Instrumental Variable Quantile Estimation of Spatial Autoregressive Models*

Liangjun Su and Zhenlin Yang†

School of Economics, Singapore Management University

May 5, 2011

Abstract

We propose a spatial quantile autoregression (SQAR) model, which allows cross-sectional dependence among the responses, unknown heteroscedasticity in the disturbances, and heterogeneous impacts of covariates on different points (quantiles) of a response distribution. The instrumental variable quantile regression (IVQR) method of Chernozhukov and Hansen (2006) is generalized to allow the data to be non-identically distributed and dependent, an IVQR estimator for the SQAR model is then defined, and its asymptotic properties are derived. Simulation results show that this estimator performs well in finite samples at various quantile points. In the special case of spatial median regression, it outperforms the conventional QML estimator without taking into account of heteroscedasticity in the errors; it also outperforms the GMM estimators with or without heteroscedasticity. An empirical illustration is provided.

JEL classifications: C13, C21, C26, C51

Key Words: Spatial Quantile Autoregression; IV Quantile Regression; Spatial Dependence; Heteroscedasticity.

---

*We thank Anil Bera, Bernard Fingleton, Jean Paelinck, Peter C. B. Phillips, Peter Robinson, and the seminar participants of the 1st World Conference of the Spatial Econometrics Association (2007), Far Eastern Econometric Society Meeting (2008), and National University of Singapore, for their helpful comments. The early version of the paper was completed when Liangjun Su was with Beijing University. He 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 in 2006, 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.

†Corresponding author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833; email: zlyang@smu.edu.sg

1 Introduction

In recent years spatial dependence among the cross-sectional units has become 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 most popular one is perhaps the *spatial autoregressive* (SAR) model of Cliff and Ord (1973, 1981), in which the outcome of a spatial unit is allowed to depend linearly on the outcomes of its neighboring units and the values of covariates, i.e.,

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

where n is the total number of spatial units, $Y_n \equiv (y_{n,1}, \cdots, y_{n,n})'$ is an $n \times 1$ vector of response values, λ_0 is the spatial lag parameter, $W_n \equiv \{w_{n,ij}\}$ is a known $n \times n$ spatial weight matrix with zero diagonal elements, $W_n Y_n$ is the spatial lagged variable, $X_n \equiv (x'_{n,1}, \cdots, x'_{n,n})'$ is an $n \times p$ matrix containing the values of the regressors, β_0 is a p-vector of regression coefficients, and $U_n \equiv (u_{n,1}, \cdots, u_{n,n})'$ denotes an n-vector of independent and identically distributed (iid) random disturbances with zero means.[1]

While the spatial models with iid innovations have been extensively studied and applied, researchers have realized that an important issue in modelling the spatial data, the heteroscedasticity, has not been adequately addressed. Spatial units are often heterogeneous in important characteristics such as size, location and area; spatial units interact with the strength and structure of social interactions changing across groups; and as a result, spatial observations are heteroscedastic, a phenomenon often observed in unemployment or crime rates data, housing prices, etc. See, e.g., Anselin (1988), Glaeser et al. (1996), LeSage (1999) for more discussions on spatial heteroscedasticity. Lin and Lee (2010) extended the GMM method to allow for heteroscedasticity in the SAR model, and Kelejian and Prucha (2010) considered the GMM estimation with heteroscedasticity for a more general spatial model. Clearly, all these models are "spatial"extensions of the usual mean regressions where model estimations are based primarily on the restriction that the error terms have zero means. As such, possible heterogeneous impacts of covariates on different points (quantiles) of a response distribution cannot be captured.

Koenker and Bassett (1978) made an important extension of the standard mean regression to the quantiles of the responses, giving what is called the *quantile regression* (QR). Since then, the QR model has been extensively studied in theoretical works and widely used in empirical applications. It has become an important tool for estimating quantile-specific effects. See Koenker (2005) for an excellent exposition of the quantile regression. If the τth conditional quantile function of $y_{n,i}$ given $x_{n,i}$ is given by $Q_y(\tau|x_{n,i}) = x'_{n,i}\beta_{0\tau}$ for $i = 1, \cdots, n$, then the standard linear QR model takes the form

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

[1] The existence of endogeneity in this model renders the ordinary least squares (OLS) estimator generally inconsistent. There are two types of estimators that have been extensively 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), and Lee (2004). The other is the generalized method of moment (GMM) estimator; see, among the others, Kelejian and Prucha (1998, 1999), and Lee (2003, 2007). Both estimators are under the assumption that the disturbances are iid. Robinson (2010) proposed an efficient estimator for the semiparametric spatial autoregressive model.

where $U_{\tau n} \equiv (u_{\tau n,1}, \cdots, u_{\tau n,n})'$ and $u_{\tau n,i} \equiv y_{n,i} - x'_{n,i}\beta_{0\tau}$ satisfying the quantile restriction

$$\Pr(u_{\tau n,i} \leq 0 \mid x_{n,i}) = \tau \ \forall i = 1, \cdots, n. \tag{1.3}$$

Here and below for notational simplicity we will suppress the dependence of $U_{\tau n}$ and $u_{\tau n,i}$ on the quantile index τ, and write $U_n \equiv U_{\tau n}$ and $u_{n,i} \equiv u_{\tau n,i}$. 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. Two most appealing features of the quantile regression are (i) its ability to allow for a separate modelling at different points of a response distribution so that the heterogeneous impacts of explanatory variables can be characterized and differentiated, and (ii) its robustness to the error distributions including outliers and heteroscedasticity.

While both SAR and QR 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 – leading to the first model of this kind which we term in this paper as the *spatial quantile autoregression* (SQAR). Indeed, the SQAR model offers an alternative way for allowing unknown heteroscedasticity in the SAR model, and gives an important method for modeling heterogeneous effects of variables on different quantiles of a response, taking into account of unobserved heterogeneity and spatial dependence. Unfortunately, the SQAR model contains an endogenous covariate (the spatial lag), rendering the ordinary QR techniques inapplicable and new methods of inference to be called for.

Having realized the limitation of the ordinary QR model in addressing typical economics problems, researchers have considered ways to "endogenize" the QR models and have developed methods for estimating them. To the best of our knowledge, Amemiya (1982) was the first to do so under the framework of a two-stage median regression. His work was then extended by Powell (1983), Chen and Portony (1996), and Kim and Muller (2004). In their seminal paper, Chernozhukov and Hansen (2005) proposed an IV model of quantile treatment effects and studied the issue of model identification. Subsequently, Chernozhukov and Hansen (2006, **CH** hereafter) proposed an *instrumental variable quantile regression* (IVQR) method for model estimation and introduced a class of tests based on it.[2] A typical linear quantile regression model with endogenous regressors can be written in the following form

$$Y_n = D_n\alpha_{0\tau} + X_n\beta_{0\tau} + U_n, \tag{1.4}$$

where $D_n \equiv (d_{n,1}, ..., d_{n,n})'$ is an $n \times k$ matrix of endogenous regressors, $\alpha_{0\tau}$ is the τ-dependent coefficients representing the structural quantile-specific effects of $d_{n,i}$ on $y_{n,i}$, and for an instrument vector $z_{n,i}$, $u_{n,i}$ satisfies the following structural quantile restriction

$$\Pr(u_{n,i} \leq 0 \mid x_{n,i}, \ z_{n,i}) = \tau \ \forall i = 1, \cdots, n. \tag{1.5}$$

(1.4) and (1.5) specify $S_Y(\tau|d, x) \equiv d'\alpha_{0\tau} + x'\beta_{0\tau}$ as the *structural quantile function* (SQF) defined by **CH**. However, previous literature focuses on the estimation of the structural quantile regression

[2] Chernozhukov and Hansen (2008) and Chernozhukov et al. (2007, 2009) proposed various alternative inference methods. Other related works on quantile regression with endogenous regressors include Abadie, Angrist and Imbens (2002), Sakata (2007), Ma and Koenker (2006), Hong and Tamer (2003), Honoré and Hu (2004), Sokbae Lee (2007), and Blundel and Powell (2007).

coefficients only under the assumption that the data are iid. Obviously, this iid assumption is not satisfied by our SQAR model due to the spatial dependence and unknown heteroscedasticity.

This paper contributes to the literature by introducing the SQAR model, which on one hand extends the conventional SAR models by allowing quantile specific effects and unknown heteroscedasticity, and on other hand extends the conventional QR models by allowing cross-sectional dependence among the responses. Such an extension seems very interesting as it allows for different degrees of spatial dependence at different quantile points of the response distribution, i.e., it allows the spatial parameter ($\lambda \equiv \lambda_\tau$) to be dependent on the quantile index τ. At the same time, it also allows, as in the ordinary quantile regression, the impact ($\beta \equiv \beta_\tau$) of the covariate $x_{n,i}$ on the response $y_{n,i}$ to be different at different quantile 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 prices of high-end houses spatially related to each other to be different to the way the prices of low-end houses related to each other, e.g., $\lambda_{0.9} \neq \lambda_{0.1}$. Interestingly, since the first version of the paper appeared, some empirical works have already been carried out using our SQAR model and the empirical evidence obtained does support the above arguments, see Kostov (2009) for agricultural land prices, Liao and Wang (2010) and Zietz et al. (2010) for housing prices. We also present in Section 4 an empirical illustration of our methodology using the popular Boston housing price data.

We propose an IVQR estimator for the SQAR model by generalizing the IVQR method for iid data of **CH** to allow for spatial dependence, heteroscedasticity, and possibly additional endogeneity (other than spatial lag). We derive the asymptotic properties of our IVQR estimator. Simulation results show that this estimator performs well in finite samples at various quantile points. Specifically, at the median point, it outperforms the conventional QML estimator without taking into account of heteroscedasticity in the errors; it also outperforms the GMM estimators with or without heteroscedasticity.

The rest of the paper is organized as follows. Section 2 introduces the SQAR 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 for the case of median regression. Also in Section 4 an empirical illustration is provided. Section 5 concludes the paper. All proofs are relegated to the appendix.

2 The Model and the Method of Estimation

In this section, we first introduce our SQAR model, and then we outline how **CH**'s IVQR method for iid data is extended to our SQAR model.

2.1 Spatial Quantile Autoregression

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, termed in this paper as the *spatial quantile autoregression* (SQAR) model. We shall motivate our SQAR model from two perspectives: the traditional quantile regression and the structural equation.

The traditional quantile regression perspective. Following the lead of Koenker and Bassett (1982), we consider the following location-scale model

$$y_{n,i} = \lambda_0 \bar{y}_{n,i} + \beta_0' x_{n,i} + \sigma_{n,i}\varepsilon_{n,i} \text{ with } \sigma_{n,i} = 1 + \underline{\lambda}\bar{y}_{n,i} + \underline{\beta}' x_{n,i} \tag{2.1}$$

where $\bar{y}_{n,i} \equiv \sum_{j=1}^{n} w_{n,ij} y_{n,j}$ denotes the ith element of $\bar{Y}_n \equiv W_n Y_n$, $x_{n,i}$ is a $p \times 1$ vector of strictly exogenous regressors, and $\varepsilon_{n,i}$ are iid unobservable error terms that are independent of X_n. Clearly, (2.1) is a fairly general class of linear location-scale models that can incorporate many classical models as special cases. First, if $\lambda_0 = \underline{\lambda} = 0$, (2.1) reduces to the classical linear location-scale model studied by Koenker and Bassett (1982) where endogeneity is ruled out. If in addition $\underline{\beta} = 0$, (2.1) becomes the location model with neither endogeneity nor heteroscedasticity. Second, if $\underline{\lambda} = 0$, (2.1) becomes the traditional SAR model with heteroscedastic error term under the assumption that $\varepsilon_{n,i}$ has mean 0. Third, if $\underline{\lambda} = 0$ and $\underline{\beta} = 0$, (2.1) becomes the classical SAR model with iid error terms where the spatial lagged dependent variable $\bar{y}_{n,i}$ only enters the location part of the model. In this paper, in addition to allowing $\underline{\beta} \neq 0$, we also allow $\underline{\lambda} \neq 0$. By doing so, we also permit $\bar{y}_{n,i}$ to enter the scale part of the model. Intuitively speaking, it is not difficult to imagine that the spatial lagged dependent variable may have an influence not only on the location of an individual's outcome but also its scale.

Let $Q_\varepsilon(\tau)$ denote the τth quantile of $\varepsilon_{n,i}$. Under the condition that $\min_{1\le i\le n} \sigma_{n,i} > c_\sigma > 0$ and that $\varepsilon_{n,i}$ is independent of X_n for all i, we have

$$\begin{aligned} \tau &= \Pr\left[\varepsilon_{n,i} \le Q_\varepsilon(\tau) \mid X_n\right] \\ &= \Pr\left[y_{n,i} - \lambda_0 \bar{y}_{n,i} - \beta_0' x_{n,i} \le \sigma_{n,i} Q_\varepsilon(\tau) \mid X_n\right] \\ &= \Pr\left[y_{n,i} - \lambda^*(\tau)\bar{y}_{n,i} - \beta^*(\tau)' x_{n,i} \le 0 \mid X_n\right], \end{aligned} \tag{2.2}$$

where $\lambda^*(\tau) = \lambda_0 + \underline{\lambda} Q_\varepsilon(\tau)$, $\beta^*(\tau) = \beta_0 + \underline{\beta}^* Q_\varepsilon(\tau)$, and $\underline{\beta}^* = \underline{\beta} + (1, 0, ..., 0)'$ with $(1, 0, ..., 0)$ being a $p \times 1$ vector. Let $u_{n,i}^* \equiv u_{\tau n,i}^* = y_{n,i} - \lambda^*(\tau)\bar{y}_{n,i} - \beta^*(\tau)' x_{n,i}$, we have

$$y_{n,i} = \lambda^*(\tau)\bar{y}_{n,i} + \beta^*(\tau)' x_{n,i} + u_{n,i}^*, \tag{2.3}$$

for $i = 1, \cdots, n$, or in matrix form

$$Y_n = \lambda^*(\tau) W_n Y_n + X_n \beta^*(\tau) + U_n^*, \tag{2.4}$$

where $U_n^* \equiv \left(u_{n,1}^*, \cdots, u_{n,n}^*\right)'$ and by (2.2) $u_{n,i}^*$ satisfies the following quantile restriction

$$\Pr\left(u_{n,i}^* \le 0 \mid X_n\right) = \tau \quad \forall i. \tag{2.5}$$

As before, we have suppressed the dependence of U_n^* and $u_{n,i}^*$ on the quantile index τ for notational simplicity. We call the model in (2.3) or (2.4) together with the quantile restriction in (2.5) as the SQAR model. It is worth mentioning in passing that (2.1) can be rewritten as

$$y_{n,i} = \left(\lambda_0 + \underline{\lambda}\varepsilon_{n,i}\right)\bar{y}_{n,i} + (\beta_0 + \underline{\beta}^*\varepsilon_{n,i})' x_{n,i}, \tag{2.6}$$

which leads to the random coefficient interpretation of our SQAR model discussed below.

The random coefficient structural model perspective. Suppose that we have a structural relationship defined by

$$y_{n,i} = \lambda(v_{n,i})\bar{y}_{n,i} + \beta(v_{n,i})' x_{n,i},\ i = 1, \cdots, n, \tag{2.7}$$

where $v_{n,i}$ is a scalar random variable that aggregates all of the unobserved factors affecting the structural outcome $y_{n,i}$ for individual i and is independent of $x_{n,j}$ for all i, j. Following the lead of **CH**, we assume that $v_{n,i}$ are independent $U(0,1)$, and that the so-called structural equation function (SQF):

$$S_y(\tau \mid \bar{y},\ x) = \lambda(\tau)\bar{y} + \beta(\tau)' x \tag{2.8}$$

is strictly increasing in τ for each $(\bar{y}, x)$ in the support of $(\bar{y}_{n,i},\ x_{n,i})$.[3]

Observing that the event $\{y_{n,i} \le \lambda(\tau)\bar{y}_{n,i} + \beta(\tau)' x_{n,i}\}$ is equivalent to $\{v_{n,i} \le \tau\}$ for any $\tau \in (0,1)$ and that $v_{n,i}$ is independent of X_n, we have the following quantile restriction

$$\Pr(y_{n,i} \le \lambda(\tau)\bar{y}_{n,i} + \beta(\tau)' x_{n,i} \mid X_n) = \Pr(v_{n,i} \le \tau) = \tau \ \ \forall i. \tag{2.9}$$

Letting $u_{n,i} \equiv u_{\tau n,i} = y_{n,i} - \lambda(\tau)\bar{y}_{n,i} - \beta(\tau)' x_{n,i}$, we can reach the SQAR model specified in (2.3)-(2.5) with $\lambda^*(\tau)$, $\beta^*(\tau)$, $u^*_{n,i}$ and U^*_n being replaced by $\lambda(\tau)$, $\beta(\tau)$, $u_{n,i}$ and U_n, respectively. In addition, by the independence of the uniform random variables $v_{n,i}$'s, we have

$$\Pr(u_{n,1} \le 0, \cdots, u_{n,n} \le 0 \mid X_n) = \Pr(v_{n,1} \le \tau, \cdots, v_{n,n} \le \tau) = \tau^n, \tag{2.10}$$

implying that conditional on X_n, the indicator functions $1(u_{n,i} \le 0)$, $i = 1, \cdots, n$, are iid Bernoulli random variables.

