$), we show in the following corollary that the choice of A does not affect the asymptotic variance of $\sqrt{n}[\hat{\theta}(\tau)-\theta_0(\tau)]$.

Corollary 3.3 *Suppose that $q=1$ and the conditions of Theorem 3.2 hold. Then*

$$\sqrt{n}[\hat{\theta}(\tau)-\theta_0(\tau)] \xrightarrow{d} N(0,\Omega_0),$$

where $\Omega_0 = J_0^{-1}S_0(J_0^{-1})'$ and $J_0 = \lim_{n\to\infty} J_n(\tau)$ with $J_n(\tau)$ being defined in (3.8).

Remark 3. In the case of over-identification ($q>1$), the choice of the weight matrix $\widehat{A}$ in the objective function $\|\hat{\gamma}(\lambda,\tau)\|_{\widehat{A}}$ generally matters. It is natural to choose $\widehat{A}$ to be a consistent estimate of the inverse of the asymptotic covariance matrix of $\sqrt{n}\hat{\gamma}(\lambda,\tau)$. In this case, A is also λ-dependent and needs to be estimated at each grid point of λ in the process of optimization.

Remark 4. Consider the method of moments estimator $\widetilde{\theta}(\tau) = \arg\min_\theta \|G_{n\tau}(\theta)\|_{\widehat{A}}$ defined in (2.10). In a separate study we have established the asymptotic normality of $\widetilde{\theta}(\tau)$ under conditions similar to those imposed in Assumptions 1-6. In particular, when we choose the optimal weight $\widehat{A} = n(\sum_{i=1}^n v_{n,i}^2 \xi_{n,i}\xi_{n,i}')^{-1}$, the asymptotic covariance of $\widetilde{\theta}(\tau)$ is equal to Ω_0. This is true regardless of the dimension of the instruments $z_{n,i}$. In other words, with the optimal choice of $\widehat{A}$ in the definition of $\widetilde{\theta}(\tau)$, it is asymptotically equivalent to $\hat{\theta}(\tau)$ in case of just identification ($q = 1$).

3.3 Estimation of VC Matrix

For statistical inferences based on our model, we need to provide a method of estimating the asymptotic variance-covariance (VC) matrix $\Omega(A)$ which depends on S_0, J_λ and J_α. Since $S_n \equiv \tau(1-\tau)\frac{1}{n}\sum_{i=1}^n v_{n,i}^2 \xi_{n,i}\xi_{n,i}'$ consistently estimates S_0, we focus on the consistent estimation of J_λ and J_α. Note that J_λ and J_α depend on the unknown densities $f_{n,i}(0)$. To estimate these quantities, we could either make some distributional assumption or use some nonparametric estimation technique as in Powell (1986) and Koenker (1994). Nevertheless, either approach will complicate the matter to a great deal. As pointed out by Pakes and Pollard (1989) and used by Honoré and Hu (2004), the derivation of the asymptotic normality implies that J_λ and J_α can be estimated by "numerical derivatives ". Recall that J_λ is the first column of $\lim_{n\to\infty} J_n(\tau) = -\partial G_\tau^0(\lambda,\beta)/\partial(\lambda,\beta')$ evaluated at $\lambda = \lambda_{0\tau}$ and $\beta = \beta_{0\tau}$ (see Eq. (3.7)), we can estimate J_λ by

$$J_{n\lambda} = \frac{G_{n\tau}^0\left(\hat{\theta}(\tau) + a_n^{(0)} e_{p+1,1}\right) - G_{n\tau}^0\left(\hat{\theta}(\tau) - a_n^{(0)} e_{p+1,1}\right)}{-2a_n^{(0)}}$$

where $\{a_n^{(0)}\}$ is a sequence of "bandwidths". Similarly, note that J_α is the limit of the derivative of $-\mathrm{E}[G_{n\tau}(\lambda;\alpha)]$ with respect to $\alpha = (\beta',\gamma')'$ evaluated at $(\lambda_{0\tau};\alpha_{0\tau})$, hence we can estimate the jth column of J_α by

$$J_{n\alpha}^{(j)} = \frac{G_{n\tau}\left(\hat{\lambda}(\tau);\hat{\alpha}(\tau) + a_n^{(j)} e_{p+q,j}\right) - G_{n\tau}\left(\hat{\lambda}(\tau);\hat{\alpha}(\tau) - a_n^{(j)} e_{p+q,j}\right)}{-2a_n^{(j)}}$$

where $\{a_n^{(j)}\}$ is a sequence of "bandwidths". These estimates are consistent provided $a_n^{(j)} = o_p(1)$ and $(\sqrt{n}a_n^{(j)})^{-1} = o_p(1)$. For example, we can take $a_n^{(j)} = n^{-\alpha}$ for some $0 < \alpha < 0.5, j = 0, 1, \cdots, p+q$.

Finally, to apply the results of Corollary 3.3 for inferences, one needs to estimate J_0. The first column of it is estimated by $J_{n\lambda}$ given above. The other columns can be estimated in a similar way by replacing $e_{p+1,1}$ by $e_{p+1,i}, i = 2, \cdots, p+1$, in the definition of $J_{n\tau}$.

4 Monte Carlo Simulations

In this section we report some results from a set of Monte Carlo experiments for the finite sample performance of our IVQR estimator of the SAR model. Also, in the special case of median regression with symmetric errors, we compare our estimator with the QMLE without taking into account of heteroscedasticity (Lee, 2004), the 2SLS and GMM estimators of Lee (2007) with iid assumption, and the robust GMM estimator of Lin and Lee (2006). The GMM estimator of Lee (2007) denoted by GMM0 and the robust GMM estimator of Lin and Lee (2006) denoted by GMMR require initial estimates of λ and β and a weighting matrix. We follow Lin and Lee (2006) and choose 2SLS estimates as initial estimates and the optimal weighting matrix for GMM under iid setting as the weighting matrix for both GMM0 and GMMR. Note that the estimator proposed by Kelejian and Prucha (2006) is essentially the 2SLS estimator when their SARAR model is reduced to SAR model. The data generating process (DGP) we employed in the Monte Carlo experiments is given below

$$Y_n = \lambda_{0\tau} W_n Y_n + X_n \beta_{0\tau} + u_n,$$

where X_n contains two columns: the first column X_{n0} is just the column of ones, and each value in the second column X_{n1} is the sum of 48 independent `Uniform(-0.25, 0.25)` random numbers. The heteroscedasticity in u_n is determined by the absolute function $|X_{n1}|$, i.e., $u_n = |X_{n1}|e_n$, where e_n is a vector of iid random variates subtracted by their τth quantiles. Three distributions for e_n are considered in the experiment: (i) standard normal, (ii) a student t distribution with two degrees of freedom, and (iii) a chi-squared distribution with two degrees of freedom. The true parameter values are taken to be $\beta_0 = \beta_1 = 1$, and $\lambda = 0.5$. The sample sizes used are 100, 200, 500, and 1000. Each set of simulation results is based 1,000 Monte Carlo samples.

The weight matrix W is generated under the two scenarios: (i) Rook contiguity, and (ii) large group interaction. The former corresponds to the case where l_n is bounded, whereas the latter corresponds to the case where l_n goes to infinity as n does but in a slower rate. To be exact, in case (i), first randomly generate n integers from 1 to n without repetition and arrange them in five rows, then form the neighborhood matrix according to the Rook contiguity and row-normalize; in case (ii) we choose the number of groups $R = n^{0.6}$, and then generate the group sizes $(m_r, r = 1, \cdots, R)$ uniformly from the interval $(m/2, 3m/2)$ where $m(\approx n/R)$ is the average group size.[2]

Note that when comparing our method with the existing QML and GMM methods, we need to restrict ourselves to the case of median with error distribution being symmetric. This is because the QML and GMM methods are applicable to the standard SAR model which is subject to zero mean (in errors) restriction. Also note that, in finding the IVQR estimate, we used the grid search method

[2]We need to make a final adjustment to make sure that $\sum_{r=1}^{R} m_r = n$. See Lin and Lee (2206) for discussions on group interactions.

(as indicated in the Remark 1 of Section 2.2) combined with an auto search. This is because a fine grid search alone may be too time consuming, and an auto search alone may lead to local minima.[3]

Table 1(a, b) summarizes the Monte Carlo bias and the root mean squared errors (RMSE) for the case of $\tau = 0.5$, where Table 1a corresponds to Rook spatial contiguity and Table 1b corresponds to large group interaction. From the results we see that the IVQR estimator (IVQRE) outperforms all other estimators, especially in the cases of nonnormal errors. Both bias and RMSE go down as the sample size goes up. In the case of normal errors (DGP1), GMM0 and GMMR estimators are comparable with IVQR estimator, but clearly not under nonnormal symmetric errors (DGP2). From Table 1a we see that QMLE behaves reasonably well in the cases of normal or t errors. This shows that as far as finite sample performance is concerned, the QMLE is not affected very much by heteroscedasticity if the spatial dependence is limited to a few neighbors. However, the results from Table 1b show that when a spatial unit depends on many others, the QMLE without taking into account of heteroscedasticity can behave quite badly in the sense of giving a large bias. The 2SLS estimator often behaves quite badly except in the case of normal errors. The effect of having more spatial neighbors is an increased variability of the estimators for λ and β_0. This seems to be true for all estimators.

It should be noted that the results given in Table 1 under DGP3 are not comparable for the estimators of the intercept parameter β_0 as the errors are generated with a zero median which is required by IVQRE, whereas the other estimators require a zero mean for the errors. Nevertheless, these results show the robustness of IVQR estimator against both excess skewness and kurtosis. Moreover, as expected, even if the errors possess zero median instead of zero mean, the GMM0 and GMMR still give very good estimates for the spatial parameter λ and the slope parameter β_1 but not the intercept parameter β_0.

Table 2 presents some further Monte Carlo results for the cases of $\tau = 0.25$ and $\tau = 0.75$. As there is no direct comparison between our IVQR estimator with the others in these cases, we only report the results corresponding to the IVQR estimator. The results indicate that the new IVQR estimator for the SAR model behaves quite well in general, and are consistent with the theoretical predictions. First, it is generally robust against nonnormality and heteroscedasticity. The IVQR estimator of the intercept has non-negligible bias for DGPs 2-3 when the sample size n is as small as 100, but the bias diminishes fast as the sample size doubles. Second, as the sample size increases, both bias and RMSE decline and the magnitude of decrease in the RMSE and standard error (not reported in the table) is generally consistent with the $\sqrt{n}$-asymptotics.

[3]We first find the interval where the global minimum lies in by the grid search method, and then do an auto search within this smaller interval. In our simulation, we have used 200 points within $[-0.99, 0.99]$.