Unifying the two perspectives. Clearly both (2.6) and (2.7) specify the data generating process (DGP) for Y_n as a system of simultaneous equations where the outcome $y_{n,i}$ for individual i is endogenously affected by the outcome $y_{n,j}$ for all $j \ne i$, and the non-observable scalar random variables $\varepsilon_{n,i}$ or $v_{n,i}$. Despite the fact that $\varepsilon_{n,i}$ enters the coefficient of $\bar{y}_{n,i}$ and $x_{n,i}$ linearly in (2.6) and $v_{n,i}$ enters the coefficient of $\bar{y}_{n,i}$ and $x_{n,i}$ nonlinearly in (2.7), we will show that the two specifications of DGPs are equivalent under some restrictions.

Let $F_n(\cdot)$ denote the distribution function of the iid variables $\varepsilon_{n,i}$ with the inverse given by $F_n^{-1}(\cdot)$. Let $v_{n,i} \equiv F_n(\varepsilon_{n,i})$. Then it is easy to see that under the restrictions:

$$\lambda(v_{n,i}) = \lambda_0 + \underline{\lambda} F_n^{-1}(v_{n,i}) \text{ and } \beta(v_{n,i}) = \beta_0 + \underline{\beta}^* F_n^{-1}(v_{n,i}), \tag{2.11}$$

the two DGPs are equivalent, and thus we can study the SQAR model by using either specification of the underlying DGP. When the above relationship holds, we will denote the population quantile residuals

[3] The strict monotonicity of $S_y(\tau \mid \bar{y},\ x)$ can easily be satisfied for economic data where both outcome and exogenous variables take only positive values. Since the elements of Y_n and the spatial weight matrix W_n only take nonnegative values, the support for the spatial lagged variable lies on the positive part of the real line. In this case, a sufficient condition for $S_y(\tau \mid \bar{y},\ x)$ to be strictly increasing in τ for each $(\bar{y}, x)$ in the support of $(\bar{y}_{n,i}, x_{n,i})$ would be $\partial\lambda(\tau)/\partial\tau > 0$ and $\partial\beta(\tau)/\partial\tau > 0$.

and coefficients simply as $u_{n,i}$ and $(\lambda(\tau), \beta(\tau))$ without starring as in (2.3)-(2.5). That is, our study will be based on the following formulation of the SQAR model

$$y_{n,i} = \lambda(\tau)\bar{y}_{n,i} + \beta(\tau)'x_{n,i} + u_{n,i},\ i = 1, \cdots, n, \tag{2.12}$$

or in matrix form,

$$Y_n = \lambda(\tau)W_nY_n + X_n\beta(\tau) + U_n, \tag{2.13}$$

where

$$\Pr(u_{n,i} \leq 0 \mid X_n) = \tau \ \ \forall\, i.^4 \tag{2.14}$$

Without assuming the uniform distribution of $v_{n,i}$, the strict exogeneity of X_n, and the strict monotonicity of the SQF, we find that it is too complicated to study (2.7). Nevertheless, under these three conditions, we can study (2.12)-(2.14) for any particular $\tau \in (0,1)$. In the following analysis, we always assume that this τ is fixed (say, $\tau = 0.5$) and then study the estimation and inferential problems associated with this SQAR model.[5]

Despite the independence of the Bernoulli random variables $1(u_{n,i} \leq 0)$, nothing in (2.7) or in the above assumptions guarantees the independence of $u_{n,i}$ across different individuals. In fact, under the restrictions in (2.11), we can show that

$$u_{n,i} = y_{n,i} - \lambda(\tau)\bar{y}_{n,i} - \beta(\tau)'x_{n,i} = \sigma_{n,i}[\varepsilon_{n,i} - Q_\varepsilon(\tau)] \tag{2.15}$$

where we recall $\sigma_{n,i} = 1 + \underline{\lambda}\bar{y}_{n,i} + \underline{\beta}'x_{n,i}$ and $\varepsilon_{n,i} = F_n^{-1}(v_{n,i})$. Noting that $\sigma_{n,i}$ are dependent of each other unless $\underline{\lambda} = 0$, $u_{n,i}$ are thus generally dependent across different individual units.

The DGP in (2.7) can be further extended to include some additional endogenous regressors without much of further technical difficulties, but to simplify our exposition we will concentrate on (2.7) and the associated SQAR model described above.

2.2 The IVQR Estimator of the SQAR Model

An important development in the literature of quantile regression is to allow endogeneity in the model and to introduce IV technique to handle the endogeneity (Chernozhukov and Hansen, 2005, 2006, 2008). If there exists an $n \times q$ instrument matrix $Z_n \equiv (z_{n,1}, ..., z_{n,n})'$ such that

$$\Pr(y_{n,i} \leq \lambda_{0\tau}\bar{y}_{n,i} + \beta_{0\tau}'x_{n,i}|x_{n,i}, z_{n,i}) = \tau \ \text{ a.s.} \tag{2.16}$$

then we can simply rely on **CH**'s idea and extend their IVQR estimation method to our SQAR model. Clearly, under the strict exogeneity of X_n and the quantile restriction in (2.15), we can simply choose,

[4] In spatial econometrics the exogenous regressor matrix X_n is often assumed to be a nonrandom matrix. In this case, (2.14) can be simply rewritten as $\Pr(u_{n,i} \leq 0) = \tau$.

[5] As kindly indicated by a referee, for median regression $Y_n = \lambda W_nY_n + X_n\beta + U_n$ can simply be treated as the DGP with U_n satisfying the median restriction $\Pr(u_{n,i} \leq 0|X_n) = 0.5, i = 1, \cdots, n$. We are interested in the consistent estimation of the structural parameters λ and β, thus it seems natural to directly impose conditions on $u_{n,i}$ in order to study the asymptotic properties of the estimators of these parameters.

as in typical GMM estimation of SAR models, the instrument matrix Z_n as the matrix consisting of linearly independent columns of W_nX_n or $[W_nX_n\ W_n^2X_n]$, such that (2.16) holds for the model specified in (2.13) and (2.15).

The IVQR idea can be made much simpler if the data $\{y_{n,i}, \bar{y}_{n,i}, x_{n,i}, z_{n,i}\}$ were iid. Note that the conditional probability $\Pr(y_{n,i} \le \lambda_{0\tau}\bar{y}_{n,i} + \beta'_{0\tau}x_{n,i}|x_{n,i}, z_{n,i})$ is a measurable function of $(x_{n,i}, z_{n,i})$. It follows from **CH** that to solve (2.16) is to find $(\lambda_{0\tau}, \beta_{0\tau})$ such that 0 is a solution to the ordinary QR of $y_{n,i} - \lambda_{0\tau}\bar{y}_{n,i} - \beta'_{0\tau}x_{n,i}$ on $(x_{n,i}, z_{n,i})$:

$$0 \in \underset{g\in\mathcal{G}}{\arg\min}\, \mathrm{E}\left[\rho_\tau(y_{n,i} - \lambda_{0\tau}\bar{y}_{n,i} - \beta'_{0\tau}x_{n,i} - g(x_{n,i}, z_{n,i}))\right], \tag{2.17}$$

where $\rho_\tau(u) \equiv [\tau - 1(u \le 0)]u$ with $1(\cdot)$ being the usual indicator function, and $\mathcal{G}$ is a class of measurable functions of (x, z) that is suitably restricted in applications. **CH** refer to this as the *instrumental variable quantile regression* (IVQR). Following **CH** we restrict $\mathcal{G}$ to the class of linear functions

$$\mathcal{G} = \{g\,(x, z) = \gamma' z : \gamma \in \Gamma\}$$

where Γ is a compact set in $\mathbb{R}^q$. In this case, the objective function in (2.17) leads immediately to the following finite sample analogue

$$Q_{n\tau}(\lambda, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \beta' x_{n,i} - \gamma' z_{n,i}). \tag{2.18}$$

That is, we restrict our attention to linear quantile regressions. Following **CH**'s arguments leading to (2.17) 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 $(\lambda_{0\tau}, \beta_{0\tau})$. Then the estimation may proceed as in **CH**.

However, as discussed earlier, our data $\{y_{n,i}, \bar{y}_{n,i}, x_{n,i}, z_{n,i}\}$ are not iid due to the spatial dependence reflected in $\{\bar{y}_{n,i}\}$ and the unknown heteroscedasticity in $\{y_{n,i}\}$. The question is whether the objective function $Q_{n\tau}(\lambda, \beta, \gamma)$ motivated by the iid data still remains valid for the spatial data. The theoretical results presented in the next section show it is still a valid objective function under certain additional regularity conditions. In this sense the IVQR estimator for our SQAR model can be defined in exactly the same way as that based on iid data. The difference is that the asymptotic properties under spatial dependence and unknown heteroscedasticity are much more involved than those under the iid assumption.

Let $\xi_{n,i} \equiv (x'_{n,i}, z'_{n,i})'$. The steps leading to the IVQR estimator of the SQAR model are summarized 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

$$(\hat{\beta}_{n\tau}(\lambda), \hat{\gamma}_{n\tau}(\lambda)) \equiv \underset{(\beta,\gamma)}{\arg\min}\, Q_{n\tau}(\lambda, \beta, \gamma); \tag{2.19}$$

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

$$\hat{\lambda}_{n\tau} = \arg\min_{\lambda} \hat{\gamma}_{n\tau}(\lambda)' \hat{A}_n \hat{\gamma}_{n\tau}(\lambda) \tag{2.20}$$

where $\hat{A}_n = A + o_p(1)$ for some positive definite matrix A; and finally

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

$$\hat{\beta}_{n\tau} \equiv \hat{\beta}_{n\tau}(\hat{\lambda}_{n\tau}). \tag{2.21}$$

Intuitively, to find $\hat{\lambda}_{n\tau}$ in step (ii), we look for a value of λ that makes the coefficient $\hat{\gamma}_{n\tau}(\lambda)$ of the instrumental variable as close to 0 as possible. The weight matrix $\hat{A}_n$ is used for asymptotic efficiency purpose. A convenient choice is to set A equal to the inverse of the asymptotic covariance matrix of $\sqrt{n}(\hat{\gamma}_{n\tau}(\lambda) - \gamma_{0\tau}(\lambda))$ where $\gamma_{0\tau}(\lambda)$ denotes the probability limit of $\hat{\gamma}_{n\tau}(\lambda)$, but that would require the consistent estimation of A at each point λ. For simplicity, we can simply set $\hat{A}_n$ to be an identity matrix; see Chernozhukov and Hansen (2006, 2008) for the case of iid data.

Remark 1. It is simple to implement the above IVQR procedure in practice: (i) for a given probability index τ of interest (e.g., $\tau = 0.5$ as for IV median regression), define a fine grid of values $\{\lambda_j, j = 1, \cdots, J\}$ that lie in a compact space (e.g., a compact subset of the interval $(-1, 1)$ when W_n is row normalized), (ii) for each j, run an ordinary QR of $y_{n,i} - \lambda_j \bar{y}_{n,i}$ on $\xi_{n,i}$ to obtain the coefficients $(\hat{\beta}_{n\tau}(\lambda_j), \hat{\gamma}_{n\tau}(\lambda_j))$, and (iii) choose $\hat{\lambda}_{n\tau}$ as the value among $\{\lambda_j, j = 1, \cdots, J\}$ that makes $\hat{\gamma}_{n\tau}(\lambda)' \hat{A}_n \hat{\gamma}_{n\tau}(\lambda)$ closest to zero.

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

$$S^0_{n\tau}(\lambda, \beta) = \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_{n,i} - \lambda \bar{y}_{n,i} - \beta' x_{n,i}) \xi_{n,i}, \tag{2.22}$$

$\psi_\tau(u) \equiv \tau - 1(u \leq 0)$ signifies the (directional) derivative of $\rho_\tau(u)$, and $\hat{P}_n$ is an estimated weight matrix. See also Abadie (1995) in a different context. Another example is to generalize the median estimator of Sakata (2007) to our spatial context. In contrast to the IVQR approach studied 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 function $S^0_{n\tau}(\cdot, \cdot)$ remains very important to the theoretical developments in this paper.

3 Asymptotic Properties of the IVQR Estimator

To study the asymptotic properties of the IVQR estimator for the SQAR model, we introduce some notation. For a matrix A_n, its Frobenius norm is denoted as $\|A_n\| = [\mathrm{tr}(A_n A_n')]^{1/2}$, and its (i, j)th element as $a_{n,ij}$. Similarly, for a vector a_n, $a_{n,i}$ denotes its ith element. Further, A_n is said to be uniformly bounded in absolute value if $\sup_{1\leq i\leq n, 1\leq j\leq n}|a_{n,ij}| < \infty$, and 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 $e_{n,i}$ be an $n\times 1$ vector with 1 in the ith place and 0 elsewhere, and I_n the $n\times n$ identity matrix. Let Λ and $\mathcal{B}$ denote the parameter spaces for λ and β, respectively, and "E" denote the expectation operator corresponding to the true parameter values $(\lambda_{0\tau}, \beta_{0\tau})$.

Let $S_n(\lambda) \equiv I_n - \lambda W_n$ and $G_n(\lambda) \equiv W_n S_n^{-1}(\lambda)$ for any value of λ. Let $\lambda_{0\tau} \equiv \lambda(\tau), \beta_{0\tau} \equiv \beta(\tau)$, $S_n \equiv S_n(\lambda_{0\tau})$, and $G_n \equiv G_n(\lambda_{0\tau})$. Noting that $S_n(I_n + \lambda_{0\tau}G_n) = I_n$, (2.13) has the reduced form

$$Y_n = S_n^{-1}(X_n\beta_{0\tau} + U_n) = X_n\beta_{0\tau} + \lambda_{0\tau}G_nX_n\beta_{0\tau} + S_n^{-1}U_n, \tag{3.1}$$

provided that S_n is nonsingular. This reduced form will be frequently used in the derivation of the asymptotic properties of the estimator proposed below.[6]

3.1 Assumptions

First we make some assumptions on the quantile residuals, the spatial weight matrix, and the parameter space for the spatial parameter.

Assumption 1. *(i)* $\Pr(u_{n,i} \leq 0) = \tau$ *for all* $i = 1, \cdots, n$. *(ii)* $sup_{n\geq 1} max_{1\leq i\leq n} E|u_{n,i}| \leq \overline{\mu} < \infty$. *(iii) The conditional distribution function* $F_{n,i}(\cdot|\bar{u}_{n,i})$ *of* $u_{n,i}$ *given* $\bar{u}_{n,i}$ *exhibits a conditional probability density function (pdf)* $f_{n,i}(\cdot|\bar{u}_{n,i})$ *that is uniformly bounded with bounded continuous first derivatives, where* $\bar{u}_{n,i} \equiv \sum_{k\neq i}^n g_{n,ik}u_{n,k}$ *and* $g_{n,ik}$ *denotes the* (i,k)*th element of* G_n.

Assumption 1 is a high level assumption because it imposes conditions on the quantile residual $u_{n,i} = u_{\tau n,i}$ directly as in the case of conditional mean or median regression. Under Assumption 3(i) below, Assumption 1(i) is equivalent to the quantile restriction in (2.14) which is implied by the DGP in (2.7) under the stated three conditions. Assumption 1(ii) is weak because it only requires the existence of the first moment of $u_{n,i}$ as in traditional quantile regressions. Even so, it is worthwhile to see some primitive conditions that ensure it to hold. Let $\bar{\Lambda}_n \equiv \mathrm{diag}\{\lambda(v_{n,1}), \cdots, \lambda(v_{n,n})\}$ and $B_{jn} \equiv \mathrm{diag}\{\beta_j(v_{n,1}), \cdots, \beta_j(v_{n,n})\}$, where $\beta_j(\cdot)$ denotes the jth component of $\beta(\cdot)$, and $j = 1, \cdots, p$. If $\|\bar{\Lambda}_nW_n\| \leq c_\lambda < 1$ almost surely (a.s.) for the Frobenius norm $\|\cdot\|$ (or any other matrix norm), then by Horn and Johnson (1985, Corollary 5.6.16) the reduced form for (2.7) exists and is given by

$$Y_n = \left(I_n - \bar{\Lambda}_nW_n\right)^{-1}\sum_{j=1}^p B_{jn}X_{jn} = \sum_{k=0}^{\infty}\left(\bar{\Lambda}_nW_n\right)^k\sum_{j=1}^p B_{jn}X_{jn},$$

where X_{jn} denotes the jth column of X_n. Under Assumptions 2-3 specified below, we can apply Lemma A.1 in the Appendix and (2.15) to show that the following three conditions are sufficient for Assumption 1(ii) to hold: a) $\mathrm{limsup}_n\|\bar{\Lambda}_nW_n\| \leq c_\lambda < 1$ a.s., b) $\beta(v_{n,i})$ are uniformly bounded a.s., and c) $\mathrm{E}(\varepsilon_{n,i}^2) = \sigma_\varepsilon^2 < \infty$.

[6] Lee (2004) showed a sufficient condition for the global identification of the SAR model given in (1.1) is that X_n and $G_n(\lambda_0)X_n\beta_0$ are not asymptotically multicolinear. Similarly, X_n and $G_nX_n\beta_{0\tau}$ in (3.1) and their relationship play a key role in the identification of the SQAR model as well.

Assumption 1(iii) specifies the conditions on the conditional density of $u_{n,i}$ given $\bar{u}_{n,i}$, which may not be straightforward to verify except for some special cases. For example, if $\lambda(v_{n,i}) = \lambda_0$ a.s. for all i in (2.11), (2.15) suggests that $u_{n,i}$ is independent of $u_{n,j}$ for all $j \neq i$ and thus of $\bar{u}_{n,i}$ under Assumptions 3(i)-(ii) below. In this case, $f_{n,i}(\cdot|\bar{u}_{n,i})$ reduces to the unconditional pdf of $u_{n,i}$, whose uniform boundedness can be easily ensured by specifying weak conditions on the marginal density of $\varepsilon_{n,i}$. This last assumption distinguishes our study significantly from that of Chernozhukov and Hansen (2006, 2008) in which iid quantile residuals are guaranteed and one only needs to specify conditions on the marginal pdf of the quantile residuals.

It is worth mentioning that despite the fact $F_{n,i}(0|\bar{u}_{n,i}) = \Pr(u_{n,i} \leq 0\ |\bar{u}_{n,i}) \neq \tau$ in general, we show in the proof of Theorem 3.2 that $\mathrm{E}[F_{n,i}(0|\bar{u}_{n,i})] = \tau$ under Assumptions 1(i) and (iii). This result plays an important role in the derivation of the asymptotic properties of our IVQR estimators.

Assumption 2. *The spatial weight matrix W_n is such that (i) its diagonal elements $w_{n,ii}$ 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 the parameter space Λ is compact with $\lambda_{0\tau}$ being an interior point, and (v) the diagonal elements $g_{n,ii}$ of G_n satisfy $\lim_{n\to\infty} \min_{1\leq i\leq n} \inf_{\lambda\in\Lambda} b_{n,i}(\lambda) = c_g > 0$ where $b_{n,i}(\lambda) = 1 + (\lambda_{0\tau} - \lambda)g_{n,ii}$.*

Like Lee (2004), Assumptions 2(i)-(iv) provide the essential features of the weight matrix for the model. Assumption 2(i) plays the normalization rule and it implies that no unit is viewed as its own neighbor. Assumption 2(ii) guarantees that the system of simultaneous equations has an equilibrium and that the reduced form is well defined. Kelejian and Prucha (1998, 1999, 2001) and Lee (2002, 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, Corollary 5.6.16), we see that the condition $\text{limsup}_n \|\lambda_{0\tau} W_n\| < 1$ is sufficient for S_n^{-1} being uniformly bounded in both row and column sums. In practice W_n is often specified to be row normalized in that $\sum_{j=1}^{n} w_{n,ij} = 1$ for all i. In many of these cases, no unit is assumed to be a neighbor to more than a given number, say, k of the other units. That is, for every j the cardinality of the set $\{w_{n,ij} \neq 0,\ i = 1, ..., n\}$ is less than or equal to k. In such cases, Assumption 2(iii) is satisfied. In the cases where the spatial weights are formulated in such a way that they decline as a function of some measure of physical or economic distance between individual units, Assumption A2(iii) will be typically satisfied. In particular, Lee (2002) demonstrates that Assumption 2(iii) is satisfied when W_n is row normalized with elements that are all nonnegative and are uniformly of order $O(1/n)$. It is worth mentioning that Assumptions 2(i)-(iii) are satisfied for the empirical models of Case (1991, 1992) and Case et al. (1993). The W_n and S_n matrices in Case (1991, 1992) are symmetric, and the row-normalization of W_n guarantees that A2(iii) is satisfied.

Assumption 2(iii) implies that $\{S_n^{-1}(\lambda)\}$ 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 as $n \to \infty$, $\ell_n \to \infty$ and $\ell_n/n \to 0$, then by Lemma A.1 in Appendix A, $g_{n,ii} = O(1/\ell_n) = o(1)$ 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 $z_{n,i}$, and weight $\hat{A}_n$, 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) $\hat{A}_n = A + o_p(1)$, where A is symmetric and positive definite.*

Assumptions 3(i)-(ii) are standard in spatial econometrics; see Kelejian and Prucha (1998, 1999). As remarked by Lee (2004), the regressors can be stochastic satisfying certain finite moment conditions. In most applications, Z_n is composed of linearly independent columns of $W_n X_n$ or $[W_n X_n, W_n^2 X_n]$, where the subset contains at least the linearly independent columns of $W_n X_n$ that are also linearly independent of the columns of X_n. The Z_n matrix chosen this way satisfies Assumption 3(ii) due to Assumptions 2(iii) and 3(i).

The formal study on the asymptotic properties (root-n consistency and asymptotic normality) relies on some additional assumptions on some important functions, all related to the sample objective function $Q_{n\tau}(\lambda, \beta, \gamma)$ defined in (2.18). First, for identification purpose, define the population counter part of $Q_{n\tau}(\lambda, \beta, \gamma)$ as

$$Q_\tau(\lambda, \beta, \gamma) = \lim_{n\to\infty} \mathrm{E}[Q_{n\tau}(\lambda, \beta, \gamma)], \tag{3.2}$$

and define, for a given λ,

$$(\beta_{0\tau}(\lambda), \gamma_{0\tau}(\lambda)) = \arg\min_{(\beta,\gamma)} Q_\tau(\lambda, \beta, \gamma), \tag{3.3}$$

which gives the population counter part of $(\hat{\beta}_{n\tau}(\lambda), \hat{\gamma}_{n\tau}(\lambda))$ defined in (2.19). Clearly, $\beta_{0\tau}(\lambda_{0\tau}) = \beta_{0\tau}$ and $\gamma_{0\tau}(\lambda_{0\tau}) = \gamma_{0\tau} = 0$. The former is obvious and the latter (saying that the true value for γ_τ is zero) follows from the arguments leading to (2.17). Next define

$$S_{n\tau}(\lambda, \beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(y_{n,i} - \lambda \bar{y}_{n,i} - \beta' x_{n,i} - \gamma' z_{n,i}) \xi_{n,i}, \tag{3.4}$$

which is the negative partial (directional) derivative of $Q_{n\tau}(\lambda, \beta, \gamma)$ with respect to $(\beta', \gamma')'$, which, when $\gamma = 0$, reduces to $S_{n\tau}^0(\lambda, \beta)$, the function introduced in (2.22). Finally, similar to $Q_\tau(\lambda, \beta, \gamma)$, define the population quantities in relation to $S_{n\tau}(\lambda, \beta, \gamma)$ and $S_{n\tau}^0(\lambda, \beta)$ as:

$$S_\tau(\lambda, \beta, \gamma) = \lim_{n\to\infty} \mathrm{E}[S_{n\tau}(\lambda, \beta, \gamma)] \;\text{ and }\; S_\tau^0(\lambda, \beta) = \lim_{n\to\infty} \mathrm{E}[S_{n\tau}^0(\lambda, \beta)]. \tag{3.5}$$

We impose the following assumption.

Assumption 4. *(i) $(\lambda_{0\tau}, \beta_{0\tau}')'$ is in the interior of a convex compact set $\Lambda \times B \subset \mathbb{R}^{1+p}$; (ii) $\partial S_\tau(\lambda, \beta, \gamma)/\partial(\beta', \gamma')$ is continuous and has full rank at $(\beta_{0\tau}(\lambda), \gamma_{0\tau}(\lambda))$ uniformly in $\lambda \in \Lambda$; (iii) $\partial S_\tau^0(\lambda, \beta)/\partial(\lambda, \beta')$ is continuous and has full column rank at $(\lambda_{0\tau}, \beta_{0\tau})$; (iv) If $S_\tau^0(\lambda^*, \beta^*) = 0$, then $\lambda^* = \lambda_{0\tau}$ and $\beta^* = \beta_{0\tau}$; and (v) $\beta_{0\tau}(\lambda)$ and $\gamma_{0\tau}(\lambda)$ are both 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 QR of $y_{n,i}-\lambda\bar{y}_{n,i}$ on $\xi_{n,i}$ and Assumption A4(iii) requires implicitly the relevance between the instruments $z_{n,i}$ and the endogenous variable $\bar{y}_{n,i}$. The former is crucial in establishing the uniform Bahadur representation for $(\hat{\beta}_{n\tau}(\lambda),\hat{\gamma}_{n\tau}(\lambda))$, and the latter is critical in establishing the asymptotic normality of our IVQR estimators $\hat{\lambda}_{n\tau}$ and $\hat{\beta}_{n\tau}$. Their importance warrants a detailed discussion.

To simplify the presentation, let $\alpha=(\beta',\gamma')'$, and similarly are $\alpha_{0\tau}$, $\alpha_{0\tau}(\lambda)$, and $\hat{\alpha}_{n\tau}(\lambda)$ defined. Also, let $\theta=(\lambda',\beta')'$ and $\theta_{0\tau}$ and $\hat{\theta}_{n\tau}$ be similarly defined. First, from (3.1), $\bar{Y}_n=G_nX_n\beta_{0\tau}+G_nU_n$. It follows that

$$y_{n,i}-\lambda\bar{y}_{n,i}=b_{n,i}(\lambda)u_{n,i}+(\lambda_{0\tau}-\lambda)c_{n,i}+\beta_{0\tau}'x_{n,i}, \tag{3.6}$$

where $b_{n,i}(\lambda)$ is defined in Assumption 2(v) and $c_{n,i}=e_{n,i}'G_nX_n\beta_{0\tau}+\sum_{k\neq i}^{n}g_{n,ik}u_{n,k}$. By the law of iterative expectations,

$$\mathrm{E}[S_{n\tau}(\lambda,\alpha)]=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\tau-F_{n,i}\left(\frac{a_{n,i}(\lambda,\alpha)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\right]\xi_{n,i}, \tag{3.7}$$

where $a_{n,i}(\lambda,\alpha)=(\lambda-\lambda_{0\tau})c_{n,i}+(\alpha-\alpha_{0\tau})'\xi_{n,i}$. Denote $J_{n\tau}(\lambda,\alpha)=-\partial\mathrm{E}[S_{n\tau}(\lambda,\alpha)]/\partial(\lambda,\alpha')$. Differentiating under the integral sign, we obtain

$$J_{n\tau}(\lambda,\alpha)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{f_{n,i}\left(\frac{a_{n,i}(\lambda,\alpha)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)\frac{\xi_{n,i}}{b_{n,i}(\lambda)}\left[c_{n,i}(\lambda,\alpha),\ \xi_{n,i}'\right]\right\}, \tag{3.8}$$

where $c_{n,i}(\lambda,\alpha)=c_{n,i}+a_{n,i}(\lambda,\alpha)g_{n,ii}/b_{n,i}(\lambda)$. This leads to two important quantities:

$$J_{n\tau}(\lambda)\equiv J_{n\tau}[\lambda,\alpha_{0\tau}(\lambda)] \;=\; \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{f_{n,i}\left(\frac{a_{n,i}(\lambda)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)\frac{\xi_{n,i}}{b_{n,i}(\lambda)}\left[c_{n,i}(\lambda),\ \xi_{n,i}'\right]\right\}, \tag{3.9}$$

$$J_{n\tau}\equiv J_{n\tau}(\lambda_{0\tau},\alpha_{0\tau}) \;=\; \frac{1}{n}\sum_{i=1}^{n}f_{n,i}(0|\bar{u}_{n,i})\xi_{n,i}\left[\mathrm{E}(c_{n,i}),\ \xi_{n,i}'\right], \tag{3.10}$$

where $a_{n,i}(\lambda)\equiv a_{n,i}[\lambda,\alpha_{0\tau}(\lambda)]$ and $c_{n,i}(\lambda)\equiv c_{n,i}[\lambda,\alpha_{0\tau}(\lambda)]$. Note that $c_{n,i}(\lambda_{0\tau},\alpha_{0\tau})=c_{n,i}$. Note that the S-quantities defined in (3.4) and (3.5) are all $(p+q)\times 1$ vectors, and the J-quantities defined in (3.8)-(3.10) are all $(p+q)\times(1+p+q)$ matrices. Partitioning $J_{n\tau}(\lambda)$ into $[J_{n\tau\lambda}(\lambda),J_{n\tau\alpha}(\lambda)]$ according to λ and α', and partitioning $J_{n\tau}$ into $[J_{n\tau\theta},J_{n\tau\gamma}]$ according to θ' and γ', we show under Assumptions 1(iii) and 2(v) that

$$(\partial/\partial\alpha')S_\tau(\lambda,\alpha)\big|_{\alpha=\alpha_{0\tau}(\lambda)}=-\lim_{n\to\infty}J_{n\tau\alpha}(\lambda)\ \text{and}\ \ (\partial/\partial\theta')S_\tau^0(\theta)\big|_{\theta=\lambda_{0\tau}}=-\lim_{n\to\infty}J_{n\tau\theta}.$$

Thus, the local identification condition of Assumption 4(ii) boils down to requiring the $(p+q)\times(p+q)$ matrix $J_{n\tau\alpha}(\lambda)$ to be positive definite for large enough n uniformly in $\lambda\in\Lambda$. Similarly, requiring $\partial S_\tau^0(\theta)/\partial\theta'$ to have full column rank at $\theta_{0\tau}$ in Assumption 4(iii) is equivalent to requiring $J_{n\tau\theta}$ to have full column rank for large enough n, or $\xi_{n,i}$ to be closely enough related to $\bar{y}_{n,i}$ as $e_{n,i}'G_nX_n\beta_{0\tau}$ is the leading term in $\mathrm{E}(c_{n,i})$ and also in $\bar{y}_{n,i}$.[7]

[7] As kindly pointed out by a referee, this assumption can be relaxed to allow for IVQR inference with weak identification. The dual approach considered in Chernozhukov and Hansen (2008) does not depend on it.

Noting that $S_\tau^0(\theta_{0\tau}) = 0$ by Assumption 1(ii), Assumption A4(iv) requires that $\theta_{0\tau}$ be the unique solution to $S_\tau^0(\theta) = 0$. This assumption is needed for the consistency of our estimator. It is weaker than the condition: $\mathrm{E}[S_{n\tau}^0(\theta^*)] = 0$ implies $\theta^* = \theta_{0\tau}$. The latter condition is usually satisfied when the data are iid or stationary time series. 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 et al. (2006) assumed a slightly weaker condition.

To proceed with our further discussions on the regularity conditions in allowing for dependence in the data, and in establishing the stochastic equicontinuity of certain functions, let $u_{n,i}(\lambda) = y_{n,i} - \lambda\bar{y}_{n,i} - \alpha_{0\tau}'(\lambda)\xi_{n,i}$. Then $u_{n,i}(\lambda_{0\tau}) = u_{n,i}$. Define

$$\eta_{n,i}(\lambda, \Delta) = -\left\{\psi_\tau[u_{n,i}(\lambda) - n^{-\frac{1}{2}}\Delta'\xi_{n,i}] - \psi_\tau[u_{n,i}(\lambda)]\right\}\xi_{n,i}.$$

Now, we state the following high level assumption.