Table 1a. Monte Carlo Bias and RMSE: $\tau = 0.5$, Rook Contiguity

			$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
DGP	Est	Par	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	QML	λ	-0.0298	0.0893	-0.0160	0.0627	-0.0034	0.0387	-0.0029	0.0283
		β_0	0.0680	0.2116	0.0332	0.1396	0.0093	0.0907	0.0075	0.0666
		β_1	0.0133	0.1823	0.0066	0.1225	0.0044	0.0800	0.0013	0.0557
	2SLS	λ	-0.0104	0.1398	-0.0024	0.1062	-0.0022	0.0731	0.0001	0.0535
		β_0	0.0223	0.3033	0.0058	0.2151	0.0067	0.1579	0.0013	0.1147
		β_1	0.0032	0.1831	-0.0001	0.1273	0.0029	0.0800	0.0001	0.0569
	GMM0	λ	-0.0272	0.0886	-0.0134	0.0622	-0.0017	0.0392	-0.0022	0.0279
		β_0	0.0626	0.2098	0.0279	0.1384	0.0057	0.0914	0.0060	0.0658
		β_1	0.0086	0.1813	0.0025	0.1221	0.0031	0.0800	0.0005	0.0556
	GMMR	λ	-0.0256	0.0879	-0.0145	0.0625	-0.0021	0.0392	-0.0021	0.0279
		β_0	0.0590	0.2082	0.0300	0.1390	0.0066	0.0914	0.0058	0.0658
		β_1	0.0082	0.1813	0.0028	0.1222	0.0031	0.0800	0.0005	0.0556
	IVQR	λ	-0.0019	0.0677	-0.0001	0.0554	-0.0009	0.0343	-0.0005	0.0191
		β_0	0.0048	0.1398	0.0007	0.1063	0.0023	0.0721	0.0018	0.0428
		β_1	-0.0024	0.1632	0.0038	0.1120	0.0033	0.0720	0.0006	0.0503
2	QML	λ	-0.0366	0.0980	-0.0168	0.0707	-0.0085	0.0430	-0.0056	0.0315
		β_0	0.0748	0.3769	0.0327	0.2715	0.0220	0.2188	0.0131	0.1261
		β_1	-0.0327	0.5918	0.0140	0.3848	0.0100	0.3983	0.0079	0.1746
	2SLS	λ	-0.0463	0.3999	-0.0481	0.3462	-0.0150	0.2537	-0.0151	0.1821
		β_0	-0.1387	4.6768	0.0423	1.4037	-0.0028	1.3501	0.0323	0.6679
		β_1	-0.0489	0.6536	-0.0112	0.3984	-0.0038	0.3796	-0.0034	0.1869
	GMM0	λ	-0.0139	0.2728	-0.0086	0.1934	-0.0022	0.0934	-0.0015	0.0668
		β_0	0.0250	0.6474	0.0149	0.4912	0.0126	0.2732	0.0033	0.1846
		β_1	-0.0418	0.5696	0.0090	0.3868	0.0062	0.4152	0.0061	0.1758
	GMMR	λ	-0.0132	0.2712	-0.0090	0.1915	-0.0013	0.1102	-0.0012	0.0661
		β_0	0.0235	0.6415	0.0151	0.4930	0.0108	0.2971	0.0028	0.1837
		β_1	-0.0421	0.5694	0.0092	0.3856	0.0062	0.4143	0.0060	0.1761
	IVQR	λ	-0.0186	0.1652	-0.0050	0.1015	-0.0025	0.0513	-0.0011	0.0267
		β_0	0.0346	0.3467	0.0073	0.1836	0.0044	0.1098	0.0028	0.0587
		β_1	-0.0078	0.2056	0.0001	0.1345	0.0017	0.0800	-0.0014	0.0553
3	QML	λ	-0.0358	0.0951	-0.0052	0.0676	-0.0066	0.0446	-0.0041	0.0317
		β_0	0.5189	0.6140	0.4910	0.5474	0.4925	0.5197	0.5026	0.5159
		β_1	0.1291	0.3953	-0.0043	0.2408	0.0301	0.1591	-0.0203	0.1120
	2SLS	λ	0.0483	0.3075	-0.0001	0.2443	-0.0031	0.1560	-0.0138	0.1124
		β_0	0.1082	1.8853	0.5046	1.2536	0.4828	0.6872	0.5325	0.6358
		β_1	0.0665	0.4169	-0.0251	0.2590	0.0223	0.1604	-0.0227	0.1154
	GMM0	λ	-0.0058	0.2176	0.0020	0.1504	-0.0053	0.0451	-0.0027	0.0322
		β_0	0.4193	0.7929	0.4701	0.6841	0.4887	0.5164	0.4985	0.5123
		β_1	0.1131	0.3995	-0.0127	0.2472	0.0285	0.1586	-0.0214	0.1122
	GMMR	λ	-0.0036	0.2186	0.0001	0.1398	-0.0059	0.0450	-0.0025	0.0323
		β_0	0.4115	0.7915	0.4766	0.6622	0.4904	0.5179	0.4980	0.5118
		β_1	0.1122	0.3991	-0.0120	0.2456	0.0286	0.1586	-0.0214	0.1122
	IVQR	λ	-0.0136	0.1566	-0.0021	0.1044	-0.0052	0.0652	-0.0013	0.0391
		β_0	0.0617	0.4925	0.0216	0.2935	0.0244	0.1952	0.0079	0.1195
		β_1	0.0023	0.2634	-0.0071	0.1838	-0.0072	0.1125	-0.0031	0.0817

Table 1b. Monte Carlo Bias and RMSE: $\tau = 0.5$, Large Group Interaction

			$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
DGP	Est	Par	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	QML	λ	-0.4602	0.4603	-0.4781	0.4782	-0.5060	0.5060	-0.5074	0.5075
		β_0	-1.2215	1.2221	-1.2027	1.2031	-1.1622	1.1623	-1.1133	1.1135
		β_1	-0.2978	0.3233	-0.1596	0.1846	-0.1034	0.1234	-0.0904	0.1040
	2SLS	λ	0.0260	0.1217	0.0235	0.1189	0.0140	0.1057	0.0114	0.0787
		β_0	0.0761	0.3856	0.0628	0.3357	0.0359	0.2694	0.0231	0.1747
		β_1	0.0043	0.1735	0.0038	0.1116	0.0025	0.0768	-0.0027	0.0567
	GMM0	λ	0.0134	0.0637	0.0165	0.0564	0.0069	0.0465	0.0081	0.0381
		β_0	0.0345	0.2001	0.0422	0.1704	0.0163	0.1155	0.0160	0.0902
		β_1	-0.0038	0.1652	0.0024	0.1077	0.0015	0.0740	-0.0030	0.0564
	GMMR	λ	0.0132	0.0638	0.0112	0.0540	0.0038	0.0459	0.0058	0.0371
		β_0	0.0345	0.2006	0.0289	0.1649	0.0094	0.1141	0.0108	0.0882
		β_1	-0.0036	0.1653	0.0010	0.1078	0.0009	0.0739	-0.0034	0.0564
	IVQR	λ	0.0102	0.0745	0.0083	0.0592	0.0055	0.0511	0.0030	0.0278
		β_0	0.0257	0.2026	0.0211	0.1659	0.0127	0.1169	0.0059	0.0602
		β_1	-0.0050	0.1525	0.0008	0.1101	0.0026	0.0710	-0.0012	0.0518
2	QML	λ	-0.4814	0.4815	-0.4933	0.4933	-0.5068	0.5068	-0.5128	0.5131
		β_0	-1.2791	1.3043	-1.2419	1.2441	-1.1634	1.1638	-1.1261	1.1269
		β_1	-0.3152	0.6002	-0.1831	0.3644	-0.0992	0.2440	-0.0855	0.2129
	2SLS	λ	-0.0229	0.3687	-0.0419	0.3812	-0.0033	0.3329	0.0350	0.2940
		β_0	4.6283	165.508	0.1462	8.5274	-0.2085	18.7344	0.2805	2.0753
		β_1	0.1852	8.7988	-0.0110	0.6028	-0.0053	0.9323	0.0133	0.2323
	GMM0	λ	-0.0340	0.2350	-0.0373	0.2025	-0.0150	0.1639	-0.0090	0.1306
		β_0	-0.0947	0.7551	-0.0967	0.5643	-0.0322	0.3996	-0.0150	0.3219
		β_1	-0.0154	0.6763	-0.0232	0.3662	0.0072	0.2448	0.0020	0.2174
	GMMR	λ	-0.0374	0.2371	-0.0495	0.1977	-0.0235	0.1559	-0.0149	0.1236
		β_0	-0.0952	0.7886	-0.1281	0.5496	-0.0526	0.3820	-0.0277	0.2986
		β_1	-0.0183	0.7095	-0.0268	0.3616	0.0050	0.2450	0.0009	0.2166
	IVQR	λ	0.0083	0.1588	0.0124	0.1177	0.0106	0.0990	0.0090	0.0617
		β_0	0.0176	0.4139	0.0399	0.3076	0.0245	0.2269	0.0187	0.1343
		β_1	0.0029	0.1987	-0.0030	0.1282	-0.0007	0.0816	0.0007	0.0570
3	QML	λ	-0.4779	0.4780	-0.4923	0.4924	-0.5064	0.5064	-0.5115	0.5118
		β_0	-1.4550	1.4580	-1.3757	1.3765	-1.2647	1.2649	-1.2036	1.2048
		β_1	-0.3129	0.4074	-0.1718	0.2556	-0.1378	0.1918	-0.0986	0.1402
	2SLS	λ	0.0236	0.2372	0.0333	0.3897	0.0079	0.2250	0.0569	0.1869
		β_0	7.3422	161.158	0.4267	22.0482	0.7478	4.8682	0.8059	1.4444
		β_1	0.1986	4.0036	-0.0097	0.8859	-0.0289	0.2003	-0.0079	0.1122
	GMM0	λ	0.0044	0.1562	0.0031	0.1866	0.0044	0.0872	0.0117	0.0459
		β_0	0.5732	0.9257	0.5328	0.9227	0.5175	0.6070	0.5368	0.5614
		β_1	0.0053	0.3426	-0.0241	0.2239	-0.0331	0.1497	-0.0142	0.1085
	GMMR	λ	-0.0034	0.1562	-0.0088	0.1862	-0.0008	0.0837	0.0080	0.0424
		β_0	0.5392	0.8997	0.4870	0.8925	0.4996	0.5855	0.5243	0.5463
		β_1	-0.0006	0.3401	-0.0282	0.2236	-0.0342	0.1500	-0.0148	0.1086
	IVQR	λ	0.0094	0.1821	-0.0016	0.2021	0.0136	0.1143	0.0171	0.0788
		β_0	0.0747	0.7230	0.0161	0.7419	0.0526	0.3727	0.0604	0.2617
		β_1	0.0164	0.2551	-0.0049	0.1813	-0.0042	0.1163	0.0024	0.0780

Table 2. More Monte Carlo Results for IVQR Estimator

		$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
DGP	Par	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
		$\tau = 0.25$, Rook Contiguity							
1	λ	-0.0002	0.1684	-0.0086	0.0499	-0.0018	0.0427	0.0012	0.0215
	β_0	-0.0183	0.4887	0.0183	0.1521	-0.0015	0.1221	-0.0066	0.0664
	β_1	-0.0058	0.1804	-0.0038	0.1177	-0.0009	0.0724	0.0015	0.0554
2	λ	-0.0191	0.3064	-0.0142	0.1295	-0.0032	0.0922	0.0007	0.0364
	β_0	-0.0227	0.9293	-0.0038	0.3713	-0.0129	0.2886	-0.0136	0.1200
	β_1	-0.0078	0.2573	-0.0104	0.1751	-0.0078	0.1082	-0.0054	0.0738
3	λ	-0.0393	0.2937	-0.0104	0.0767	-0.0013	0.0565	0.0019	0.0202
	β_0	0.1368	1.1144	0.0374	0.3160	0.0027	0.2254	-0.0092	0.0833
	β_1	-0.0185	0.1589	-0.0038	0.0980	-0.0042	0.0637	-0.0017	0.0469
		$\tau = 0.75$, Rook Contiguity							
1	λ	-0.0101	0.1543	0.0009	0.0695	0.0038	0.0371	0.0001	0.0207
	β_0	0.0235	0.1784	0.0097	0.0848	-0.0003	0.0431	0.0032	0.0237
	β_1	-0.0043	0.1701	0.0012	0.1272	-0.0021	0.0803	0.0001	0.0567
2	λ	-0.0554	0.2919	-0.0007	0.1503	0.0037	0.0620	-0.0037	0.0357
	β_0	0.1380	0.4204	0.0525	0.2121	0.0157	0.0646	0.0136	0.0403
	β_1	0.0020	0.2600	-0.0129	0.1885	0.0041	0.1098	-0.0025	0.0772
3	λ	-0.0753	0.3787	-0.0260	0.2605	0.0025	0.1172	-0.0059	0.0606
	β_0	0.1862	0.5380	0.1211	0.3783	0.0336	0.1397	0.0245	0.0779
	β_1	-0.0135	0.4292	-0.0132	0.3212	0.0083	0.1959	0.0017	0.1392
		$\tau = 0.25$, Large Group Interaction							
1	λ	0.0099	0.0947	0.0081	0.0777	0.0059	0.0518	0.0030	0.0265
	β_0	0.0196	0.3550	0.0210	0.2896	0.0152	0.1782	0.0070	0.0922
	β_1	0.0053	0.1706	-0.0004	0.1225	0.0021	0.0787	0.0017	0.0546
2	λ	-0.0219	0.2375	0.0062	0.1625	0.0110	0.1398	0.0105	0.0635
	β_0	-0.1950	1.1515	-0.0269	0.7181	0.0066	0.5370	0.0248	0.2218
	β_1	-0.0152	0.2653	-0.0041	0.1753	-0.0069	0.1076	0.0108	0.0761
3	λ	-0.0233	0.1997	-0.0027	0.1265	-0.0008	0.1161	0.0019	0.0291
	β_0	-0.1421	1.1568	-0.0216	0.7082	-0.0069	0.5827	0.0078	0.1378
	β_1	-0.0199	0.1752	-0.0016	0.1113	0.0016	0.0680	0.0001	0.0459
		$\tau = 0.75$, Large Group Interaction							
1	λ	0.0238	0.0987	0.0062	0.0607	0.0028	0.0317	0.0019	0.0267
	β_0	0.0547	0.1941	0.0207	0.1220	0.0082	0.0471	0.0043	0.0351
	β_1	0.0087	0.1815	-0.0071	0.1284	-0.0065	0.0767	0.0020	0.0548
2	λ	0.0180	0.1785	0.0012	0.1571	0.0046	0.1219	0.0088	0.0652
	β_0	0.1041	0.2934	0.0655	0.2509	0.0344	0.1337	0.0227	0.0821
	β_1	0.0086	0.2593	-0.0070	0.1925	-0.0050	0.1176	-0.0040	0.0750
3	λ	0.0214	0.2128	0.0228	0.1600	0.0175	0.1228	0.0188	0.0978
	β_0	0.1330	0.3700	0.1079	0.3225	0.0511	0.1722	0.0398	0.1348
	β_1	0.0146	0.4484	0.0082	0.3355	-0.0059	0.1923	-0.0129	0.1409

5 Concluding Remarks

We proposed a spatial autoregressive (SAR) model under quantile restrictions, and an instrumental variable quantile regression (IVQR) method for the model estimation. Large sample properties of the IVQR estimator for the SAR model under quantile restrictions are examined. Monte Carlo evidence is provided for the good finite sample performance of the IVQR estimator. In the special case of median restriction with symmetric error distributions, the IVQR estimator compares favorably against the existing GMM estimators with or without taking into account of the heteroscedasticity. Furthermore, the IVQR method is less demanding on the moments of the error and is quite robust against nonnormality and heteroscedasticity of the errors.

The new model and estimation method give important extensions to both the standard spatial regression models and the standard quantile regression models. These extensions should be very useful for the applied researchers.

Appendix

A Proof of the Main Results

Let C signify a generic constant whose exact value may vary from case to case. Let B_{1n} and B_{2n} be two $n \times n$ matrices that are uniformly bounded in both row and column sums. Let B_{3n} be a conformable matrix whose elements are uniformly $O(a_n)$ for a certain sequence a_n. Frequently we will use the following two evident facts (see, e.g., Kelejian and Prucha, 1999; Lee, 2002a, 2002b):

Fact 1: $B_{1n}B_{2n}$ is also uniformly bounded in both row and column sums.
Fact 2: The elements of $B_{1n}B_{3n}$ and $B_{3n}B_{1n}$ are uniformly $O(a_n)$.

Noting that both W and S_n^{-1} are uniformly bounded in both row and column sums under our assumption. It is easy to apply the above facts to prove the following lemma.

Lemma A.1 *1) $H_n \equiv W_n S_n^{-1}$ is uniformly bounded in both row and column sums. 2) The elements $h_{n,ij}$ of H_n are uniformly $O(1/l_n)$, where the notation l_n is defined in Assumption 5*.*

Proof. 1) follows straightforwardly from Fact 1 and Assumption 2(iii) which states that W_n and S_n^{-1} are uniformly bounded in both row and column sums. 2) follows from Fact 2 and Assumptions 2(iii) and 5*. ■

A.1 Proof of Theorem 3.1

Recall $\alpha_{0\tau} = (\beta_{0\tau}', 0')'$ introduced after Assumption 4, $\alpha_0(\lambda,\tau) = (\beta_0(\lambda,\tau)', \gamma_0(\lambda,\tau)')'$ defined below (3.1), $\alpha(\lambda,\tau) = (\beta(\lambda,\tau)', \gamma(\lambda,\tau)')'$ introduced before Assumption 5, and $\hat{\alpha}(\lambda,\tau) = (\hat{\beta}(\lambda,\tau)', \hat{\gamma}(\lambda,\tau)')'$ defined by (2.7). We frequently write $\alpha_0(\lambda,\tau)$ as $\alpha_{0\lambda\tau}$. Recall $\Delta \equiv \Delta(\lambda,\tau) = \sqrt{n}(\alpha(\lambda,\tau) - \alpha_0(\lambda,\tau))$, and let

$$\widehat{\Delta}(\lambda,\tau) = \sqrt{n}(\hat{\alpha}(\lambda,\tau) - \alpha_0(\lambda,\tau)).$$

Recall further $u_{n,i}(\lambda) = y_{n,i} - \lambda \bar{y}_{n,i} - \alpha_{0\lambda\tau}'\xi_{n,i}$ and $u_{n,i}^*(\lambda, \Delta(\lambda,\tau)) = u_{n,i}(\lambda) - n^{-1/2}\Delta(\lambda,\tau)'\xi_{n,i} = y_{n,i} - \lambda\bar{y}_{n,i} - \alpha(\lambda,\tau)'\xi_{n,i}$, both defined right above Assumption 5. It follows from Step (iii) leading to (2.9) that

$$\widehat{\Delta}(\lambda,\tau) = \arg \min_{\Delta \in \mathbb{R}^{p+q}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(u_{n,i}^*(\lambda,\Delta))v_{n,i}. \tag{A.1}$$

Set

$$V_n(\tau,\lambda;\Delta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\tau(u_{n,i}^*(\lambda,\Delta))\xi_{n,i}v_{n,i},$$

which is $\sqrt{n}G_{n\tau}(\lambda,\beta,\gamma)$, written in terms of Δ. Noting that $-\Delta' V_n(\tau,\lambda;\kappa\Delta)$ is an increasing function of $\kappa \geq 1$, the result of Theorem 3.1 then follows from the following three lemmas, according to Lemma A.4 of Koenker and Zhao (1996).

Lemma A.2 *Suppose Assumptions 1-6 hold. Then, with $M < \infty$,*

$$\sup_{\lambda \in \Lambda} \sup_{\|\Delta\| \leq M} \| V_n(\tau, \lambda; \Delta) - V_n(\tau, \lambda; 0) - E[V_n(\tau, \lambda; \Delta) - V_n(\tau, \lambda; 0)] \| = o_p(1).$$

Proof. We first establish a pointwise convergence result. Let

$$\begin{aligned} S_n(\tau, \lambda; \Delta) &= -\{V_n(\tau, \lambda; \Delta) - V_n(\tau, \lambda; 0) - \mathrm{E}[V_n(\tau, \lambda; \Delta) - V_n(\tau, \lambda; 0)]\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \widetilde{s}_{ni}(\tau, \lambda; \Delta), \end{aligned} \tag{A.2}$$

where $s_{ni}(\tau, \lambda; \Delta) = [1(y_{n,i} - \lambda \bar{y}_{n,i} \leq (\alpha_{0\lambda\tau} + n^{-1/2}\Delta)'\xi_{n,i}) - 1(y_{n,i} - \lambda \bar{y}_{n,i} \leq \alpha'_{0\lambda\tau}\xi_{n,i})]\xi_{n,i}v_{n,i}$, and $\widetilde{s}_{ni}(\tau, \lambda; \Delta) = s_{ni}(\tau, \lambda; \Delta) - \mathrm{E}[s_{ni}(\tau, \lambda; \Delta)]$. We need to show

$$\|S_n(\tau, \lambda; \Delta)\| = o_p(1), \text{ for each fixed } \lambda \text{ and } \Delta, \tag{A.3}$$

which holds if

$$S_{nk}(\tau, \lambda; \Delta) = o_p(1), \text{ for each fixed } \lambda,\ \Delta, \text{ and } k = 1, \cdots, p+q, \tag{A.4}$$

where S_{nk} is the kth component of the $(p+q) \times 1$ vector $S_n(\tau, \lambda; \Delta)$. By construction, $\mathrm{E}S_{nk}(\tau, \lambda; \Delta) = 0$. By Assumption 5, $\mathrm{Var}(S_{nk}(\tau, \lambda; \Delta)) = o(1)$. Thus (A.4) holds by Chebyshev inequality. Define

$$a_{ni}(\lambda, \Delta) = (\lambda - \lambda_{0\tau}) \sum_{l \neq i}^{n} h_{n,il}u_{n,l} + (\lambda - \lambda_{0\tau})e'_{n,i}H_nX_n\beta_{0\tau} + (\alpha_{0\lambda\tau} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i}, \tag{A.5}$$

Clearly, $\Delta = 0$ leads to $a_{ni}(\lambda, 0) = a_{ni}(\lambda)$ defined in (3.6). When we plug $\Delta = \sqrt{n}(\alpha(\lambda, \tau) - \alpha_{0\lambda\tau})$ into (A.5), $a_{ni}(\lambda, \Delta)$ becomes $\overline{a}_{ni}(\lambda, \alpha(\lambda, \tau))$, where

$$\overline{a}_{ni}(\lambda, \alpha) = (\lambda - \lambda_{0\tau}) \sum_{l \neq i}^{n} h_{n,il}u_{n,l} + (\lambda - \lambda_{0\tau})e'_{n,i}H_nX_n\beta_{0\tau} + (\alpha - \alpha_{0\tau})'\xi_{n,i} \tag{A.6}$$

Recall $b_{ni}(\lambda) = 1 - ((\lambda - \lambda_{0\tau})h_{n,ii}$. Then

$$\begin{aligned} & y_{n,i} - \lambda \bar{y}_{n,i} - (\alpha_{0\lambda\tau} + n^{-1/2}\Delta)'\xi_{n,i} \\ = \; & u_{n,i} - (\lambda - \lambda_{0\tau})\bar{y}_{n,i} - (\alpha_{0\lambda\tau} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i} \\ = \; & (1 - ((\lambda - \lambda_{0\tau})h_{n,ii})u_{n,i} - (\lambda - \lambda_{0\tau}) \sum_{l \neq i}^{n} h_{n,il}u_{n,l} \\ & - \left\{ (\lambda - \lambda_{0\tau})e'_{n,i}H_nX_n\beta_{0\tau} + (\alpha_{0\lambda\tau} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i} \right\} \\ = \; & b_{ni}(\lambda)u_{n,i} - a_{ni}(\lambda, \Delta), \end{aligned}$$

so that

$$1\{y_{n,i} - \lambda \bar{y}_{n,i} \leq (\alpha_{0\lambda\tau} + n^{-1/2}\Delta)'\xi_{n,i}\} = 1\{b_{ni}(\lambda)u_{n,i} \leq a_{ni}(\lambda, \Delta)\}. \tag{A.7}$$