Assumption 5. *$Var[n^{-\frac{1}{2}}\sum_{i=1}^n \eta_{n,i}(\lambda, \Delta)] = o(1)$ for each $\lambda \in \Lambda$* and $\|\Delta\| \leq M < \infty$.

Assumption 5 restricts the degree of dependence in the data. If $u_{n,i} \equiv u_{\tau n,i}$ are independent across i (say, when $\underline{\lambda}= 0$ in the definition of $\sigma_{n,i}$), we verify in an early version of the paper that under Assumptions 1-3, the following conditions are sufficient for Assumption 5 to hold: (i) the elements $w_{n,ij}$ of W_n are uniformly at most of order ℓ_n^{-1} such that as $n \to \infty$, $\ell_n \to \infty$ and $\ell_n/n \to 0$, (ii) $\sup_n \max_{1\leq i\leq n} \mathrm{E}u_{n,i}^2 \leq c_u < \infty$. (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. Nevertheless, it 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. In addition, the above conditions become insufficient if $u_{n,i}$'s are dependent across i.

Following Pinkse et al. (2007), we can control the variance of $n^{-\frac{1}{2}}\sum_{i=1}^n \eta_{n,i}(\lambda, \Delta)$ 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 et al. (2007).

Assumption 5*. *(i) Let $\eta_{n,ik}(\lambda, \Delta)$ denote the kth element of $\eta_{n,i}(\lambda, \Delta)$, $k = 1, \cdots, p+q$. For any $j = 1, \cdots, J$, let G_n^*, $G_n^* \subset G_{nj}$ be any sets for which $\forall t = 1, \cdots, m_{nj}$, if $G_{njt} \cap G_n^* \neq \varnothing$ then $G_{njt} \cap G_n^{**} = \varnothing$. Let $S_{nk}^*(\lambda) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{G}_n^*} \eta_{n,sk}(\lambda, \Delta)$ and $S_{nk}^{**}(\lambda) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{G}_n^{**}} \eta_{n,sk}(\lambda, \Delta)$. Then for each $\lambda \in \Lambda$,*

$$\|\mathrm{Cov}\,(\mathcal{S}_{nk}^*(\lambda),\ \mathcal{S}_{nk}^{**}(\lambda))\| \leq \sqrt{\mathrm{Var}\,(\mathcal{S}_{nk}^*(\lambda))\,\mathrm{Var}\,(\mathcal{S}_{nk}^{**}(\lambda))}\ \alpha_{m_{nj}},\ \ k = 1, \cdots, p+q,$$

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 between two quantities, corresponding to two different sets of subgroups of the same group. It is weaker than Assumption A in Pinkse et al. (2007). See that paper for a discussion on the need of dividing observations into finite J groups. 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 the appendix that Assumption 5* suffices to ensure Assumption 5.

Finally, define $\upsilon_{n\tau}(\lambda) = -\sqrt{n}\,[S_{n\tau}(\lambda) - \mathrm{E}S_{n\tau}(\lambda)]$, where $S_{n\tau}(\lambda) \equiv S_{n\tau}[\lambda, \alpha_{0\tau}(\lambda)]$. Define $S_{\tau}(\lambda) \equiv S_{\tau}[\lambda, \alpha_{0\tau}(\lambda)]$. We make the following assumption.

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

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

$$\mathcal{M} = \{g(y, \bar{y}, \xi; \lambda) = 1(y - \lambda\bar{y} - \alpha_{0\tau}'(\lambda)\xi \leq 0)\xi : \lambda \in \Lambda\}.$$

If $(y_{n,i}, \bar{y}_{n,i}, \xi_{n,i})$ are iid with probability law P_n, it is easy to verify that $\{g(\cdot;\lambda) : \lambda \in \Lambda\}$ is an Euclidean class with envelope $\overline{g}$ such that $\overline{g}(y, \bar{y}, \xi) \equiv \|\xi\|$ and $\int \overline{g}(y, \bar{y}, \xi) dP_n = \mathrm{E}\|\xi\| < \infty$. 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)]. For spatial data, we can show that Assumption 6 holds provided $\lim_{n\to\infty} \ell_n/\sqrt{n} = c \in (0, \infty]$. This latter condition with $c = \infty$ has been assumed in Lee (2002) for the consistency of least squares estimation of SAR models and in Robinson (2010) for the adaptive estimation of SAR models. Nevertheless, it is not necessary here because there may exist other cases where Assumption 6 holds.

3.2 Asymptotic Distribution

We are now ready to state the asymptotic property of the IVQR estimators defined in (2.19)-(2.21) above. The following theorem shows that the QR estimator $\hat{\alpha}_{n\tau}(\lambda)$ has a Bahadur representation uniformly in λ.

Theorem 3.1 *Suppose Assumptions 1-6 hold. Then, we have*

$$\sqrt{n}[\hat{\alpha}_{n\tau}(\lambda) - \alpha_{0\tau}(\lambda)] = J_{n\tau\alpha}^{-1}(\lambda)\sqrt{n}S_{n\tau}(\lambda) + o_p(1) \textit{ uniformly in } \lambda \in \Lambda.$$

Note that $\sup_{\lambda} |\sqrt{n}S_{n\tau}(\lambda)| = O_p(1)$ by Lemma A.4 and $\sup_{\lambda}|J_{n\tau}(\lambda)| = O(1)$ by Assumptions 1-3 and Lemma A.1. An immediate consequence of Theorem 3.1 is that $\|\hat{\alpha}_{n\tau}(\lambda) - \alpha_{0\tau}(\lambda)\| = O_p(n^{-1/2})$ uniformly in $\lambda \in \Lambda$. This uniform $\sqrt{n}$-consistency for $\hat{\alpha}_{n\tau}(\lambda)$ is crucial in proving the $\sqrt{n}$-consistency of $\hat{\theta}_{n\tau}$ presented in the next theorem.

Let $J_\tau = \lim_{n\to\infty} J_{n\tau}$ with $J_{n\tau}$ being defined in (3.10), partitioned as $J_\tau = [J_{\tau\lambda}, J_{\tau\beta}, J_{\tau\gamma}]$ according to λ, β' and γ'. Partition conformably $[J_{\tau\beta}, J_{\tau\gamma}]^{-1} = [\bar{J}'_{\tau\beta}, \bar{J}'_{\tau\gamma}]'$, where $\bar{J}_{\tau\beta}$ is $p \times (p+q)$ and $\bar{J}_{\tau\gamma}$ is $q \times (p+q)$. We have the main results for the asymptotic normality of our IVQR estimator.

Theorem 3.2 *Suppose that $J_{n\tau\alpha}$ is of full rank and Assumptions 1-6 hold. Then, we have*

$$\sqrt{n}(\hat{\theta}_{n\tau} - \theta_{0\tau}) \xrightarrow{d} N[0, \ \ \Omega_\tau(A)\Sigma_{0\tau}\Omega'_\tau(A)],$$

where $\Sigma_{0\tau} \equiv \lim_{n\to\infty} \mathrm{Var}[\sqrt{n}S_{n\tau}(\lambda_{0\tau})] = \tau(1-\tau)\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \xi_{n,i}\xi'_{n,i}$,

$$\Omega_\tau(A) \equiv \left[H_\tau(A)J_{\tau\lambda}[J'_{\tau\lambda}H_\tau(A)J_{\tau\lambda}]^{-1}, \ \ \{I_{p+q} - J_{\tau\lambda}[J'_{\tau\lambda}H_\tau(A)J_{\tau\lambda}]^{-1}J'_{\tau\lambda}H_\tau(A)\}\bar{J}'_{\tau\beta}\right]',$$

and $H_\tau(A) = \bar{J}'_{\tau\gamma}A\bar{J}_{\tau\gamma}$.

The asymptotic variance of $\sqrt{n}(\hat{\theta}_{n\tau} - \theta_{0\tau})$ depends on the choice of the weight matrix A in the case of over-identification ($q > 1$). In the case of just identification ($q = 1$), however, A becomes a scalar which is canceled out in the variance formula, and hence the choice of A does not affect the asymptotic variance of $\sqrt{n}(\hat{\theta}_{n\tau} - \theta_{0\tau})$. A much simpler result thus follows, writing $J_{\tau\theta} = [J_{\tau\lambda}, J_{\tau\beta}]$,

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

$$\sqrt{n}(\hat{\theta}_{n\tau} - \theta_{0\tau}) \xrightarrow{d} N[0, \ \ J^{-1}_{\tau\theta}\Sigma_{0\tau}(J^{-1}_{\tau\theta})'].$$

Remark 3. In the case of over-identification ($q > 1$), the choice of the weight matrix $\hat{A}_n$ in the objective function $\hat{\gamma}_{n\tau}(\lambda)'\hat{A}_n\hat{\gamma}_{n\tau}(\lambda)$ generally matters. As discussed in Section 2, it is natural to choose $\hat{A}_n$ to be a consistent estimator of the inverse of the asymptotic covariance matrix of $\sqrt{n}[\hat{\gamma}_{n\tau}(\lambda) - \gamma_{0\tau}(\lambda)]$. In this case, since A is generally λ-dependent, it needs to be estimated at each grid point of λ in the process of optimization.

Remark 4. Consider $\tilde{\theta}_{n\tau} = \arg\min_\theta S^0_{n\tau}(\theta)'\hat{P}_n S^0_{n\tau}(\theta)$, the method of moments (MM) estimator with $S^0_{n\tau}(\theta)$ being defined in (2.22). Under conditions similar to those imposed in Assumptions 1-6, we can establish the asymptotic normality of $\tilde{\theta}_{n\tau}$ In particular, when we choose the optimal weight $\hat{P}_n = (n^{-1}\sum_{i=1}^n \xi_{n,i}\xi'_{n,i})^{-1}$, the asymptotic covariance of $\tilde{\theta}_{n\tau}$ is equal to $(J'_{\tau\theta}\Sigma^{-1}_{0\tau}J_{\tau\theta})^{-1}$, which becomes $J^{-1}_{\tau\theta}\Sigma_{0\tau}(J^{-1}_{\tau\theta})'$ in the case of just identification ($q = 1$). Consequently, if we restrict our attention to the MM estimator $\tilde{\theta}_{n\tau}$ by choosing only one IV for $\bar{y}_{n,i}$ with the above optimal weight, and the IVQR estimator $\hat{\theta}_{n\tau}$ based on the same IV, then the two estimators are asymptotically equivalent. Nevertheless, $\hat{\theta}_{n\tau}$ is usually less efficient than the MM estimator $\tilde{\theta}_{n\tau}$ that uses more than one IV in the optimal way.

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 matrix $\Omega_\tau(A)\Sigma_{0\tau}\Omega'_\tau(A)$. The definition of $\Sigma_{0\tau}$ leads naturally to a consistent estimator of it as $\Sigma_{n\tau} \equiv \tau(1-\tau)\frac{1}{n}\sum_{i=1}^n \xi_{n,i}\xi'_{n,i}$. The estimation of $\Omega_\tau(A)$ depends on the estimation of A and J_τ. The former is discussed in Remark 3 above. For the latter, we follow Powell (1991) and estimate $J_\tau = [J_{\tau\lambda}, J_{\tau\alpha}]$ by $\hat{J}_\tau = [\hat{J}_{\tau\lambda}, \hat{J}_{\tau\alpha}]$, where

$$\begin{aligned}\hat{J}_{\tau\lambda} &\equiv (2nh)^{-1}\sum_{i=1}^{n} 1\left\{|\hat{u}_{n,i}| \leq h\right\}\xi_{n,i}\hat{c}_{n,i},\\ \hat{J}_{\tau\alpha} &\equiv (2nh)^{-1}\sum_{i=1}^{n} 1\left\{|\hat{u}_{n,i}| \leq h\right\}\xi_{n,i}\xi_{n,i}',\end{aligned}$$

$\hat{c}_{n,i} = e_{n,i}' G_n X_n \hat{\beta}_\tau + \sum_{k\neq i}^{n} g_{n,ik}\hat{u}_{n,k}$, $\hat{u}_{n,i} = y_{n,i} - \hat{\lambda}_\tau \bar{y}_{n,i} - x_{n,i}'\hat{\beta}_\tau$, and $h \equiv h(n)$ is a bandwidth parameter such that as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$. Following Koenker (2005, pp. 80-81), we set

$$h = \hat{\kappa}\left[\Phi^{-1}(\tau + cn^{-1/3}) - \Phi^{-1}(\tau - cn^{-1/3})\right],$$

where Φ^{-1} is the inverse of the standard normal CDF, $\hat{\kappa}$ is a robust estimate of scale/standard deviation, e.g., $\hat{\kappa}$ = median$|\hat{u}_{n,i}$-median$(\hat{u}_{n,i})|/0.6745$ (Hogg and Craig, 1995, p. 390), and c is a proportional constant. In the application below we set $c = 0.5$. [8]

4 Simulation and Application

In this section we report some simulation results to investigate the finite sample performance of our IVQR estimator of the SQAR 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 (2010). The GMM estimator of Lee (2007) denoted by GMM0 and the robust GMM estimator of Lin and Lee (2010) denoted by GMMR require initial estimates of λ and β and a weight matrix. We follow exactly Lin and Lee (2010, p. 40, paragraph 3) for the definitions of 2SLS, GMM0 and GMMR. In particular, GMM0 uses $(G_n - \frac{1}{n}\text{tr}G_n I_n)$ as the weight matrix and $(G_n X_n \beta, X_n)$ as the IV matrix, GMMR uses the same IV but a different weight $G_n - \text{Diag}(G_n)$, and 2SLS is what they called the simple 2SLS which uses the linearly independent columns of $(W_n X_n,\ X_n)$ as the IV matrix. This 2SLS also serves as initial estimates for GMM0 and GMMR. Note that the estimator proposed by Kelejian and Prucha (2010) is essentially the 2SLS estimator when their SARAR model is reduced to SAR model.

The DGP employed in the simulations takes the form of (2.7): $y_{n,i} = \lambda(v_{n,i})\bar{y}_{n,i} + \beta(v_{n,i})' x_{n,i}$, where $x_{n,i} = \left(1,\ x_i^0\right)$, $\left\{x_i^0\right\}$ are iid $N(0,1)$,

$$\begin{aligned}\lambda(v_{n,i}) &= 0.5 + 0.1\ F_n^{-1}(v_{n,i}),\ \text{and}\\ \beta(v_{n,i}) &= (2.0,\ 1.0)' + (0.5,\ 0.5)' F_n^{-1}(v_{n,i}),\end{aligned}$$

for $i = 1, \cdots, n$, where $\{v_{n,i}\}$ are iid $U(0,1)$, and F_n is chosen to be (i) standard normal, (ii) standardized t_3, and (iii) standardized χ_3^2. With these specifications, the values for $\lambda(\tau)$ and $\beta(\tau)' = \{\beta_1(\tau), \beta_2(\tau)\}$ under different τ and F_n are summarized as follows.

[8] Alternatively one could follow Pakes and Pollard (1989) and Honoré and Hu (2004) and estimate the J_τ-quantities using numerical derivatives.

Table 1. Summary of True Quantile Parameters used in Simulations

	Standard normal			Standardized t_3			Standardized χ_3^2		
τ	$\lambda(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$	$\lambda(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$	$\lambda(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$
0.25	0.4326	1.6628	0.6628	0.4558	1.7792	0.7792	0.4270	1.6351	0.6351
0.50	0.5000	2.0000	1.0000	0.5000	2.0000	1.0000	0.4741	1.8706	0.8706
0.75	0.5674	2.3372	1.3372	0.5442	2.2208	1.2208	0.5452	2.2262	1.2262

The weight matrix W_n is generated under 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 at a slower rate. To be exact, in case (i) we 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 $R = \lfloor n^{0.6} \rfloor$, 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.[9] The sample sizes used are 100, 200, 500, and 1000. Each set of simulation results is based 1,000 Monte Carlo samples.

The first set of Monte Carlo experiments is carried out at $\tau = 0.5$, which allows comparisons of our method with the existing QMLE and GMM methods, in particular when the errors are symmetrically distributed. This is because the latter methods are applicable to the standard SAR model with the zero mean (in errors) restriction. Also note that, in finding the IVQR estimate, we used the grid search method (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.[10] Tables 2 and 3 summarize the Monte Carlo bias, the standard deviation (StDev) and the root mean squared errors (RMSE) of the various estimators, where Table 2 corresponds to the spatial parameter $\lambda(0.5)$, and Table 3 the slope parameter $\beta_2(0.5)$. From Table 2 we see that all estimators (except 2SLS) of $\lambda(0.5)$ perform similarly, although a slight edge may go to our IVQR estimator when errors are nonnormal. However, the results in Table 3 clearly show that our IVQR estimator of $\beta_2(0.5)$ outperforms all others, in particular when the errors are positively skewed. These results show the robustness of IVQR estimator against both excess skewness and kurtosis.