We next show that (A.3) holds uniformly over $(\lambda, \Delta) \in \Lambda \times \Gamma$, where $\Gamma \equiv \{\Delta : \|\Delta \leq M\|\}$, and $M \in (0, \infty)$. This will hold by the triangle inequality provided

$$\sup_{\lambda \in \Lambda} \sup_{\|\Delta\| \leq M} \left| S_{nk}^{+}(\tau, \lambda; \Delta) \right| = o_p(1), \text{ and } \sup_{\lambda \in \Lambda} \sup_{\|\Delta\| \leq M} \left| S_{nk}^{-}(\tau, \lambda; \Delta) \right| = o_p(1), \tag{A.8}$$

where S_{nk}^{+} and S_{nk}^{-} are defined analogously to S_{nk} but with the kth element $\xi_{n,ik}$ of $\xi_{n,i}$ in the term $\xi_{n,i} v_{n,i}$ replaced by $\xi_{n,ik}^{+} \equiv \max(\xi_{n,ik}, 0)$ and $\xi_{n,ik}^{-} \equiv \max(-\xi_{n,ik}, 0)$, respectively. We will only show the first part of (A.8) since the other case is similar. Define for every $\kappa \in \mathbb{R}$, $a_{ni}(\lambda, \Delta, \kappa) = a_{ni}(\lambda, \Delta) + \kappa \|n^{-1/2} \xi_{n,i}\|$, and

$$\begin{aligned} \widetilde{S}_{nk}^{+}(\tau, \lambda; \Delta, \kappa) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ 1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta, \kappa)) - \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta, \kappa))] \\ &\quad -1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, 0)) + \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, 0))] \} \xi_{n,ik}^{+} v_{n,i}. \end{aligned}$$

Note that $\widetilde{S}_{nk}^{+}(\tau, \lambda; \Delta, 0) = S_{nk}^{+}(\tau, \lambda; \Delta)$. We follow Koul (1991) and Bai (1994) to show that the first part of (A.8) is a consequence of the following result

$$\sup_{\lambda \in \Lambda} \left| \widetilde{S}_{nk}^{+}(\tau, \lambda; \Delta, \kappa) \right| = o_p(1) \text{ for every given } \Delta \text{ and } \kappa. \tag{A.9}$$

Since Γ is compact, we can partition it into a finite number $N(\sigma)$ of subsets $\{\Gamma_1, \cdots, \Gamma_{N(\sigma)}\}$ such that the diameter of each subset is not greater than σ. Fix $s \in \{1, \cdots, N(\sigma)\}$ and $\Delta_s \in \Gamma_s$. Noting that $\Delta' \xi_{n,i} \leq \Delta_s' \xi_{n,i} + \sigma \|\xi_{n,i}\|$ for any $\Delta \in \Gamma_s$, it follows from the monotonicity of the indicator function and the nonnegativity of $\xi_{n,ik}^{+} v_{n,i}$ that for any $\Delta \in \Gamma_s$,

$$\begin{aligned} & S_{nk}^{+}(\tau, \lambda; \Delta) - \widetilde{S}_{nk}^{+}(\tau, \lambda; \Delta_s, \sigma) \\ = \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta_s, \sigma))] - \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta))] \} \xi_{n,ik}^{+} v_{n,i} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ 1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta)) - 1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta_s, \sigma)) \} \xi_{n,ik}^{+} v_{n,i} \\ \leq \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta_s, \sigma))] - \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta))] \} \xi_{n,ik}^{+} v_{n,i}. \end{aligned}$$

A reverse inequality holds with σ replaced by $-\sigma$ for all $\Delta \in \Gamma_s$. By the triangle inequality and Taylor expansions, we have for sufficiently large n,

$$\begin{aligned} & \sup_{\Delta \in \Gamma_s} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta_s, \sigma))] - \mathrm{E} [1 (b_{ni}(\lambda) u_{n,i} \leq a_{ni}(\lambda, \Delta))] \} \xi_{n,ik}^{+} v_{n,i} \right| \\ \leq \; & \sup_{\Delta \in \Gamma_s} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \mathrm{E} \left[F_{n,i} \left(\frac{a_{ni}(\lambda, \Delta_s, \sigma)}{b_{ni}(\lambda)} \right) - F_{n,i} \left(\frac{a_{ni}(\lambda, \Delta)}{b_{ni}(\lambda)} \right) \right] \right| \xi_{n,ik}^{+} v_{n,i} \\ \leq \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \mathrm{E} \left[F_{n,i} \left(\frac{a_{ni}(\lambda, \Delta_s, \sigma)}{b_{ni}(\lambda)} \right) - F_{n,i} \left(\frac{a_{ni}(\lambda, \Delta_s)}{b_{ni}(\lambda)} \right) \right] \right| \xi_{n,ik}^{+} v_{n,i} \end{aligned}$$

$$+\sup_{\Delta\in\Gamma_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\mathrm{E}\left[F_{n,i}\left(\frac{a_{ni}(\lambda,\Delta_s)}{b_{ni}(\lambda)}\right)-F_{n,i}\left(\frac{a_{ni}(\lambda,\Delta)}{b_{ni}(\lambda)}\right)\right]\right|\xi_{n,ik}^{+}v_{n,i}$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}f_{n,i}\left(\frac{a_{ni}(\lambda,\Delta_s,c_i^*\sigma)}{b_{ni}(\lambda)}\right)\frac{\sigma\left\|\xi_{n,i}\right\|}{|b_{ni}(\lambda)|}\xi_{n,ik}^{+}v_{n,i}$$
$$+\sup_{\Delta_s\in\Gamma_s}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}f_{n,i}\left(\frac{a_{ni}(\lambda,\Delta_s^*)}{b_{ni}(\lambda)}\right)\frac{\sigma\left\|\xi_{n,i}\right\|}{|b_{ni}(\lambda)|}\xi_{n,ik}^{+}v_{n,i}$$
$$\leq \frac{C\sigma}{n}\sum_{i=1}^{n}\frac{\left\|\xi_{n,i}\right\|\xi_{n,ik}^{+}v_{n,i}}{|b_{ni}(\lambda)|}=\sigma O(1),$$

where c_i^* lies between 0 and 1 and Δ_s^* lies between Δ_s and Δ. Consequently,

$$\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|S_n^{+}(\tau,\lambda;\Delta)\right\|\leq\sup_{s\leq N(\sigma)}\sup_{\lambda\in\Lambda}\left|\widetilde{S}_n^{+}(\tau,\lambda;\Delta_s,\sigma)\right|+\sup_{s\leq N(\sigma)}\sup_{\lambda\in\Lambda}\left|\widetilde{S}_n^{+}(\tau,\lambda;\Delta_s,-\sigma)\right|+\sigma O_p(1).$$

By the compactness of Γ, the term σ can be made arbitrarily small and $N(\sigma)$ is finite. So we can prove (A.8) by proving (A.9).

To show (A.9), we also use a chaining argument. Let Δ and κ be fixed. Without loss of generality, assume the support of λ can be written as $\Lambda=[c_1,c_2]$. Partition the interval Λ into $N(\delta^*)$ subintervals at the points $c_1=\lambda_0<\lambda_1<\cdots<\lambda_{N_1}=c_2$, where δ^* denotes the length of each interval. Let $d_{ni}(\lambda,\Delta)\equiv[a_{ni}(\lambda,\Delta)-a_{ni}(\lambda,0)]/b_{ni}(\lambda)$. Then, $d_{ni}(\lambda,\Delta)=n^{-1/2}\Delta'\xi_{n,i}/b_{ni}(\lambda)$ and for $\lambda,\lambda^*\in\Lambda$,

$$\sup_{|\lambda-\lambda^*|\leq\delta^*}|d_{ni}(\lambda,\Delta)-d_{ni}(\lambda^*,\Delta)|=\sup_{|\lambda-\lambda^*|\leq\delta^*}\left|\frac{(\lambda-\lambda^*)\,n^{-1/2}\Delta'\xi_{n,i}}{[1-(\lambda-\lambda_{0\tau})\,h_{n,ii}]\,[1-(\lambda^*-\lambda_{0\tau})\,h_{n,ii}]}\right|\leq n^{-1/2}C\delta^*,$$

for some large enough C and sufficiently large n. Define

$$\overline{S}_{nk}^{+}(\tau,\lambda;\Delta,\kappa,\varsigma)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(u_{n,i}\leq\frac{a_{ni}(\lambda,\Delta,\kappa)}{b_{ni}(\lambda)}+\varsigma n^{-1/2}C\delta^*\right)-\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda,\Delta,\kappa)}{b_{ni}(\lambda)}+\varsigma n^{-1/2}C\delta^*\right)\right.$$
$$\left.-1\left(u_{n,i}\leq\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)+\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)\right\}\xi_{n,ik}^{+}v_{n,i}.$$

Then $\overline{S}_{nk}^{+}(\tau,\lambda;\Delta,\kappa,0)=\widetilde{S}_{nk}^{+}(\tau,\lambda;\Delta,\kappa)$ for sufficiently large n. By the monotonicity of the indicator function and cdf and the nonnegativity of $\xi_{n,ik}^{+}v_{n,i}$, we have that for all λ with $|\lambda-\lambda_s|\leq\delta^*$ and sufficiently large n,

$$\widetilde{S}_{nk}^{+}(\tau,\lambda;\Delta,\kappa)-\overline{S}_{nk}^{+}(\tau,\lambda_s;\Delta,\kappa,1)$$
$$\leq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_s,\Delta,\kappa)}{b_{ni}(\lambda_s)}+Cn^{-1/2}\delta^*\right)-\mathrm{E}F_{n,ii}\left(\frac{a_{ni}(\lambda_s,\Delta,\kappa)}{b_{ni}(\lambda_s)}\right)\right\}\xi_{n,ik}^{+}v_{n,i}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(u_{n,i}\leq\frac{a_{ni}(\lambda_s,0)}{b_{ni}(\lambda_s)}\right)-\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_s,0)}{b_{ni}(\lambda_s)}\right)\right.$$
$$\left.-1\left(u_{n,i}\leq\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)+\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)\right\}\xi_{n,ik}^{+}v_{n,i},$$

and a reverse inequality holds with C replaced by $-C$. By the monotonicity of cdf, for sufficiently large n, we have

$$
\begin{aligned}
& \sup_{\lambda\in\Lambda}\left|\widetilde{S}^{+}_{nk}(\tau,\lambda;\Delta,\kappa)\right| \\
\leq\; & \max_{s}\left|\overline{S}^{+}_{nk}(\tau,\lambda_s;\Delta,\kappa,1)\right| + \max_{s}\left|\overline{S}^{+}_{nk}(\tau,\lambda_s;\Delta,\kappa,-1)\right| \\
& + \max_{s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_s,\Delta,\kappa)}{b_{ni}(\lambda_s)}+\frac{C\delta^{*}}{\sqrt{n}}\right)-\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_s,\Delta,\kappa)}{b_{ni}(\lambda_s)}-\frac{C\delta^{*}}{\sqrt{n}}\right)\right\}\xi^{+}_{n,ik}v_{n,i} \\
& + \sup_{\substack{\lambda_l,\lambda_m\in\Lambda,\\ |\lambda_l-\lambda_m|\leq\delta^*}}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left[1\left(u_{n,i}\leq\frac{a_{ni}(\lambda_l,0)}{b_{ni}(\lambda_l)}\right)-\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_l,0)}{b_{ni}(\lambda_l)}\right)\right]\right.\right. \\
& \left.\left. -\left[1\left(u_{n,i}\leq\frac{a_{ni}(\lambda_m,0)}{b_{ni}(\lambda_m)}\right)-\mathrm{E}F_{n,i}\left(\frac{a_{ni}(\lambda_m,0)}{b_{ni}(\lambda_m)}\right)\right]\right\}\xi^{+}_{n,ik}v_{n,i}\right|. \qquad \text{(A.10)}
\end{aligned}
$$

The first two terms on the right hand side of (A.10) are $o_p(1)$ because $\left\|\overline{S}^{+}_{nk}(\tau,\lambda;\Delta,\kappa,\varsigma)\right\| = o_p(1)$ for every given ς due to an argument similar to the proof of (A.4). They are in fact the maximum of finite number of $o_p(1)$ terms. The third term is $o_p(1)$ as it is no greater than $Cc_f\delta^*\frac{1}{n}\sum_{i=1}^{n}\xi^{+}_{n,ik}v_{n,i} = O(\delta^*)$ with $c_f \equiv \sup_n \max_{1\leq i\leq n}\sup_x f_{n,i}(x)$, which can be made arbitrarily small by choosing small enough δ^* and large enough n. The last term in (A.10) is ensured to be small due to the stochastic equicontinuity property by Assumption 6. ■

Lemma A.3 *Suppose Assumptions 1-6 hold. Then*

$$\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\|E[V_n(\tau,\lambda;\Delta)-V_n(\tau,\lambda;0)]+J_{n\alpha}(\lambda,\tau)\Delta\| = o_p(1).$$

Proof. Let $a_{ni}(\lambda,\Delta)$ and $b_{ni}(\lambda)$ be defined in (A.5) and Assumption A2(v), respectively. By Taylor expansions and the dominated convergence theorem, we have for sufficiently large n:

$$
\begin{aligned}
& \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\|\mathrm{E}[V_n(\tau,\lambda;\Delta)-V_n(\tau,\lambda;0)]+J_{n\alpha}(\lambda,\tau)\Delta\| \\
=\; & \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{1\left(u_{n,i}\leq\frac{a_{ni}(\lambda,\Delta)}{b_{ni}(\lambda)}\right)-1\left(u_{n,i}\leq\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)\right\}\xi_{n,i}v_{n,i}-J_{n\alpha}(\lambda,\tau)\Delta\right\| \\
=\; & \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left[F_{n,i}\left(\frac{a_{ni}(\lambda,\Delta)}{b_{ni}(\lambda)}\right)-F_{n,i}\left(\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)\right]\xi_{n,i}v_{n,i}-J_{n\alpha}(\lambda,\tau)\Delta\right\| \\
=\; & \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{n}\sum_{i=1}^{n}\int_0^1\mathrm{E}\left[f_{n,i}\left(\frac{a_{ni}(\lambda,0)+sn^{-1/2}\Delta'\xi_{n,i}}{b_{ni}(\lambda)}\right)-f_{n,i}\left(\frac{a_{ni}(\lambda,0)}{b_{ni}(\lambda)}\right)\right]ds\frac{\xi_{n,i}\xi'_{n,i}\Delta}{b_{ni}(\lambda)}\right\| \\
=\; & o_p(1).\ ■
\end{aligned}
$$

Lemma A.4 *Suppose Assumptions 1-6 hold. Then*

$$\sup_{\lambda\in\Lambda}\left\|V_n(\tau,\lambda;\widehat{\Delta}_\tau)\right\| = O(n^{-1/2}),\ \ \textit{and}\ \sup_{\lambda\in\Lambda}\|V_n(\tau,\lambda;0)\| = O_p(1).$$

Proof. By Theorem 3.3 of Koenker and Bassett (1978) or the proof of Lemma A2 in Ruppert and Carroll (1980),

$$\begin{aligned}
\sup_{\lambda\in\Lambda}\left\|V_n(\tau,\lambda;\widehat{\Delta}_{\lambda\tau})\right\| &= \sup_{\lambda\in\Lambda}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\hat{\alpha}'_{\lambda\tau}\xi_{n,i})\xi_{n,i}v_{n,i}\right\| \\
&\leq \sup_{\lambda\in\Lambda}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}1(y_{n,i}-\lambda\bar{y}_{n,i}-\hat{\alpha}'_{\lambda\tau}\xi_{n,i}=0)\left\|\xi_{n,i}v_{n,i}\right\| \\
&\leq 2(p+q)n^{-1/2}\max_{1\leq i\leq n}\left\|\xi_{n,i}v_{n,i}\right\| = O(n^{-1/2}).
\end{aligned}$$

By the definition of $\alpha_0(\lambda,\tau)$, $G_\tau(\lambda,\alpha_0(\lambda,\tau))=0$. It follows that

$$\begin{aligned}
\sup_{\lambda\in\Lambda}\left\|V_n(\tau,\lambda;0)\right\| &= \sup_{\lambda\in\Lambda}\left\|\sqrt{n}G_{n\tau}(\lambda;\alpha_0(\lambda,\tau))\right\| \\
&\leq \sup_{\lambda\in\Lambda}\left\|\sqrt{n}\{G_{n\tau}(\lambda;\alpha_0(\lambda,\tau))-\mathrm{E}[G_{n\tau}(\lambda,\alpha_0(\lambda,\tau))]\}\right\| \\
&\quad +\sup_{\lambda\in\Lambda}\left\|\sqrt{n}\{\mathrm{E}[G_{n\tau}(\lambda,\alpha_0(\lambda,\tau))]-G_\tau(\lambda,\alpha_0(\lambda,\tau))\}\right\| \\
&= O_p(1)+O_p(1)=O_p(1) \text{ by Assumption 6. } \blacksquare
\end{aligned}$$

A.2 Proof of Theorem 3.2

For convenience we collect some important notation. Let $\alpha\equiv(\beta',\gamma')'$. Recall $\alpha_0(\tau)=(\beta_0(\tau)',0')'$, $\alpha(\lambda,\tau)\equiv(\beta(\lambda,\tau)',\gamma(\lambda,\tau)')'$, $\rho_\tau(u)=(\tau-1(u\leq 0))u$,

$$\begin{aligned}
Q_{n\tau}(\lambda,\beta,\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\alpha'\xi_{n,i})v_{n,i}, \\
Q_\tau(\lambda,\beta,\gamma) &\equiv \lim_{n\to\infty}\mathrm{E}[Q_{n\tau}(\lambda,\beta,\gamma)],
\end{aligned}$$

$(\hat{\beta}(\lambda,\tau),\hat{\gamma}(\lambda,\tau))\equiv\arg\min_{(\beta,\gamma)}Q_{n\tau}(\lambda,\beta,\gamma)$, and $(\beta_0(\lambda,\tau),\gamma_0(\lambda,\tau))\equiv\arg\min_{(\beta,\gamma)}Q_\tau(\lambda,\beta,\gamma)$. Define

$$\begin{aligned}
&\hat{\lambda}(\tau)\equiv\arg\min_{\lambda}\left\|\hat{\gamma}(\lambda,\tau)\right\|_{\widehat{A}}, \quad \lambda^*(\tau)\equiv\arg\min_{\lambda}\left\|\gamma_0(\lambda,\tau)\right\|_{A}, \\
&\hat{\beta}(\tau)\equiv\hat{\beta}(\hat{\lambda}(\tau),\tau), \quad \beta^*(\tau)\equiv\beta_0(\lambda^*(\tau),\tau), \\
&\hat{\gamma}(\tau)\equiv\hat{\gamma}(\hat{\lambda}(\tau),\tau), \quad \gamma^*(\tau)\equiv\gamma_0(\lambda^*(\tau),\tau).
\end{aligned}$$

Let τ be fixed. Following Chernozhukov and Hansen (2006), we prove the theorem in three steps: (i) Show that $\theta_{0\tau}=(\lambda_{0\tau},\beta'_{0\tau})'$ uniquely solves the limit problem, i.e., $\lambda^*(\tau)=\lambda_0(\tau)$ and $\beta^*(\tau)=\beta_0(\tau)$; (ii) $\hat{\lambda}(\tau)\xrightarrow{p}\lambda_0(\tau)$ and $\hat{\alpha}(\tau)\xrightarrow{p}\alpha_0(\tau)$; (iii) $\sqrt{n}(\hat{\theta}(\tau)-\theta_0(\tau))\xrightarrow{d}N(0,\Omega(A))$.

Step (i): By Assumptions 1(ii) and 4(iv), $\theta_{0\tau}=(\lambda_{0\tau},\beta'_{0\tau})'$ is the unique solution to $G^0_\tau(\theta)=0$, which implies that it uniquely solves the equation

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\psi_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\beta'x_{n,i}-0'z_{n,i})]\xi_{n,i}v_{n,i}=0. \tag{A.11}$$

By the global convexity of $Q_\tau(\lambda,\alpha)$ in α for each λ, and the fact that $\alpha_0(\lambda,\tau)=(\beta_0(\lambda,\tau)',\gamma_0(\lambda,\tau)')'$ is in the interior of $\mathcal{B}\times\mathcal{G}$, $\alpha_0(\lambda,\tau)$ uniquely solves the first order condition of minimizing $Q_\tau(\lambda,\alpha)$ over α:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\psi_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\alpha_0(\lambda,\tau)'\xi_{n,i})]\xi_{n,i}v_{n,i}=0. \tag{A.12}$$

We now show that $\lambda^*(\tau)=\lambda_0(\tau)$ uniquely minimizes $\|\gamma_0(\lambda,\tau)\|_A$ over λ subject to the constraint in (A.12). Clearly, $\|\gamma_0(\lambda_0(\tau),\tau)\|=0$ by (A.11) and $\lambda_0(\tau)$ satisfies (A.12). That is, $\lambda_0(\tau)\in\arg\min_\lambda\|\gamma_0(\lambda,\tau)\|_A$ subject to the constraint in (A.12). It is also the unique solution by (A.11). Now $\beta(\lambda^*(\tau),\tau)=\beta(\lambda_0(\tau),\tau)=\beta_0(\tau)$ by (A.12).

Step (ii): Recall $\hat{\alpha}(\lambda,\tau)\equiv(\hat{\beta}(\lambda,\tau)',\hat{\gamma}(\lambda,\tau)')'$, $\alpha_0(\lambda,\tau)\equiv(\beta_0(\lambda,\tau)',\gamma_0(\lambda,\tau)')'$ and $\alpha_0(\tau)\equiv(\beta_{0\tau}',\gamma_{0\tau}')'$. Let $o_p^*(1)$ denote $o_p(1)$ uniformly in $\lambda\in\Lambda$. By the remark after Theorem 3.1,

$$\|\hat{\alpha}(\lambda,\tau)-\alpha_0(\lambda,\tau)\|=o_p^*(1),\ \text{and}\ \|\hat{\gamma}(\lambda,\tau)-\gamma_0(\lambda,\tau)\|=o_p^*(1)\ \text{in particular.} \tag{A.13}$$

By Assumption 3(iv), $\widehat{A}=A+o_p(1)$. It follows that $\|\hat{\gamma}(\lambda,\tau)\|_{\widehat{A}}-\|\gamma_0(\lambda,\tau)\|_A=o_p^*(1)$. By Assumption 4(v), $\|\gamma_0(\lambda,\tau)\|_A$ is continuous in λ; it is uniquely minimized at $\lambda^*(\tau)=\lambda_0(\tau)$ by Step (i). It follows that $\hat{\lambda}(\tau)\xrightarrow{p}\lambda_0(\tau)$. Now let $\lambda_n(\tau)\xrightarrow{p}\lambda_0(\tau)$. By (A.13) and the continuity of $\alpha_0(\lambda,\tau)$ in λ,

$$\hat{\alpha}(\lambda_n(\tau),\tau)\xrightarrow{p}\alpha_0(\lambda_0(\tau),\tau)=\alpha_0(\tau).$$

In particular, $\hat{\alpha}(\tau)=\hat{\alpha}(\hat{\lambda}(\tau),\tau)\xrightarrow{p}\alpha_0(\tau)$ as desired.