The second set of Monte Carlo experiments focuses on the behavior of IVQR estimator at other quantile points. Tables 4 and 5 present results for $\tau = 0.25$ and 0.75, where Table 4 corresponds to $\lambda(\tau)$, and Table 5 corresponds to $\beta_2(\tau)$. The results indicate that the IVQR estimator for the SQAR model behaves quite well in general, and are consistent with the theoretical predictions. It is generally robust against nonnormality and heteroscedasticity, and as the sample size increases, both StDev and RMSE decline and the magnitude of decrease is generally consistent with the $\sqrt{n}$-asymptotics.

[9] Under Rook contiguity, spatial units are considered as the neighbors of a certain spatial unit if they fall above, below, left or right of this spatial unit. Under group interaction, one needs to make a final adjustment to make sure that $\sum_{r=1}^{R} m_r = n$. Note that this spatial layout generalizes that considered in Case (1991) and used in Lee (2004). See these two papers for more discussions on group interactions.

[10] 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 2. Empirical Bias [StDev]{RMSE} for Estimators of $\lambda(\tau)$, $\tau = 0.5$

n	Estimator	Standard Normal	Standized t_3	Standardized χ^2_3
		`(a) Spatial Layout: Rook Contiguity`		
100	QML	-0.0191 [.0826]{.0848}	-0.0179 [.0767]{.0788}	0.0046 [.0818]{.0819}
	2SLS	-0.0051 [.1223]{.1224}	-0.0040 [.1230]{.1231}	0.0163 [.1246]{.1257}
	GMM0	-0.0159 [.0822]{.0837}	-0.0129 [.0941]{.0950}	0.0083 [.0817]{.0821}
	GMMR	-0.0182 [.0820]{.0840}	-0.0150 [.0924]{.0936}	0.0064 [.0813]{.0815}
	IVQR	0.0006 [.0948]{.0948}	0.0001 [.0604]{.0604}	0.0019 [.0776]{.0776}
200	QML	-0.0103 [.0665]{.0673}	-0.0115 [.0627]{.0638}	0.0131 [.0694]{.0706}
	2SLS	-0.0047 [.1177]{.1178}	-0.0057 [.1173]{.1174}	0.0197 [.1223]{.1238}
	GMM0	-0.0066 [.0671]{.0674}	-0.0074 [.0671]{.0675}	0.0166 [.0698]{.0718}
	GMMR	-0.0066 [.0671]{.0674}	-0.0076 [.0659]{.0663}	0.0165 [.0698]{.0717}
	IVQR	-0.0006 [.0852]{.0852}	-0.0006 [.0543]{.0543}	-0.0010 [.0711]{.0711}
500	QML	-0.0040 [.0430]{.0432}	-0.0046 [.0423]{.0425}	0.0221 [.0452]{.0503}
	2SLS	-0.0051 [.0764]{.0766}	-0.0044 [.0774]{.0775}	0.0208 [.0779]{.0807}
	GMM0	-0.0028 [.0433]{.0434}	-0.0033 [.0423]{.0425}	0.0239 [.0456]{.0515}
	GMMR	-0.0019 [.0433]{.0434}	-0.0025 [.0423]{.0423}	0.0247 [.0455]{.0518}
	IVQR	-0.0034 [.0534]{.0535}	-0.0022 [.0335]{.0336}	-0.0028 [.0453]{.0453}
1000	QML	0.0017 [.0296]{.0297}	0.0006 [.0310]{.0310}	0.0268 [.0333]{.0427}
	2SLS	0.0018 [.0516]{.0517}	0.0011 [.0522]{.0522}	0.0274 [.0544]{.0609}
	GMM0	0.0027 [.0297]{.0298}	0.0019 [.0309]{.0309}	0.0280 [.0336]{.0437}
	GMMR	0.0026 [.0297]{.0298}	0.0017 [.0307]{.0308}	0.0278 [.0336]{.0436}
	IVQR	0.0001 [.0385]{.0385}	0.0001 [.0240]{.0240}	0.0008 [.0331]{.0331}
		`(b) Spatial Layout: Group Interaction`		
100	QML	-0.0278 [.0844]{.0889}	-0.0263 [.0804]{.0846}	-0.0065 [.0842]{.0845}
	2SLS	-0.0301 [.1592]{.1621}	-0.0306 [.1640]{.1668}	-0.0066 [.1631]{.1632}
	GMM0	-0.0203 [.0833]{.0857}	-0.0192 [.0789]{.0812}	0.0015 [.0829]{.0829}
	GMMR	-0.0154 [.0832]{.0846}	-0.0149 [.0787]{.0801}	0.0061 [.0820]{.0823}
	IVQR	-0.0131 [.1141]{.1149}	-0.0061 [.0699]{.0702}	-0.0136 [.1052]{.1060}
200	QML	-0.0123 [.0717]{.0728}	-0.0116 [.0681]{.0691}	0.0116 [.0702]{.0711}
	2SLS	-0.0180 [.1204]{.1217}	-0.0198 [.1271]{.1286}	0.0074 [.1223]{.1226}
	GMM0	-0.0078 [.0707]{.0712}	-0.0079 [.0675]{.0679}	0.0167 [.0696]{.0715}
	GMMR	-0.0120 [.0707]{.0717}	-0.0114 [.0668]{.0677}	0.0121 [.0679]{.0690}
	IVQR	-0.0067 [.0764]{.0767}	-0.0039 [.0497]{.0499}	-0.0031 [.0652]{.0653}
500	QML	-0.0077 [.0521]{.0527}	-0.0073 [.0511]{.0517}	0.0169 [.0533]{.0559}
	2SLS	-0.0202 [.1126]{.1144}	-0.0204 [.1159]{.1176}	0.0071 [.1095]{.1098}
	GMM0	-0.0040 [.0520]{.0522}	-0.0042 [.0510]{.0512}	0.0204 [.0530]{.0568}
	GMMR	-0.0063 [.0518]{.0522}	-0.0060 [.0503]{.0506}	0.0178 [.0521]{.0551}
	IVQR	-0.0125 [.0914]{.0922}	-0.0056 [.0551]{.0554}	-0.0080 [.0740]{.0744}
1000	QML	-0.0063 [.0405]{.0410}	-0.0074 [.0399]{.0406}	0.0181 [.0422]{.0459}
	2SLS	-0.0090 [.0719]{.0725}	-0.0102 [.0706]{.0713}	0.0148 [.0745]{.0760}
	GMM0	-0.0044 [.0401]{.0404}	-0.0054 [.0395]{.0399}	0.0202 [.0419]{.0465}
	GMMR	-0.0036 [.0401]{.0402}	-0.0043 [.0392]{.0395}	0.0209 [.0418]{.0467}
	IVQR	-0.0030 [.0471]{.0472}	-0.0013 [.0295]{.0295}	-0.0018 [.0399]{.0400}

Table 3. Empirical Bias [StDev]{RMSE} for Estimators of $\beta_2(\tau)$, $\tau = 0.5$

n	Estimator	Standard Normal	Standardized t_3	Standardized χ^2_3
		`(a) Spatial Layout: Rook Contiguity`		
100	QML	0.0139 [.1325]{.1332}	0.0450 [.8882]{.8893}	0.1490 [.1398]{.2043}
	2SLS	0.0068 [.1381]{.1383}	0.0534 [1.5992]{1.6001}	0.1416 [.1440]{.2020}
	GMM0	0.0073 [.1326]{.1328}	0.0626 [1.6636]{1.6648}	0.1418 [.1399]{.1993}
	GMMR	0.0077 [.1327]{.1329}	0.0631 [1.6664]{1.6676}	0.1422 [.1402]{.1997}
	IVQR	0.0007 [.1028]{.1028}	0.0000 [.0653]{.0653}	0.0055 [.0864]{.0865}
200	QML	0.0060 [.0914]{.0916}	0.0081 [.0931]{.0934}	0.1354 [.0951]{.1655}
	2SLS	0.0013 [.0918]{.0918}	0.0033 [.0926]{.0926}	0.1302 [.0951]{.1612}
	GMM0	0.0017 [.0911]{.0911}	0.0039 [.0929]{.0930}	0.1309 [.0947]{.1616}
	GMMR	0.0017 [.0911]{.0911}	0.0039 [.0929]{.0930}	0.1309 [.0947]{.1616}
	IVQR	-0.0036 [.0684]{.0685}	-0.0020 [.0438]{.0438}	-0.0011 [.0584]{.0584}
500	QML	0.0105 [.0556]{.0566}	0.0107 [.0597]{.0607}	0.1395 [.0580]{.1511}
	2SLS	0.0092 [.0570]{.0578}	0.0090 [.0605]{.0611}	0.1381 [.0593]{.1503}
	GMM0	0.0089 [.0557]{.0564}	0.0091 [.0595]{.0602}	0.1376 [.0579]{.1493}
	GMMR	0.0087 [.0557]{.0564}	0.0089 [.0596]{.0602}	0.1374 [.0579]{.1491}
	IVQR	0.0019 [.0402]{.0403}	0.0015 [.0252]{.0253}	0.0023 [.0341]{.0342}
1000	QML	0.0105 [.0404]{.0418}	0.0103 [.0411]{.0423}	0.1403 [.0416]{.1463}
	2SLS	0.0097 [.0417]{.0428}	0.0094 [.0427]{.0438}	0.1394 [.0433]{.1459}
	GMM0	0.0096 [.0404]{.0416}	0.0093 [.0410]{.0421}	0.1393 [.0416]{.1454}
	GMMR	0.0096 [.0404]{.0416}	0.0094 [.0410]{.0420}	0.1393 [.0416]{.1454}
	IVQR	0.0002 [.0276]{.0276}	0.0002 [.0174]{.0174}	0.0005 [.0235]{.0235}
		`(b) Spatial Layout: Group Interaction`		
100	QML	0.0139 [.1169]{.1177}	0.0112 [.1146]{.1151}	0.1421 [.1188]{.1852}
	2SLS	0.0068 [.1192]{.1194}	0.0041 [.1215]{.1216}	0.1342 [.1199]{.1800}
	GMM0	0.0071 [.1168]{.1170}	0.0047 [.1159]{.1160}	0.1349 [.1181]{.1793}
	GMMR	0.0067 [.1167]{.1169}	0.0043 [.1167]{.1168}	0.1346 [.1180]{.1790}
	IVQR	0.0031 [.1002]{.1003}	0.0024 [.0633]{.0634}	0.0080 [.0855]{.0859}
200	QML	0.0075 [.0798]{.0802}	0.0076 [.0814]{.0818}	0.1370 [.0820]{.1597}
	2SLS	0.0040 [.0804]{.0805}	0.0045 [.0827]{.0829}	0.1330 [.0823]{.1564}
	GMM0	0.0020 [.0796]{.0797}	0.0028 [.0814]{.0815}	0.1313 [.0815]{.1546}
	GMMR	0.0029 [.0797]{.0797}	0.0036 [.0816]{.0816}	0.1323 [.0817]{.1555}
	IVQR	-0.0002 [.0655]{.0655}	0.0001 [.0421]{.0421}	0.0016 [.0565]{.0565}
500	QML	0.0036 [.0572]{.0573}	0.0028 [.0576]{.0576}	0.1337 [.0589]{.1461}
	2SLS	0.0022 [.0577]{.0577}	0.0014 [.0581]{.0581}	0.1322 [.0592]{.1448}
	GMM0	0.0021 [.0572]{.0572}	0.0014 [.0576]{.0576}	0.1322 [.0588]{.1447}
	GMMR	0.0022 [.0572]{.0572}	0.0015 [.0576]{.0576}	0.1323 [.0589]{.1448}
	IVQR	0.0005 [.0390]{.0390}	0.0003 [.0245]{.0245}	0.0008 [.0329]{.0329}
1000	QML	0.0023 [.0390]{.0390}	0.0018 [.0394]{.0395}	0.1314 [.0392]{.1371}
	2SLS	0.0017 [.0392]{.0392}	0.0013 [.0396]{.0397}	0.1308 [.0393]{.1366}
	GMM0	0.0015 [.0390]{.0390}	0.0011 [.0395]{.0395}	0.1307 [.0392]{.1364}
	GMMR	0.0015 [.0390]{.0390}	0.0011 [.0395]{.0395}	0.1306 [.0391]{.1364}
	IVQR	-0.0006 [.0271]{.0271}	-0.0004 [.0171]{.0171}	0.0001 [.0234]{.0234}

Table 4. Empirical Bias [StDev]{RMSE} for IVQR Estimate of $\lambda(\tau)$, $\tau = 0.25, 0.75$

τ	n	Standard Normal	Standardized t_3	Standardized χ_3^2
		(a) Spatial Layout: Rook Contiguity		
0.25	100	-0.0084 [.1613]{.1615}	0.0213 [.1298]{.1316}	0.0006 [.0894]{.0894}
	200	-0.0067 [.0999]{.1001}	0.0042 [.1010]{.1011}	-0.0066 [.0737]{.0740}
	500	0.0095 [.0579]{.0587}	-0.0023 [.0460]{.0461}	0.0052 [.0321]{.0325}
	1000	0.0059 [.0420]{.0424}	-0.0022 [.0325]{.0326}	0.0026 [.0235]{.0236}
0.75	100	0.0185 [.1675]{.1685}	-0.0022 [.1229]{.1229}	-0.0059 [.2214]{.2214}
	200	-0.0214 [.1166]{.1185}	-0.0138 [.0693]{.0707}	-0.0420 [.1108]{.1184}
	500	-0.0256 [.0596]{.0649}	-0.0131 [.0449]{.0468}	-0.0319 [.0806]{.0867}
	1000	-0.0130 [.0419]{.0439}	-0.0017 [.0324]{.0324}	-0.0064 [.0526]{.0530}
		(b) Spatial Layout: Group Interaction		
0.25	100	-0.0162 [.1666]{.1674}	-0.0270 [.1282]{.1310}	-0.0158 [.1316]{.1325}
	200	0.0080 [.1117]{.1120}	-0.0330 [.1280]{.1322}	-0.0116 [.0588]{.0599}
	500	-0.0201 [.1117]{.1135}	0.0097 [.0550]{.0559}	-0.0015 [.0391]{.0391}
	1000	0.0051 [.0415]{.0418}	-0.0021 [.0383]{.0384}	0.0025 [.0290]{.0291}
0.75	100	-0.0360 [.1606]{.1646}	-0.0490 [.1946]{.2006}	-0.0294 [.2015]{.2036}
	200	-0.0203 [.0761]{.0788}	-0.0214 [.1358]{.1375}	-0.0308 [.1241]{.1279}
	500	-0.0208 [.0590]{.0626}	-0.0147 [.0665]{.0681}	0.0023 [.0849]{.0850}
	1000	-0.0125 [.0544]{.0558}	-0.0054 [.0541]{.0543}	-0.0227 [.0569]{.0612}

Table 5. Empirical Bias [StDev]{RMSE} for IVQR Estimate of $\beta_2(\tau)$, $\tau = 0.25, 0.75$

τ	n	Standard Normal	Standardized t_3	Standardized χ_3^2
		(a) Spatial Layout: Rook Contiguity		
0.25	100	0.0633 [.1138]{.1302}	0.0648 [.0863]{.1080}	0.0195 [.0592]{.0623}
	200	0.0809 [.0762]{.1111}	0.0359 [.0573]{.0677}	0.0151 [.0444]{.0469}
	500	0.0469 [.0405]{.0620}	0.0457 [.0306]{.0550}	0.0231 [.0258]{.0346}
	1000	0.0585 [.0296]{.0656}	0.0284 [.0223]{.0361}	0.0211 [.0184]{.0280}
0.75	100	-0.1007 [.1089]{.1483}	-0.0316 [.0856]{.0912}	-0.0436 [.1564]{.1624}
	200	-0.0406 [.0843]{.0936}	-0.0408 [.0571]{.0702}	-0.0567 [.0880]{.1047}
	500	-0.0353 [.0476]{.0593}	-0.0298 [.0365]{.0471}	-0.0515 [.0557]{.0758}
	1000	-0.0531 [.0331]{.0625}	-0.0352 [.0226]{.0419}	-0.0527 [.0384]{.0652}
		(b) Spatial Layout: Group Interaction		
0.25	100	0.0702 [.1158]{.1354}	0.0485 [.0698]{.0850}	0.0076 [.0786]{.0790}
	200	0.0452 [.0729]{.0858}	0.0509 [.0487]{.0704}	0.0414 [.0454]{.0614}
	500	0.0401 [.0496]{.0638}	0.0240 [.0317]{.0398}	0.0226 [.0266]{.0349}
	1000	0.0398 [.0313]{.0506}	0.0283 [.0210]{.0352}	0.0183 [.0181]{.0257}
0.75	100	-0.0354 [.1196]{.1247}	-0.0510 [.1034]{.1153}	-0.0536 [.1346]{.1449}
	200	-0.0307 [.0869]{.0921}	-0.0354 [.0507]{.0618}	-0.0647 [.0805]{.1033}
	500	-0.0528 [.0458]{.0699}	-0.0379 [.0331]{.0503}	-0.0552 [.0535]{.0768}
	1000	-0.0371 [.0331]{.0497}	-0.0267 [.0226]{.0349}	-0.0431 [.0417]{.0600}

An empirical illustration. The popular Boston house price data of Harrison and Rubinfeld (1978), corrected and augmented with longitude and latitude by Gilley and Pace (1996), is used to illustrate the application of our model and method. The data contains 506 observations (1 observation per census tract) from Boston Metropolitan Statistical Area, and is now freely available through the `sedep` package of the open source software R. The response variable is the median value (corrected) of owner-occupied homes in 1000's (`MEDV`), and the thirteen explanatory variables are: per capita crime rate by town (`crime`); proportion of residential land zoned for lots over 25,000 square feet (`zoning`); proportion of non-retail business acres per town (`industry`); Charles River dummy variable (= 1 if tract bounds river; 0 otherwise) (`charlesr`); nitric oxides concentration (parts per 10 million)(`nox`); average number of rooms per dwelling (`room`); proportion of owner-occupied units built prior to 1940 (`houseage`); weighted distances to five Boston employment centres (`distance`); index of accessibility to radial highways (`access`); full-value property-tax rate per 10,000 (`taxrate`); pupil-teacher ratio by town (`ptratio`); $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town (`blackpop`); and lower status of the population proportion (`lowclass`).