Step (iii): Consider a small ball $B_{\epsilon_n}(\lambda_{0\tau})$ of radius ϵ_n centered at $\lambda_{0\tau}\equiv\lambda_0(\tau)$. Let $\lambda_n\equiv\lambda_n(\tau)\in B_{\epsilon_n}(\lambda_{0\tau})$ where $\epsilon_n\to 0$ slowly enough. Let $g_{n,i}(\lambda,\alpha)=\psi_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\alpha'\xi_{n,i})\xi_{n,i}v_{n,i}$, $\mathrm{E}g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\equiv\mathrm{E}[g_{n,i}(\lambda,\alpha)]_{(\lambda,\alpha)=(\lambda_n,\hat{\alpha}_{\lambda_n\tau})}$, and $G_{n0}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_{n,i}(\lambda_{0\tau},\alpha_0(\lambda_{0\tau},\tau))-\mathrm{E}g_{n,i}(\lambda_{0\tau},\alpha_0(\lambda_{0\tau},\tau))]$. By Lemma A.4 and the stochastic equicontinuity condition in Assumption 6(ii),

$$\begin{aligned}
O(n^{-1/2}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\tau(y_{n,i}-\lambda_n\bar{y}_{n,i}-\hat{\alpha}_{\lambda_n\tau}'\xi_{n,i})\xi_{n,i}v_{n,i}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})-\mathrm{E}g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})]+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\\
&= G_{n0}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})+o_p(1).
\end{aligned} \tag{A.14}$$

By mean value theorem and dominated convergence arguments, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}g_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\\
=\ &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\psi_\tau(y_{n,i}-\lambda_n\overline{y}_{n,i}-\hat{\alpha}_{\lambda_n\tau}'\xi_{n,i})\xi_{n,i}v_{n,i}\\
=\ &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left[F_{n,i}(0)-F_{n,i}\left(c_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\right)\right]\xi_{n,i}v_{n,i}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{f_{n,i}\left(s_i^* c_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\right)v_{n,i}\xi_{n,i}\left[\frac{\sum_{l\neq i}^{n}h_{n,il}u_{n,l}}{b_{ni}(\lambda_n)}+\frac{e_{n,i}'H_n\xi_n\hat{\alpha}_{\lambda_n\tau}}{b_{ni}(\lambda_n)}\right]\right\}\sqrt{n}(\lambda_n-\lambda_{0\tau}) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{f_{n,i}(s_i^* c_{n,i}(\lambda_n,\hat{\alpha}_{\lambda_n\tau}))\right\}\frac{v_{n,i}\xi_{n,i}\xi_{n,i}'}{b_{ni}(\lambda_n)}\sqrt{n}(\hat{\alpha}_{\lambda_n\tau}-\alpha_{0\tau}) \\
&= -(J_\lambda+o_p(1))\sqrt{n}(\lambda_n-\lambda_{0\tau})-(J_\alpha+o_p(1))\sqrt{n}(\hat{\alpha}_{\lambda_n\tau}-\alpha_{0\tau}), \qquad (A.15)
\end{aligned}
$$

where $c_{n,i}(\lambda,\alpha)=\overline{a}_{ni}(\lambda,\alpha)/b_{ni}(\lambda)$ with $\overline{a}_{ni}(\cdot,\cdot)$ and $b_{ni}(\cdot)$ being defined, respectively, in (A.6) and Assumption 2, and s_i^* lies between 0 and 1. The last line follows from the definitions of J_λ and J_α and the fact that $\overline{a}_{ni}(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\to 0$ and $b_{ni}(\lambda_n)\to 1$ as $\epsilon_n\to 0$. This is because $h_{n,ii}=O(1/l_n)$, $\sum_{l\neq i}^{n}h_{n,il}\mathrm{E}u_{n,l}\leq\overline{\mu}\sum_{l=1}^{n}|h_{n,il}|=O(1)$, $e_{n,i}'H_n\xi_n=O(1)$, $\xi_{n,i}=O(1)$ and $(\lambda_n,\hat{\alpha}_{\lambda_n\tau})\to(\lambda_{0\tau},\alpha_{0\tau})$ as $\epsilon_n\to 0$. Putting (A.14) and (A.15) together, we have

$$
O(n^{-1/2})=G_{n0}-(J_\lambda+o_p(1))\sqrt{n}(\lambda_n-\lambda_{0\tau})-(J_\alpha+o_p(1))\sqrt{n}(\hat{\alpha}_{\lambda_n\tau}-\alpha_{0\tau}), \qquad (A.16)
$$

which implies that

$$
\sqrt{n}(\hat{\alpha}_{\lambda_n\tau}-\alpha_{0\tau})=J_\alpha^{-1}G_{n0}-J_\alpha^{-1}J_\lambda(1+o_p(1))\sqrt{n}(\lambda_n-\lambda_{0\tau})+o_p(1). \qquad (A.17)
$$

Partition conformably $J_\alpha^{-1}=[\overline{J}_\beta',\overline{J}_\gamma']'$, where $\overline{J}_\beta$ and $\overline{J}_\gamma$ are $p\times(p+q)$ and $q\times(p+q)$ matrices, respectively. Then

$$
\sqrt{n}(\hat{\beta}(\lambda_n,\tau)-\beta_0(\tau))=\overline{J}_\beta G_{n0}-\overline{J}_\beta J_\lambda(1+o_p(1))\sqrt{n}(\lambda_n-\lambda_{0\tau})+o_p(1), \qquad (A.18)
$$

and

$$
\sqrt{n}(\hat{\gamma}(\lambda_n,\tau)-0)=\overline{J}_\gamma G_{n0}-\overline{J}_\gamma J_\lambda(1+o_p(1))\sqrt{n}(\lambda_n-\lambda_{0\tau})+o_p(1). \qquad (A.19)
$$

By Step (ii), with probability approaching one,

$$
\hat{\lambda}(\tau)=\underset{\lambda_n\in B_{\epsilon_n}(\lambda_{0\tau})}{\arg\min}\ \|\hat{\gamma}(\lambda_n,\tau)\|_{\widehat{A}}. \qquad (A.20)
$$

By Liapounov central limit theorem, $G_{n0}\xrightarrow{d}N(0,S_0)$. Hence

$$
\sqrt{n}\|\hat{\gamma}(\lambda_n,\tau)\|_{\widehat{A}}=\left\|O_p(1)-\overline{J}_\gamma J_\lambda(1+o_p(1))\sqrt{n}(\lambda_n(\tau)-\lambda_{0\tau})\right\|_{A+o_p(1)} \qquad (A.21)
$$

It follows from (A.20) and (A.21) that $\sqrt{n}(\hat{\lambda}(\tau)-\lambda_{0\tau})=O_p(1)$ by the full rank properties of $\overline{J}_\gamma J_\lambda$ and A. Consequently,

$$
\begin{aligned}
\sqrt{n}(\hat{\lambda}(\tau)-\lambda_{0\tau}) &= \underset{s\in\mathbb{R}}{\arg\min}\left\|\overline{J}_\gamma G_{n0}-\overline{J}_\gamma J_\lambda s\right\|_A+o_p(1) \\
&= (J_\lambda'\overline{J}_\gamma'A\overline{J}_\gamma J_\lambda)^{-1}J_\lambda'\overline{J}_\gamma'A\overline{J}_\gamma G_{n0}+o_p(1). \qquad (A.22)
\end{aligned}
$$

Simple algebra shows that

$$
\sqrt{n}\left(\hat{\alpha}(\hat{\lambda}(\tau),\tau)-\alpha_{0\tau}\right)=J_\alpha^{-1}\left[I_{p+q}-J_\lambda(J_\lambda'\overline{J}_\gamma'A\overline{J}_\gamma J_\lambda)^{-1}J_\lambda'\overline{J}_\gamma'A\overline{J}_\gamma\right]G_{n0}+o_p(1), \qquad (A.23)
$$

and

$$\begin{pmatrix} \sqrt{n}(\hat{\lambda}(\tau)-\lambda_{0\tau}) \\ \sqrt{n}(\hat{\beta}(\tau)-\beta_{0\tau}) \end{pmatrix} = \begin{pmatrix} (J_\lambda'\overline{J}_\gamma' A \overline{J}_\gamma J_\lambda)^{-1} J_\lambda' \overline{J}_\gamma' A \overline{J}_\gamma \\ \overline{J}_\beta \left[I_{p+q} - J_\lambda (J_\lambda'\overline{J}_\gamma' A \overline{J}_\gamma J_\lambda)^{-1} J_\lambda' \overline{J}_\gamma' A \overline{J}_\gamma \right] \end{pmatrix} G_{n0} + o_p(1). \qquad \text{(A.24)}$$

The conclusion then follows from the fact that $G_{n0} \xrightarrow{d} N(0, S_0)$. ■

A.3 Proof of Corollary 3.3

When $q = 1$, $\overline{J}_\gamma J_\lambda$ is a nonzero scalar. By (A.23) and the fact that $G_{n0} = O_p(1)$, we have

$$\sqrt{n}(\hat{\gamma}(\hat{\lambda}(\tau),\tau) - 0) = \overline{J}_\gamma [I_{p+1} - J_\lambda (\overline{J}_\gamma J_\lambda)^{-1} \overline{J}_\gamma] G_{n0} + o_p(1) = o_p(1). \qquad \text{(A.25)}$$

By (A.16) and (A.25) and the fact that $\hat{\lambda}(\tau) \xrightarrow{p} \lambda_{0\tau}$, we have

$$[J_\lambda \; J_{\alpha,1:p}] \begin{pmatrix} \sqrt{n}(\hat{\lambda}(\tau)-\lambda_{0\tau}) \\ \sqrt{n}(\hat{\beta}(\hat{\lambda}(\tau),\tau)-\beta_{0\tau}) \end{pmatrix} = G_{n0} + o_p(1), \qquad \text{(A.26)}$$

where $J_{\alpha,1:p}$ is the first p columns of J_α. The result then follows from the fact that $J_0 = [J_\lambda \; J_{\alpha,1:p}]$ and $G_{n0} \xrightarrow{d} N(0, S_0)$. ■

B Proof of Some Ancillary Results

In this section we first prove that together with Assumptions 1-3, Assumption 5* implies Assumption 5, and then prove that Assumption 5** implies Assumption 5.

B.1 Assumptions 1, 2, 3, 5* ⇒ Assumption 5

For notational simplicity, we will suppress the dependence of $\eta_{n,i}(\lambda)$ on λ and write it as $\eta_{n,i}$. Write

$$\mathrm{Var}[\mathcal{S}_n(\lambda)] \equiv \frac{1}{n}\sum_{i=1}^n c_{ni}^2 \mathrm{Var}(\eta_{n,i}) + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_{ni}c_{nj}\mathrm{Cov}(\eta_{n,i},\eta_{nj}) \equiv \mathcal{S}_{n1} + \mathcal{S}_{n2}.$$

It suffices to show that $\mathrm{Var}(\eta_{n,i}) = o(1)$ and $\mathrm{Cov}(\eta_{n,i},\eta_{n,j}) = o(n^{-1})$ under Assumptions 1-3 and 5*.