We use the Euclidean distance in terms of longitude and latitude to set up the spatial weight matrix. We choose the threshold distance to be 0.05 which gives a W_n matrix with 19.08% non-zero elements. The instrumental variables are $W_n X_n^*$, where X_n^* contains variables `access, taxrate, ptratio, blackpop`, and `lowclass`, chosen by excluding the explanatory variables that are not significant in the OLS regression, dummy variables, and the variables causing strong correlation among the columns of $(W_n X_n, X_n)$.

The results are summarized in Table 6 where all regressors are standardized.[11] From the results we see that while the regression coefficients under the SQAR model all have the same sign as those under OLS regression, their magnitude do change across the quantile points. Thus, the way that the explanatory variable affect the house price is different at different points of the distribution of house price. More interestingly, we observe that the spatial effect also changes across the quantile points, confirming our arguments in motivating our SQAR model given in the introduction.

5 Concluding Remarks

We proposed a SAR model under quantile restrictions, referred to as the *spatial quantile autoregression* (SQAR) in this article. The IVQR method of Chernozhukov and Hansen (2005, 2006, 2008) is extended to allow for heteroscedasticity and dependence in data for estimating the proposed SQAR model. Large sample properties of the IVQR estimator for the SQAR 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.

[11] Note that in calculating these standard errors based on the method introduced in Sec. 3.3, it is important to standardize the exogenous regressors for numerical stability.

Table 6. Summary of IVQR Estimates (SEs) for Boston House Price Data

	$\tau = .1$	$\tau = .25$	$\tau = .5$	$\tau = .75$	$\tau = .9$
constant	10.4743 (0.128)	14.9486 (0.206)	18.7180 (0.099)	20.1739 (0.143)	19.6130 (0.807)
crime	-0.7244 (0.323)	-0.8797 (0.091)	-1.2385 (0.083)	-0.5911 (0.244)	-1.3094 (0.137)
zoning	0.0994 (0.121)	0.4723 (1.063)	0.7923 (0.304)	1.5696 (0.753)	1.1969 (0.448)
industry	-0.0985 (0.582)	-0.0749 (0.838)	-0.0111 (0.509)	-0.4280 (0.194)	-0.4731 (0.237)
charlesrv	-0.0099 (0.829)	0.0851 (1.287)	0.1949 (0.989)	0.6660 (1.215)	1.5411 (0.239)
noxsq	-0.3783 (2.703)	-0.2933 (1.805)	-0.7690 (1.618)	-1.3224 (1.502)	-2.6972 (0.425)
rooms2	2.3250 (1.146)	2.8882 (2.203)	3.2511 (1.723)	3.7818 (1.404)	3.6638 (0.354)
houseage	-0.6176 (0.749)	-0.6554 (0.909)	-0.6162 (0.238)	-0.3174 (0.332)	-0.1104 (0.268)
distance	-1.2191 (2.237)	-1.5622 (0.426)	-2.1310 (1.282)	-2.7019 (0.827)	-2.7668 (0.551)
access	-0.1021 (0.587)	0.8546 (0.168)	1.5911 (0.145)	2.6215 (0.288)	3.9468 (0.235)
taxrate	-1.6277 (0.739)	-2.1223 (0.440)	-1.7878 (0.254)	-1.9147 (0.382)	-1.0539 (0.125)
ptratio	-0.5355 (1.014)	-0.8589 (2.049)	-1.3303 (1.852)	-1.8904 (1.484)	-2.6259 (0.341)
blackpop	0.7180 (0.201)	0.8437 (0.225)	1.0651 (0.304)	1.2466 (0.160)	1.2037 (0.162)
lowclass	-2.8720 (0.422)	-2.2938 (0.472)	-2.4619 (0.477)	-2.4131 (0.423)	-2.8703 (0.293)
spatial	0.3512 (0.054)	0.2177 (0.051)	0.1282 (0.050)	0.1812 (0.098)	0.3464 (0.116)

The new model and estimation method give important extensions to both the standard spatial regression models and the standard quantile regression models. It also extends the IVQR technique to allow for dependence and heteroscedasticity. These extensions prove to be very useful to the applied researchers.

Appendix: Proof of the Main Results

Lemma A.1 *(Kelejian and Prucha, 1999; Lee, 2002): Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are uniformly bounded in both row and column sums. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(\ell_n^{-1})$. Then*
(i) the sequence $\{A_nB_n\}$ are uniformly bounded in both row and column sums,
(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and
(iii) the elements of A_nC_n and C_nA_n are uniformly $O(\ell_n^{-1})$.

To proceed with the proofs of our main results, it is helpful to review the frequently used notation and functions. Recall that λ and β denote generically the spatial lag parameter and the coefficients of the covariates $x_{n,i}$, and γ the coefficients of the instruments $z_{n,i}$.

Notation:

$\theta_{0\tau}$	$= (\lambda_{0\tau}, \beta'_{0\tau})'$:	true value of $\theta = (\lambda, \beta')'$ at a fixed τ point,
$\hat{\theta}_{n\tau}$	$= (\hat{\lambda}_{n\tau}, \hat{\beta}'_{n\tau})'$:	IVQR estimator of $\theta_{0\tau}$,
$\alpha_{0\tau}$	$= (\beta'_{0\tau}, 0')'$:	true value of $\alpha = (\beta', \gamma')'$ at a fixed τ point,
$\alpha_{0\tau}(\lambda)$	$= (\beta'_{0\tau}(\lambda), \gamma'_{0\tau}(\lambda))'$:	true value of α given τ and λ, defined in (3.3),
$\hat{\alpha}_{n\tau}(\lambda)$	$= (\hat{\beta}'_{n\tau}(\lambda), \hat{\gamma}'_{n\tau}(\lambda))'$:	QR estimator of $\alpha_{0\tau}(\lambda)$ given λ, defined in (2.19),
$\sigma_{n,i}$	$= 1 + \underline{\lambda}\bar{y}_{n,i} + \underline{\beta}'x_{n,i}$:	used in the expression $u_{n,i} = \sigma_{n,i}[\varepsilon_{n,i} - Q_\varepsilon(\tau)]$,
$\bar{u}_{n,i}$	$= \sum_{k\neq i}^n g_{n,ik}u_{n,k}$	independent of $u_{n,i}$ if the $u_{n,k}$'s are independent,
$c_{n,i}$	$= e'_{n,i}G_nX_n\beta_{0\tau} + \bar{u}_{n,i}$:	defined below (3.6).

Functions:

$\rho_\tau(u)$	$= [\tau - 1(u \le 0)]\,u$, where $1(\cdot)$ is the usual indicator function,
$\psi_\tau(u)$	$= \tau - 1(u \le 0)$, which is the directional derivative of $\rho_\tau(u)$,
$b_{n,i}(\lambda)$	$= 1 + (\lambda_{0\tau} - \lambda)g_{n,ii}$, as in Assumption 2(v), $g_{n,ii}$ is the (i,i) element of G_n,
$u_{n,i}(\lambda)$	$= y_{n,i} - \lambda\bar{y}_{n,i} - \alpha'_{0\tau}(\lambda)\xi_{n,i}$, where $\xi_{n,i} = (x'_{n,i}, z'_{n,i})'$,
$a_{n,i}(\lambda, \alpha)$	$= (\lambda - \lambda_{0\tau})c_{n,i} + (\alpha - \alpha_{0\tau})'\xi_{n,i}$, defined below (3.7),
$c_{n,i}(\lambda, \alpha)$	$= c_{n,i} + a_{n,i}(\lambda, \alpha)g_{n,ii}/b_{n,i}(\lambda)$, defined below (3.8),
$Q_{n\tau}(\lambda, \alpha)$	$= \frac{1}{n}\sum_{i=1}^n \rho_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \alpha'\xi_{n,i})$, as in (2.18), expectation has limit $Q_\tau(\lambda, \alpha)$,
$S_{n\tau}(\lambda, \alpha)$	$= \frac{1}{n}\sum_{i=1}^n \psi_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \alpha'\xi_{n,i})\xi_{n,i}$, as in (3.4), expectation has limit $S_\tau(\lambda, \alpha)$,
$J_{n\tau}(\lambda, \alpha)$	$= \frac{1}{n}\sum_{i=1}^n \mathrm{E}\left\{f_{n,i}\left(\frac{a_{n,i}(\lambda,\alpha)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)\frac{\xi_{n,i}}{b_{n,i}(\lambda)}\left[c_{n,i}(\lambda, \alpha), \xi'_{n,i}\right]\right\}$, as in (3.8).

A notational convention is followed for a *concentrated function*, i.e., $a_{n,i}(\lambda) \equiv a_{n,i}[\lambda, \alpha_{0\tau}(\lambda)]$, and similarly defined for $c_{n,i}(\lambda)$, $J_{n\tau}(\lambda)$, and $S_{n\tau}(\lambda)$. Also, a function evaluated at the true parameter value is denoted by dropping the parentheses, e.g., $J_{n\tau} \equiv J_{n\tau}(\lambda_{0\tau}, \alpha_{0\tau})$.

Proof of Theorem 3.1.

Let $\Delta \equiv \Delta(\lambda, \tau) = \sqrt{n}[\alpha(\lambda, \tau) - \alpha_{0\tau}(\lambda)]$ and $\hat{\Delta}_{n\tau}(\lambda) = \sqrt{n}[\hat{\alpha}_{n\tau}(\lambda) - \alpha_{0\tau}(\lambda)]$. Let $u^*_{n,i}(\lambda, \Delta(\lambda, \tau)) = u_{n,i}(\lambda) - n^{-1/2}\Delta(\lambda, \tau)'\xi_{n,i} = y_{n,i} - \lambda\bar{y}_{n,i} - \alpha(\lambda, \tau)'\xi_{n,i}$. It follows from Step (iii) leading to (2.21) that

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

Set

$$V_{n\tau}(\lambda;\Delta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\tau(u^*_{n,i}(\lambda,\Delta))\xi_{n,i}, \tag{A.2}$$

which is $\sqrt{n}S_{n\tau}(\lambda,\alpha(\lambda,\tau))$, rewritten 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 *Let $M < \infty$ and $\mathcal{V}_{n\tau}(\lambda;\Delta) \equiv -\{V_{n\tau}(\lambda;\Delta) - V_{n\tau}(\lambda;0) - E[V_{n\tau}(\lambda;\Delta) - V_{n\tau}(\lambda;0)]\}$. Suppose Assumptions 1-6 hold. Then*

$$\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\|\mathcal{V}_{n\tau}(\lambda;\Delta)\| = o_p(1).$$

Proof. We first establish a pointwise convergence result. We need to show

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

which holds if the kth component of the $(p+q)\times 1$ vector $\mathcal{V}_{n\tau}(\lambda;\Delta)$,

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

By construction, $\mathrm{E}\mathcal{V}_{n\tau k}(\lambda;\Delta) = 0$. By Assumption 5, $\mathrm{Var}(\mathcal{V}_{n\tau k}(\lambda;\Delta)) = o(1)$. Thus (A.4) holds by the Chebyshev's inequality.

Let $\alpha_{0\tau\lambda} \equiv \alpha_{0\tau}(\lambda)$ and

$$\bar{a}_{n,i}(\lambda,\Delta) \equiv (\lambda - \lambda_{0\tau})c_{n,i} + (\alpha_{0\tau\lambda} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i}. \tag{A.5}$$

Clearly, $\bar{a}_{n,i}(\lambda,0) = a_{n,i}(\lambda,\alpha_{0\tau}(\lambda)) = a_{n,i}(\lambda)$, and $\bar{a}_{n,i}(\lambda,\Delta) = a_{n,i}(\lambda,\alpha(\lambda,\tau))$ for $\Delta = \sqrt{n}(\alpha(\lambda,\tau) - \alpha_{0\tau\lambda})$. Noting that

$$\begin{aligned}
& y_{n,i} - \lambda\bar{y}_{n,i} - (\alpha_{0\tau\lambda} + n^{-1/2}\Delta)'\xi_{n,i} \\
&= u_{n,i} - (\lambda - \lambda_{0\tau})\bar{y}_{n,i} - (\alpha_{0\tau\lambda} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i} \\
&= [1 + (\lambda_{0\tau} - \lambda)g_{n,ii}]\,u_{n,i} - (\lambda - \lambda_{0\tau})\sum_{l\neq i}^{n} g_{n,il}u_{n,l} \\
&\quad -[(\lambda - \lambda_{0\tau})e'_{n,i}G_nX_n\beta_{0\tau} + (\alpha_{0\tau\lambda} + n^{-1/2}\Delta - \alpha_{0\tau})'\xi_{n,i}] \\
&= b_{n,i}(\lambda)u_{n,i} - \bar{a}_{n,i}(\lambda,\Delta),
\end{aligned}$$

we have

$$1[y_{n,i} - \lambda\bar{y}_{n,i} \leq (\alpha_{0\tau\lambda} + n^{-1/2}\Delta)'\xi_{n,i}] = 1[b_{n,i}(\lambda)u_{n,i} \leq \bar{a}_{n,i}(\lambda,\Delta)]. \tag{A.6}$$

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

$$\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left|\mathcal{V}^+_{n\tau k}(\lambda;\Delta)\right| = o_p(1) \text{ and } \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left|\mathcal{V}^-_{n\tau k}(\lambda;\Delta)\right| = o_p(1), \tag{A.7}$$

where $\mathcal{V}^+_{n\tau k}$ and $\mathcal{V}^-_{n\tau k}$ are defined analogously to $\mathcal{V}_{n\tau k}$ but with the kth element $\xi_{n,ik}$ of $\xi_{n,i}$ being 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.7) since the other case is similar. Define for every $\kappa \in \mathbb{R}$, $\bar{a}_{n,i}(\lambda,\Delta,\kappa) = \bar{a}_{n,i}(\lambda,\Delta) + \kappa \| n^{-1/2}\xi_{n,i}\|$, and

$$\begin{aligned}\tilde{\mathcal{V}}^{+}_{n\tau k}(\lambda;\Delta,\kappa) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta,\kappa)\right] - \mathrm{E}1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta,\kappa)\right] \\ &\quad -1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,0)\right] + \mathrm{E}1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,0)\right]\}\,\xi^{+}_{n,ik}.\end{aligned}$$

Note that $\tilde{\mathcal{V}}^{+}_{n\tau k}(\lambda;\Delta,0) = \mathcal{V}^{+}_{n\tau k}(\lambda;\Delta)$. We follow Koul (1991) and Bai (1994) to show that the first part of (A.7) is a consequence of the following result

$$\sup_{\lambda\in\Lambda}\left|\tilde{\mathcal{V}}^{+}_{n\tau k}(\lambda;\Delta,\kappa)\right| = o_p(1) \text{ for every given } \Delta \text{ and } \kappa. \tag{A.8}$$

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 $\Delta'\xi_{n,i} \le \Delta'_s\xi_{n,i} + \sigma\|\xi_{n,i}\|$ for any $\Delta \in \Gamma_s$, it follows from the monotonicity of the indicator function and the nonnegativity of $\xi^{+}_{n,ik}$ that for any $\Delta\in\Gamma_s$,

$$\begin{aligned}&\mathcal{V}^{+}_{n\tau k}(\lambda;\Delta) - \tilde{\mathcal{V}}^{+}_{n\tau k}(\lambda;\Delta_s,\sigma) \\ &\quad= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta_s,\sigma)\right] - 1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta)\right]\right\}\xi^{+}_{n,ik} \\ &\qquad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta)\right] - 1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta_s,\sigma)\right]\right\}\xi^{+}_{n,ik} \\ &\quad\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta_s,\sigma)\right] - 1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta)\right]\right\}\xi^{+}_{n,ik}.\end{aligned}$$

A reverse inequality holds with σ replaced by $-\sigma$ for all $\Delta\in\Gamma_s$. By the triangle inequality, Taylor expansions, and Assumptions 1(iii), 2(v), and 3(i)-(ii), we have for sufficiently large n,

$$\begin{aligned}&\sup_{\Delta\in\Gamma_s}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta_s,\sigma)\right] - 1\left[b_{n,i}(\lambda)u_{n,i} \le \bar{a}_{n,i}(\lambda,\Delta)\right]\right\}\xi^{+}_{n,ik}\right| \\ &\quad\le \sup_{\Delta\in\Gamma_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\mathrm{E}\left[F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s,\sigma)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right) - F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\right]\right|\xi^{+}_{n,ik} \\ &\quad\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\mathrm{E}\left[F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s,\sigma)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right) - F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\right]\right|\xi^{+}_{n,ik} \\ &\qquad+ \sup_{\Delta\in\Gamma_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\mathrm{E}\left[F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right) - F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\right]\right|\xi^{+}_{n,ik} \\ &\quad\le \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}f_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s,c_i^{*}\sigma)}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\frac{\sigma\left\|\xi_{n,i}\right\|}{|b_{n,i}(\lambda)|}\xi^{+}_{n,ik} \\ &\qquad+ \sup_{\Delta_s\in\Gamma_s}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}f_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta_s^{*})}{b_{n,i}(\lambda)}\,\middle|\,\bar{u}_{n,i}\right)\frac{\sigma\left\|\xi_{n,i}\right\|}{|b_{n,i}(\lambda)|}\xi^{+}_{n,ik} \\ &\quad\le \frac{C\sigma}{n}\sum_{i=1}^{n}\frac{\left\|\xi_{n,i}\right\|\xi^{+}_{n,ik}}{|b_{n,i}(\lambda)|} = \sigma O(1),\end{aligned}$$

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

$$\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\mathcal{V}_{n\tau k}^{+}(\lambda;\Delta)\right\| \leq \sup_{s\leq N(\sigma)}\sup_{\lambda\in\Lambda}\left|\tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta_s,\sigma)\right| + \sup_{s\leq N(\sigma)}\sup_{\lambda\in\Lambda}\left|\tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta_s,-\sigma)\right| + \sigma O_p(1).$$

By the compactness of Γ, the term σ can be made arbitrarily small and $N(\sigma)$ is finite. So we can prove (A.7) by proving (A.8).

To show (A.8), we use similar arguments. 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 $\alpha_{ni}(\lambda,\Delta) \equiv [\bar{a}_{n,i}(\lambda,\Delta) - \bar{a}_{n,i}(\lambda,0)]/b_{n,i}(\lambda)$. Then $\alpha_{ni}(\lambda,\Delta) = n^{-1/2}\Delta'\xi_{n,i}/b_{n,i}(\lambda)$. By Assumption 2(v), $\min_{1\leq i\leq n}\inf_{\lambda\in\Lambda}|1-(\lambda-\lambda_{0\tau})\,g_{n,ii}| \geq c_g/2$ for sufficiently large n. It follows that for any $\lambda,\lambda^*\in\lambda$ and sufficiently large n,

$$\begin{aligned}\sup_{|\lambda-\lambda^*|\leq\delta^*}|\alpha_{ni}(\lambda,\Delta)-\alpha_{ni}(\lambda^*,\Delta)| &= \sup_{|\lambda-\lambda^*|\leq\delta^*}\left|\frac{(\lambda-\lambda^*)\,n^{-1/2}\Delta'\xi_{n,i}}{[1-(\lambda-\lambda_{0\tau})\,g_{n,ii}]\,[1-(\lambda^*-\lambda_{0\tau})\,g_{n,ii}]}\right| \\ &\leq 4n^{-1/2}\delta^* c_g^{-2}\,\|\Delta\|\max_{1\leq i\leq n}\left\|\xi_{n,i}\right\| \leq Cn^{-1/2}\delta^*,\end{aligned}$$

where C is a large finite constant and the last inequality holds because $\max_{1\leq i\leq n}\left\|\xi_{n,i}\right\|$ is finite by Assumptions 3(i)-(ii). Define

$$\begin{aligned}&\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa,\varsigma)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda,\Delta,\kappa)}{b_{n,i}(\lambda)}+\varsigma n^{-1/2}C\delta^*\right) - \mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta,\kappa)}{b_{n,i}(\lambda)}+\varsigma n^{-1/2}C\delta^*\;\middle|\;\bar{u}_{n,i}\right)\\ &\qquad -1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\right)+\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\;\middle|\;\bar{u}_{n,i}\right)\Bigg\}\xi_{n,ik}^{+}.\end{aligned}$$

Then $\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa,0) = \tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa)$ for sufficiently large n. By the monotonicity of the indicator function and cumulative distribution function (cdf) and the nonnegativity of $\xi_{n,ik}^{+}$, we have that for all λ with $|\lambda-\lambda_s|\leq\delta^*$ and sufficiently large n,

$$\begin{aligned}&\tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa)-\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda_s;\Delta,\kappa,1)\\ &\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_s,\Delta,\kappa)}{b_{n,i}(\lambda_s)}+Cn^{-1/2}\delta^*\;\middle|\;\bar{u}_{n,i}\right)-\mathrm{E}F_{n,ii}\left(\frac{\bar{a}_{n,i}(\lambda_s,\Delta,\kappa)}{b_{n,i}(\lambda_s)}\;\middle|\;\bar{u}_{n,i}\right)\Bigg\}\xi_{n,ik}^{+}\\ &\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda_s,0)}{b_{n,i}(\lambda_s)}\right)-\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_s,0)}{b_{n,i}(\lambda_s)}\;\middle|\;\bar{u}_{n,i}\right)\\ &\qquad -1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\right)+\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\;\middle|\;\bar{u}_{n,i}\right)\Bigg\}\xi_{n,ik}^{+},\end{aligned}$$

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

n, we have

$$\begin{aligned}
&\sup_{\lambda\in\Lambda}\left|\tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa)\right| \\
&\quad\leq \max_{s}\left|\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda_s;\Delta,\kappa,1)\right| + \max_{s}\left|\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda_s;\Delta,\kappa,-1)\right| \\
&\quad + \max_{s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left[F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_s,\Delta,\kappa)}{b_{n,i}(\lambda_s)}+\frac{C\delta^*}{\sqrt{n}}\middle|\bar{u}_{n,i}\right)-F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_s,\Delta,\kappa)}{b_{n,i}(\lambda_s)}-\frac{C\delta^*}{\sqrt{n}}\middle|\bar{u}_{n,i}\right)\right]\xi_{n,ik}^{+} \\
&\quad + \sup_{\substack{\lambda_l,\lambda_m\in\lambda,\\ |\lambda_l-\lambda_m|\leq\delta^*}}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left[1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda_l,0)}{b_{n,i}(\lambda_l)}\right)-\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_l,0)}{b_{n,i}(\lambda_l)}\middle|\bar{u}_{n,i}\right)\right]\right.\right. \\
&\qquad\qquad\left.\left.-\left[1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda_m,0)}{b_{n,i}(\lambda_m)}\right)-\mathrm{E}F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda_m,0)}{b_{n,i}(\lambda_m)}\middle|\bar{u}_{n,i}\right)\right]\right\}\xi_{n,ik}^{+}\right|. \qquad (\mathrm{A.9})
\end{aligned}$$

The first two terms on the right hand side of (A.9) are $o_p(1)$ because $\left\|\bar{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa,\varsigma)\right\| = o_p(1)$ 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 no greater than $2Cc_f c_\xi\delta^*$ with $c_f \equiv \sup_{n\geq 1}\max_{1\leq i\leq n}\sup_{(u,\bar{u})} f_{n,i}(u|\bar{u}) < \infty$ by Assumption 1(iii) and $c_\xi \equiv \sup_{n\geq 1}\frac{1}{n}\sum_{i=1}^{n}\xi_{n,ik}^{+} < \infty$ by Assumptions 3(i)-(ii), which can be made arbitrarily small by choosing small enough δ^*. The last term in (A.9) is ensured to be small due to the stochastic equicontinuity property by Assumption 6. Hence $\sup_{\lambda\in\Lambda}|\tilde{\mathcal{V}}_{n\tau k}^{+}(\lambda;\Delta,\kappa)| = o_p(1)$ as $n\to\infty$ and $\delta^*\to 0$. ■

Lemma A.3 *Recall $J_{n\tau\alpha}(\lambda) = n^{-1}\sum_{i=1}^{n}E\Big[f_{n,i}\Big(\frac{a_{n,i}(\lambda)}{b_{n,i}(\lambda)}\Big|\bar{u}_{n,i}\Big)\Big]\frac{\xi_{n,i}\xi_{n,i}'}{b_{n,i}(\lambda)}$ by (3.9) and the remarks after (3.10). Suppose Assumptions 1-6 hold. Then*

$$\mathit{sup}_{\lambda\in\Lambda}\ \mathit{sup}_{\|\Delta\|\leq M}\|E[V_{n\tau}(\lambda;\Delta)-V_{n\tau}(\lambda;0)]+J_{n\tau\alpha}(\lambda)\Delta\| = o(1).$$

Proof. Let $\bar{a}_{n,i}(\lambda,\Delta)$ and $b_{n,i}(\lambda)$ be defined in (A.5) and Assumption A2(v), respectively. By Taylor expansions, we have for sufficiently large n:

$$\begin{aligned}
&\sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\|\mathrm{E}[V_{n\tau}(\lambda;\Delta)-V_{n\tau}(\lambda;0)]+J_{n\tau\alpha}(\lambda)\Delta\| \\
&\quad= \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left\{1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda,\Delta)}{b_{n,i}(\lambda)}\right)-1\left(u_{n,i}\leq\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\right)\right\}\xi_{n,i}-J_{n\tau\alpha}(\lambda)\Delta\right\| \\
&\quad= \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}\left[F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,\Delta)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)-F_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)\right]\xi_{n,i}-J_{n\tau\alpha}(\lambda)\Delta\right\| \\
&\quad= \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\left\|\frac{1}{n}\sum_{i=1}^{n}\int_0^1\mathrm{E}\left[f_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,0)+sn^{-1/2}\Delta'\xi_{n,i}}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)-f_{n,i}\left(\frac{\bar{a}_{n,i}(\lambda,0)}{b_{n,i}(\lambda)}\middle|\bar{u}_{n,i}\right)\right]ds\right. \\
&\qquad\qquad\left.\times\frac{\xi_{n,i}\xi_{n,i}'\Delta}{b_{n,i}(\lambda)}\right\| \\
&\quad\leq \sup_{\lambda\in\Lambda}\sup_{\|\Delta\|\leq M}\frac{\bar{c}_f}{n}\sum_{i=1}^{n}\left\|\int_0^1\frac{n^{-1/2}\Delta'\xi_{n,i}}{b_{n,i}(\lambda)}sds\right\|\left\|\frac{\xi_{n,i}\xi_{n,i}'\Delta}{b_{n,i}(\lambda)}\right\| \\
&\quad\leq \sup_{\lambda\in\Lambda}\frac{\bar{c}_f M^2}{2n^{3/2}}\sum_{i=1}^{n}\frac{\left\|\xi_{n,i}\right\|^3}{b_{n,i}(\lambda)^2}\leq\frac{c_f M^2}{2n^{3/2}}\sum_{i=1}^{n}\frac{\left\|\xi_{n,i}\right\|^3}{\inf_{\lambda\in\Lambda}b_{n,i}(\lambda)^2} = o(1),
\end{aligned}$$

where $\bar{c}_f = \sup_{n\geq 1}\max_{1\leq i\leq n}\sup_{(u,\bar{u})}|f_{n,i}^{(1)}(u|\bar{u})|$ with $f_{n,i}^{(1)}(\cdot|\bar{u})$ denoting the first derivative of $f_{n,i}(\cdot|\bar{u})$, the first inequality follows from the Taylor expansion and Assumption 1(iii), and the last inequality follows from the fact that $\xi_{n,i}$ is uniformly bounded under Assumptions 3(i)-(ii) and the fact that $\lim_{n\to\infty}\min_{1\leq i\leq n}\inf_{\lambda\in\Lambda}|b_{n,i}(\lambda)| > c_g > 0$ by Assumption 2(v). ■

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

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

Proof. By the computational properties of quantile regression [e.g., Theorem 3.3 of Koenker and Bassett (1978), Lemma A2 in Ruppert and Carroll (1980)] and Assumptions 3(i),

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

By the definition of $\alpha_{0\tau}(\lambda)\equiv(\beta_{0\tau}(\lambda)',\gamma_{0\tau}(\lambda)')'$ below (3.2), $S_\tau(\lambda,\alpha_{0\tau}(\lambda))=0$. It follows that

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

Proof of Theorem 3.2.

Refer to the notation listed at the beginning of the appendix, and define

$$\lambda_\tau^* \equiv \arg\min_\lambda\|\gamma_{0\tau}(\lambda)\|_A \ \beta_\tau^*\equiv\beta_{0\tau}(\lambda_\tau^*), \text{ and } \gamma_\tau^*\equiv\gamma_{0\tau}(\lambda_\tau^*),$$

where $\|B\|_A=\{B'AB\}^{1/2}$. Following **CH**, 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) $\hat{\lambda}_{n\tau}\xrightarrow{p}\lambda_{0\tau}$ and $\hat{\alpha}_{n\tau}\xrightarrow{p}\alpha_{0\tau}$; (iii) $\sqrt{n}(\hat{\theta}_{n\tau}-\theta_{0\tau})\xrightarrow{d}N(0,\Omega_\tau(A)\Sigma_{0\tau}\Omega_\tau(A)')$.

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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\psi_\tau(y_{n,i}-\lambda\bar{y}_{n,i}-\beta'x_{n,i}-0'z_{n,i})]\xi_{n,i}=0. \tag{A.10}$$

By the global convexity of $Q_\tau(\lambda,\alpha)$ in α for each λ, and the fact that $\alpha_{0\tau}(\lambda)=(\beta_{0\tau}(\lambda)',\gamma_{0\tau}(\lambda)')'$ is in the interior of $\mathcal{B}\times\mathbb{R}^q$, $\alpha_{0\tau}(\lambda)$ uniquely solves the first order condition of minimizing $Q_\tau(\lambda,\alpha)$ over α:

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

We now show that $\lambda_\tau^* = \lambda_{0\tau}$ uniquely minimizes $\|\gamma_{0\tau}(\lambda)\|_A$ over λ subject to the constraint in (A.11). Clearly, $\|\gamma_{0\tau}(\lambda_{0\tau})\| = 0$ by (A.10) and $\lambda_{0\tau}$ satisfies (A.11). That is, $\lambda_{0\tau} \in \arg\min_\lambda \|\gamma_{0\tau}(\lambda)\|_A$ subject to the constraint in (A.11). It is also the unique solution by (A.10). Now $\beta_\tau(\lambda_\tau^*) = \beta_\tau(\lambda_{0\tau}) = \beta_{0\tau}$ by (A.11).