Define $a_{nij}^\Delta(\lambda) = (\lambda-\lambda_{0\tau})\sum_{l\neq i,j}^n h_{n,il}u_{n,l} + (\lambda-\lambda_{0\tau})e_{n,i}'H_nX_n\beta_{0\tau} + (\alpha_{0\lambda\tau} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i}$. Note that $a_{nij}^\Delta(\lambda)$ differs from $a_{ni}(\lambda,\Delta)$ in (A.5) only in the term $(\lambda-\lambda_{0\tau})h_{n,ij}u_{n,j}$. So we can write

$$\begin{aligned} y_{n,i} - \lambda\bar{y}_{n,i} - \left(\alpha_{0\lambda\tau} + n^{-1/2}\Delta\right)'\xi_{n,i} &= b_{ni}(\lambda)u_{n,i} - a_{ni}(\lambda,\Delta) \\ &= b_{ni}(\lambda)u_{n,i} - (\lambda-\lambda_{0\tau})h_{n,ij}u_{n,j} - a_{nij}^\Delta(\lambda). \end{aligned}$$

$a_{nij}^0(\lambda)$ is defined as $a_{nij}^\Delta(\lambda)$ with 0 in place of Δ. Since $h_{n,ii} = O(1/l_n) = o(1)$ by Lemma A.1 and Assumption A5*, we have $b_{ni}(\lambda) > 0$ for any $\lambda \in \Lambda$ for sufficiently large n and $b_{ni}(\lambda) \to 1$ as $n \to \infty$.

Noting that $|\eta_{n,i}| \le 1$, we have

$$\begin{aligned}
\mathrm{Var}(\eta_{n,i}) &\le \mathrm{E}\left|1(y_{n,i} - \lambda\bar{y}_{n,i} \le (\alpha_{0\lambda\tau} + n^{-1/2}\Delta)'\xi_{n,i}) - 1(y_{n,i} - \lambda\bar{y}_{n,i} \le \alpha'_{0\lambda\tau}\xi_{n,i})\right| \\
&= \mathrm{E}\left|1\left\{b_{ni}(\lambda)u_{n,i} - a_{ni}(\lambda,0) \le n^{-1/2}\Delta'\xi_{n,i}\right\} - 1\left\{b_{ni}(\lambda)u_{n,i} - a_{ni}(\lambda,\Delta) \le 0\right\}\right| \\
&\le \mathrm{E}1\left\{|b_{ni}(\lambda)u_{n,i} - a_{ni}(\lambda,0)| \le n^{-1/2}\left|\Delta'\xi_{n,i}\right|\right\} \\
&\le Cn^{-1/2}\left|\Delta'\xi_{n,i}\right| = O(n^{-1/2}), \qquad \text{(B.1)}
\end{aligned}$$

where the second inequality is due to the fact that $|1(u \le s) - 1(u \le 0)| \le 1(|u| \le |s|)$ and the last equality follows from Assumption 1.

Let $\mathcal{F}_{-ij}$ be the σ-field formed by u_n excluding $u_{n,i}$ and $u_{n,j}$. By Assumption 1(i), for all $i \neq j$, the joint density of $u_{n,i}$ and $u_{n,j}$ conditional on $\mathcal{F}_{-ij}$ is given by $f_{n,i}(u_i)f_{n,j}(u_j)$. Let $\varsigma_{n,1} \equiv b_{ni}(\lambda)u_{n,i} - (\lambda - \lambda_{0\tau})h_{n,ij}u_{n,j}$ and $\varsigma_{n,2} = b_{nj}(\lambda)u_{n,j} - (\lambda - \lambda_{0\tau})h_{n,ji}u_{n,i}$. It is easy to see that the joint density of $\varsigma_{n,1}$ and $\varsigma_{n,2}$ conditional on $\mathcal{F}_{-ij}$ is given by

$$\frac{1}{|\pi_{ni}(\lambda)|}f_{n,i}\left(\frac{b_{nj}(\lambda)\varsigma_1 + (\lambda - \lambda_{0\tau})h_{n,ij}\varsigma_2}{\pi_{nij}(\lambda)}\right)f_{n,j}\left(\frac{(\lambda - \lambda_{0\tau})h_{n,ji}\varsigma_1 + b_{ni}(\lambda)\varsigma_2}{\pi_{nij}(\lambda)}\right)$$

provided $\pi_{nij}(\lambda) \equiv b_{ni}(\lambda)b_{nj}(\lambda) - (\lambda - \lambda_{0\tau})^2 h_{n,ij}h_{n,ji} \neq 0$. The last condition is satisfied for sufficiently large n because in this case, $(\lambda - \lambda_{0\tau})^2 h_{n,ij}h_{n,ji} = O(1/l_n^2) = o(1)$ and $b_{ni}(\lambda)b_{nj}(\lambda) \to 1$ as $n \to \infty$. In addition, it is easy to verify that the marginal density of $\varsigma_{n,1}$ is given by

$$\frac{1}{|b_{ni}(\lambda)|}\int f_{n,i}\left(\frac{\varsigma_1 + (\lambda - \lambda_{0\tau})h_{n,ij}\varsigma_2}{b_{ni}(\lambda)}\right)f_{n,j}\left(\varsigma_2\right)d\varsigma_2.$$

Let $s_{ni} \equiv \mathrm{sign}(\Delta'\xi_{n,i})$, which takes value 1 if $\Delta'\xi_{n,i} \ge 0$ and -1 otherwise. Then

$$\begin{aligned}
&\mathrm{E}[\eta_{n,i}\eta_{n,j}] = \mathrm{E}[\mathrm{E}(\eta_{n,i}\eta_{n,j}|\mathcal{F}_{-ij})] \\
=\;& \mathrm{E}\{[1(\varsigma_{n,1} \le a^{\Delta}_{nij}(\lambda)) - 1(\varsigma_{n,1} \le a^{0}_{nij}(\lambda))] \cdot [1(\varsigma_{n,2} \le a^{\Delta}_{nji}(\lambda)) - 1(\varsigma_{n,2} \le a^{0}_{nji}(\lambda))]|\mathcal{F}_{-ij}\} \\
=\;& s_{ni}s_{nj}\mathrm{E}\left\{\frac{1}{|\pi_{nij}(\lambda)|}\int_{a^0_{nij}(\lambda)}^{a^{\Delta}_{nij}(\lambda)}\int_{a^0_{nji}(\lambda)}^{a^{\Delta}_{nji}(\lambda)} f_{n,i}\left(\frac{b_{nj}(\lambda)\varsigma_1 + (\lambda - \lambda_{0\tau})h_{n,ij}\varsigma_2}{\pi_{nij}(\lambda)}\right)\right. \\
&\left.\times f_{n,j}\left(\frac{(\lambda - \lambda_{0\tau})h_{n,ji}\varsigma_1 + b_{ni}(\lambda)\varsigma_2}{\pi_{nij}(\lambda)}\right)d\varsigma_2 d\varsigma_1\right\} \\
\le\;& \frac{1}{n\,|\pi_{nij}(\lambda)|}c_f^2\left|\Delta'\xi_{n,i}\right|\left|\Delta'\xi_{n,j}\right| = O(n^{-1}), \qquad \text{(B.2)}
\end{aligned}$$

where $c_f \equiv \sup_n \max_{1 \le i \le n}\sup_u f_{n,i}(u)$. Similarly,

$$\begin{aligned}
\mathrm{E}\left[\eta_{n,i}\right] &= \mathrm{E}\left[\mathrm{E}(\eta_{n,i}|\mathcal{F}_{-ij})\right] \\
&= \mathrm{E}\left\{\left[1\left(\varsigma_{n,1} \le a^{\Delta}_{nij}(\lambda)\right) - 1\left(\varsigma_{n,1} \le a^{0}_{nij}(\lambda)\right)\right]|\mathcal{F}_{-ij}\right\} \\
&= s_{ni}\mathrm{E}\left\{\int_{a^0_{nij}(\lambda)}^{a^{\Delta}_{nij}(\lambda)}\frac{1}{|b_{ni}(\lambda)|}\int f_{n,i}\left(\frac{\varsigma_1 + (\lambda - \lambda_{0\tau})h_{n,ij}\varsigma_2}{b_{ni}(\lambda)}\right)f_{n,j}\left(\varsigma_2\right)d\varsigma_2 d\varsigma_1\right\} \\
&\le \frac{1}{\sqrt{n}\,|b_{ni}(\lambda)|}c_f\left|\Delta'\xi_{n,i}\right| = O\left(n^{-1/2}\right). \qquad \text{(B.3)}
\end{aligned}$$

When $l_n \to \infty$, $h_{n,ij} = O(1/l_n) = o(1)$ and $\sum_{l=1}^n |h_{n,il}| = O(1)$ by Lemma A.1. Let $S_{nij} \equiv \sum_{l\neq i,j}^n h_{n,il}[u_{n,l} - \mathrm{E}u_{n,l}]$. Then $\mathrm{E}S_{nij} = 0$ and $\mathrm{Var}(S_{nij}) = \sum_{l\neq i,j}^n h_{n,il}^2 \sigma_{n,l}^2 \leq \max_{1\leq i,j\leq n} |h_{n,ij}| \overline{\sigma}^2 \sum_{l=1}^n |h_{n,il}| = O(1/l_n) = o(1)$. Hence $S_{nij} \xrightarrow{p} 0$ by the Chebyshev inequality. It follows from the uniform boundedness of $f_{n,i}(\cdot)$ and the dominated convergence theorem that

$$\mathrm{E}\left[f_{n,i}\left(\frac{(\lambda-\lambda_{0\tau})S_{nij}+t_{nij}}{b_{ni}(\lambda)}\right) - f_{n,i}\left(\frac{t_{nij}}{b_{ni}(\lambda)}\right)\right] \to 0.$$

Now we can apply Taylor expansions to obtain

$$\begin{aligned}
&\mathrm{E}[\eta_{n,i}] \\
=\;& s_{ni}\mathrm{E}\left\{\int_{a_{nij}^0(\lambda)}^{a_{nij}^\Delta(\lambda)} \frac{1}{b_{ni}(\lambda)} \int f_{n,i}\left(\frac{\varsigma_1}{b_{ni}(\lambda)}\right) f_{n,j}(\varsigma_2)\, d\varsigma_2 d\varsigma_1\right\} \\
&+ s_{ni}\mathrm{E}\left\{\int_{a_{nij}^0(\lambda)}^{a_{nij}^\Delta(\lambda)} \frac{1}{b_{ni}(\lambda)} \int f_{n,i}^{(1)}\left(\frac{\varsigma_1 + s_{ij}^*\varsigma_2}{b_{ni}(\lambda)}\right) \frac{(\lambda-\lambda_{0\tau})h_{n,ij}\varsigma_2}{b_{ni}(\lambda)} f_{n,j}(\varsigma_2)\, d\varsigma_2 d\varsigma_1\right\} \\
=\;& s_{ni}\mathrm{E}\left[F_{n,i}\left(\frac{a_{nij}^\Delta(\lambda)}{b_{ni}(\lambda)}\right) - F_{n,i}\left(\frac{a_{nij}^0(\lambda)}{b_{ni}(\lambda)}\right)\right] + O\left(n^{-1/2} l_n^{-1}\right) \\
=\;& \frac{s_{ni}\Delta'\xi_{n,i}}{\sqrt{n}b_{ni}(\lambda)} \mathrm{E}\left[f_{n,i}\left(\frac{a_{nij}^0(\lambda)}{b_{ni}(\lambda)}\right)\right] + o(n^{-1/2}) \\
=\;& \frac{s_{ni}\Delta'\xi_{n,i}}{\sqrt{n}b_{ni}(\lambda)} \left\{f_{n,i}\left(\frac{t_{nij}}{b_{ni}(\lambda)}\right) + \mathrm{E}\left[f_{n,i}\left(\frac{(\lambda-\lambda_{0\tau})S_{nij}+t_{nij}}{b_{ni}(\lambda)}\right) - f_{n,i}\left(\frac{t_{nij}}{b_{ni}(\lambda)}\right)\right]\right\} + o(n^{-1/2}) \\
=\;& \frac{s_{ni}\Delta'\xi_{n,i}}{\sqrt{n}b_{ni}(\lambda)} \left\{f_{n,i}\left(\frac{t_{nij}}{b_{ni}(\lambda)}\right) + o(1)\right\} + o(n^{-1/2}) \\
=\;& \frac{s_{ni}\Delta'\xi_{n,i}}{\sqrt{n}b_{ni}(\lambda)} f_{n,i}(t_{nij}) + o(n^{-1/2}),
\end{aligned}$$

where s_{ij}^* lies between 0 and $(\lambda-\lambda_{0\tau})h_{n,ij}$, $t_{nij} = (\lambda-\lambda_{0\tau})\sum_{l\neq i,j}^n h_{n,il}\mathrm{E}u_{n,l} + (\lambda-\lambda_{0\tau})e_{n,i}' H_n X_n \beta_{0\tau}$ $+(\alpha_{0\lambda\tau} - \alpha_{0\tau})'\xi_{n,i}$, and $f_{n,i}^{(1)}(\cdot)$ is the first derivative of $f_{n,i}(\cdot)$. Similarly,

$$\begin{aligned}
&\mathrm{E}[\eta_{n,i}\eta_{nj}] \\
=\;& s_{ni}s_{nj}\mathrm{E}\left\{\frac{1}{|\pi_{ni}(\lambda)|} \int_{a_{nij}^0(\lambda)}^{a_{nij}^\Delta(\lambda)} \int_{a_{nji}^0(\lambda)}^{a_{nji}^\Delta(\lambda)} \left[f_{n,i}\left(\frac{\varsigma_1}{b_{ni}(\lambda)}\right) + O(h_{n,ij})\right]\right. \\
&\times \left.\left[f_{n,j}\left(\frac{\varsigma_2}{b_{nj}(\lambda)}\right) + O(h_{n,ji})\right] d\varsigma_2 d\varsigma_1\right\} \\
=\;& s_{ni}s_{nj}\mathrm{E}\left\{\frac{1}{|\pi_{ni}(\lambda)|} \int_{a_{nij}^0(\lambda)}^{a_{nij}^\Delta(\lambda)} \int_{a_{nji}^0(\lambda)}^{a_{nji}^\Delta(\lambda)} \left[f_{n,i}\left(\frac{\varsigma_1}{b_{ni}(\lambda)}\right) f_{n,j}\left(\frac{\varsigma_2}{b_{nj}(\lambda)}\right) + o(1)\right] d\varsigma_2 d\varsigma_1\right\} \\
=\;& s_{ni}s_{nj}\mathrm{E}\left\{\left[F_{n,i}\left(\frac{a_{nij}^\Delta(\lambda)}{b_{ni}(\lambda)}\right) - F_{n,i}\left(\frac{a_{nij}^0(\lambda)}{b_{ni}(\lambda)}\right)\right]\left[F_{n,j}\left(\frac{a_{nji}^\Delta(\lambda)}{b_{nj}(\lambda)}\right) - F_{n,j}\left(\frac{a_{nji}^0(\lambda)}{b_{nj}(\lambda)}\right)\right]\right\} + o(n^{-1}) \\
=\;& \frac{s_{ni}s_{nj}\Delta'\xi_{n,i}\xi_{n,j}'\Delta}{nb_{ni}(\lambda)b_{nj}(\lambda)} \mathrm{E}\left[f_{n,i}\left(\frac{a_{nij}^0(\lambda)}{b_{ni}(\lambda)}\right) f_{n,j}\left(\frac{a_{nji}^0(\lambda)}{b_{nj}(\lambda)}\right)\right] + o(n^{-1}) \\
=\;& \frac{s_{ni}s_{nj}\Delta'\xi_{n,i}\xi_{n,j}'\Delta}{nb_{ni}(\lambda)b_{nj}(\lambda)} f_{n,i}(t_{nij}) f_{n,j}(t_{nji}) + o(n^{-1}).
\end{aligned}$$

Consequently,

$$\text{Cov}\left(\eta_{n,i},\eta_{nj}\right)=\text{E}\left[\eta_{n,i}\eta_{nj}\right]-\text{E}\left[\eta_{n,i}\right]E\left[\eta_{nj}\right]=o(n^{-1}).$$

B.2 Assumption 5** $\Rightarrow$ Assumption 5

Let $\mathcal{S}_{nj}(\lambda)$ and $\mathcal{S}_{njt}(\lambda)$ denote the partial sums of $n^{-1/2}\eta_{n,i}(\lambda)$ over observations in group j and subgroup t of group j, respectively, i.e.,

$$\mathcal{S}_{n,j}(\lambda)=\sum_{t=1}^{m_{nj}}\mathcal{S}_{njt}(\lambda),\text{ where }\mathcal{S}_{njt}(\lambda)=\sum_{s\in\mathcal{G}_{njt}}n^{-1/2}\eta_{n,s}(\lambda).$$

Then $\mathcal{S}_n(\lambda)=\sum_{j=1}^J\mathcal{S}_{nj}(\lambda)=\sum_{j=1}^J\sum_{t=1}^{m_{nj}}\mathcal{S}_{njt}(\lambda)$. Because J is finite, by Cauchy-Schwartz inequality it suffices to show that $\text{Var}(\mathcal{S}_{nj}(\lambda))=o(1)$ for each $j=1,\cdots,J$ and $\lambda\in\Lambda$.

Fix $j\in\{1,\cdots,J\}$ and $\lambda\in\Lambda$. Write

$$\text{Var}(\mathcal{S}_{nj}(\lambda))=\sum_{t=1}^{m_{nj}}\text{Var}(\mathcal{S}_{njt}(\lambda))+2\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}}\text{Cov}(\mathcal{S}_{njl}(\lambda),\mathcal{S}_{njt}(\lambda))\equiv I_{n1}+I_{n2}$$

By arguments analogous to those used in the last subsection ((B.1)-(B.3) in particular), one can show that $\text{Var}(\eta_{n,s}(\lambda))\leq C_1n^{-1/2}$ and $\text{Cov}(\eta_{n,i}(\lambda),\eta_{n,s}(\lambda))\leq C_2n^{-1}$ for $i\neq s$ and finite constants C_1,C_2. Hence,

$$\begin{aligned}\text{Var}(\mathcal{S}_{njt}(\lambda)) &= \frac{1}{n}\sum_{s\in\mathcal{G}_{njt}}\text{Var}(\eta_{n,s}(\lambda))+\frac{1}{n}\sum_{i\in\mathcal{G}_{njt}}\sum_{s\in\mathcal{G}_{njt},s\neq i}\text{Cov}(\eta_{n,i}(\lambda),\eta_{n,s}(\lambda))\\ &\leq C_1n^{-3/2}n_{jt}+C_2n^{-2}n_{jt}^2,\end{aligned}$$

which implies that $I_{n1}\leq C_1n^{-3/2}\sum_{t=1}^{m_{nj}}n_{jt}+C_2n^{-2}\sum_{t=1}^{m_{nj}}n_{jt}^2\leq C_1n^{-1}+C_2n_{jt}/n=o(1)$ by Assumption 5**(ii). Now by Assumption 5**(i),

$$\begin{aligned}I_{n2} &\leq 2\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}}\sqrt{\text{Var}(\mathcal{S}_{njl}(\lambda))\text{Var}(\mathcal{S}_{njt}(\lambda))}\alpha_{m_{nj}}\\ &= o(1)\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}}\alpha_{m_{nj}}=o(m_{nj}^2\alpha_{m_{nj}})=o(1).\end{aligned}$$

Consequently, $\text{Var}(\mathcal{S}_{nj}(\lambda))=o(1)$. This completes the proof.

REFERENCES

Abadie, A., 1995. Changes in Spanish labor income structure during the 1980's: a quantile-regression approach. Working paper 9521, CEMFI, Madrid, Spain.

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variable estimates of the effects of subsidized training on the quantiles of trainee earnings. *Econometrica* 70, 91-117.

Amemiya, T., 1982. Two stage least absolute deviations estimators. *Econometrica* 50, 689-711.

Andrews, D. W. K., 1994. Empirical process methods in econometrics. In R. F. Engle & D. L. McFadden (eds.), *Handbook of Econometrics*, vol 4, pp. 2247-2294. New York: North-Holland.

Anselin, L., 1988. *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic Press.

Bai, J., 1994. Weak convergence of the sequential empirical processes of residuals in ARMA models. *Annals of Statistics* 22, 2051-2061.

Chen, L., Portony, S., 1996. Two-staged regression quantile and two-staged trimmed least-squares estimators for structural equation models. *Communication in Statistics: Theory and Method* 25, 1005-1032.

Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. *Econometrica* 73, 245-261.

Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 127, 491-525.

Hansen, B. E., 1996. Stochastic equicontinuity for unbounded dependent heterogenous arrays. *Econometric Theory* 12, 347-359.

Hong, H., Tamer, E., 2003. Inference in censored models with endogenous regressors. *Econometrica* 71, 905-932.

Honoré, B. E., Hu, L., 2004. On the performance of some robust instrumental variable estimators. *Journal of Business & Economic Statistics* 22, 30-39.

Horn, R., Johnson, C., 1985. *Matrix Analysis.* New York: John Wiley & Sons.

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., 2006. Specification and estimation of spatial autoregressive models with autoregressive and heteroscedasticity disturbances. Working paper, University of Maryland, College Park.

Koenker, R., 1994. Confidence intervals for regression quantiles, in P. Mandl and M. Huskova (eds.), *Asymptotic Statistics*, pp. 349–359, New York: Springer-Verlag.

Koenker, R., 2005. *Quantile Regression.* Cambridge: Cambridge University Press.

Koenker, R., Bassett, G., 1978. Regression Quantiles. *Econometrica* 46, 33-50.

Koenker, R., Zhao, Q., 1996. Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12, 793-813.

Koul, H. L., 1991. A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference* 29, 1291-1308.

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

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

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

Lin, X., Lee, L-F., 2006. GMM estimation of spatial autoregressive models with unknown heteroscedasticity. Working paper, Department of Economics, Ohio State University.

Liu, X., Lee, L-F., Bollinger, C. R., 2007. Improved efficient quasi maximum likelihood estimator of spatial autoregressive models. Working paper. Department of Economics, Ohio State University.

Ma, L., Koenker, R., 2006. Quantile regression methods for recursive structural equation models. *Journal of Econometrics* 134, 471-506.

Ord, J. K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027-1057.

Pinkse, J., Slade, M., 1998. Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125-154.

Pinkse, J., Slade, M., Shen, L., 2006. Dynamic spatial discrete choice using one-step GMM: an application to mine operating decisions. *Spatial Economic Analysis* 1, 53–99.

Pinkse, J., Shen, L., Slade, M., 2007. A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics* 140, 215-225.

Powell, J. L., 1983. The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica* 51, 1569-1575.

Powell, J. L., 1986. Censored regression quantiles. *Journal of Econometrics* 32, 143-155.

Robinson, P. M., 2007. Efficient estimation of the semiparametric spatial autoregressive model. Forthcoming in *Journal of Econometrics*.

Ruppert, D., Carroll, R. J., 1980. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828-838.

Sakata, S., 2006. Instrumental variable estimation based on conditional median restriction. Forthcoming in *Journal of Econometrics*.

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.