Step (ii): Let $o_p^*(1)$ denote $o_p(1)$ uniformly in $\lambda \in \Lambda$. By the remark after Theorem 3.1,

$$\|\hat{\alpha}_{n\tau}(\lambda) - \alpha_{0\tau}(\lambda)\| = o_p^*(1), \text{ and } \|\hat{\gamma}_{n\tau}(\lambda) - \gamma_{0\tau}(\lambda)\| = o_p^*(1) \text{ in particular.} \tag{A.12}$$

By Assumption 3(iii), $\hat{A}_n = A + o_p(1)$. It follows that $\|\hat{\gamma}_{n\tau}(\lambda)\|_{\hat{A}_n} - \|\gamma_{0\tau}(\lambda)\|_A = o_p^*(1)$. By Assumption 4(v), $\|\gamma_{0\tau}(\lambda)\|_A$ is continuous in λ; it is uniquely minimized at $\lambda_\tau^* = \lambda_{0\tau}$ by Step (i). It follows that $\hat{\lambda}_{n\tau} \xrightarrow{p} \lambda_{0\tau}$. Now let $\lambda_{n\tau} \xrightarrow{p} \lambda_{0\tau}$. By (A.12) and the continuity of $\alpha_{0\tau}(\lambda)$ in λ, $\hat{\alpha}_{n\tau}(\lambda_{n\tau}) \xrightarrow{p} \alpha_{0\tau}(\lambda_{0\tau}) = \alpha_{0\tau}$. In particular, $\hat{\alpha}_{n\tau} = \hat{\alpha}_{n\tau}(\hat{\lambda}_{n\tau}) \xrightarrow{p} \alpha_{0\tau}$ as desired.

Step (iii): Consider a small ball $B_{\epsilon_n}(\lambda_{0\tau})$ of radius ϵ_n centered at $\lambda_{0\tau}$. Let $\lambda_n \in B_{\epsilon_n}(\lambda_{0\tau})$ where $\epsilon_n \to 0$ slowly enough. Let $m_{ni}(\lambda, \alpha) \equiv \psi_\tau(y_{n,i} - \lambda\bar{y}_{n,i} - \alpha'\xi_{n,i})\xi_{n,i}$, $\mathrm{E}m_{ni}(\lambda_n,\ \hat{\alpha}_{n\tau}(\lambda_n)) \equiv \mathrm{E}[m_{ni}(\lambda,\alpha)]_{(\lambda,\alpha)=(\lambda_n,\hat{\alpha}_{n\tau}(\lambda_n))}$, and $M_n \equiv n^{-1/2}\sum_{i=1}^n [m_{ni}(\lambda_{0\tau}, \alpha_{0\tau}(\lambda_{0\tau})) - \mathrm{E}m_{ni}(\lambda_{0\tau},\ \alpha_{0\tau}(\lambda_{0\tau}))]$. By Lemma A.4 and the stochastic equicontinuity condition in Assumption 6(ii),

$$\begin{aligned}
O(n^{-1/2}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau(y_{n,i} - \lambda_n\bar{y}_{n,i} - \hat{\alpha}_{n\tau}(\lambda_n)'\,\xi_{n,i})\xi_{n,i} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n [m_{ni}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) - \mathrm{E}m_{ni}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n))] + \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}m_{ni}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) \\
&= M_n + \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}m_{ni}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) + o_p(1).
\end{aligned} \tag{A.13}$$

By Assumptions 1(i) and (iii) and the Fubini's theorem,

$$\begin{aligned}
\mathrm{E}\left[F_{n,i}(0\ |\bar{u}_{n,i})\right] &= \mathrm{E}\left[\int_{-\infty}^0 f_{n,i}(u\ |\bar{u}_{n,i})du\right] = \int_{-\infty}^0 \int_{-\infty}^\infty f_{n,i}(u\ |\bar{u}) f_{\bar{u}_{n,i}}(\bar{u})\, d\bar{u}du \\
&= \int_{-\infty}^0 f_{u_{n,i}}(u)du = \Pr(u_{n,i} \le 0) = \tau,
\end{aligned}$$

where $f_{\bar{u}_{n,i}}$ and $f_{u_{n,i}}$ denotes the marginal pdf's of $\bar{u}_{n,i}$ and $u_{n,i}$, respectively. With this, by Assumptions 1(iii) and 3(i)-(ii), the mean value theorem, and dominated convergence arguments, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}m_{ni}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}\left[F_{n,i}(0\ |\bar{u}_{n,i}) - F_{n,i}\left(\chi_{n,i}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n))\ |\bar{u}_{n,i}\right)\right]\xi_{n,i} \\
&= -\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[f_{n,i}\left(s_i^*\chi_{n,i}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n))\ |\bar{u}_{n,i}\right)\xi_{n,i}\frac{\bar{u}_{n,i} + e_{n,i}'G_n\xi_n\hat{\alpha}_{n\tau}(\lambda_n)}{b_{n,i}(\lambda_n)}\right]\sqrt{n}(\lambda_n - \lambda_{0\tau}) \\
&\quad -\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[f_{n,i}(s_i^*\chi_{n,i}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n))\ |\bar{u}_{n,i})\right]\frac{\xi_{n,i}\xi_{n,i}'}{b_{n,i}(\lambda_n)}\sqrt{n}(\hat{\alpha}_{n\tau}(\lambda_n) - \alpha_{0\tau}) \\
&= -(J_{\tau\lambda} + o_p(1))\sqrt{n}(\lambda_n - \lambda_{0\tau}) - (J_{\tau\alpha} + o_p(1))\sqrt{n}(\hat{\alpha}_{n\tau}(\lambda_n) - \alpha_{0\tau}),
\end{aligned} \tag{A.14}$$

where $\chi_{n,i}(\lambda,\alpha) \equiv \bar{a}_{n,i}(\lambda,\alpha)/b_{n,i}(\lambda)$ with $\bar{a}_{n,i}(\cdot,\cdot)$ and $b_{n,i}(\cdot)$ being defined, respectively, in (A.5) and Assumption 2(v), and s_i^* lies between 0 and 1. The last line follows from the definitions of $J_{\tau\lambda}$ and $J_{\tau\alpha}$ and the fact that $\bar{a}_{n,i}(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) \to 0$ and $b_{n,i}(\lambda_n) \to 1$ as $\epsilon_n \to 0$. This is because $\sum_{l\neq i}^n g_{n,il}\mathrm{E}u_{n,l} \leq \overline{\mu}\sum_{l=1}^n |g_{n,il}| = O(1)$ by Assumptions 1(ii), 2(iii) and Lemma A.2, $e_{n,i}'G_n\xi_n = O(1)$ by Assumptions 2(iii) and 3(i)-(ii) and Lemma A.2, and $(\lambda_n, \hat{\alpha}_{n\tau}(\lambda_n)) \to (\lambda_{0\tau}, \alpha_{0\tau})$ as $\epsilon_n \to 0$. Putting (A.13) and (A.14) together, we have

$$O(n^{-1/2}) = M_n - (J_{\tau\lambda} + o_p(1))\sqrt{n}(\lambda_n - \lambda_{0\tau}) - (J_{\tau\alpha} + o_p(1))\sqrt{n}(\hat{\alpha}_{n\tau}(\lambda_n) - \alpha_{0\tau}), \tag{A.15}$$

which implies that

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

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

$$\sqrt{n}(\hat{\beta}_{n\tau}(\lambda_n) - \beta_{0\tau}) = \bar{J}_{\tau\beta}M_n - \bar{J}_{\tau\beta}J_{\tau\lambda}(1 + o_p(1))\sqrt{n}(\lambda_n - \lambda_{0\tau}) + o_p(1),$$

and

$$\sqrt{n}(\hat{\gamma}_{n\tau}(\lambda_n) - 0) = \bar{J}_{\tau\gamma}M_n - \bar{J}_{\tau\gamma}J_{\tau\lambda}(1 + o_p(1))\sqrt{n}(\lambda_n - \lambda_{0\tau}) + o_p(1).$$

By Step (ii), with probability approaching one,

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

By Liapounov's central limit theorem, $M_n \xrightarrow{d} N(0, S_0)$. Hence

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

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

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

Simple algebra shows that

$$\sqrt{n}(\hat{\alpha}_{n\tau}(\hat{\lambda}_{n\tau}) - \alpha_{0\tau}) = J_{\tau\alpha}^{-1}[I_{p+q} - J_{\tau\lambda}(J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma}J_{\tau\lambda})^{-1}J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma}]M_n + o_p(1), \tag{A.16}$$

and

$$\begin{pmatrix} \sqrt{n}(\hat{\lambda}_{n\tau} - \lambda_{0\tau}) \\ \sqrt{n}(\hat{\beta}_{n\tau} - \beta_{0\tau}) \end{pmatrix} = \begin{pmatrix} (J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma}J_{\tau\lambda})^{-1}J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma} \\ \bar{J}_{\tau\beta}[I_{p+q} - J_{\tau\lambda}(J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma}J_{\tau\lambda})^{-1}J_{\tau\lambda}'\bar{J}_{\tau\gamma}'A\bar{J}_{\tau\gamma}] \end{pmatrix} M_n + o_p(1).$$

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

Proof of Corollary 3.3.

When $q = 1$, $\bar{J}_{\tau\gamma}J_{\tau\lambda}$ is a nonzero scalar. By (A.16) and the fact that $M_n = O_p(1)$, we have

$$\sqrt{n}(\hat{\gamma}_{n\tau}(\hat{\lambda}_\tau) - 0) = \bar{J}_{\tau\gamma}[I_{p+1} - J_{\tau\lambda}(\bar{J}_{\tau\gamma}J_{\tau\lambda})^{-1}\bar{J}_{\tau\gamma}]M_n + o_p(1) = o_p(1). \tag{A.17}$$

By (A.15) and (A.17) and the fact that $\hat{\lambda}_\tau \xrightarrow{p} \lambda_{0\tau}$, we have

$$[J_{\tau\lambda}\ J_{\tau\alpha,1:p}]\begin{pmatrix} \sqrt{n}(\hat{\lambda}_{n\tau} - \lambda_{0\tau}) \\ \sqrt{n}(\hat{\beta}_{n\tau}(\hat{\lambda}_{n\tau}) - \beta_{0\tau}) \end{pmatrix} = M_n + o_p(1),$$

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

Proof of the Result: Assumption 5* $\Rightarrow$ Assumption 5.

Let $\mathcal{S}_{nk,j}$ and $\mathcal{S}_{nk,jt}$ denote the partial sums of $n^{-1/2}\eta_{n,ik}$ over observations in group j and subgroup t of group j, respectively, i.e., $\mathcal{S}_{nk,j} = \sum_{t=1}^{m_{nj}} \mathcal{S}_{nk,jt}$, where $\mathcal{S}_{nk,jt} = \sum_{s\in\mathcal{G}_{njt}} n^{-1/2}\eta_{n,sk}$ and we suppress the dependence of the $\mathcal{S}$-quantities and $\eta_{n,sk}$ on (λ, Δ). Let $\mathcal{S}_{nk} = \sum_{j=1}^{J} \mathcal{S}_{nk,j}$. Because J and $p+q$ are finite, by Cauchy-Schwarz inequality it suffices to show that $\mathrm{Var}(\mathcal{S}_{nk,j}) = o(1)$ for each $j = 1, \cdots, J$ and $k = 1, \cdots, p+q$. Fix $j \in \{1, \cdots, J\}$ and $k \in \{1, \cdots, p+q\}$. Write

$$\mathrm{Var}(\mathcal{S}_{nk,j}) = \sum_{t=1}^{m_{nj}} \mathrm{Var}(\mathcal{S}_{nk,jt}) + 2\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}} \mathrm{Cov}(\mathcal{S}_{nk,jl}, \mathcal{S}_{nk,jt}) \equiv I_{n1} + I_{n2}$$

First we can show that $\mathrm{Var}(\eta_{n,sk}(\lambda)) \le C_1 n^{-1/2}$ and $\mathrm{Cov}(\eta_{n,ik}(\lambda), \eta_{n,sk}(\lambda)) \le C_2 n^{-1}$ for $i \ne s$ and finite constants C_1, C_2. It follows that

$$\mathrm{Var}(\mathcal{S}_{nk,jt}) = \frac{1}{n}\sum_{s\in\mathcal{G}_{njt}} \mathrm{Var}(\eta_{n,sk}) + \frac{1}{n}\sum_{i\in\mathcal{G}_{njt}}\sum_{s\in\mathcal{G}_{njt}, s\ne i} \mathrm{Cov}(\eta_{n,ik}, \eta_{n,sk}) \le C_1 n^{-3/2} n_{jt} + C_2 n^{-2} n_{jt}^2,$$

which implies that $I_{n1} \le C_1 n^{-3/2}\sum_{t=1}^{m_{nj}} n_{jt} + C_2 n^{-2}\sum_{t=1}^{m_{nj}} n_{jt}^2 \le C_1 n^{-1} + C_2 n_{jt}/n = o(1)$ by Assumption 5**(ii). Now by Assumption 5**(i),

$$I_{n2} \le 2\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}} \sqrt{\mathrm{Var}(\mathcal{S}_{nk,jl})\mathrm{Var}(\mathcal{S}_{nk,jt})}\alpha_{m_{nj}} = o(1)\sum_{l=1}^{m_{nj}-1}\sum_{t=l+1}^{m_{nj}} \alpha_{m_{nj}} = o(m_{nj}^2\alpha_{m_{nj}}) = o(1).$$

Consequently, $\mathrm{Var}(\mathcal{S}_{nk,j}) = o(1)$. 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.

Blundell, R., Powell, J. L., 2007. Censored regression quantiles with endogenous regressors. Journal of Econometrics 141, 68-83.

Case, A. C., 1991. Spatial patterns in household demand. Econometrica 59, 935-965.

Case, A. C., 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22, 491-508.

Case, A. C., H. S. Rosen, J. R. Hines, 1993. Budget spillovers and fiscal policy interdependence: evidence from the states. Journal of Public Economics 52, 285-307.

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.

Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: a robust inference approach. Journal of Econometrics 142, 379-398.

Chernozhukov, V., Hansen, C., Jansson, M., 2007. Inference approach for instrumental variable quantile regression. Economics Letters 95, 272-277.

Chernozhukov, V., Hansen, C., Jansson, M., 2009. Finite sample inference for quantile regression models. Journal of Econometrics 152, 93-103.

Cliff, A. D. and Ord, J. K., 1973. Spatial Autocorrelation. London: Pion.

Cliff, A. D. and Ord, J. K., 1981. Spatial Process, Models and Applications. London: Pion.

Gilley, O. W., Pace, R. K., 1996. On the Harrison and Rubinfeld Data. Journal of Environmental Economics and Management 31, 403-405.

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

Harrison, D., Rubinfeld, D. L., 1978. Hedonic housing prices and he demand for clear air. Journal of Environmental Economics and Management 5, 81-102.

Hogg, R. V., Craig, A. T., 1995. Introduction to Mathematical Statistics. 5th ed., Prentice Hall, New Jersey.

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, 99-121.

Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model, International Economic Review 40, 509-533.

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

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

Kim, T. H., Muller, C., 2004. Two-stage quantile regression when the first stage is based on quantile regression. Econometrics Journal 7, 218-231.

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

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

Koenker, R., Bassett, G., 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50, 43–61.

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

Kostov, P., 2009. A spatial quantile regression hedonic model of agricultural land prices. Spatial Economic Analysis 4, 53-72.

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

LeSage, J. P., 1999. The theory and practice of spatial econometrics. www.spatial-econometrics.com.

Lee, L-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252-277.

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.

Lee, S., 2007. Endogeneity in quantile regression models: a control function approach. Journal of Econometrics 141, 1131-1158.

Liao, W-C., Wang, X., 2010. Hedonic house prices and spatial quantile regression. IRES Working Paper, Institute of Real Estate Studies, National University of Singapore.

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

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

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. E., 1998. Contracting in space: an application of spatial statistics to discrete-choice models. Journal of Econometrics 85, 125-154.

Pinkse, J., Slade, M. E., 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. E., 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.

Powell, J. L., 1991. Estimation of monotonic regression models under quantile restrictions, in Nonparametric and Semiparametric Methods in Econometrics, pp. 357-384, W. Barnett, J. Powell, and G. Tauchen (eds), Cambridge: Cambridge University Press.

Robinson, P. M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6-17.

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., 2007. Instrumental variable estimation based on conditional median restriction. Journal of Econometrics 141, 350-382.

Zietz, J., Zietz, E. N., Sirmans, G. S., 2008. Determinants of house prices: a quantile regression approach. Journal of Real Estate Finance and Economics 37, 317-333.